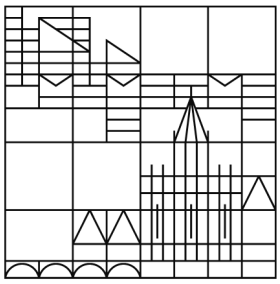
Trait and State Academic Emotions: Two Sides of the Same Coin?

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Madeleine Bieg

an der

Universität Konstanz



Mathematisch-Naturwissenschaftliche Sektion
Fachbereich Psychologie

Tag der mündlichen Prüfung: 9. Dezember 2013
1. Referent: Prof. Dr. Thomas Götz
Universität Konstanz
2. Referent: Prof. Dr. Markus Dresel
Universität Augsburg

Vorveröffentlichungen der Dissertation

Teilergebnisse dieser Dissertation wurden bereits in folgenden Beiträgen vorgestellt:

Publikationen

Bieg, M., Goetz, T., & Hubbard, K. (2013). Can I master it and does it matter? An intraindividual analysis on control-value antecedents of trait and state academic emotions. *Learning and Individual Differences*, 28, 102-108. doi: 10.1016/j.lindif.2013.09.006

Bieg, M., Goetz, T., & Lipnevich, A. A. (submitted). What students think they feel differs from what they really feel - Academic self-concept moderates the discrepancy between students' trait and state emotional self-reports.

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

Konferenzbeiträge

Bieg, M., Götz, T., Lüdtke, O., Pekrun,R. & Hall, N.C. (2013, September). *Geschlechtsunterschiede in Mathematik – Mädchen denken nur sie hätten mehr Angst.* Paper presented at the 14. Fachgruppentagung Pädagogische Psychologie der Deutschen Gesellschaft für Psychologie, Hildesheim.

Bieg, M., Goetz, T., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013, April). *The gender gap in math anxiety: Contradictory findings from trait vs. state assessments*. Paper presented at the Annual Meeting of the American Educational Reserach Associtation 2013, San Francisco.

Bieg, M., Götz, T., & Lipnevich, A. A. (2012, September). *Wenn du denkst du fühlst, dann denkst du nur du fühlst - Der Einfluss des akademischen Selbstkonzepts auf die Diskrepanz von trait- und state-basierten Emotionseinschätzungen*. Paper presented at the 48. Kongress der Deutschen Gesellschaft für Psychologie, Bielefeld.

Bieg, M., Goetz, T., & Lipnevich, A. A. (2012, July). *What students think they feel is different from what they really feel*. Paper presented at the 30th International Congress of Psychology 2012, Cape Town.

Goetz, T., Bieg, M., & Hall, N.C. (2012, July). *Do girls really experience more math anxiety than boys?* Paper presented at the 30th International Congress of Psychology 2012, Cape Town.

Bieg, M., Goetz, T., Hubbard, K., & Keller, M. (2012, April). *Control and Value as Antecedents of State-based and Trait-based Achievement Emotions*. Paper presented at the AERA 2012 conference, Vancouver.

Bieg, M., Keller, M., & Götz, T. (2011, September). *Antezedenzien von State-Emotionen im Lern- und Leistungskontext*. Paper presented at the 13. Fachgruppentagung Pädagogische Psychologie der DGPs, Erfurt.

Eigenabgrenzung

Wie der entsprechenden Angabe der Co-Autoren zu entnehmen ist (siehe Vorveröffentlichungen der Dissertation), ist die vorliegende Arbeit unter Mithilfe verschiedener Personen entstanden.

Im ersten und zweiten Artikel habe ich hauptverantwortlich die Fragestellung und den theoretischen Hintergrund erarbeitet, die Analyse der Daten, Interpretation der Ergebnisse und das Verfassen des Manuskripts übernommen. Im dritten Artikel bestand meine Eigenleistung in der Idee der Fragestellung und der Mitwirkung in ihrer theoretischen Einbettung. Überdies war ich für die Analyse der Daten verantwortlich und habe bei der Interpretation der Ergebnisse sowie beim Verfassen des Manuskripts mitgearbeitet.

Contents

Summary

Emotions in the school setting are gaining increasing attention among educational researchers but also among practitioners and policy makers. Emotions in achievement contexts, referred to as academic emotions, are of high importance with regard to students' self-regulated learning, academic achievement, life-long learning, and career choices but are also valuable outcomes themselves. Yet, what do we mean when we are talking about emotions? An important distinction needs to be made, namely the one between trait and state emotions. Trait emotions are seen as habitual tendencies whereas state emotions are emotions experienced in a specific situation. When studying academic emotions, researchers usually rely on the assessment of emotions via self-reports from study participants, and a large proportion of previous studies have investigated emotions through the use of generalized self-reports ("How much enjoyment do you experience in general?"; i.e., trait emotions). However, momentary assessments examining actual emotions in achievement and learning situations ("How much enjoyment are you experiencing right now?"; i.e., state emotions) are becoming more popular as they are believed to be more ecological valid. It is assumed that state emotions are directly assessed and thus influenced by situational cues, whereas in trait assessments, individuals' beliefs and semantic knowledge affect outcomes of the assessment (accessibility model of emotional self-report; Robinson & Clore, 2002). Thus, there may be a discrepancy between trait and state emotions. Research that explicitly compares trait and state emotions in the academic context is lacking, however, this appears to be a promising enterprise for determining whether it is justifiable to draw conclusions about trait emotions from state emotions and vice versa. In order to close this gap in educational research on emotions, the present dissertation comprises three empirical studies that aimed at comparing trait and state emotions and their assessments with regard to structural (Study 1) as well as mean-level differences (Study 2 and Study 3).

The first study explored structural relations between cognitive appraisal antecedents and academic emotions as stated in Pekrun's control-value theory (2006). The appraisals of control and value, and the interaction of the two as predictors of emotions, were studied while using multiple trait and state assessments in one sample of 120 students in grades 8 and 11. Participants were asked about their control and value appraisals, and the discrete emotions of pride, anxiety, and boredom, in four subject domains. The appraisal antecedents as well as the emotions were assessed trait-based and state-based. In line with the hypotheses, results showed that control positively predicted pride and negatively predicted anxiety and boredom.

Value positively predicted pride and anxiety and negatively predicted boredom. Furthermore, the interaction between control and value predicted emotions over and above the single main effects. An intraindividual approach was utilized, meaning data were analyzed within persons (multiple trait and state measurement points per person) rather than between persons. The analyses revealed that appraisal-emotion relationships were quite similar in trait and state data.

In the second study, trait and state assessments of academic emotions were compared with regard to mean-level differences to investigate whether there was a discrepancy between the two types of academic emotions and whether self-concept of ability moderated this discrepancy. A total of 225 secondary school students from two different countries enrolled in grades 8 and 11 (German sample; $n = 94$) and grade 9 (Swiss sample; $n = 131$) participated. Students' trait academic emotions of enjoyment, pride, anger, and anxiety in mathematics were assessed with a self-report questionnaire. Furthermore, state academic emotions were assessed through the use of the experience-sampling method while participants were in class. The results revealed that students' scores on the trait assessment of emotions were generally higher than their scores on the state assessment. Further, as expected, students' academic self-concept in the domain of mathematics was shown to partly explain the discrepancy between scores on trait and state emotions. Results indicated that there was a belief-driven discrepancy between what students think they feel (trait emotion) and what they actually feel (state emotion). Thus, the two methods are quite different and trait emotions generally being rated higher than state emotions, which has important implications for future studies that use self-reports to assess academic emotions.

Study 3 sought to examine gender differences in trait (habitual) versus state (momentary) mathematics anxiety in two study samples. In line with the accessibility model of emotional self-report (Robinson & Clore, 2002), it was assumed that the frequently reported difference in trait mathematics anxiety between boys and girls would not emerge in state emotions. In the first study, 584 students were recruited from grades 5 to 10, and in the second study, 111 high school students from grades 8 and 11 participated. For trait math anxiety, the findings from both studies replicated previous research showing female students to report higher levels of anxiety than male students. However, no gender differences were observed for state anxiety as assessed by experience-sampling during a math test (first study) and when attending math classes (second study). The discrepant findings for trait versus state math anxiety were partly accounted for by students' competence beliefs in mathematics, with

female students showing lower perceived competence than male students despite having the same average math grades.

The three studies included in the present dissertation found that, although the structural relations between appraisal antecedents and emotions were found to be similar in trait and state data (Study 1), there were clear discrepancies between trait and state emotions with regard to mean-levels (Study 2 and Study 3). This discrepancy can be explained by students' gender (Study 3) but also by subjective control beliefs that students hold (Study 2 and Study 3). The results of the present studies will hopefully encourage future researchers of academic emotions to clearly operationalize and differentiate between emotions as traits or states as both seem to be of value depending on the respective research question. For example, trait emotions have a stronger relation to future behavior and choices (Wirtz, Kruger, Napa Scollon, & Diener, 2003) but are unable to capture situational fluctuations of emotions. Findings from the present dissertation also strengthen ongoing endeavors to positively influence students' subjective control conceptualized from either a trait (e.g., students' academic self-concept) or state (e.g., subjective situational control) perspective. Implications for future research and practice are discussed, especially with regard to the importance of subjective beliefs and emotions in the achievement context.

Zusammenfassung

Emotionen im schulischen Kontext wird sowohl in der Forschung als auch unter Praktikern und in der Politik zunehmende Bedeutung zugesprochen. Emotionen im Lern- und Leistungskontext spielen in Bezug auf selbstreguliertes Lernen, akademische Leistung, lebenslanges Lernen sowie Fächer- und Berufs- oder Studienwahl eine bedeutsame Rolle, sind aber auch an sich wertvolle Produkte des Lernprozesses. Was meinen wir jedoch genau damit, wenn wir über Emotionen sprechen? Eine wichtige Unterscheidung wird zwischen Trait- und State-Emotionen getroffen. Trait-Emotionen werden als habituelle Tendenzen, mit einer bestimmten Emotion zu reagieren, definiert. State-Emotionen hingegen beziehen sich auf Emotionen, so wie sie im Moment in einer spezifischen Situation erlebt werden. Zur Untersuchung von Lern- und Leistungsemotionen werden in Studien meist Selbstberichte der Teilnehmenden eingeholt. Eine Vielzahl bisheriger Studien untersuchte Emotionen mittels generalisierter Einschätzung der Emotionsintensitäten im Selbstbericht („Wie viel Freude erlebst du im Allgemeinen?“; d.h. Trait-Emotionen). State-Erhebungen, die Emotionen in der tatsächlichen Lern- und Leistungssituation erfassen („Wie viel Freude erlebst du in diesem Moment?“; d.h. State-Emotionen), werden jedoch zunehmend beliebter aufgrund ihrer erwarteten höheren ökologischen Validität. Es wird angenommen, dass State-Emotionen direkt erfasst werden können und folglich durch die konkrete Situation beeinflusst werden, während bei Trait-Erhebungen subjektive Überzeugungen und semantisches Wissen die Erhebung beeinflussen (Modell zur Zugänglichkeit emotionaler Selbstberichte; accessibility model of emotional self-report; Robinson & Clore, 2002). Folglich könnte eine Diskrepanz zwischen Trait- und State-Emotionen bestehen. Forschung, die explizit und systematisch Trait- und State-Emotionen im akademischen Kontext vergleicht, fehlt bislang. Dies stellt jedoch ein vielversprechendes und dringend notwendiges Unterfangen dar, um zu bestimmen, inwieweit es gerechtfertigt ist Schlüsse über Trait-Emotionen auf der Grundlage von State-Emotionen und umgekehrt zu ziehen. Um diese Lücke im Bereich der Emotionsforschung in der Pädagogischen Psychologie zu schließen, vereint die vorliegende Dissertation drei empirische Studien, die darauf abzielten, Trait- und State-Emotionen und deren Erfassung im Hinblick auf strukturelle (Studie 1) und Mittelwerts-Unterschiede (Studie 2 und Studie 3) zu vergleichen.

In der ersten Studie wurden strukturelle Beziehungen zwischen kognitiven Appraisal-Antezedenzien und Emotionen entsprechend Pekruns Kontroll-Wert-Theorie (Pekrun, 2006) untersucht. Die Appraisals (Einschätzungen) Kontrolle und Wert sowie ihre Interaktion

wurden als Prädiktoren von Emotionen untersucht, wobei sowohl mehrere Trait-Erhebungen als auch mehrere State-Erhebungen in einer Stichprobe von 120 Schülerinnen und Schülern der Klassenstufen 8 und 11 herangezogen wurden. Studienteilnehmerinnen und –teilnehmer wurden bezüglich ihrer Kontroll- und Wertappraisals sowie der diskreten Emotionen Stolz, Angst und Langeweile in vier Fächern befragt. Appraisal-Antezedenzien und Emotionen wurden trait- und state-basiert erfragt. Entsprechend der Hypothesen zeigten die Ergebnisse, dass Kontrolle positiv mit Stolz zusammenhängt und negativ mit Angst und Langeweile. Die Einschätzung des subjektiven Werts sagte Stolz und Angst positiv vorher und Langeweile negativ. Des Weiteren war auch die Interaktion zwischen Kontrolle und Wert ein Prädiktor der Emotionen zusätzlich zu den einzelnen Haupteffekten. Die Daten wurden mittels intraindividuellem Ansatz ausgewertet, d.h. die Relationen wurden innerhalb von Personen (mehrere Trait- und State-Messzeitpunkte pro Person) analysiert und nicht zwischen Personen. Die Analysen ergaben, dass die Appraisal-Emotions-Beziehungen bei Trait- und State-Erhebungen relativ ähnlich waren.

In der zweiten Studie wurden Trait- und State-Erhebungen von Lern- und Leistungsemotionen in Bezug auf Mittelwertsunterschiede verglichen. Es wurde untersucht, ob eine Diskrepanz zwischen Trait- und State-Emotionen besteht und ob das akademische Fähigkeitsselbstkonzept diese Diskrepanz moderiert. Insgesamt 225 Gymnasiasten aus zwei verschiedenen Ländern aus Klassenstufe 8 und 11 (deutsche Stichprobe; n = 94) und Klassenstufe 9 (schweizerische Stichprobe; n = 131) nahmen an der Studie teil. Mittels Selbstberichtsfragebogen wurden die Trait-Emotionen Freude, Stolz, Ärger und Angst der Schülerinnen und Schüler in Mathematik erhoben. Des Weiteren wurden State-Emotionen mittels Experience-Sampling Methode während des Unterrichts erfasst. Die Ergebnisse zeigten, dass die Werte der Schüler bei Trait-Emotionserhebungen allgemein höher waren als die Werte bei den State-Erhebungen. Überdies konnte erwartungsgemäß das Selbstkonzept in Mathematik teilweise die Diskrepanz zwischen den beiden Werten von Trait- und State-Emotionen erklären. Die Ergebnisse deuten darauf hin, dass eine durch Überzeugungen getriebene Diskrepanz zwischen dem besteht, was Schüler denken, was sie fühlen (Trait-Emotion) und dem, was sie wirklich fühlen (State-Emotion). Folglich sind die zwei Methoden, die beide Emotionen erfassen sollen, deutlich unterschiedlich, wobei Trait-Emotionen allgemein höher eingeschätzt werden als State-Emotionen. Dies hat wichtige Implikationen zur Folge für zukünftige Studien, die Selbstberichte nutzen, um akademische Emotionen zu erfassen.

Studie drei hatte zum Ziel Geschlechterunterschiede in Trait- versus State-Mathematikangst in zwei Stichproben zu untersuchen. Entsprechend des „Modells zur Zugänglichkeit emotionaler Selbstberichte“ (accessibility model of emotional self-report; Robinson & Clore, 2002) wurde angenommen, dass die vielberichtete höhere Ausprägung des Angstempfindens in Mathematik bei Mädchen im Vergleich zu Jungen nicht bei State-Emotionen zu finden ist. Zwei Studien wurden durchgeführt, um Geschlechterunterschiede in Trait-Angst und State-Angst zu untersuchen. In der ersten Studie nahmen 584 Gymnasiasten der Klassenstufen 5 bis 10 und in der zweiten Studie 111 Gymnasiasten der Klassenstufen 8 und 11 teil. In Bezug auf Trait-Angst replizierten beide Studien bisherige Forschungsergebnisse, die zeigten, dass Mädchen höhere Angstwerte berichten als Jungen. Es wurden allerdings keine Geschlechterunterschiede in Bezug auf State-Angst (erfasst mittels Experience-Sampling) während eines Mathematiktests (erste Studie) und während des Mathematikunterrichts (zweite Studie) gefunden. Die unterschiedlichen Befunde für Trait- und State-Mathematikangst konnten teilweise durch die Kompetenzüberzeugungen der Schülerinnen und Schüler erklärt werden, wobei Schülerinnen trotz durchschnittlich gleicher Mathematiknoten niedrigere wahrgenommene Kompetenz angaben als Jungen.

Die drei Studien der vorliegenden Dissertation ergaben, dass trotz gleicher struktureller Beziehungen zwischen Appraisal-Antezedenzien und Emotionen bei Trait- und State-Daten (Studie 1), klare Diskrepanzen zwischen Trait- und State-Emotionen in Bezug auf Mittelwerte bestehen (Studie 2 und 3). Diese Diskrepanz konnte sowohl durch das Geschlecht der Schüler (Studie 3) als auch durch subjektive Kontrollüberzeugungen, die Schülerinnen und Schüler haben (Studie 2 und Studie 3), erklärt werden. Die Ergebnisse der vorliegenden Studien ermutigen Forschende hoffentlich dazu, klar zwischen der Konzeptualisierung von Emotionen als Traits und States zu unterscheiden, da beide je nach Fragestellung von Bedeutung zu sein scheinen. So scheinen Trait-Emotionen beispielsweise stärker zukünftige Entscheidungen und zukünftiges Verhalten vorherzusagen (Wirtz et al., 2003), sind aber nicht in der Lage situationale Fluktuationen im emotionalen Erleben zu erfassen. Die Befunde der vorliegenden Dissertation bestärken überdies aktuelle Bemühungen die subjektive Kontrolle bei Schülern und Schülerinnen zu fördern – konzeptualisiert als Trait (z.B. akademisches Selbstkonzept) und State (z.B. subjektive situationale Kontrolle). Implikationen für zukünftige Forschung und Praxis werden diskutiert, speziell in Bezug auf die Wichtigkeit subjektiver Überzeugungen und Emotionen im Lern- und Leistungskontext.

It is a basic fact of the human condition that

memories are what we get to keep from our experience […].

(Kahneman & Riis, 2005, p. 286)

1 General Introduction

1.1 The Relevance of Emotions in the Academic Context

In the academic context, teachers and students are constantly engaged in an exchange of knowledge with a critical focus on cognitive abilities. Indeed, intelligence is able to predict approximately 25% ($r \approx .50$) of the variance in students' academic achievement and achievement in general (Strenze, 2007), making this connection one of the strongest and most robust in the psychological literature. However, there is still a lot of unexplained variance that needs to be accounted for, and as a result, psychosocial variables such as motivation, interest, self-concept, and emotions are gaining increasing attention in the school setting among educational researchers but also among practitioners and policy makers. These psychosocial variables play an important role in the school setting as they influence learning outcomes and achievement to a notable extent (Allen, Robbins, & Sawyer, 2010; Robbins et al., 2004). In addition to motivation, self-concept, and interest, a growing amount of research has emerged in the field of emotions in learning and achievement settings in recent years, which is reflected in the publication of one edited volume (Schutz & Pekrun, 2007) and several special issues published in the *Educational Psychologist* (Schutz & Lanehart, 2002), *Learning and Instruction* (Efklides & Volet, 2005), the *Educational Psychology Review* (Linnenbrink, 2006), *Contemporary Educational Psychology* (Linnenbrink-Garcia & Pekrun, 2011), *Learning and Individual Differences* (Lipnevich & Roberts, 2012), and the *Journal of Psychoeducational Assessment* (MacCann, Lipnevich, & Roberts, 2012). Emotions (e.g., enjoyment and anxiety) are important as they have a profound influence on learning strategies, motivation, school outcomes, and domain and career choices but are also valuable outcomes in and of themselves in the school context. Increasing positive affect towards learning and reducing negative emotions are important for helping to foster students' interest in learning and gaining new knowledge. Thus, exploring emotions in the learning and achievement context seems to be a worthwhile enterprise as emotions are important agents in the learning process but also crucial outcome variables in need of further examination.

1.2 Trait and State Emotions

Emotions in the academic context are generally defined as multidimensional constructs in which different components are distinguished (Pekrun et al., 2004): (1) the affective component describes the core feeling of an emotion, (2) the motivational component implies

that behavioral reactions are related to emotions, (3) the cognitive component describes the thoughts that are related to the emotion, and (4) the physiological component includes all bodily reactions that accompany an emotion. Sometimes a fifth component, namely the expressive component, is included (Nett, Goetz, & Hall, 2011; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). It describes the accompanying facial expressions and body postures and is the component from which individuals can draw conclusions about the emotional experiences of others.

In the present dissertation, trait and state academic emotions and their assessment are the topics of interest. Generally, definitions and conceptualizations within the research literature of trait and state constructs are largely uniform although the findings surrounding these constructs are highly variable and it is therefore difficult to draw a strict and concise picture of these two conceptualizations. The general introduction of the present dissertation will attempt to summarize some theoretical approaches and conceptualizations. In the following sections, definitions and conceptualizations of trait and state emotions are detailed and different ways to operationalize traits and states will be presented. Further, some information about the assessment of and basic assumptions about trait and state emotions in the present dissertation are provided before summarizing previous research with regard to the core objectives of the dissertation.

1.2.1 Defining Trait and State – Two Different Conceptualizations for Emotions

Schemas help people to structure and simplify a complex world. This holds true not only for understanding the non-animate world but also for the social world. Consistent with this idea, individuals attempt to establish rudimentary guidelines to characterize other people in terms of their personality in order to predict their behavior. This is where psychological traits come into play. Thus, it is no surprise that the first descriptions and definitions of these psychological traits, mainly in the form of personality traits, date back to ancient times and philosophers such as Aristotle (Matthews, Deary, & Whiteman, 2009) who attempted to describe characteristics of individuals such as modesty and vanity. In the psychological context, Allport and Odbert (1936) were one of the first to systematically investigate personality traits and did so using a lexical approach, meaning they identified words in the English language that describe personality traits. A longstanding tradition among trait psychologists consisted of attempting to describe the stable characteristics of individuals (see McCrae & Costa, 1995; Wiggins & Pincus, 1992 for reviews). However, researchers soon realized that traits were not as helpful in describing concrete behavior as once hoped because

they ignore situation-specific influences and the variability within persons. Thus, researchers also started to examine what can be thought of as the complementary component of traits, namely psychological states. The first distinction between trait and state dates back to Cattell and Scheier (1961) and resulted in the development of the widely known State-Trait-Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970), which explicitly focuses on the emotion of anxiety.

Emotional traits are defined as enduring individual tendencies to react in a specific way (e.g., experience a specific emotion) in identical (stability) or similar (consistency) situations (Amelang, Bartussek, Stemmler, & Hagemann, 2006). Another conceptualization defines emotional traits as repeatedly occurring emotional states when specific types of situations are encountered (i.e., habitual emotions). According to Titz (2001), the second conceptualization is preferable from a pragmatic standpoint because repeatedly occurring states are quantifiable whereas reaction tendencies are much harder to measure. States, in contrast, are defined as momentary occurrences and describe a transient emotional experience (e.g., Eid, Schneider, & Schwenkmezger, 1999). Thus, trait emotions comprise a longer time frame compared to state emotions and imply a certain level of stability. They are assumed to be stable characteristics of persons or describe individual tendencies to experience a certain emotional state. In contrast, state emotions are defined as momentarily occurring emotions that are more strongly influenced by situational variables (Eid et al., 1999). The different conceptualizations of trait and state often lead to conventional inquires into interindividual differences, which are more strongly related to traits, whereas intraindividual differences typically correspond with states (Hertzog & Nesselroade, 1987). Thus, the differentiation is grounded in stable differences between persons versus change due to situational differences.

1.2.2 Different Operationalizations of Trait and State Conceptualizations

From a research perspective, it is vital to find ways to operationalize the two different conceptualizations of emotions (see Table 1.1). When studying trait and state, researchers usually rely on the convenient and cost-effective method of assessing the two different conceptualizations of a construct, i.e., trait and state, via self-reports of study participants[1]. Whereas it seems relatively straightforward to assess states by asking participants about their experiences in a specific situation (e.g., “How much enjoyment are you experiencing right now?”), some debate has emerged about the best method for assessing traits. One possibility

[1] It should be noted that for the construct of emotion there are physiological and imaging techniques available that are usually not as specific as self-reports and focus strongly on assessment of arousal.

for measuring a trait is to directly ask participants about their general behavior or experiences (e.g., "How much enjoyment do you experience in general?"). However, as the validity of direct trait assessments is questionable, researchers started to investigate other ways of operationalizing traits.

Apart from directly asking people about their traits, other methods for operationalization are united by a commonly held belief that multiple state measures in a specific class of situations are the basis for determining a trait (see Table 1.1). One idea is that averaging several state assessments should represent a trait (Epstein, 1983; Zuckerman, 1976). This embodies the assumption that traits reflect habitual behavior or experiences and therefore are obtained by averaging over several situations of the same class. Another idea was proposed in latent state-trait models (Hagemann & Meyerhoff, 2008; Steyer, Schmitt, & Eid, 1999). The aim of this approach was to extract trait facets from variable states. One state measurement point is assumed to contain trait and state components (Steyer, Ferring, & Schmitt, 1992), and thus an underlying trait can be extracted from state measures. In this version of latent state-trait theory, a specific measurement of a variable is decomposed into (1) a trait, (2) a state or an interaction between situation and person, and (3) a measurement error. Fleeson (2001), in contrast, tried to operationalize traits as density distributions of states. He argued that as the intraindividual variability in states is high, it is not sufficient to only take the mean into account when operationalizing a trait. Additionally, the shape of the distribution, which he assumes to be unique for every individual, guarantees a comprehensive description of a person's trait. This assumption had already been expressed in Zuckerman et al.'s proposition to take mean and variance into account when describing traits (Zuckerman, Persky, & Link, 1967).

Table 1.1. *Overview of the conceptualization and operationalization of constructs as traits and states*

Construct	e.g., emotions	
Conceptualization	**Trait** • habitual • tendency to react	**State** • momentary experience • transient
Operationalization	Trait assessment (generalized self-reports)	State assessment (experience-sampling method)
	Measuring a trait by the use of multiple state assessments	

1.2.3 Trait and State Emotions in the Present Dissertation

Irrespective of the other possibilities for operationalizing traits that were detailed in the previous section, in the present dissertation a forthright, methodologically-centered approach was used such that trait emotions were assessed via generalized emotional self-reports. In contrast, state emotions were directly assessed in real-time and therefore comprise a much shorter time frame (see Robinson & Clore, 2002). The idea of differentiating traits and states by asking students about their emotions *'right now'* versus *'in general'* goes back to Zuckerman (1960) and is closely linked to the approach used with Spielberger's STAI (Spielberger et al., 1970). For the present dissertation, it was assumed that previous research intended to capture emotions while using trait and state emotional assessments (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011; Pekrun, Goetz, Titz, & Perry, 2002). To summarize and clarify once more, when referring to trait emotions or trait assessments in the following pages, what is meant are emotions that are assessed via generalized self-reports. When discussing state emotions or state assessments, these terms refer to emotions as assessed directly in a specific context (i.e., perceived anxiety in mathematics class at the present moment the questionnaire is administered).

1.2.4 Comparing Trait and State Assessments – Previous Research

When studying academic emotions, researchers usually rely on the assessment of emotions via self-reports of study participants. A large proportion of previous studies have investigated emotions through the use of general self-reports (i.e., to assess trait emotions) due largely to their long standing history in the literature as well the relative ease with which they can be administered. However, momentary assessments examining actual emotions in ecologically valid achievement and learning situations (i.e., state emotions) are becoming more and more popular (e.g., Ahmed, Minnaert, van der Werf, & Kuyper, 2010a; Ahmed, van der Werf, Minnaert, & Kuyper, 2010b; Goetz, Frenzel, Stoeger, & Hall, 2010; Nett et al., 2011). Although researchers who use state assessments of emotions are convinced of the propriety of this method (see Schwarz, 2012 for a rationale for using state assessments), I know of no study explicitly aimed at comparing trait-based and state-based emotional assessments in the academic context. However, this appears to be a promising and important enterprise regarding methodological as well as theoretical advances in the field. As both assessment methods are accepted procedures, knowing the extent to which they are similar or different should allow for the possibility to infer whether conclusions from trait emotions on actual state emotions or vice versa are justified.

Trait and state scales of anxiety tend to be moderately correlated (see Schwenkmezger, 1985 for an overview of correlations between trait and state anxiety measured with the STAI). Interestingly, even trait assessments and aggregated state assessments are only weakly or moderately correlated (e.g., Steptoe, Gibson, Hamer, & Wardle, 2007), which implies that the two do not capture the same construct. Furthermore, mean-level differences were found between trait and state assessments (Buehler & McFarland, 2001; Wilson & Gilbert, 2005). Different approaches in the literature thus far attempt to explain the discrepancy between trait-based and state-based assessments. Generally, it is assumed that generalized trait assessments of emotions are influenced by variables such as memory biases and global heuristics (see Scollon, Kim-Prieto, & Diener, 2003). Social desirability, cognitive biases and cultural norms are possible impact factors as well (Scollon et al., 2003). Conversely, state evaluations capturing momentarily occurring emotional experience seem to be less 'contaminated' by other constructs. These assumptions are embraced in the *accessibility model of emotional self-report* as proposed by Robinson and Clore (2002). In their model, the authors clearly distinguish between emotional self-reports that are assessed online and those that are not. The model identifies four different sources of information for emotional self-report: experiential information, episodic memory, situation-specific belief, and identity-related belief (see Figure 1.1). As long as self-reports are momentary or online, the direct experiential information is accessible to the individual, but as self-reports become increasingly distant from the actual events they are intended to cover, the more beliefs come into play while making globalized or trait ratings. Thus, according to this theory, the major difference between trait and state assessments of emotions is the type of knowledge that is used to answer the self-report questionnaires. State assessments only require the individual to recall experiences that just occurred and therefore episodic memory can be used. Trait assessments, however, ask individuals to recall or summarize memories that are not readily available, and thus, participants may not attempt to recall single episodes or events but instead utilize semantic knowledge to answer the questionnaire, which leads to a higher consistency of trait emotional self-reports with subjective beliefs.

Kahneman's conception of the 'remembering' and 'experiencing' selves (Kahneman, 2011; Kahneman & Riis, 2005) reflects a similar notion regarding the distinction between constructs conceptualized as traits and states. The two selves are used differently when constructs are examined by global self-report assessments versus online assessments (Conner & Barrett, 2012), with the experiencing self being used to answer state assessments and the remembering self being used to answer trait-based questions.

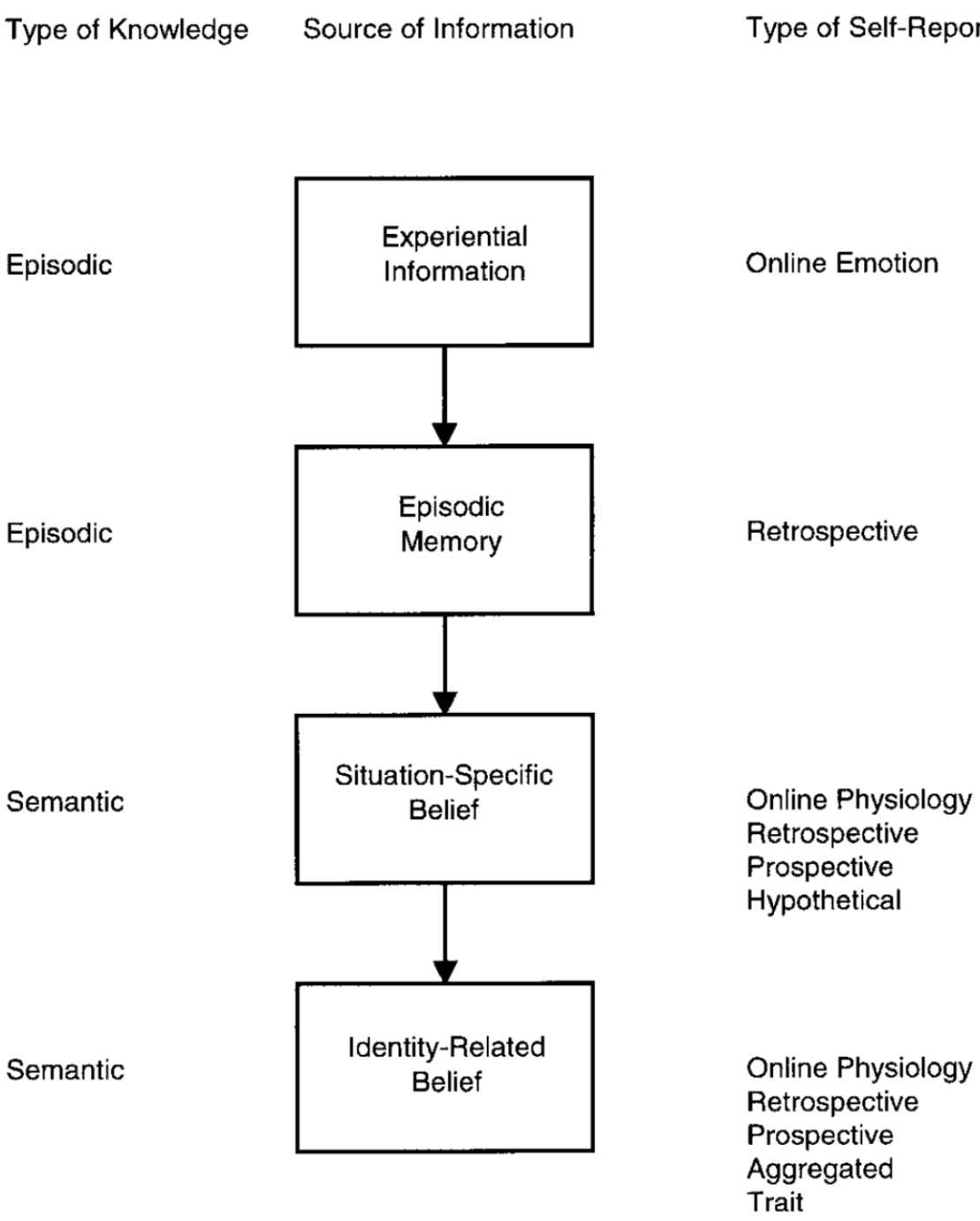


Figure 1.1. *Accessibility model of emotional self-report (in Robinson & Clore, 2002, p. 937)*

Other researchers present slightly different explanations for the discrepancy between trait and state assessments. Wilson and colleagues, for example, proposed an effect named *focalism* that mainly occurs in the forecast of emotions related to specific events: people are exclusively focused on a single event in the future (e.g., earning a doctoral degree) while neglecting how other events and personal circumstances will influence their mood during that time (Comerford, 2011; Gilbert & Wilson, 2007; Wilson, Wheatley, Meyers, Gilbert, &

Axsom, 2000). Other research that compared retrospective trait and state assessments found that peak effects (meaning the moments with highest emotional intensity) more strongly influence trait assessments. Further, experiences towards the end (i.e., end effects) of a certain time period but prior to the retrospective report also seem to more strongly influence the respective trait assessments (Kahneman, 2011).

Several researchers claim that trait and state assessments capture different constructs. Indeed, the predictive validity of trait and state assessments differs. Empirical studies consistently found that trait assessments are more predictive of future behavior and choices (Hsee & Hastie, 2006; Levine, Lench, & Safer, 2009; Wirtz et al., 2003) than the actual state assessments. Furthermore, it was found that state assessments and trait assessments account for different aspects of the variance in a set of dependent variables, which is once again an indicator that the two assessment methods capture different constructs (Augustine & Larsen, 2012).

Irrespective of which explanation is used, researchers generally find a gap between memory and experience (Miron-Shatz, Stone, & Kahneman, 2009). With regard to the construct of emotions, many different studies in various contexts found trait emotional assessments to be rated higher than state emotional assessments (e.g., Ben-Zeev, McHugo, Xie, Dobbins, & Young, 2012; Buehler & McFarland, 2001). In summary, previous research has theoretically assumed and empirically determined that there are differences between trait and state emotional assessments. However, the field of educational psychology is lacking research that illuminates the relations between trait and state emotional assessments.

1.3 The Present Dissertation – Objectives and Outline

The present dissertation links to previous research that attempted to illuminate the relations between trait and state emotions and reveal causes for the discrepancy between trait and state emotional assessments. It takes structural as well as mean-level differences into account. First, the different relations between emotions and their antecedents are investigated with emotions being conceptualized as trait and state emotions. Second, differences regarding trait and mean state intensities are examined. Third, this expected trait-state discrepancy is further explained by moderating variables. Although there is a paucity of research investigating how trait and state emotions are related, it is a promising enterprise for gaining further insight into the relations between the two different conceptualizations and assessment methods of emotions as both are used regularly (i.e., trait assessments) or gaining more attention (i.e., state assessments) among educational psychology researchers.

The aim of the present dissertation was to investigate differences between trait and state emotions pertaining to antecedent-emotion relations and mean-level differences with regard to trait and state emotional assessments. Further, moderators of the expected discrepancy between trait and state assessments were to be identified, meaning variables were investigated that contributed to an intensification of the discrepancy between trait and state emotional assessments. In order to answer the research questions, three different studies were conducted.

1.3.1 Research Questions for the Present Dissertation

1.3.1.1 Structural similarity among antecedents of trait- and state-based emotions

Most of the previous research on trait and state emotional assessments has focused on mean-level differences. However, it is important to also investigate structural similarities and differences between trait and state emotions, namely the antecedent-emotion relationship. Antecedents of emotions have gained heightened attention in recent years likely because it is possible to directly influence them via environmental variables. Appraisals are one type of emotional antecedent and refer to a person's subjective judgments of the situation. The prominent control-value theory of achievement emotions (Pekrun, 2006) proposes that control and value are especially important appraisal antecedents of emotions in learning and achievement situations. Further, the interaction of control and value is assumed to be able to predict emotions over and above their single main effects, meaning that the two variables interact and the relation between one appraisal antecedent and the emotion is dependent on

the level of the other variable. However, the interaction was mostly neglected in previous research.

There are a number of published studies that have investigated the validity of the control-value theory when applied to trait emotions (Pekrun, 2000, 2006; Pekrun, Frenzel, Goetz, & Perry, 2007a; Pekrun & Stephens, 2010). For state emotions, there is supporting evidence for the appropriateness of the control-value theory, predominantly in the subject of mathematics and for few selected emotions (Ahmed et al., 2010b; Frenzel, Pekrun, & Goetz, 2007; Goetz et al., 2010; Pekrun et al., 2010), but further testing is needed, especially in other subject domains and with a wider range of emotions. Therefore, it seems reasonable to explicitly investigate structural differences between appraisal antecedent-emotion relations of trait versus state emotions. Further, by using an intraindividual approach in the analysis of multiple trait and state emotional assessments, which is recommended when testing the assumptions of the theory (Pekrun, 2006), it will be possible to gain insight into the intraindividual functioning of appraisal-emotion relations.

Research questions:

(1) Are the assumptions of the control-value theory supported when analyzing trait and state data intraindividually?
(2) Is the interaction of control and value able to predict emotions over and above the single main effects?
(3) Are there structural differences in the antecedent-emotion relations (control, value, and Control × Value) between trait and state emotional assessments?

1.3.1.2 Mean-level differences between trait and state emotions

Previous research has indicated that differences exist between trait and state emotional assessments. One common finding concerns the difference between the intensity rating of trait and state emotions in which trait emotions were found to be rated more intensely than the corresponding aggregated state emotions (sometimes referred to as intensity bias; Barrett, 1997; Wirtz et al., 2003). However, educational psychologists have yet to investigate whether these finding also hold for emotions assessed in the academic context with students.

1.3.1.3 Possible moderators of the trait-state discrepancy

According to Robinson and Clore's accessibility model of emotional self-report (2002), the assessment of trait emotions is influenced by subjective beliefs (semantic knowledge as opposed to episodic knowledge). It is important to note that the term 'belief'

does not necessarily imply inconsistency with objective reality. Possible beliefs that are associated with emotional self-reports are manifold and not specified in the article by Robinson and Clore (2002), however, some examples include beliefs about the self and social and gender stereotypes. While analyzing differences between trait and state academic emotions, subjective beliefs and other possible moderating variables that are important in the school context should be identified. This was also a goal of the present dissertation especially since there has yet to be research explicitly investigating such variables.

The second of three studies in the present dissertation sought to examine whether there exists a discrepancy between trait and state emotional assessments across a broad range of academic emotions. Further, a moderator of the trait-state discrepancy, namely academic self-concept, was investigated as an important subjective belief which is seen as a crucial antecedent of emotions according to the control-value theory (Pekrun, 2006).

Research questions:

(4) Is there a discrepancy between trait and state emotional assessments in the academic context?
(5) Given there is a discrepancy between trait and state emotional assessments, is it possible to explain the discrepancy via students' academic self-concept?

1.3.1.4 The gender gap in mathematics anxiety

Anxiety is one of the most important emotions in the learning and achievement context as its influence on academic achievement can be detrimental (Zeidner, 2007). Mathematics is one of the most prominent domains in which it is assumed that girls have higher anxiety ratings than boys (Frenzel et al., 2007), commonly referred to as a gender gap. However, also stemming from the accessibility model of emotional self-report (Robinson & Clore, 2002), the question arises once again as to whether previous findings from trait assessments (as in Frenzel et al., 2007) also hold for state assessments and whether or not subjective beliefs are responsible for the emergence of gender differences in trait versus state anxiety.

Hence, in the third study, the focus is on the important and well-researched emotion of anxiety in mathematics and the study seeks to determine whether the gender gap in math anxiety, meaning that girls give higher anxiety ratings than boys, which is regularly found in trait assessments, also holds for state assessments. Further, based on the accessibility model,

the study aimed at investigating whether girls' assumed lower self-concept moderates the trait-state relation for this emotion.

Research questions:

(6) Does gender moderate the magnitude of the trait-state discrepancy?
(7) Is it possible to find other variables (i.e., control or competence beliefs) that contribute to the trait-state discrepancy, and thus the gender gap in math anxiety?

To summarize, whereas the first set of research questions are more concerned with structural similarities between trait and state emotions in terms of their antecedents as proposed by the control-value theory, the subsequent questions refer to mean-level differences and possible moderators of the expected discrepancy between trait and state emotions (see Figure 1.2 for a graphical overview).

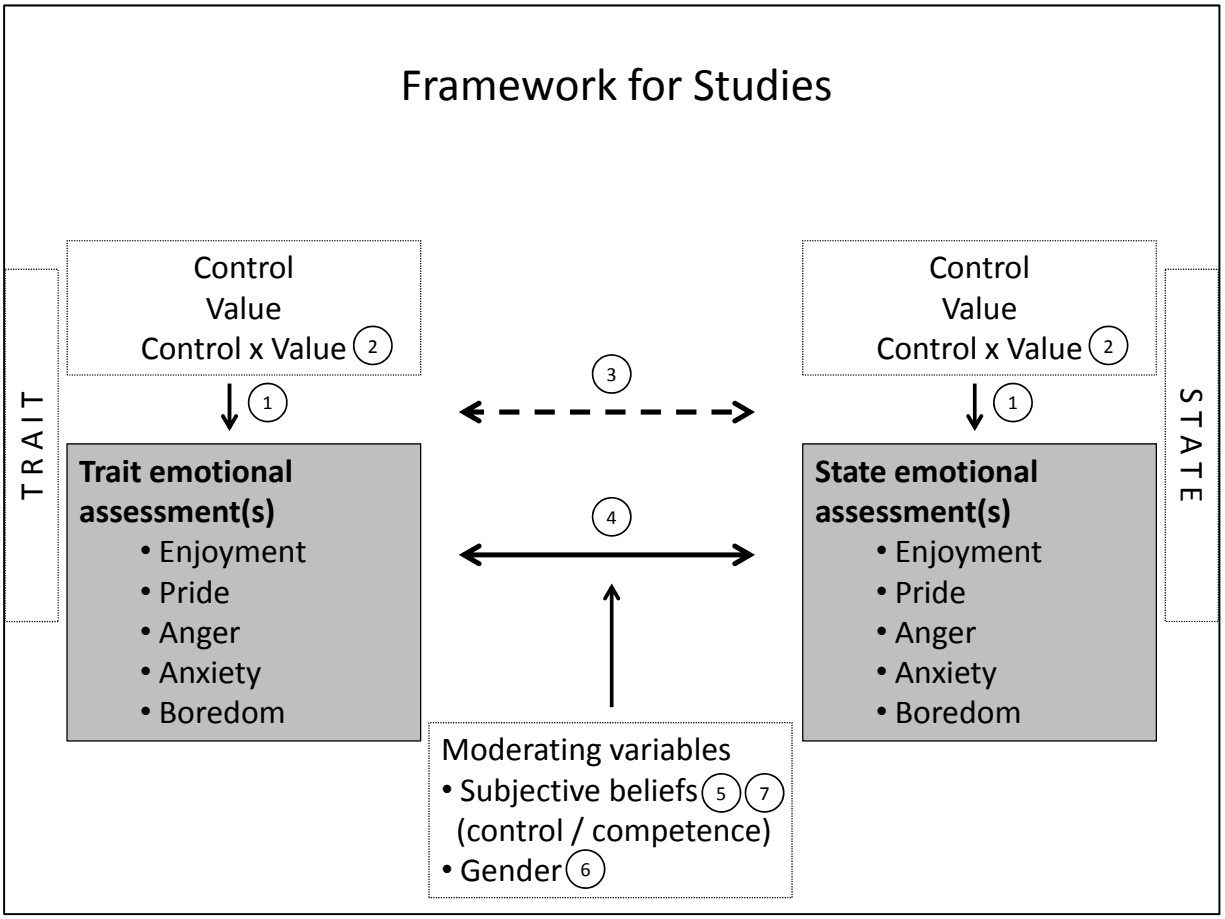


Figure 1.2. *Overview of relations between variables and respective research questions (1)-(7) in the different study projects*

1.3.2 Summarizing the Significance of the Present Dissertation

As research on emotions continues to gain prominence in the academic context, investigating conceptualizations of emotions and their operationalization becomes exceedingly important. The vast majority of previous research utilized trait assessments to investigate students' academic emotions. Unfortunately, these reports may not have captured the emotions students actually felt but rather students' thoughts about their emotions (Robinson & Clore, 2002). Thus, gaining insight into trait and state emotions and their relations may prove valuable for the planning of future research and interpreting previous findings.

Emotion is one variable that is assumed to guide students' future domain and career choices (cf. Eccles, 1985). Differentiating between trait and state academic emotions and understanding the interrelations of trait and state assessment is therefore crucial when considering the practical implications of the present dissertation. According to previous research, trait emotions are more predictive of future choices than their actual state experiences (Wirtz et al., 2003) and this could also be the case for academic emotions. Thus, knowing about the discrepancy between trait and state emotions could enable us to inform students about their biased ratings (at least in cases of unfavorable ratings of trait emotions) and therefore could be a first step in changing their perceptions and helping them base their future choices on 'real' experiences instead of stereotypic beliefs. Further, identifying variables that contribute to the discrepancy may provide us with initial evidence that can be used in determining where to implement intervention programs (e.g., to attract more female students into natural science domains).

1.3.3 Dissertation Outline – Three Studies

The present dissertation combines three empirical studies that were aimed at comparing structural (Study 1) and mean-level differences (Study 2 and Study 3) in trait and state emotions. Further, control beliefs (Study 2 and 3) and gender (Study 3) were investigated as moderators of the discrepancy between trait and state emotional assessments. The three studies are presented in the following chapters (Chapter 2 to 4) and can be read and understood independently of each other.

In Study 1 (Chapter 2), trait and state assessments of students' pride, anxiety, and boredom and their appraisal antecedents of control and value were assessed. Multiple trait and state assessments for each participant were analyzed intraindividually with control, value, and

the control-value interaction as predictors of the academic emotions. The intraindividual approach made it possible to analyze functioning within individuals. Results showed that control, value, and the interaction between the two were able to predict students' emotions in the expected directions. Further, results from trait and state assessments suggest structural similarity between the two assessment methods.

In Study 2 (Chapter 3), the discrepancy between students' trait and state emotions of enjoyment, pride, anger, and anxiety in mathematics were examined in German and Swiss samples. Further, self-concept was investigated as a moderator of the discrepancy between trait and state academic emotions. Results showed that in general, trait ratings were higher than state emotional ratings. Further, self-concept beliefs were found to moderate the discrepancy between trait and state emotions such that positive emotions were overestimated in students with higher self-concept in mathematics and negative emotions were overestimated in students with lower self-concepts compared to their actual state emotions.

In Study 3 (Chapter 4), which is closely related to Study 2, an established finding in the educational psychology context was questioned. Data from two studies were analyzed with regard to the gender gap (girls report higher levels of math anxiety than boys) in math test anxiety and math classroom anxiety. It was assumed that the gender gap likely only exists in trait but not state anxiety. In fact, it was found that girls rated their trait anxiety higher than boys, but there was no gender difference when state test anxiety and state classroom anxiety were assessed. Further, the discrepancy between trait and state anxiety ratings could be explained by girls' lower control beliefs.

Chapter 5 summarizes the findings of the three empirical studies. Strengths and weaknesses as well as implications for research and practice are discussed.

2 Can I Master It and Does It Matter? An Intraindividual Analysis on Control-Value Antecedents of Trait and State Academic Emotions

2.1 Summary

The present study explored the relations between cognitive appraisal antecedents and academic emotions as stated in Pekrun's control-value theory (2006). The appraisals of control and value, and the interaction of the two as predictors of emotions, were studied while using both trait and state (via experience-sampling) assessments in one sample. Control and value appraisals, and the discrete emotions of pride, anxiety, and boredom, were assessed in four subject domains in a sample of $N = 120$ students in grades 8 and 11. Multilevel analyses showed that control, value, and their interaction predict the respective emotions in the expected direction while using an intraindividual approach in analyzing the data. Furthermore, results revealed that appraisal-emotion relationships are quite similar in trait and state data. Implications for future research are outlined regarding the use of intraindividual approaches and for educational practice with respect to the promotion of control and value appraisals.

2.2 Introduction

Emotions in achievement contexts, referred to as academic emotions, have long been neglected despite a growing body of research that clearly documents their importance with regard to learning (for example, self-regulated learning: Op't Eynde, De Corte, & Verschaffel, 2007), academic achievement (e.g., Pekrun et al., 2002), lifelong learning (Goetz, Zirngibl, Pekrun, & Hall, 2003), and career choices (Wigfield, Battle, Keller, & Eccles, 2002). Irrespective of these findings, research on academic emotions did not begin to receive much empirical attention until the early 1990s, with the exception of test anxiety (Zeidner, 2007) and Weiner's attributional theory of academic emotions (Weiner, 1985). One important aspect of research on academic emotions is the investigation of their possible antecedents. In addition to the scientific importance of conducting this research, it is especially relevant from a practical perspective as knowledge concerning the antecedents of students' emotional experiences is required to inform the development of effective intervention programs and instructional techniques.

In the research literature there are various theoretical perspectives on emotions, each with specific ideas about how emotions emerge (Gross & Barrett, 2011). The appraisal perspective is a fundamental approach that explains the variability in peoples' emotional reactions in identical situations due to different evaluations of the situation. In the context of learning and achievement, subjective control and value are assumed to be particularly important appraisal antecedents as stated in the control-value theory of achievement emotions (Pekrun, 2006). According to this theory, it is assumed that a person's subjective evaluations of control and value influence their subsequent emotions. In order to understand how appraisals influence peoples' emotions, it is important to study appraisal-emotion relationships from an intraindividual perspective, meaning how the different appraisals within a person are related to the emotions experienced by this person.

In research on academic emotions, students are typically asked to give trait self-reports of their emotions (global or 'in general' ratings), which can be problematic as there is empirical evidence that trait assessments, unlike state assessments, are prone to retrospective biases (Robinson & Clore, 2002). Thus, it is recommended that the results of trait assessments of emotions be interpreted with caution as it remains unclear the extent to which they reflect actual emotions or rather beliefs about emotions.

The goal of the present study is twofold. First, the assumptions of the control-value theory regarding appraisal-emotion relationships will be tested using an intraindividual approach (multiple measurement points per person both in trait and state assessments). Second, we compare the two assessment methods of trait and state with regard to potential structural differences in the relations between appraisals and emotions in one sample.

2.3 Theoretical Background

2.3.1 Control and Value Appraisal Antecedents of Emotions

In the academic context, Pekrun's control-value theory of achievement emotions represents a prominent appraisal theory that describes control and value as especially important appraisal antecedents of emotions (Pekrun, 2000, 2006). According to Pekrun's control-value theory (Pekrun, 2000, 2006), control refers to the appraisal of the possibility to personally influence activities and outcomes and may include perceptions such as competence beliefs and causal attributions. Value refers to one's appraisal of the significance or importance of an outcome.

In accordance with the theory's assumptions, empirical findings consistently demonstrate that control is positively related to positive emotions such as enjoyment or pride and negatively related to negative emotions such as anger and anxiety (e.g., Frenzel et al., 2007; Goetz, Pekrun, Hall, & Haag, 2006). For value appraisals, the relation is different. According to the theory, high value appraisals intensify positive as well as negative emotions. This means if the outcome of a task or activity is evaluated as particularly important, stronger positive and negative emotions should be experienced compared to when value is low with the sole exception of boredom. Lower levels of boredom are expected to be experienced when a task or outcome is perceived as high in value (Pekrun et al., 2010). Studies consistently find the expected positive association between value and positive emotions, however, the correlation between value and negative emotions has been found to be both positive (e.g., Pekrun, 2000) as well as negative (negative correlations but positive associations when using structural equation modeling: Goetz et al., 2006).

Beyond the independent effects of control and value on emotions, the control-value theory explicitly proposes that control and value should interact to produce a combined effect when predicting achievement emotions (Pekrun, 2006). Depending on the subjective value of the activity or outcome, the magnitude of the effect of perceived control on emotions is expected to differ. Alternatively, the effect of perceived value on emotions would be expected

to differ as a function of the level of perceived control. For example, compared to students who have low control and low value appraisals, students with low control appraisals but high value appraisals for an outcome (e.g., an important final exam) will likely experience more anxiety. However, it seems as if interaction effects have been largely neglected in previous research on appraisal-emotion relations despite their importance (see Nagengast et al., 2011; Trautwein et al., 2012 with respect to motivational constructs). Only one recently published experience-sampling study by (Goetz et al., 2010) explored the influence of an interactive effect in predicting positive state emotions. Findings from this study indicated that the relation between control appraisals and enjoyment, pride, and contentment, was stronger in situations where high value appraisals were reported.

2.3.2 Trait and State – Different Ways of Assessing Academic Emotions

In the present study, a methodologically-centered definition will be used such that trait emotions are considered to be global emotion reports that entail judgments over lengthy periods of time, whereas state emotion assessments are direct or 'on-line' assessments of the current situation (see Robinson & Clore, 2002). Trait emotions are derived from memory and potentially impacted by subjective beliefs, whereas for state emotions, memory biases are assumed to play a less significant role (see Kahneman, 2011; Robinson & Clore, 2002).

Empirically, the distinction between trait and state assessments is reflected in differences between the means of trait and state emotions in which traits are consistently rated higher than states; a finding often referred to as 'intensity bias' (see Buehler & McFarland, 2001; Robinson & Clore, 2002). However, beyond mean-level analyses, it is important to also investigate the structural similarities and differences of trait and state emotions in order to clarify how these two assessment methods differ.

2.3.3 Using an Intraindividual Approach to Study Appraisal-Emotion Relationships

It is vital to use an intraindividual approach when studying how the appraisals of control and value are connected to emotions. An intraindividual approach involves investigating the variation of variables *within* persons. This approach is explicitly encouraged in Pekrun's control-value theory (2006), however, the majority of previous studies employing trait emotion assessments have done so using an interindividual approach such that the variation of variables *between* individuals was analyzed. This is likely the result of only assessing appraisals and emotions once per person. Unfortunately, evaluating interindividual differences can become problematic when attempting to draw conclusions about

intraindividual functioning. This is referred to as an ecological fallacy and involves interpreting data on a lower or intraindividual level that are in fact aggregated on a higher level (Hox, 2010; Krapp, 2002; Valsiner, 1986). For example, it was found that at the group level anxiety and motivation to learn were uncorrelated, however, when analyzed at the intraindividual level, motivation to learn and anxiety were positively related for some students and negatively related for others (Pekrun et al., 2002; for a classic example see Robinson, 1950). As this example highlights, analyses conducted at the interindividual or population level do not necessarily provide accurate information regarding intraindividual functioning. To draw a valid conclusion from the population level to the individual level stringent conditions must be met, however, in psychological research this rarely occurs (for a discussion see Molenaar & Campbell, 2009).

At present, there are a few studies on the control-value theory that utilize an intraindividual approach. These studies have focused solely on measuring state emotions and are limited regarding the range of subject domains addressed (mathematics: Ahmed et al., 2010b) and emotions examined (positive emotions: Goetz et al., 2010; boredom: Pekrun et al., 2010). We believe that it is imperative to adopt an intraindividual approach with state and trait data, which requires multiple trait as well as state assessments per person.

2.3.4 Aim of the Present Study

The aim of the present study was to investigate the validity of the control-value theory for trait and state emotions within a single sample while using an intraindividual approach (multiple trait and state measures within persons). In the present study, trait questionnaire measures of appraisals and emotions were assessed four times from each student in four different subject domains in order to capture a broad sample of emotional experiences and related appraisals in the school context. Additionally, state measures from the same students were assessed in the same four domains during school lessons using an experience-sampling method. Through the use of an intraindividual approach, we examined the influences of control and value appraisals as well as their combined interactive effect. The selection of emotions was based on the two dimensions of valence and activation as highlighted in Watson and Tellegen's (1985) circumplex model. We focused specifically on pride and anxiety as typical positive and negative activating academic emotions. Furthermore, boredom was chosen as it is a frequently experienced negative deactivating emotion in academic contexts (Larson & Richards, 1991). We did not measure positive deactivating emotions (such as relief or relaxation) in our study as they are typically experienced after an event (rather than during

the event) and therefore are not especially suitable for state assessments. In summary, beyond testing the assumptions of the control-value theory through the use of an intraindividual approach, we also aimed to investigate the structural similarities of the appraisal-emotion relationship in trait and state emotions within one sample in order to more clearly evaluate the validity of former findings on trait data. More importantly, we sought to determine if the two different assessment methods lead to similar conclusions.

2.4 Research Questions and Hypotheses

2.4.1 Hypothesis 1

It is assumed that control positively predicts pride and negatively predicts anxiety and boredom (1a). Value should be positively associated with pride and anxiety (1b) yet negatively correlated with boredom (1c). We assume that the relations between appraisals and emotions are the same for trait and state emotions. Although there may be slight differences concerning the strength of the effects, there is no plausible reason to expect structural differences concerning the direction of the effects of control and value on trait and state emotions both analyzed intraindividually.

2.4.2 Hypothesis 2

As proposed by the control-value theory (Pekrun, 2006), we expect to find that control and value interact in predicting trait and state emotions over and above their independent effects, meaning that in addition to the additive effects we also expect to find a multiplicative effect.

2.5 Method

2.5.1 Sample and Data Collection

The sample consisted of N = 120 students (grade 8: 48.3%, M_{age} = 14.32 years; grade 11: 51.7%, M_{age} = 17.55) from 44 different classes (two to three randomly chosen students per class) of the top track of the state school system in Germany (i.e., Gymnasium), which incorporates approximately one third of students in secondary schools across the country. The gender of the participants was balanced with 60 males (grade 8: 28; grade 11: 32) and females. All participants first provided trait data via paper and pencil questionnaires and then state data via a computer-based experience-sampling method (see Hektner, Schmidt, & Csikszentmihalyi, 2007).

2.5.2 Assessment of Trait Data

Trait data was obtained using a paper and pencil questionnaire administered to students by trained experimenters. Appraisals of control and value were assessed for each of the four subject domains of mathematics, physics, German, and English (first foreign language). In order to be able to directly compare trait and state data, the same single items for trait and state assessments were used with one item each for control (i.e., academic self-concept) and value (i.e., perceived importance). The response format consisted of a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

The item for control was adapted from the Self Description Questionnaire (SDQ; Marsh, 1990; "I have always done well in [DOMAIN]."). The item for value (i.e., achievement value) was adapted from Frenzel et al. (2007; "It is very important for me to get a good grade in [DOMAIN].").[2] Finally, single-item measures on a 5-point Likert scale, each adapted to the four subject domains, were used to assess the trait emotions of pride, anxiety, and boredom (e.g., "How much pride do you generally experience during [DOMAIN] classes?"; see Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013).

2.5.3 Assessment of State Data

After the assessment of trait data, the experience-sampling period began. In a design that combined event-based and random sampling, the students were asked to activate a personal digital assistant (PDA) at the beginning of mathematics, physics, German, and English classes for a period of two weeks. Once activated, the PDA randomly signaled within the next 40 minutes and asked students to answer an electronic questionnaire about their current emotions and control and value appraisals in that specific class. For practical reasons, and to avoid confounding the state assessment by distracting participants with lengthy self-report questionnaires (see Goetz et al., 2010), we used a single-item measure with a 5-point Likert scale each for control, value, and the three emotions (parallel wording with trait assessments were adjusted for the class; e.g., "How much pride are you experiencing during this class?"; see Goetz et al., 2013). In total, this procedure resulted in N = 1510 state measures with a mean of 12.58 state assessments per participant.

[2] With regard to comparability, we used the same single items for control and value in state-based and trait-based data for our analyses, although in the trait questionnaire whole scales were assessed. The scale for control consisted of three items and the scale for value consisted of four items. Each parallel formulated single-item was highly correlated with the respective scale (r = .88 to r = .91 for control and r = .91 to r = .92 for value), indicating high validity of the single items. All alphas ranged between .87 and .91 for control and .88 and .92 for value.

2.5.4 Statistical Analyses

The main focus of our analyses was on intraindividual functioning in trait and state data. For trait data, assessments of each of the four domains were nested within persons (N = 4 x 120 = 480). Multiple state measures (N = ~ 12.58 x 120 = 1510) were also nested within persons. The resulting data reflected a two-level structure with measurement points for trait questionnaire data and state experience-sampling data (N = 1990) nested within participants (N = 120).[3] A graphical depiction of the data structure can be found in Figure 2.1. To account for the nested data structure, analyses were conducted via multilevel modeling using HLM 6.08 (Hierarchical Linear Modeling; Raudenbush, Bryk, & Congdon, 2009).

[3] It should be noted that our data represents a three-level structure with measurement points nested within students who are nested within classes. However, as the design effect was below two, we did not account for the third level in our analyses (Muthén & Satorra, 1995).

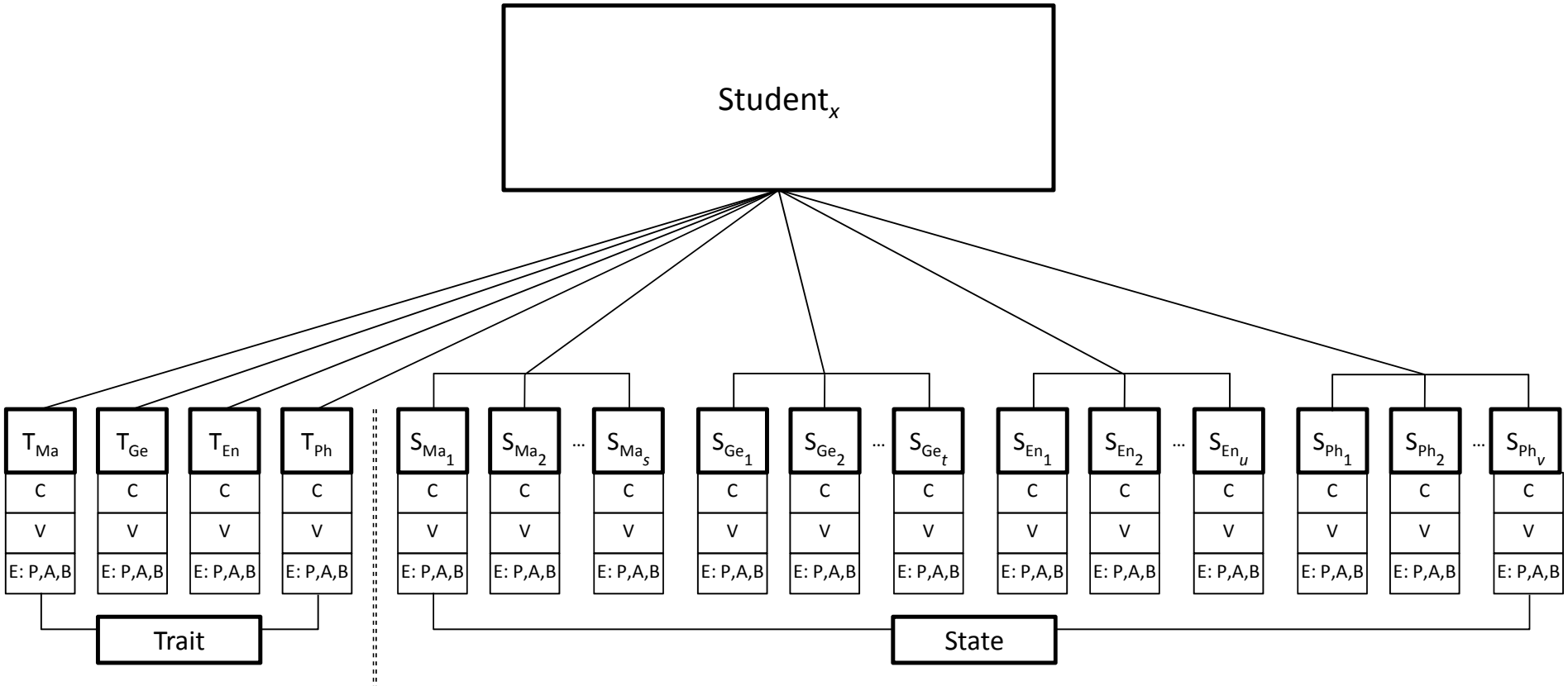


Figure 2.1. *Graphical depiction of assessed data and data structuring*

Note. Four trait assessments and *s*, *t*, *u* or *v* state assessments per person in the subject domains of mathematics (Ma), German (Ge), English (En), and physics (Ph) were assessed with a maximum of 30 state assessments per person. The same data structure was used for each emotion within each student *x*. C = Control; V = Value; E = Emotion (pride (P), anxiety (A), and boredom (B))

In our hierarchical linear regression model, control, value, and the Control × Value interaction were introduced to predict the academic emotions. Furthermore, we used a dummy variable for differentiating between trait vs. state assessments (1 = *trait*, 0 = *state*). All variables (with the exception of the dummy variable) were *z*-standardized across the whole sample prior to performing the multilevel analyses with the product terms used to test for interaction effects. The interaction terms were not restandardized (Aiken & West, 1991). As our analyses focused on the intraindividual level, meaning how control and value are associated with emotions within persons, all variables (including the dummy variable) and interactions were introduced into the model group-mean centered, that is, for each student on their mean (Enders & Tofighi, 2007).

To investigate possible differences between trait and state assessments concerning the effects of appraisal antecedents on emotions, we also introduced interaction terms combining control (C), value (V), and Control × Value (C × V) with the trait dummy (T) variable (C × T; V × T; C × V × T) resulting in the following multilevel equations[4]:

Level 1:

$$EMOTION_{ij} = \beta_{0j} + \beta_{1j}(CONTROL\text{-}\bar{X}_{.j}) + \beta_{2j}(VALUE\text{-}\bar{X}_{.j}) + \beta_{3j}(C\times V\text{-}\bar{X}_{.j}) + \beta_{4j}(TRAIT\text{-}\bar{X}_{.j}) + \beta_{5j}(TRAIT\times CONTROL\text{-}\bar{X}_{.j}) + \beta_{6j}(TRAIT\times VALUE\text{-}\bar{X}_{.j}) + \beta_{7j}(TRAIT\times C\times V\text{-}\bar{X}_{.j}) + r_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + u_{0j;}$$

$$\beta_{fj} = \gamma_{f0} \text{ with } f = 1, 2, \ldots, 7$$

2.6 Results

Table 2.1 shows the descriptive statistics and within-subject correlations of variables for trait and state data. In both trait and state data, boredom was the most intensely rated emotion and anxiety the least intense. Within-subject correlations were low to medium in size with correlations between appraisals and emotions going in the expected direction.

[4] Because we were only interested in within-group effects, we used the analysis of covariance model with fixed effects (Snijders & Bosker, 2012).

Table 2.1. *Descriptive Statistics and Correlations*

	Trait		State		Correlations				
	M	*SD*	*M*	*SD*	Pride	Anxiety	Boredom	Control	Value
Pride	2.45	1.23	1.66	1.07	--	-.14*	-.27***	.33***	.32***
Anxiety	1.74	1.16	1.42	0.95	.01	--	.14**	-.29***	-.06
Boredom	3.40	1.28	3.05	1.45	-.14***	.08*	--	-.18***	-.30***
Control	3.08	1.25	2.87	1.33	.32***	-.09**	-.08	--	.42***
Value	3.61	1.20	2.74	1.52	.23***	.10*	-.11**	.29***	--

Note. Descriptive statistics and within-subject correlations of variables are displayed. The hierarchical structure of data was accounted for. Trait correlations are shown above the diagonal (Level 1: $N = 480$; Level 2: $N = 120$) and state correlations are shown below the diagonal (Level 1: $N = 1510$; Level 2: $N = 120$).
* $p < .05$. ** $p < .01$. *** $p < .001$.

The results of the hierarchical linear regressions of the three models are shown in Table 2.2. In the upper part of Table 2.2 the results of the regression analyses for the predictors in state data are shown. Furthermore, the differences in the effects between trait and state data are modeled by introducing the dummy variable for trait vs. state (trait dummy; 1 = *trait*, 0 = *state*) and the respective interaction terms (C x T, V x T, C x V x T). These results are shown in the middle part of Table 2.2. Standard errors of the coefficients and residual variances of level 1 and level 2 for every model are indicated in the lowest part of Table 2.2.

Table 2.2. *Hierarchical Linear Models*

	Pride		Anxiety		Boredom	
	b	*SE*	*b*	*SE*	*b*	*SE*
Intercept	0.03	0.04	-0.01	0.04	-0.00	0.04
Control (C)	0.24***	0.02	-0.13***	0.03	-0.07*	0.03
Value (V)	0.15***	0.03	0.11***	0.03	-0.11***	0.03
C x V	0.05*	0.02	-0.05*	0.02	-0.13***	0.02
Trait Dummy (T)	0.50***	0.05	0.32***	0.06	0.38***	0.06
C x T	-0.04	0.06	-0.18**	0.06	0.01	0.06
V x T	0.13*	0.06	0.03	0.07	-0.20**	0.07
C x V x T	-0.01	0.05	0.00	0.06	0.15*	0.06
Residual variance						
Level 2	0.16		0.09		0.12	
Level 1	0.70		0.86		0.84	

Note. All outcome and predictor variables, except for dummy variables, were first z-standardized ($M = 0$, $SD = 1$) across the entire sample. In constructing the product terms, the product of individual (*z*-scored) standardized variables was used. The product terms were not restandardized. All predictors were group-mean centered, that is, for each individual on their mean.
* $p < .05$. ** $p < .01$. *** $p < .001$.

2.6.1 Hypotheses 1 and 2 - Control, Value, and Control × Value Effects

In line with our hypotheses, the conditional effects[5] of control (b = .24, p < .001) and value (b = .15, p < .001) and the Control × Value interaction effect (b = .05, p < .05) were positive and significant for state pride. The significant interaction can be interpreted such that the relation between control and pride was stronger in situations with higher value appraisals. In the first panel of Figure 2.2 there is a graphical depiction of the interaction.

As expected, control negatively predicted state anxiety (b = -.13, p < .001) whereas it was positively predicted by value (b = .11, p < .001). The Control × Value interaction also significantly predicted anxiety (b = -.05, p < .05). As shown in the second panel of Figure 2.2, the interaction can be interpreted such that there was a stronger negative association between control and anxiety in cases of high value.

With respect to state boredom, the conditional effects of control (b = -.07, p < .05) and value (b = -.11, p < .001) were significant as was the Control × Value interaction effect (b = -.13, p < .001). The appraisal antecedents predicted boredom in the assumed negative direction. As depicted in the third panel of Figure 2.2, the significant interaction effect showed the relation between control and boredom to be different depending on the value appraisal.

[5] It should be noted that "conditional effect" is the precise notation for the effects of control and value when an interaction is present in the equation (Aiken & West, 1991). In the present study, additional analyses without interaction terms showed that the independent effects of control and value (i.e., main effects) were also significant.

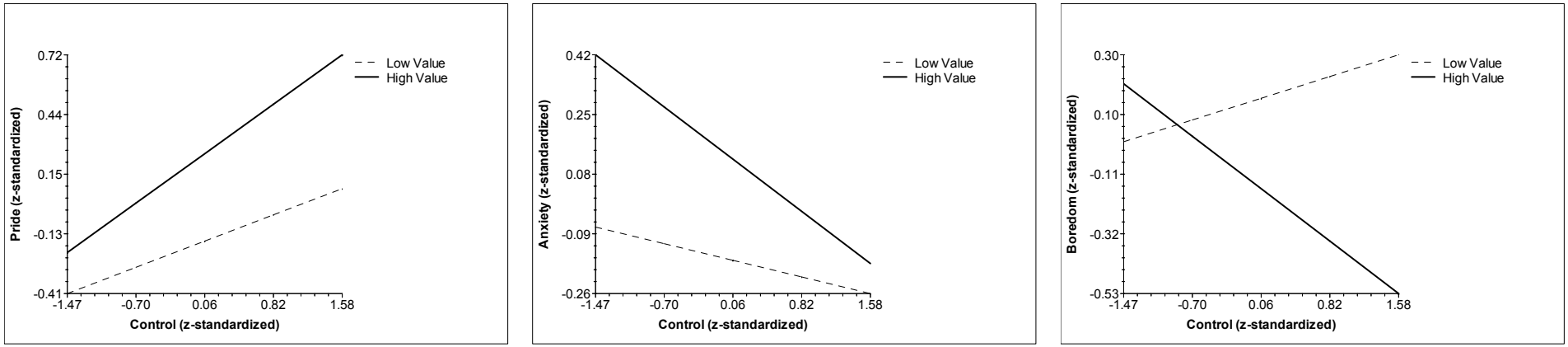


Figure 2.2. *Graphical display of significant Control × Value interactions in predicting pride, anxiety, and boredom*

Note. The graphs were created using the graph modeling function of HLM 6.08 (Hierarchical Linear Modeling; Raudenbush et al., 2009). On the x- and y-axes, *z*-values are shown. On the x-axis, the 5th to 95th percentiles are depicted in order to exclude outliers. The achievement value variable was dichotomized based on the averaged lower (Low Value) and upper (High Value) quartiles. Post-hoc simple slope analyses (Sibley, 2008) were conducted for each low and high value group (-1 *SD* / +1 *SD*; pride: $p < .001$ / $p < .001$; anxiety: $p = .017$ /$p < .001$; boredom: $p = .070$ / $p < .001$).

2.6.2 Structural Differences between Trait and State

As can be seen in the coefficients of the interactions with the trait dummy, which can be interpreted as reflecting the differences between the predictors in trait versus state assessments, the effects of control, value, and the Control × Value interaction were similar for trait and state emotions concerning the direction of the effects. This is completely in line with our hypotheses, however, the strength of the effects for some appraisals was found to differ between trait and state emotions. For pride, the significant Value x Trait interaction showed that for trait reports, the influence of value was stronger than for state reports (b = .13, p < .05). For anxiety, there was a significant Control × Trait interaction indicating that for trait reports, the negative influence of control on anxiety was significantly stronger than for state ratings (b = -.18, p < .01). Unexpectedly, the three-way-interaction (C x V x T) for boredom was significant and showed that the Control × Value interaction for trait ratings was significantly lower than for state ratings (b = .15, p < .05). Further analysis revealed that the multiplicative effect of control and value on boredom was not significant in trait data. Furthermore, the Value × Trait interaction effect on boredom was significant (b = -.20, p < .01), indicating a stronger negative effect of value on trait boredom.

2.7 Discussion and Implications

2.7.1 Control and Value as Appraisal Antecedents

In line with our first hypothesis (1a), control positively predicted pride and negatively predicted anxiety and boredom. This finding is in accordance with the theoretical assumption that perceptions of control are positively associated with positive emotions and negatively related to negative emotions (Pekrun, 2000). For state data, the negative relationship between control and boredom was relatively weak. One possible explanation for this weak effect is that the relationship between control and boredom is not linear but perhaps curvilinear. It is possible that students feel bored not only in situations of very low control but also in situations of very high control (see Pekrun et al., 2010). Results also revealed that value appraisals were positively correlated with pride and anxiety, which is consistent with hypothesis 1b. If the outcome of the task was deemed to be important, the intensity of the positively and negatively valenced emotions was higher. As expected (hypothesis 1c), the relation between value and boredom was negative (see Pekrun et al., 2010). In situations that were perceived as subjectively unimportant, more intense boredom was experienced.

Compared with previous studies that utilized an interindividual approach, we also found positive associations between control and pride and negative associations between control and anxiety and control and boredom. Interestingly, the relation between value and anxiety was positive, which is in line with the assumptions of the control-value theory. However, previous studies (e.g., Goetz et al., 2006) reported negative relations between value appraisals and negative emotions. One possible explanation for these contradictory findings may be the different analytical methods that were used, with interindividual approaches sometimes resulting in negative relations and intraindividual approaches resulting in positive relations. Future research is needed to investigate this discrepancy.

2.7.2 Interactions between Control and Value

In addition to the main effects of control and value on emotions, the present study revealed significant Control × Value interactions in the prediction of pride (see Goetz et al., 2010) as stated in our second hypothesis. For negative emotions, the combined effect of low control and high value resulted in more intense feelings of anxiety. For boredom, there was an interaction such that the relationship between control and boredom was different in cases of high versus low value. It is important to note that we chose to consistently depict the interaction effects with control as the predictor and value as the moderator for the sake of clarity, however, the reverse relationship could exist as well. It was beyond the scope of the current study to determine which appraisal antecedent was the true moderator and there are no theoretical assumptions in the control-value theory to help clarify the true nature of this relationship. To gain further insight into the moderating role of these two appraisals and their relationship with emotions, future experimental studies are necessary. Regardless, our results showed that control and value do indeed interact when predicting emotions. As interaction terms were mostly neglected in previous research on analyses of classical appraisal theory (Nagengast et al., 2011), our results are an important addition to the literature.

2.7.3 Structural Differences in Appraisal-Emotion Relationships between Trait and State Data

Consistent with our assumptions, trait and state data revealed a similar pattern of relations involving control, value, and the Control × Value interaction as predictors of the emotions assessed. There were no structural differences regarding the direction of the association between appraisals and emotions. This finding is noteworthy in showing that the results from two assessment methods underscore the generalizability of the assumptions of the control-value theory. Furthermore, this finding supports the results from previous studies that

used trait assessments of emotions but perhaps drew inappropriate conclusions about intraindividual functioning from interindividual analyses. However, there were differences with respect to some appraisals regarding the strength of association. On trait-based measures, value appraisals had a stronger influence on pride and boredom, and control was a stronger predictor of anxiety. Unexpectedly, the Control × Value interaction in boredom was only significant in state data, thus implying that there were only independent effects of control and value on trait boredom.

The differential influence of some appraisal antecedents concerning the strength of the effects in trait versus state emotions could possibly be due to the participants' subjective beliefs having a greater influence on the rating of trait emotions compared to state emotions. It could be the case that in memory-based trait assessments, the evaluation of trait emotions and antecedent control or value beliefs reciprocally influence each other over time resulting in the formation of a coherent belief system. The formation of this coherent belief system could result in stronger relations between appraisal antecedents and emotions in trait but not in state assessments in which appraisals and emotions are assessed situationally and are therefore devoid of any reciprocal feedback processes. This idea reasonably fits into the assumptions of Robinson and Clore's (2002) accessibility model insofar as trait assessments are assumed to be biased due to selective recall and partly influenced by subjective beliefs, which is less likely to occur with state assessments. As this is just one preliminary explanation, the possible influences of memory and beliefs on trait and state emotions needs to be examined in future studies.

2.7.4 Limitations, Strengths, and Implications

It should be noted that although it is quite reasonable to assume that the direction of influence is from control and value appraisals to emotions, the present data structure was correlational in nature thus precluding causal conclusions. However, the control-value theory is also explicit in identifying the critical importance of feedback loops between emotions and control and value appraisals. Furthermore, we chose single-item measures to assess the relevant constructs in order to keep the state questionnaire brief and to be able to compare trait and state assessments. Although this approach may not be as optimal as using full scales, past studies have successfully utilized single-item measures and found them to be advantageous and reliable (Goetz et al., 2010; Wanous, Reichers, & Hudy, 1997). Our student sample was also limited to the upper track, warranting further research with students of different age groups and in different achievement settings.

One strength of the present study was the use of an intraindividual design. As proposed in the control-value theory, conclusions on intraindividual functioning from appraisals to emotions are only justified when the intraindividual variation of emotions are considered. In future research, it is important to carefully consider the choice of data collection method and the strategy of analysis while evaluating intraindividual functioning. The control-value theory also states that control and value should interact in predicting emotions. In former studies the interaction was oftentimes disregarded, however, the present study can confirm the importance of including interaction terms. An additional strength of our study was that both trait and state assessment methods were administered. Capturing emotions and appraisals on-line in the daily life of students in class enables researchers to gain insight into students' emotions without (or with fewer) memory distortions and on an intraindividual basis. Future studies in which state and trait assessments are evaluated within a single sample are warranted in order to capitalize on recent developments in the field of emotion research in educational psychology.

From a practical standpoint, the present study supports ongoing research on intervention programs that promote perceptions of personal control (e.g., Hall, Chipperfield, Perry, Ruthig, & Goetz, 2006; Hall et al., 2007). Due to the inherent differences between trait and state emotions, it is important to distinguish between approaches taken by parents and teachers that attempt to influence these two classes of emotions. For example, in order to influence students' trait emotions, which are more enduring and stable than state emotions, it would be necessary to target their control and value beliefs via constructs such as academic self-concept. In contrast, students' state emotions are temporary and unstable, and attempts to manipulate control and value appraisals during class should consist of situation-specific strategies, such as providing more autonomy support or making the learning materials more relevant to students' lives.

In line with the observed interaction effects of control and value, our findings should encourage researchers who are developing programs aimed at enhancing control perceptions to also consider the role of value appraisals. Nevertheless, whereas the enhancement of students' control perceptions in class may lead to more positive emotional experiences, an enhancement of value may be a double-edged sword that could result in the undesirable intensification of negative emotions (e.g., anxiety). Thus, it may not always be reasonable to emphasize the value of a test, for example – especially when students' control beliefs are low, as this could lead to unwanted anxiety.

Our findings suggest that teachers should be explicitly informed of the importance of control and value in connection with students' emotions. Teachers should be encouraged to develop classroom environments that facilitate emotional experiences via control and value perceptions. These experiences can be influenced, for example, through autonomy support, value induction, clear goal structures, or expectations (Pekrun, 2006) with the ultimate aim of promoting learning and achievement. Finally, control and value are relatively well-researched appraisal antecedents of academic emotions, but the present recommendations for educational practice are in need of further empirical support. More studies are needed that focus on intervention programs aimed at influencing students' appraisals via environmental factors (i.e., instructional methods) and therefore positively influence subsequent emotional experiences.

3 What Students Think They Feel Differs From What They Really Feel – Academic Self-Concept Moderates the Discrepancy Between Students' Trait and State Emotional Self-Reports

3.1 Summary

This study investigated whether there is a discrepancy pertaining to trait and state academic emotions and whether self-concept of ability moderates this discrepancy. A total of 225 secondary school students from two different countries enrolled in grades 8 and 11 (German sample; n = 94) and grade 9 (Swiss sample; n = 131) participated. Students' trait academic emotions of enjoyment, pride, anger, and anxiety in mathematics were assessed with a self-report questionnaire, whereas to assess their state academic emotions experience-sampling method was employed. The results revealed that students' scores on the trait assessment of emotions were generally higher than their scores on the state assessment. Further, as expected, students' academic self-concept in the domain of mathematics was shown to partly explain the discrepancy between scores on trait and state emotions. Our results indicate that there is a belief-driven discrepancy between what students think they feel (trait assessment) and what they really feel (state assessment). Implications with regard to the assessment of self-reported emotions in future studies and practical implications for the school context are discussed.

3.2 Theoretical Background

Much of what we call emotion is nothing more nor less than
a certain kind—a biased, prejudiced, or strongly evaluative kind—of thought.
But emotions and behaviors significantly influence and affect thinking,
just as thinking significantly influences what we call emotions and behaviors.
(Ellis, 1999, p.71)

The emotions that a student experiences whenever learning in school is involved has become a growing area of research in education and psychology and a focus of attention for scholars, policy-makers, and the public. A number of special issues in leading journals have been dedicated to the study of academic emotions (Efklides & Volet, 2005; Linnenbrink-Garcia & Pekrun, 2011; Linnenbrink, 2006; MacCann et al., 2012; Schutz & Lanehart, 2002), and emotional and social skills have moved the center of current standards movements and legislation (e.g., Partnership for 21st Century Skills, 2006). Further, a number of large-scale international assessments have integrated emotions and related constructs into their programs (e.g., PISA; Naemi et al., 2013). The increased interest in the study of emotions is not surprising, as students' affect has been shown to relate to a wide range of important process and outcome variables in the academic context and emotions are seen as important outcome variables themselves. Variables related to emotions include learning strategies (for example self-regulated learning: Op't Eynde et al., 2007), academic achievement (Goetz & Hall, 2013; Pekrun et al., 2002), lifelong learning (Goetz et al., 2003), and domain and career choices (Wigfield et al., 2002). Overall, beyond intelligence and domain-specific skills emotions have been consistently shown to be important predictors of learning and achievement (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011) and are considered to be valued educational outcomes.

When investigating students' emotions, most of the studies rely on questionnaires to capture students' academic emotions 'in general', or habitual emotions (trait). Recently, however, the focus of research has been shifting toward assessing students' state emotions in real-life context via the experience-sampling method (see Linnenbrink-Garcia & Pekrun, 2011). The advantage of real-life assessments is in their higher ecological validity (e.g., Shiffman, Stone, & Hufford, 2008) as study participants are asked during their daily routines and not outside the context in question. In line with this assumption, initial investigations consistently reveal a discrepancy with regard to mean-level differences between trait emotions, and emotions that are currently experienced, or state emotions (see Robinson &

Clore, 2002). These theoretical developments coupled with empirical findings call for further investigation of differences between trait and state assessment methods.

3.2.1 Assessing Trait and State Emotions: The Accessibility Model of Emotional Self-Report

One reason for the found discrepancy between trait and state assessments of emotions may be attributable to the fact that trait emotions seem to be more strongly influenced by semantic knowledge than state emotions are. Robinson and Clore (2002) synthesize the findings with respect to the discrepancy between trait and state emotional assessment and propose an accessibility model of emotional self-report. The authors distinguish between trait and state emotional self-reports by classifying them according to the respective memory systems. Trait emotions are semantic, conceptual, and decontextualized, whereas state emotions are episodic, experiential, and contextual (Robinson & Clore, 2002). It is further suggested that state emotions are directly assessed and thus influenced by situational cues, whereas in trait assessments it is individuals' beliefs and semantic knowledge that affect outcomes of the assessment. As a result, there is an expected discrepancy between trait and state emotional assessments with traits relating stronger to subjective beliefs.

A number of studies have examined mean-level differences between trait- and state-based assessments of mood or emotions, sometimes reporting inconsistent findings (i.e., trait ratings being higher and lower than state ratings; e.g., Barrett, 1997; Levine et al., 2009). However, the vast majority of investigations show higher intensities of trait as compared to state emotions (e.g., Buehler & McFarland, 2001; Goetz et al., 2013; Wilson & Gilbert, 2005). This discrepancy between trait and state emotional assessments has been termed *intensity bias* (Buehler & McFarland, 2001) or *impact bias* (Wilson & Gilbert, 2005).

The study of differences between trait and state self-reports has not been limited to the area of emotions. Additional empirical support for the discrepancy between trait and state self-reports comes from a variety of branches of psychology with studies investigating how semantic knowledge influences this discrepancy. For example, one of the earlier studies revealed that recalled and actual symptoms of women's menstrual cycle significantly differed, with women overestimating the severity of symptoms upon recall, as compared to their real-time ratings (McFarland, Ross, & DeCourville, 1989). The authors found that the more female participants believed in the influence of menstruation on well-being the more they overestimated their recalled symptoms as compared to their state-rated symptoms. Similarly, Porter et al. (2000) investigated how assessment of trait coping strategies was biased

according to gender stereotypes compared to momentary assessment of coping strategies. Another example comes from van den Brink and colleagues' study that compared individuals' recalled and diary ratings of the severity of headaches (van den Brink, Bandell-Hoekstra, & Huijer Abu-Saad, 2001). In it, study participants reported higher intensity and duration of their headaches in the retrospective assessment, as compared to their ratings captured by diaries (real-time, state assessments). The results of these studies are relatively consistent: Trait assessments appear to be more strongly influenced by subjective beliefs as compared to state assessments, with traits being rated higher than states. Further, some studies provide initial evidence that this discrepancy can be explained by subjective theories that people hold.

These empirical findings indicate that trait emotions do not appear to be a good indicator of actual state emotions. Trait emotions are assumed to be influenced by subjective beliefs and are generally overestimated, as compared to state assessments (Robinson & Clore, 2002). The reported tendency for the individuals to rate trait emotions higher makes scientists question trait assessments' ecological validity. The review of literature on emotions in educational psychology, however, shows that there is a clear preponderance of studies that employ trait-based emotional assessments. Critical remarks about trait assessments considered, one may wonder why trait measures are still used at all to assess emotions. In addition to favorable economic considerations, with trait assessments being far less costly than state assessments, various studies demonstrate that traits are stronger predictors of future behavior and future choices (Safer, Levine, & Drapalski, 2002; Wilson & Gilbert, 2005; Wirtz et al., 2003). In the educational context these future choices could represent domain and/or career choices (Eccles, 1985; Wigfield et al., 2002). Thus, the aforementioned findings indicate that trait and state assessments may not be used interchangeably and should be selected depending on a research question that researchers are attempting to answer (Conner & Barrett, 2012). The current study will provide additional evidence and offer further insight into the discrepancy between trait and state emotional assessments.

3.2.2 Academic Self-Concept as a Possible Moderator of the Trait-State Discrepancy

Researchers have been trying to identify variables that may explain the discrepancy between ratings of trait and state emotions, and found subjective beliefs to be particularly relevant (Robinson & Clore, 2002). For academic emotions, it is assumed that students' self-concept belief is an important moderating variable. The importance of self-concept can be inferred from Pekrun's control-value theory of achievement emotions (2006), which stipulates that the component of control, commonly represented by academic self-concept, is one of the

most prominent antecedents of academic emotions. Academic self-concept represents an important control belief in the school context and is defined as memory structure and representation of the abilities and competencies a person has (Nagengast & Marsh, 2012). It has been shown to be positively associated with positive emotions and negatively with negative emotions. Due to its prominent role in academic emotions, self-concept belief should be particularly effective in explaining the discrepancy between trait and state emotional assessments in a way that this belief more strongly influences trait emotional assessments but does not bias state emotional assessments.

To our knowledge, there is only one study that investigated the role of self-concept in explaining the discrepancy between trait and state emotions (Goetz et al., 2013). This study examined gender differences in trait and state mathematics anxiety and showed that despite similar state mathematics anxiety ratings girls report higher trait mathematics anxiety ratings as compared to boys. The discrepancy between trait and state mathematics anxiety in girls was partly explained by girls' lower self-concepts thus showing the significant role that self-concept plays in clarifying existing differences between the two approaches to assessment.

Several other studies investigated the influence of self-esteem, a construct that is closely related to self-concept, on emotional ratings. Robinson and Barrett (2010) conducted three studies examining links between self-esteem and emotional judgments. The authors found that people with high self-esteem tended to more positively rate their trait emotional experiences. State emotional assessments, however, were found to be unrelated to self-esteem. Another study showed that self-esteem influences recall of emotional experiences in a way that high self-esteem more strongly biases positive emotional recalls (Christensen, Wood, & Barrett, 2003). In sum, in line with the accessibility model of emotional self-report (Robinson & Clore, 2002) the results of these studies found that trait reports were more strongly influenced by semantic knowledge as compared to state self-reports. Further, self-concept and self-esteem were shown to be potential moderators of the discrepancy between the two approaches to emotional assessment. The current study will attempt to further extend our understanding of the role that self-concept plays in explaining these trait-state differences.

3.3 Aims of the Present Study and Hypotheses

The aim of this study was to compare students' trait and state emotional self-reports with respect to a possible discrepancy between the two approaches. We also wanted to

investigate whether academic self-concept impacts the magnitude of the discrepancy between self-reported trait and state emotions.

Based on the findings of prior empirical studies (e.g., Buehler & McFarland, 2001; Goetz et al., 2013), we expect to find a discrepancy between the rated intensity of trait academic emotions and state academic emotions. We expect trait emotions to be rated higher than state emotions (intensity bias; Hypothesis 1). Beyond our attempt to replicate previous findings of the intensity bias in the academic context, we intend to explain the discrepancy between trait and state emotional assessments with students' academic self-concept. We expect self-concept to positively predict the discrepancy between trait and state emotions in positive emotions and negatively predict it in negative emotions (Hypothesis 2). That is, as control positively relates to positive emotions and negatively relates to negative emotions we expect students' self-concept beliefs to influence trait emotional assessments in the same direction.

Our study hypotheses were investigated in two samples from two different countries. Four emotions were examined: Two positive, activating emotions of enjoyment and pride, and two negative, activating emotions of anger and anxiety. These were chosen based on their high importance and frequently occurrence in the school context (Pekrun et al., 2002). We investigated our hypotheses in the context of mathematics because several studies found that academic emotions are organized in a domain-specific way (Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007a; Goetz et al., 2006). As mathematics is one major domain in the school context, e.g., because of its importance for a wide range of professions, we assume that this is a good starting point to investigate the hypotheses.

To summarize, we were interested in examining differences between students' trait and state emotional assessments of enjoyment, pride, anger, and anxiety in mathematics. We expected trait emotions to be rated higher than their respective state emotions. Furthermore, we investigated whether self-concept of ability can explain this discrepancy between the two assessment methods.

3.4 Method

3.4.1 Sample

Two samples were included in the current study. The first sample consisted of $N = 94$ German students of grade 8 (54.8 %, $M_{age} = 14.30$ years, $SD = 0.51$; 24 males) and grade 11

(M_{age} = 17.57, SD = 0.58; 19 males) of 39 different classes (two to three randomly chosen students per class) from the upper track of the state school system in Germany (Gymnasium). The second sample included N = 131 9th -graders from German-speaking Switzerland enrolled in 41 classes (three to four randomly chosen students per class; 44.3 % male, M_{age} = 15.67 years, SD = 0.64).

Although Germany and Switzerland are neighboring countries, there are several differences in their school systems that stem from rather unique educational traditions. One major difference is that students in Switzerland are separated according to ability tracks at a later point in time (usually after six years as compared to four years in Germany). Another notable difference has to do with the class size, which is usually smaller in Switzerland.

3.4.2 Procedure

Students' trait emotions and self-concept in mathematics were assessed via paper-and-pencil questionnaire that was administered by trained experimenters. The same items were used in the German and the Swiss sample. The procedure for the assessment of students' state emotions was highly similar in the German and the Swiss sample and started right after the trait assessment. State data were assessed by employing a computer-based experience-sampling method (see Hektner et al., 2007). In the German sample two to three randomly chosen students from each classroom were provided with a personal digital assistant (PDA). In the Swiss sample three to four students per classroom were provided with a PDA. The participants were asked to activate PDAs at the beginning of every mathematics class for a period of two weeks. The PDA randomly signaled within 40 minutes from the start of a lesson, prompting students to answer questions about their momentary emotions during that specific class. Therefore, our research design combines event-based and random sampling (Shiffman et al., 2008). Students who took fewer than two assessments were excluded from the analyses. In total, this procedure resulted in N = 415 state measures with a mean number of 4.41 state assessments per student in the German sample and N = 749 state measures with a mean number of 5.72 state assessments per student in the Swiss sample.

3.4.2.1 Assessment of trait emotions

In both samples single items were used to assess the four trait emotions of enjoyment, pride, anger, and anxiety: 'How much [EMOTION] do you generally experience during mathematics classes?' The response format consisted of a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

3.4.2.2 Assessment of state emotions

State emotions were assessed using single items for each of the four emotions (parallel wording to trait assessment adjusted for the lesson: 'How much [EMOTION] are you experiencing during this class?'; see Goetz et al., 2013). The decision to use single items was due to practical reasons (e.g., minimizing lesson disruptions) and to avoid unintentionally evoking or changing emotions by the emotional assessment itself (see Goetz et al., 2010). Responses ranged from *strongly disagree* to *strongly agree* (5-point Likert scale). In the Swiss sample students were asked to report emotions they are experiencing *'right now'* as compared to *'during this class.'* This was the only difference in the assessment between the two samples.

3.4.2.3 Assessment of self-concept

Similarly to trait emotions, students' academic self-concept was assessed via paper-and-pencil questionnaire. Three items for academic self-concept were adapted from the Self-Description Questionnaire (SDQ; Marsh, 1990). Sample item includes: 'I have always done well in mathematics.'. The total score was calculated by taking an average of the three self-concept items.

3.4.3 Statistical Analysis

The main focus of our analyses was on the discrepancy between trait and state emotions and how this discrepancy is moderated by self-concept[6]. For that reason, we combined trait and state emotion measures for each emotion into one variable and separated them in our analyses by introducing a dummy called "Trait" with state measures being coded as 0 (reference group) and trait measures coded as 1. As trait and state emotion measures are nested within students, and students are nested within classes, our data reflect a three-level structure with measurement points nested within students and students nested within classes. Thus, the analyses were conducted via multilevel statistics using HLM 6.08 (Hierarchical Linear Modeling; Raudenbush et al., 2009).

The advantages of the multilevel statistical approach, as compared to other analytical strategies that have been used to study differences between trait and state emotions (e.g., Christensen et al., 2003; mean-level differences and moderator analysis) is that we can

[6] In our analyses, we did not report gender as a possible moderator of the trait-state discrepancy. Significant gender differences in math trait anxiety but not in state math anxiety were found (as reported in the study of Goetz et al., 2013). However, as gender differences in emotions were not a major concern in the present study, we decided not to include it in our analyses.

account for the nested data structure (multiple measurement points per person and persons nested within classes) and for different numbers of measurement points per person (one trait measure but different number of state measures per person). This results in more adequate standard errors in statistical testing. Furthermore, while using this intraindividual analysis (trait-state discrepancy within each student), we assure that we do not commit an ecological fallacy and draw conclusions on the wrong level of analysis (Molenaar & Campbell, 2009).

3.4.3.1 Level 1 variable

In order to test Hypothesis 1 (discrepancy between trait and state emotional assessments; 0 = *state*, 1 = *trait*), we introduced the Trait dummy into all of our hierarchical linear regression models. Due to the coding of this variable, the intercept evaluated as γ_{000} describes the mean state emotion (i.e. the value if all predictors are zero). The effect of the Trait dummy (γ_{100}) in our models can be interpreted as an indicator of the discrepancy between state and trait emotions. Significant positive effects of the Trait dummy indicate significantly higher trait ratings as compared to state ratings.

3.4.3.2 Level 2 variable

We further examined whether the discrepancy between trait and state assessments can be predicted by students' academic self-concept in mathematics (Hypothesis 2). Therefore, we added self-concept as a *z*-standardized variable into our multilevel analyses as a predictor of the slope of the Trait dummy (slope-as-outcome model), which results in a cross-level interaction between Level 1 and Level 2 (Self-concept × Trait interaction; γ_{110}). This interaction term represents the effect of self-concept on the amount of difference between trait and state emotion scores. Positive effects indicate that higher self-concept values are associated with higher discrepancies between trait and state assessments, whereas negative effects for the self-concept variable indicate smaller discrepancies. For the sake of completeness, self-concept was also introduced into the model to predict the intercept (γ_{010}). This 'main effect', however, was not of importance for testing our hypotheses.

The mixed model regression equation for Model 1 (combined model), used for each of the four emotions, is as follows:

$$Y_{ijk}[\text{Emotion value } i \text{ of student } j \text{ in class } k] = \gamma_{000} + \gamma_{100}(\text{Trait}) + \gamma_{010}(z\text{Self-concept}) + \gamma_{110}(z\text{Self-concept*Trait}) + r_0 + r_1(\text{Trait}) + u_{00} + e$$

3.4.3.3 Level 3 variable

In addition to the Trait dummy and the self-concept variable, a dummy for either Switzerland (CH_Dummy; German model, Model 2) or Germany (DE_Dummy; Swiss model, Model 3) was introduced on the third level into the analyses to account for possible differences between the two samples. The difference between the two samples may be twofold. On the one hand, the samples were assessed in different countries (Germany vs. Switzerland) and on the other hand, slightly different instructions for state emotions assessment were used ('in this class' vs. 'right now'). Thus, we present our analyses for the combined sample as well as for each of the two countries as a reference group (including a dummy variable for the other country, respectively). Coefficients for the interaction of each variable with the respective country dummy (i.e. Trait × CH_Dummy, γ_{101}; Self-concept × CH_Dummy; γ_{011}; Trait × Self-concept × CH_Dummy; γ_{111}) indicate differences between the effect for the country as compared to the reference group, e.g. in the German model the dummy for Switzerland indicates differences between the effect for the Swiss sample compared to the German sample (reference group).

Hierarchical linear modeling, regression equations for Models 2 and 3:

Model 2 – German model (German sample is reference group)

$$Y_{ijk}[\text{Emotion value } i \text{ of student } j \text{ in class } k] = \gamma_{000} + \gamma_{100}(\text{Trait}) + \gamma_{010}(z\text{Self-concept}) + \gamma_{001}(\text{CH\_dummy}) + \gamma_{110}(z\text{Self-concept*Trait}) + \gamma_{011}(z\text{Self-concept*CH\_dummy}) + \gamma_{101}(\text{Trait*CH\_dummy}) + \gamma_{111}(\text{Trait}*z\text{Self-concept*CH\_dummy}) + r_0 + r_1(\text{Trait}) + u_{00} + e$$

Model 3 – Swiss model (Swiss sample is reference group)

$$Y_{ijk}[\text{Emotion value } i \text{ of student } j \text{ in class } k] = \gamma_{000} + \gamma_{100}(\text{Trait}) + \gamma_{010}(z\text{Self-concept}) + \gamma_{001}(\text{DE\_dummy}) + \gamma_{110}(z\text{Self-concept*Trait}) + \gamma_{011}(z\text{Self-concept*DE\_dummy}) + \gamma_{101}(\text{Trait*DE\_dummy}) + \gamma_{111}(\text{Trait}*z\text{Self-concept*DE\_dummy}) + r_0 + r_1(\text{Trait}) + u_{00} + e$$

3.5 Results

3.5.1 Descriptive Statistics

The reliability of the self-concept scale was satisfying (German sample: $\alpha = .91$; Swiss sample: $\alpha = .86$). Table 3.1 shows means and standard deviations of variables for the combined sample and the German and Swiss samples separately. As expected, trait ratings are

higher than state ratings for every emotion in both samples. The only exception is that state *enjoyment* in the Swiss sample was rated higher than trait *enjoyment*.

Table 3.1. *Descriptive statistics*

	Combined sample ($N = 225$)				**German sample** ($n = 94$)				**Swiss sample** ($n = 131$)			
	Trait		State		Trait		State		Trait		State	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Enjoyment	2.57	1.14	2.52	1.27	2.65	1.16	2.20	1.30	2.51	1.12	2.71	1.21
Pride	2.51	1.13	1.95	1.22	2.72	1.18	1.67	1.09	2.35	1.07	2.12	1.27
Anger	2.76	1.28	1.97	1.26	2.91	1.21	1.84	1.24	2.64	1.32	2.04	1.27
Anxiety	1.79	1.12	1.51	1.03	2.00	1.25	1.52	1.06	1.63	0.99	1.51	1.00
Self-concept	3.03	1.15	--	--	2.96	1.21	--	--	3.09	1.10	--	--

3.5.2 Hierarchical Linear Regression

The results of the hierarchical linear regression for the four emotions of *enjoyment*, *pride*, *anger*, and *anxiety* are shown in Table 3.2. Further, the variance components are depicted in this table. We also calculated the explanatory power of self-concept with regard to the slope variance, that is, as a predictor of the trait-state discrepancy (see Aguinis, Gottfredson, & Culpepper, 2013).

Table 3.2. *Predicting Emotions: Results from Multilevel Modeling*

	Enjoyment			Pride			Anger			Anxiety		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Level 1												
Intercept (γ_{000})	2.52***	2.25***	2.71***	1.95***	1.74***	2.10***	1.97***	1.79***	2.10***	1.50***	1.49***	1.51***
Trait (γ_{100})	0.05	0.42**	-0.22*	0.56***	1.02***	0.22*	0.77***	1.09***	0.54***	0.28**	0.48***	0.12
Level 2												
Self-concept (γ_{010})	0.19***	0.17**	0.17*	0.17**	0.19*	0.12	-0.14**	-0.22**	-0.10	-0.11*	-0.09	-0.13
Cross-level interactions L1-L2												
Self-concept × Trait (γ_{110})	0.23*	0.13	0.35**	0.30***	0.28**	0.36**	-0.34***	-0.32**	-0.33**	-0.33***	-0.41***	-0.24*
Level 3												
CH_dummy (γ_{001})		0.46***			0.37**			0.31*			0.02	
DE_dummy (γ_{001})			-0.46***			-0.37**			-0.31*			-0.02
Cross-level interactions L1-L3												
Trait × CH_dummy (γ_{101})		-0.63***			-0.80***			-0.55**			-0.36*	
Trait × DE_dummy (γ_{101})			0.63***			0.80***			0.55**			0.36*
Cross-level interactions L2-L3												
Self-concept × CH_dummy (γ_{011})		-0.01			-0.07			0.12			-0.04	
Self-concept × DE_dummy (γ_{011})			0.01			0.07			-0.12			0.04
Cross-level interactions L1-L2-L3												
Trait × Self-concept × CH_dummy/ (γ_{111})		0.22			0.08			-0.01			0.17	
Trait × Self-concept × /DE_dummy (γ_{111})			-0.22			-0.08			0.01			-0.17
Variance components												
Within-student (L1) variance (σ^2)	1.103	1.096		0.924	0.916		1.120	1.118		0.735	0.735	
Intercept (L2) variance (τ_{00})	0.419	0.391		0.460	0.429		0.419	0.404		0.254	0.255	
Slope (L2) variance (τ_{11})	0.174	0.122		0.252	0.128		0.243	0.173		0.258	0.217	
Intercept-slope (L2) covariance (τ_{01})	-0.269	-0.215		-0.320	-0.234		-0.230	-0.196		-0.109	-0.104	
Intercept (L3) variance	0.034	0.014		0.059	0.051		0.046	0.042		0.021	0.020	
Explanatory power	0.022	0.090		0.261	0.369		0.314	0.357		0.295	0.332	

Note. Model 1 = combined model; Model 2 = combined dataset with German sample as reference group; Model 3 = combined dataset with Swiss sample as reference group.
Description of variables: Trait = Trait dummy (0 = state, 1 = trait); CH_dummy = Swiss dummy (0 = German sample, 1 = Swiss sample); DE_dummy = German dummy (0 = Swiss sample, 1 = German sample).

German sample: $N_{Level\ 1} = 509$; $N_{Level\ 2} = 94$; $N_{Level\ 3} = 39$; Swiss sample: $N_{Level\ 1} = 880$; $N_{Level\ 2} = 131$; $N_{Level\ 3} = 41$.
Explanatory power refers to the proportion of slope variance explained by the Level 2 and Level 3 predictors. The slope variance of the models in which no cross-level interaction is included was: $\tau_{11} = 0.178$ for enjoyment in the combined model and $\tau_{11} = 0.134$ in the German/Swiss model; $\tau_{11} = 0.341$ for pride in the combined model and $\tau_{11} = 0.203$ in the German/Swiss model; $\tau_{11} = 0.354$ for anger in the combined model and $\tau_{11} = 0.269$ in the German/Swiss model; $\tau_{11} = 0.366$ for anxiety in the combined model and $\tau_{11} = 0.325$ in the German/Swiss model.
* $p < .05$, ** $p < .01$, *** $p < .001$.

3.5.2.1 Model 1 – Combined model

In Model 1 the coefficient for the intercept (γ_{000}) is to be interpreted as the mean emotion score when all other variables in the model are equal to zero. Thus, this represents the respective mean state emotion for a student who has a mean self-concept. The mean state score for the emotion of *enjoyment* was 2.52, 1.95 for *pride*, 1.97 for *anger*, and 1.50 for *anxiety*. The coefficient for the Trait dummy (γ_{100}) is positive and significant for every emotion with the exception of *enjoyment*, for which no significant difference was found. Thus, with one exception, trait emotions are rated higher than state emotions (Hypothesis 1).

Regression weights for the Self-concept × Trait (γ_{110}) interaction indicate the influence of self-concept on the discrepancy between trait and state emotional assessments. As expected, for *enjoyment* (.23) and *pride* (.30) the coefficients were positive, whereas for *anger* (-.34) and *anxiety* (-.33) the coefficients were negative. This suggests that high self-concept in mathematics is associated with higher discrepancies between trait and state *enjoyment* and *pride* and smaller discrepancies between trait and state *anger* and *anxiety* (Hypothesis 2).

3.5.2.2 Models 2 and 3 – German model and Swiss model

In order to account for differences between German and Swiss samples, we calculated models for each of the four emotions with a country dummy for Switzerland (CH_dummy; Model 2) and Germany (DE_dummy; Model 3). With regard to our first hypothesis, trait emotions were rated significantly higher than state emotions. In the German model (Model 2), coefficients for the Trait dummy (γ_{100}) were 0.42 for *enjoyment*, 1.02 for *pride*, 1.09 for *anger*, and 0.48 for *anxiety*. Coefficients in the Swiss model were -0.22 for *enjoyment*, 0.22 for *pride*, 0.54 for *anger*, and 0.12 (*n.s.*) for *anxiety*. Hence, trait ratings were once again higher with the exception of *enjoyment* in the Swiss sample. Here, unexpectedly, the mean trait enjoyment was lower than the mean state enjoyment. Further, the discrepancy between trait and state *anxiety* was not significant in the Swiss sample. For each emotion, the discrepancy between trait and state ratings was found to be significantly lower in the Swiss sample (negative coefficient for Trait × CH_Dummy, γ_{101}).

With regard to self-concept as a moderator of the discrepancy between trait and state assessments, the coefficients for the Trait × Self-concept interaction (γ_{110}) were 0.13 for *enjoyment* (*n.s.*), 0.28 for *pride*, -0.32 for *anger*, and -0.41 for *anxiety* in the German sample (Model 2). In the Swiss sample (Model 3), the coefficients for the Trait x self-concept interaction (γ_{110}) were 0.35 for *enjoyment*, 0.36 for *pride*, -0.33 for *anger*, and -0.24 for

anxiety. The strength of the moderation effect of self-concept on the trait-state discrepancy did not differ significantly between the two countries (all coefficients for Trait × Self-concept × Country dummy (γ_{111}) were non-significant).

3.6 Discussion

The aim of the present study was to investigate whether there is a discrepancy between trait and state academic emotions, and whether this discrepancy could be explained by students' academic self-concept. The results of our study revealed a significant discrepancy between trait and state emotions in mathematics in a way that trait emotions were generally rated higher than state emotions with the exception of enjoyment and anxiety in the Swiss sample. Thus, our hypothesis about the discrepancy between trait and state mathematics emotions was generally supported (Hypothesis 1). This finding appears to be consistent with previous studies that have demonstrated an intensity bias in the prediction, recall, and evaluation of emotions in general (e.g., Buehler & McFarland, 2001; Wirtz et al., 2003). Due to the fact that we used parallel item formulations for trait and state emotional assessments, directly comparing mean-level differences was justified in our study. Despite the fact that both methods (i.e., trait and state) are routinely employed to assess students' emotions, they obviously index different aspects of this construct. Thus, researchers and practitioners alike should refrain from drawing conclusions from mean-levels in trait assessments to mean-levels in state assessments and the other way around.

As predicted, self-concept moderated the magnitude of the discrepancy between trait and state emotional assessments (Hypothesis 2) with the exception of enjoyment in the German sample. Specifically, students with lower self-concept tended to more strongly overestimate their negative trait emotions (anger and anxiety) as compared to their actual state emotions. Conversely, students with higher self-concept tended to more strongly overestimate their positive trait emotions (enjoyment and pride) as compared to their actual state emotions in mathematics. Overall, trait emotional assessments seem to be influenced by subjective beliefs, and academic self-concept represents one of the most important beliefs in school. Our finding that self-concept moderates the magnitude of the difference between trait and state emotions is consistent with the view that trait emotions are more strongly biased by subjective beliefs and therefore capture *beliefs* about emotions and not necessarily actual emotions (Robinson & Clore, 2002).

We just argued that it is not possible to draw conclusions from mean trait emotions to mean state emotions. However, knowing students' academic self-concept should allow us to make a rough estimate of the similarity of trait and state emotional assessments and therefore the possibility to predict mean trait emotions from mean state emotions and vice versa. As trait emotions can be easily gauged, an estimate of the extent to which trait emotions reflect actual mean state emotions can be helpful, especially when more costly state assessments are not available. When talking about positive trait emotions, students with lower self-concepts seem to have a more 'realistic' estimate of their trait emotions, when state emotions are viewed as a benchmark for the 'actual' or 'real' emotions. The other way around, students with higher self-concepts seem to less strongly overestimate their negative trait emotions. Furthermore, it might be possible to find the self-concept threshold where the intensity of the respective trait and state emotion is estimated equally.

The explanatory power of self-concept in the prediction of the discrepancy between trait and state emotional assessments was .02 for enjoyment and .26 for pride, .31 for anger, and .30 for anxiety in the combined model. Overall, self-concept explained a substantial amount of variance in the discrepancy between trait and state assessments; however, it is only one of the beliefs which is important with regard to academic emotions. According to Pekrun's control-value theory (2006), value is another important appraisal antecedent that relates to the subsequent emotions. Intrinsic value reflects the value of an activity independent of the results. The lower explanatory power of self-concept in the trait-state discrepancy for the emotion of enjoyment may be attributable to the fact that enjoyment is one emotion, for which value appraisal may be more important than self-concept appraisal and thus, intrinsic value beliefs may be more predictive of the discrepancy between trait and state.

Related to this idea is a possible explanation of the finding that in the German sample, surprisingly, self-concept was not found to significantly moderate the magnitude of the trait-state discrepancy. And this although the analyses with the country dummies showed that the finding of self-concept being a moderator of the discrepancy across the two samples were rather consistent. Thus, one reason for this unexpected finding could be the aforementioned importance of intrinsic value beliefs with regard to enjoyment. It is possible that value beliefs contribute much more to the trait-state discrepancy for enjoyment than does self-concept.

Another difference between the two samples was that in the Swiss sample average ratings of state enjoyment were higher than average ratings of trait enjoyment. In general, the discrepancy between trait and state emotional assessments was in all cases stronger in the

German sample. The reason for this difference may be manifold. It is possible that cultural differences may lead to the difference. Another explanation may come from the different state item wording as enjoyment is a rather situation-specific emotion. Thus, the wording 'How do you feel *right now*' may lead to a stronger focus on the situation as compared to the specific math lesson. Future studies should employ identical items to compare results across samples and may use anchoring vignettes (e.g., Guindon & Boyle, 2012) when assessing differences in emotion self-reports across different countries. To summarize, despite several unexpected results, our study revealed quite consistent findings with trait emotions being rated higher than state emotions and self-concept being a moderator of the trait-state discrepancy.

3.6.1 Limitations and Future Directions

Our sample is limited to the upper track of the school system and only includes students from grade levels 8, 9, and 11. Future research may downward or upward extend our study and explore whether our findings generalize to students of different ages. Further, we only investigated our hypotheses in one specific domain, namely the domain of mathematics. This is justified given that academic emotions were found to be domain-specific with regard to mean-level differences (Goetz et al., 2007a). Future research could test whether the findings of the present study generalize to other academic disciplines, which we assume should be the case as similar results were found in different contexts before (Robinson & Clore, 2002).

Additionally, we only investigated the trait-state discrepancy with the emotions of enjoyment, pride, anger, and anxiety. Future research could include other emotions that are of high importance in the learning and achievement context. For example, boredom and relief are other relevant and frequently occurring emotions in school (Nett et al., 2011; Pekrun et al., 2011).

We also used two different wordings for the state items in the two samples. Future studies should pay attention to the different formulation of items and investigate how this perhaps results in different outcomes, as manifested in larger or smaller discrepancies between trait and state emotional assessments.

Finally, our study investigated self-concept as a moderator of the trait-state discrepancy. As self-concept was shown to predict a significant amount of variance in the discrepancy between trait and state emotional assessments, it seems that self-concept is one of the most important variables with a high explanatory power. However, future studies may

examine other possible moderators, such as value (e.g., intrinsic value for enjoyment) or stereotypic beliefs about emotions. It is possible that the effect of different moderators on the trait-state discrepancy may vary depending on the emotion being studied. For example, value beliefs could be more important in one emotion (e.g., enjoyment) than in another emotion (e.g., pride). Hence, investigating different combinations of discrete emotions and variables that may serve as moderators of trait-state discrepancy may prove to be a fruitful avenue for research.

The results of the present study raise questions about the ecological validity of trait assessments as they seem to be strongly related to subjective beliefs and memory biases. In other words, they do not assess actual emotions. We would like to encourage researchers to differentiate between the two assessment methods and bear in mind that they cannot be used interchangeably. Hence, we implore investigators to choose one approach versus the other depending on a research question.

3.6.2 Implications for Educational Practice

Explicating our findings from a practical perspective is particularly important: Students' emotional beliefs seem to have strong impact on their future career choices more than their actual emotions. As traits affect future behavior (Wirtz et al., 2003) and domain and career choices in the school context (Wigfield et al., 2002), it is important to keep in mind that subjective beliefs may influence these choices, too. This may prevent students from proceeding careers in the respective domain.

Thus, when one is interested in far-reaching consequences of emotional beliefs, trait emotions are the assessment method one should use. In this way possible interventions can be derived. Students could be made aware of the possible discrepancy between their actual emotions and what they think about their emotions and how their beliefs may influence their career choices. Encouraging them to check whether their beliefs are consistent with their actual emotions can be a promising way to help students to go into mathematics careers (Goetz et al., 2013). In order to change subjective beliefs, cognitive interventions such as attributional retraining seem promising (Hall et al., 2007; Peyton et al., 2008; Robbins, Oh, Le, & Button, 2009). By prompting students to closely monitor their emotions we may help them to realize that they are not as anxious or angry as they believe they are.

With regard to this, teachers play a key role and they could be informed of the important influence of student self-concept on trait emotions and therefore its possible effects

on individuals' domain and career choices. From an intervention perspective, there are multiple programs aimed at fostering students' self-concept (O'Mara, Marsh, Craven, & Debus, 2006). It could be expected that a change in self-concept beliefs comes along with changes in emotional beliefs and may therefore contribute to basing future decisions on more realistic estimates of how one feels.

3.6.3 Conclusion

The results of our study show that although trait and state assessments are intended to gauge the same construct, they are different. According to Robinson and Clore (2002), state emotions refer to actual emotions (episodic, experiential, and contextual) whereas trait emotions refer to *beliefs* about emotions (semantic, conceptual, and decontextualized). As Ellis (1999, p. 71) noted in the initial quotation "[…] thinking significantly influences what we call emotions […]" seems to hold true at least for trait emotional assessments. This leads to the recommendation that researchers should clearly differentiate between the two assessment methods and assess emotions according to the main research question. Further, we found that the discrepancy between trait and state emotions is in part explained by students' self-concept beliefs, with higher self-concept being associated with a stronger discrepancy of positive emotions and lower self-concept beliefs being associated with overestimation of negative emotions each compared to actual state emotions. In sum, it appears that what students think they feel (trait assessment) is not necessarily what they really feel (state assessment).

4 Do Girls Really Experience More Anxiety in Mathematics?

4.1 Summary

Two studies were conducted to examine gender differences in trait (habitual) versus state (momentary) mathematics anxiety (Study 1: N = 584, high school students, grades 5 to 10; Study 2: N = 111, high school students, grades 8 and 11). For trait math anxiety, the findings of both studies replicated previous research showing female students to report higher levels of anxiety than male students. However, no gender differences were observed for state anxiety as assessed by experience sampling during taking a math test (Study 1) and when attending math classes (Study 2). The discrepant findings for trait versus state math anxiety were partly accounted for by students' competence beliefs in mathematics, with female students showing lower perceived competence than male students despite having the same average math grades. Implications for educational practice and the assessment of anxiety are discussed.

4.2 Introduction

Female students report higher levels of mathematics anxiety than male students, as documented in meta-analyses of studies with secondary school students from around the globe (Hyde, Fennema, Ryan, Frost, & Hopp, 1990; see also Else-Quest, Hyde, & Linn, 2010, for data from the Programme for International Student Assessment, PISA; OECD 2004). These findings are discouraging given the negative effects of anxiety on psychological health, learning behaviors, self-regulation, and academic achievement (Diener, 2000; Pekrun et al., 2002; Zeidner, 1998). Research further shows math anxiety to negatively predict course enrollment, career choices, and life-long learning in mathematics-related fields, thus contributing to the underrepresentation of females in many domains of science, technology, engineering, and mathematics (STEM; Eccles, 2012; Halpern et al., 2007; National Academy of Sciences, 2006; Wigfield et al., 2002; Wirtz et al., 2003). This gender gap in math anxiety stands in marked contrast to the fact that female students typically obtain similar, or only slightly lower, levels of achievement in mathematics as compared to their male counterparts (Else-Quest et al., 2010; Hyde, Lindberg, Linn, Ellis, & Williams, 2008).

However, existing research on mathematics anxiety is almost exclusively based on self-reports of trait-like (habitual) anxiety as opposed to state (momentary) anxiety assessed during real-life experiences. As trait versus state self-report assessments can lead to very different results (e.g., Porter et al., 2000), this notable omission of state-based measures raises the issue of whether differences in math anxiety actually exist between male versus female students in everyday life. By evaluating both trait and state-based measures of math anxiety in students of various ages, the present study aimed to directly address this intriguing research question.

4.3 The Gender Gap in Math Anxiety: The Issue of Perceived Competence

There is considerable empirical support for the idea that self-report measures of trait anxiety are significantly impacted by subjective beliefs (Robinson & Clore, 2002). In contrast, such beliefs are much less likely to bias real-time reports of anxiety as experienced in a given situation (state anxiety). This assumption is in line with the accessibility model of emotional self-reports (Robinson & Clore, 2002) in which state measures are assumed to evaluate emotions, whereas trait measures better reflect *beliefs* about emotions.

With respect to anxiety, subjective beliefs involving personal competence represent a critical antecedent of this emotion and play a central role in self-reports of trait emotions more

generally (Pekrun, 2006). Research has shown that girls typically report significantly lower levels of perceived competence, as compared with boys, on measures of math-related self-efficacy and self-concepts of ability (Goetz, Frenzel, Hall, & Pekrun, 2008; Hyde et al., 1990). Given the relative lack of achievement differences in mathematics between boys and girls, findings further suggest that gender stereotypes about mathematics may be largely responsible for girls having lower levels of perceived competence in this domain, as evidenced by statements such as "Girls and mathematics are a bad fit" or "Mathematics is clearly a male domain" (Keller, 2002; Steele & Aronson, 1995; see also mathematics-related stereotypes within the internal/external frame of reference model; Marsh, 1986).

Given these findings, we propose that the gender gap in trait mathematics anxiety may be due to the use of trait self-report methods that allow personal competence beliefs to bias reports of anxiety. Moreover, we propose that measures of anxiety completed by students while actually learning about math or being tested in math content will be less impacted by their personal beliefs and show weaker gender differences than trait measures. Although girls may report more trait math anxiety than boys due to lower levels of perceived competence, such gender differences should be less pronounced on state self-reports of math anxiety. This assertion is consistent with a few prior studies showing gender differences to be observed on trait but not state self-report measures of related variables (e.g., coping strategies; Porter et al., 2000). However, to the best of our knowledge, there to date exist no empirical studies in which this research question is explored with respect to math anxiety.

4.4 The Present Research

The present research evaluates the assertion that girls report higher levels of anxiety in mathematics on trait-oriented self-report measures but that this gender difference is less pronounced in state self-reports. By implication, we expected girls to show a greater discrepancy in levels of trait versus state math anxiety than boys (Hypothesis 1). We further expected that the stronger discrepancy in reports of trait versus state math anxiety for girls could be explained by girls' lower competence beliefs, given the importance of such appraisals as antecedents of self-reported trait emotions (Hypothesis 2). Although not the primary focus of the present study, it was further anticipated that our findings would replicate previous research showing girls and boys to have similar grades in mathematics.

Two studies were conducted with each evaluating both trait and state self-report measures of anxiety, self-reports of perceived competence, as well as math achievement. To

evaluate the generalizability of the study findings, test-related as well as class-related math anxiety were assessed, and students of various age groups were included in the study samples. In Study 1 (5^{th} to 10^{th} graders), *test anxiety* in mathematics was assessed using both trait and state measures, with the latter completed during a math test. In Study 2 (8^{th} and 11^{th} graders), *class-related* mathematics anxiety was assessed using trait and state measures, with the latter administered during regular math classes. In both studies, the state measures of anxiety involved experience-sampling methods (ESM; Csikszentmihalyi & Larson, 1987; Hektner et al., 2007).

4.5 Method

4.5.1 Samples and Procedure

The samples included students from multiple grade levels in the top track of the education system in Germany (i.e., Gymnasium; approximately one third of the total student cohort). The Study 1 sample consisted of 584 students (24 classrooms from 6 schools) from grades 5 through 10 (45% female; M_{age} = 13.67 years, SD_{age} = 1.84). This study was part of the *Project for the Analysis of Learning and Achievement in Mathematics* (PALMA; Pekrun et al., 2007b). The Study 2 sample included 111 students (two to four students randomly selected from each of 41 classrooms across seven schools) from grades 8 and 11 (53% female; M_{age} = 15.96 years, SD_{age} = 1.71).

In both studies, trait and demographic data were assessed using a standardized questionnaire at the beginning of the study, after which state self-report measures were administered. In Study 1, state *mathematics test anxiety* was assessed immediately prior to a mathematics test and twice during the test (after approximately one third and two thirds of the test had been completed). The self-report questions were integrated into the answer sheet for the test. In Study 2, state *class-related mathematics anxiety* was assessed via a digital questionnaire presented on a personal digital assistant (PDA) following a randomized audible signal. The signal occurred once during each math class over a two-week period. Students activated the PDA at the start of class and the signal occurred at randomized times over the next 40 minutes (five assessments per student, on average).

4.5.2 Study Measures

4.5.2.1 Anxiety

In Study 1, trait mathematics test anxiety was assessed using the Achievement Emotions Questionnaire – Mathematics (AEQ-M; see Pekrun et al., 2011). Participants were instructed to answer how they typically felt when taking tests in mathematics (four items; e.g., "*When taking the math test, I am tense and nervous*"; α = .83). State mathematics test anxiety was assessed with the item "*I am anxious*" (see Goetz, Preckel, Pekrun, & Hall, 2007b). The answer format for the trait and state measures was a five-point Likert scale ranging from (1) "*strongly disagree*" to (5) "*strongly agree*." In Study 2, trait and state mathematics anxiety were assessed with the following items: "*How much anxiety do you generally experience during mathematics classes*" (trait) and "*How much anxiety are you experiencing during this class?*" (state). The response format for both items was a five-point Likert scale ranging from (1) "*not at all*" to (5) "*very strongly*".

4.5.2.2 Perceived competence

Subjective perceptions of competence were operationalized as self-efficacy and self-concept beliefs (cf., Skinner, 1996) and assessed using established scales. In Study 1, trait mathematics self-efficacy was measured using a four-item scale utilized in PISA assessments (Organisation for Economic Co-operation and Development, 2003, 2004; sample item: "I am confident that I can understand even the most difficult content in mathematics"; α = .89). The response format consisted of a five-point Likert scale ranging from (1) "*almost never*" to (5) "*almost always*." In Study 2, academic self-concept was assessed using three items of the Self Description Questionnaire (SDQ; Marsh, 1990; German version, Kunter et al., 2002; sample item: "*Mathematics is one of my best subjects*"; α = .89). The response format was a 5-point Likert scale ranging from (1) "*strongly disagree*" to (5) "*strongly agree*."

4.5.2.3 Achievement

In both studies, academic performance was operationalized as students' midterm grades in mathematics that, in the German school system, are typically based on a single written exam combined with scores on course-specific oral exams. Grades range from 1 (*very good*) to 6 (*insufficient*), with higher numbers representing poorer performance. To have achievement values be interpreted more intuitively, grade values were inverted such that higher numbers indicated better performance.

4.5.3 Data Analysis

To evaluate the main study hypotheses, a multi-level, intraindividual modeling approach was adopted to account for the nested structure of the data in both studies. For each of the two studies, HLM 6.08 software (Hierarchical Linear Modeling; Raudenbush et al., 2009) was used to conduct multi-level analyses comprising three levels (measures nested within students, and students nested within classrooms).

4.5.3.1 Level 1 (measures within students)

Students' anxiety scores served as the outcome variable and included two types of measures within each person, namely one trait measure (Study 1: trait anxiety score divided by the number of items; Study 2: the score on the single trait anxiety item) and multiple state measures (Study 1: three ratings – one before and two during the test; Study 2: ESM assessments during class – four ratings per participant, on average). The *Trait/State* variable (uncentered) differentiated between the type of measure used (0 = *state*, 1 = *trait*). Due to the coding of this variable, the intercept evaluated as γ_{000} describes overall mean state anxiety when other linear terms' coding values also are zero (e.g., mean state anxiety for males, mean state anxiety for students with average self-rated competence). This variable's effect (γ_{100}) can be interpreted as the difference between trait and state anxiety scores, with positive values indicating that trait scores were higher than state scores.

4.5.3.2 Level 2 (student level)

Two Level 2 variables, as well as their interaction term, were included in our models, namely *Gender* (0 = *male*, 1 = *female*; γ_{010}, uncentered), *Competence* (Study 1: self-efficacy; Study 2: self-concept; γ_{020}, *z*-standardized across persons), and *Gender* × *Competence* (γ_{030}, multiplicative term).

4.5.3.3 Level 3 (class level)

The classes in which students were nested constituted the third level. The class level was included to take into account the clustering of students within classes when estimating standard errors.

4.5.3.4 Cross-level interactions Level 1-Level 2

Three cross-level multiplicative interaction terms were included in our models, namely *Trait/State* × *Gender* (γ_{110}), *Trait/State* × *Competence* (γ_{120}), and *Trait/State* × *Gender* × *Competence* (γ_{130}). These interaction terms represent the effects of gender, competence, and the Gender × Competence interaction on the difference between trait and state anxiety scores.

A number of different models were calculated to test the study hypotheses, with each constructed as a "slopes-as-outcome model" (Raudenbush & Bryk, 2002).[7] Model 1 examined the effect of the Trait/State × Gender interaction (γ_{110}), thus testing if gender was a predictor of the effect of the Trait/State variable. As such, Model 1 assessed whether boys and girls differed in terms of the discrepancy between their trait and state anxiety scores (Hypothesis 1). Model 2 examined the effect of the Trait/State × Competence interaction (γ_{120}), thus testing if competence was a predictor of the discrepancy between trait and state anxiety scores. In Model 3, both gender and competence were included as predictors of the trait-state discrepancy (γ_{110}, γ_{120}). As such, Model 3 tested if gender effects on the trait-state discrepancy were reduced when competence was included, and thus examined competence as a mediator of gender effects. Model 4 additionally included the three-way interaction between the Trait/State variable, gender, and competence (γ_{130}), thus testing if the effects of competence differed by gender. In all of the models, the corresponding main effects were also included (γ_{010}, γ_{020}, γ_{030}). By constructing our models in this manner, it can be inferred whether gender differences in the discrepancy between trait and state math anxiety can be explained by gender-linked differences in competence beliefs (Hypothesis 2).

[7] For a formalized description, the mixed equation for Model 4 was as follows:
$\text{Anxiety}_{ijk} = \gamma_{000} + \gamma_{010}\times\text{Gender} + \gamma_{020}\times\text{Competence} + \gamma_{030}\times\text{Gender}\times\text{Competence} + \gamma_{100}\times\text{Trait/State} + \gamma_{110}\times\text{Trait/State}\times\text{Gender} + \gamma_{120}\times\text{Trait/State}\times\text{Competence} + \gamma_{130}\times\text{Trait/State}\times\text{Gender}\times\text{Competence} + r_{0jk} + r_{1jk}\times\text{Trait/State} + u_{00k} + e_{ijk}$. The indices *i*, *j*, and *k* refer to measures, persons, and classrooms, respectively.

4.6 Results

4.6.1 Preliminary Analyses

Table 4.1 provides the results of *t*-tests for mean level difference tests as a function of gender, as well as corresponding effect sizes (Cohen's *d*; Cohen, 1988), for both studies (also see Figure 4.1). The pattern of results was as anticipated: In both studies, girls reported significantly higher trait anxiety and lower competence beliefs than boys. The size of these effects was medium to large. However, girls and boys did not significantly differ with respect to mathematics achievement or state anxiety (concerning math tests in Study 1, and math class in Study 2).[8] In Study 1, separate analyses for each of the three single state test anxiety items also revealed no significant gender differences.

Table 4.1. *Descriptive Statistics and Mean Level Differences*

		Boys		Girls			
Scales	Study	*M*	*SD*	*M*	*SD*	*t-value*	Effect size *d*
Anxiety-Trait	1	2.63	1.02	3.11	1.12	–5.40***	–0.32
	2	1.62	1.05	2.39	1.35	–3.39***	–0.45
Anxiety-State	1	1.35	0.58	1.34	0.53	0.27	0.01
	2	1.55	0.80	1.48	0.61	0.50	0.07
Competence	1	2.95	0.73	2.48	0.82	7.04***	0.43
	2	3.29	1.19	2.53	1.10	3.52**	0.47
Achievement	1	4.22	0.94	4.14	1.01	1.03	0.06
	2	3.98	1.09	3.76	0.97	1.11	0.15

Note. Positive *t*-values reflect higher scores for boys. For multi-item measures, scale values were divided by the number of items. Study 1: 316 boys, 268 girls; Study 2: 52 boys, 59 girls.
** $p < .01$. *** $p < .001$.

[8] Pearson Product-Moment correlations were used to evaluate the relations between math anxiety, perceived competence, and math achievement. Trait math anxiety correlated negatively with math achievement in Studies 1 and 2 (*r*s = –.35, $p < .01$, and –.15, *ns*, for boys, and –.42, $p < .01$ and –.27, $p < .05$, for girls, respectively). Trait math anxiety also correlated negatively with competence beliefs in Studies 1 and 2 (*r*s = –.46, $p < .01$, and –.12, *ns*, for boys, and –.51 and –.44, both *p*s < .01, for girls, respectively). Correlations between state math anxiety and math achievement were not significant. Finally, competence beliefs correlated significantly positively with math achievement in Studies 1/2 (*r*s = 43/.78 for boys and .50/78 for girls, respectively; all *p*s < .01). These links between trait math anxiety and math achievement, and between competence beliefs and math achievement, are in line with numerous previous studies (Goetz et al., 2007a; Ma, 1999; Valentine, DuBois, & Cooper, 2004), supporting the validity of our study measures.

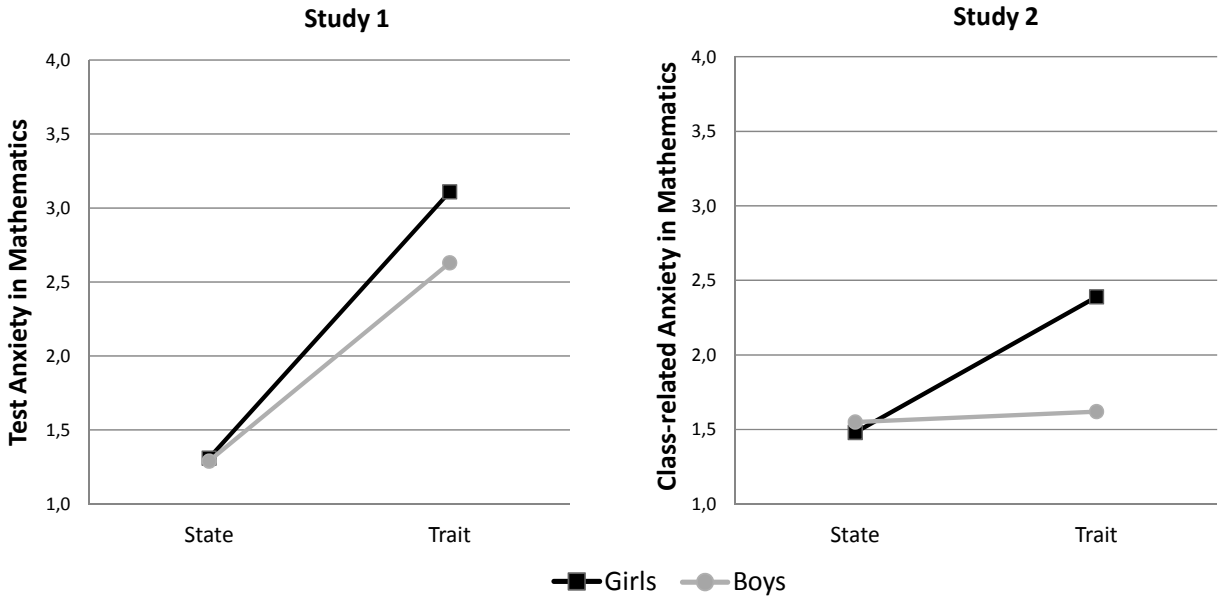


Figure 4.1. *Mean levels in trait and state anxiety by gender (Study 1: test anxiety; Study 2: class-related anxiety)*

4.6.2 Main Analyses

The results of the main analyses are outlined in Table 4.2.

4.6.2.1 Model 1

The main effect of the type of measure (Trait/State variable; γ_{100}) on the anxiety scores was significant for Study 1 but not for Study 2.[9] The main effect of gender on the anxiety scores (γ_{010}) was not significant in either study. By contrast, the effect of the Trait/State × Gender interaction (γ_{110}) was significant in both studies (Study 1: .47; Study 2: .77). This finding strongly supports Hypothesis 1 in showing that gender predicted differences between trait and state self-reports of math anxiety, with the discrepancy being significantly greater for girls than for boys.

4.6.2.2 Model 2

The effect of the Trait/State × Competence interaction (γ_{120}) was significantly negative in both studies (Study 1: –.55; Study 2: –.37). This effect showed higher competence beliefs to correspond with notably weaker trait-state differences in anxiety.

[9] Trait and state scores for anxiety can be directly compared in Study 2 due to parallel item wordings. They cannot directly be compared in Study 1 due to the use of different measures in the trait versus state assessments (multi-item scale vs. single item). In other words, in Study 1 the main effect for the Trait/State variable confounds the trait versus state framing with item wording. However, this confound does not inherently imperil this variable's interactions, as with gender, which are more central to the study's aims.

Table 4.2. *Predicting Mathematics Anxiety: Results from Multilevel Modeling*

	Study 1				Study 2			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Level 1								
Intercept (γ_{000})	1.34***	1.35***	1.34***	1.34***	1.53***	1.52***	1.59***	1.56***
	(0.05)	(0.04)	(0.05)	(0.05)	(0.10)	(0.06)	(0.10)	(0.11)
Trait/State (γ_{100})	1.29***	1.50***	1.42***	1.41***	0.10	0.51***	0.18	0.12
	(0.08)	(0.05)	(0.07)	(0.07)	(0.15)	(0.11)	(0.16)	(0.17)
Level 2								
Gender (γ_{010})	0.02		0.02	0.02	–0.02		–0.12	–0.10
	(0.05)		(0.05)	(0.05)	(0.11)		(0.12)	(0.12)
Competence (γ_{020})		–0.02	–0.02	–0.02		–0.10	–0.13	–0.13*
		(0.02)	(0.03)	(0.03)		(0.06)	(0.06)	(0.06)
Gender × Competence (γ_{030})				0.02				–0.07
				(0.03)				(0.06)
Cross-level interactions L1-L2								
Trait/State × Gender (γ_{110})	0.47***		0.18	0.18	0.77**		0.61*	0.61**
	(0.11)		(0.09)	(0.09)	(0.23)		(0.23)	(0.22)
Trait/State × Competence (γ_{120})		–0.55***	–0.53***	–0.52***		–0.37***	–0.26**	–0.28**
		(0.04)	(0.04)	(0.04)		(0.10)	(0.09)	(0.09)
Trait/State × Gender × Competence (γ_{130})				–0.04				–0.18
				(0.04)				(0.10)
Variance components								
Within-student (L1) variance (σ^2)	0.307	0.307	0.307	0.307	1.037	1.050	1.039	1.039
Intercept (L2) variance (τ_{00})	0.181	0.180	0.180	0.179	0.050	0.031	0.044	0.039
Slope (L2) variance (τ_{11})	0.737	0.491	0.483	0.483	0.151	0.134	0.090	0.060
Intercept-slope (L2) covariance (τ_{01})	–0.047	–0.055	–0.055	–0.055	0.084	0.064	0.061	0.048
Intercept (L3) variance	0.014	0.014	0.015	0.015	0.054	0.045	0.040	0.041
Explanatory power	0.068	0.379	0.389	0.389	0.485	0.546	0.695	0.797

Note. Trait/State: 0 = state, 1 = trait; Gender: 0 = male, 1 = female; Study 1: $N_{\text{Level 1}}$ = 2,336; $N_{\text{Level 2}}$ = 584; $N_{\text{Level 3}}$ = 24; Study 2: $N_{\text{Level 1}}$ = 543; $N_{\text{Level 2}}$ = 111; $N_{\text{Level 3}}$ = 41. Explanatory power refers to the proportion of slope variance explained by the L2 predictors. The slope variance of the model in which no cross-level interactions are included was τ_{11} = 0.791 for Study 1 and τ_{11} = 0.295 for Study 2.
* $p < .05$. ** $p < .01$. *** $p < .001$.

4.6.2.3 Model 3

The effect of the Trait/State × Competence interaction (γ_{120}) continued to be significant in both studies. The effect of the Trait/State × Gender interaction (γ_{110}) was significant in Study 2 but no longer significant in Study 1. As compared with Model 1, this effect of gender on the trait-state discrepancy was reduced in both studies due to inclusion of the Trait/State × Competence interaction term (Study 1: from .47 to .18; Study 2: from .77 to .61).

4.6.2.4 Model 4

Findings revealed that the effects of gender on the trait-state discrepancy (γ_{110}) and competence on the trait-state discrepancy (γ_{120}) were additive, as γ_{130} did not reach statistical significance in either study.

In Models 1 through 4, the effect of gender on the trait-state discrepancy (γ_{110}) can be interpreted as a moderator effect: Gender played a significant role in predicting how large the differences between trait and state anxiety were. The findings further suggest that this moderating effect of gender (γ_{110}) was partly mediated by perceived competence, given the reduction seen for the Trait/State × Gender coefficient (γ_{110}) when the Trait/State × Competence interaction was added (γ_{120}). The pattern here is of mediated moderation (Preacher, Rucker, & Hayes, 2007). Thus, the results support Hypothesis 2 by showing that girls' Trait/State discrepancies were associated with their lower levels of perceived competence compared with boys' levels.[10]

4.7 Discussion and Conclusion

The present findings are consistent with previous research documenting the well-known gender gap in self-reports of trait mathematics anxiety, but expand upon previous results in showing girls to report higher levels of anxiety than boys on trait self-reports but not on state-based measures. Put simply, these findings suggest that girls do not in fact experience more anxiety than boys during mathematics instruction and testing situations, despite reporting higher levels of habitual math anxiety. Moreover, the study findings also indicate that girls' competence beliefs, that are lower than those of boys despite similar achievement

[10] In supplementary analyses, we included academic achievement and grade level (Study 1: 5/6/7 vs. 8/9/10; Study 2: 8 vs. 11) as additional predictors in all four models. Controlling for achievement and grade level in this way led to a pattern of results that was equivalent to the findings of the main analyses and left the conclusions of the studies unaffected. Results of these analyses are available in Table 4.3.

outcomes, may be partly responsible for girls reporting higher levels of habitual mathematics anxiety.

The present finding that trait-oriented self-reports of anxiety are impacted by competence beliefs is in line with the accessibility model of emotional self-reports (Robinson & Clore, 2002) in which state measures are understood to evaluate individuals' emotions (i.e., actual experiences), whereas trait measures are understood to reflect individuals' *beliefs* about emotions. Competence judgments represent perhaps the most critical cognitive appraisal with respect to students' emotions, as reflected by their observed power (37.9% and 54.6% in Studies 1 and 2, respectively) for explaining the discrepancy in levels of trait versus state math anxiety. However, other mathematics-related cognitions also warrant investigation in this regard (e.g., perceived value, content difficulty, achievement expectations; Pekrun, 2006) to further elucidate the specific cognitive processes responsible for gender differences on trait self-reports of anxiety. Moreover, research on the role of gender stereotypes about mathematics as potential antecedents of the gender bias in these anxiety-arousing cognitions would also be an intriguing area for future investigation (cf., Keller, 2002; Wheeler & Petty, 2001).

When comparing levels of trait vs. state self-reports (Study 2 allows for such a comparison due to the use of parallel item wordings), our findings suggest that girls do indeed tend to overestimate their habitual mathematics anxiety, whereas boys do not. Our results also confirm that competence beliefs play an important role with respect to girls' overestimation of trait math anxiety (cf., research on the intensity bias in trait vs. state measures; Buehler & McFarland, 2001). The assertion that reflective cognitive processes may be responsible for gender differences in reports of trait math anxiety is further supported by the lack of gender differences in math achievement, suggesting that psychological constructs, over and above performance, merit attention as antecedent variables.

From a practical perspective, the effect of gender on reported, self-perceived trait math anxiety being largely due to stereotyped cognitions (as opposed to ability) is troubling given the negative impact of perceived trait anxiety on subjective well-being, motivation, and learning behavior. As self-reports of trait mathematics anxiety have also been empirically linked to decision-making processes (cf., Wirtz et al., 2003), it is possible that girls' unfounded beliefs about their math anxiety may contribute to the underrepresentation of females in math-intensive domains such as the physical sciences, technology, and engineering.

To reiterate, our findings suggest that whereas girls may report greater habitual anxiety in mathematics than do boys, they do not in fact experience greater anxiety than boys when learning about, or being tested on, math content. Our study samples consisted of students from the highest track of the German school system (Gymnasium; approximately one third of the total student cohort), a large proportion of whom are high achievers and expected to assume positions of leadership in society. Thus, even among these high achievers, a sizeable number of female students can be expected to not pursue further study or employment in math-intensive domains (Eccles, 2012), simply because of lower subjective evaluations of their math abilities and, consequently, higher levels of perceived habitual math anxiety relative to boys.

Although these findings depict a troubling scenario in which girls may opt out of entire occupational domains due to unjustified biases and perceived anxiety levels, they are also encouraging in suggesting this situation can be improved by directly addressing girls' self-defeating cognitions and emotions in mathematics. Educators could help girls facilitate their well-being and engagement in math-related domains by explicitly informing them that their achievement and anxiety in actual math classes do not significantly differ from those of the boys, despite persistent beliefs to the contrary. Similarly, cognitive interventions (e.g., Hall et al., 2007) could be used to reduce the gender gap in trait math anxiety. Such measures can be expected to have far reaching economic implications by potentially increasing returns on societal investments in STEM education and redressing the present international shortage of expertise in math-intensive fields (e.g., engineers, scientists). By encouraging girls to not shortchange their potential for success in these domains, it is anticipated that the gender gap in perceptions of math anxiety, and the detrimental consequences of girls believing they experience more anxiety than they actually do, can be substantially reduced.

4.8 Supplementary Material

The findings of the multi-level analyses controlling for student achievement and grade level are reported in Table 4.3. The gamma specifications in the table refer to Model 4, with achievement included as an additional student level variable, and grade level as an additional class level variable. The single mixed equation for this model is as follows:

$$\text{Anxiety}_{ijk} = \gamma_{000} + \gamma_{001}\times\text{Grade level} + \gamma_{010}\times\text{Gender} + \gamma_{011}\times\text{Gender}\times\text{Grade level} + \gamma_{020}\times\text{Competence} + \gamma_{021}\times\text{Competence}\times\text{Grade level} + \gamma_{030}\times\text{Gender}\times\text{Competence} + \gamma_{031}\times\text{Gender}\times\text{Competence}\times\text{Grade level} + \gamma_{040}\times\text{Achievement} + \gamma_{041}\times\text{Achievement}\times\text{Grade level} + \gamma_{100}\times\text{Trait/State} + \gamma_{101}\times\text{Trait/State}\times\text{Grade level} + \gamma_{110}\times\text{Trait/State}\times\text{Gender} + \gamma_{111}\times\text{Trait/State}\times\text{Gender}\times\text{Grade level} + \gamma_{120}\times\text{Trait/State}\times\text{Competence} + \gamma_{121}\times\text{Trait/State}\times\text{Competence}\times\text{Grade level} + \gamma_{130}\times\text{Trait/State}\times\text{Gender}\times\text{Competence} + \gamma_{131}\times\text{Trait/State}\times\text{Gender}\times\text{Competence}\times\text{Grade level} + \gamma_{140}\times\text{Trait/State}\times\text{Achievement} + r_{0jk} + r_{1jk}\times\text{Trait/State} + u_{00k} + e_{ijk}$$

The indices i, j, and k refer to measures, students, and classrooms, respectively.

The results of these additional analyses show that the findings of the main analysis are replicated when controlling for achievement and grade level: Again, for both studies, there is a significant effect for γ_{110} in Model 1, in line with the proposed effect of gender on the trait-state discrepancy (Hypothesis 1). In both studies, this effect is reduced after including γ_{120} in the model (Models 3 and 4; in line with the mediation assumption, Hypothesis 2). The effects for γ_{120} are significant across Models 2 to 4 in both studies. Thus, including achievement and grade level in the models leads to a pattern of results that leaves the study conclusions unaffected.

Table 4.3 *Predicting Mathematics Anxiety: Results from Multilevel Modeling*

	Study 1			
	Model 1	Model 2	Model 3	Model 4
Level 1				
Intercept (γ_{000})	1.44***	1.45***	1.46***	1.46***
	(.07)	(.05)	(.07)	(.07)
Trait/State (γ_{100})	1.25***	1.43***	1.32***	1.29***
	(.09)	(.06)	(.10)	(.10)
Level 2				
Gender (γ_{010})	0.00		-0.02	-0.03
	(.09)		(.09)	(.10)
Competence (γ_{020})		-0.05	-0.06	-0.05
		(.05)	(.05)	(.06)
Gender × Competence (γ_{030})				0.01
				(.05)
Achievement (γ_{040})	0.03	0.05	0.05	0.05
	(.05)	(.05)	(.05)	(.05)
Cross-level interactions L1-L2				
Trait/State × Gender (γ_{110})	0.42**		0.28*	0.30*
	(.12)		(.12)	(.12)
Trait/State × Competence (γ_{120})		–0.39***	–0.35***	–0.36***
		(.05)	(.06)	(.06)
Trait/State × Gender × Competence (γ_{130})				–.10
				(.06)
Trait/State × Achievement (γ_{140})	–0.43***	–0.24***	–0.25***	–0.24***
	(.05)	(.06)	(.06)	(.06)
Level 3				
Grade level (γ_{001})	–0.23**	–0.20**	–0.23**	–0.24**
	(.08)	(.05)	(.08)	(.08)
Cross-level interactions L1-L3				
Trait/State × Grade level (γ_{101})	0.11	0.12	0.16	0.19
	(.14)	(.08)	(.13)	(.13)
Cross-level interactions L2-L3				
Gender × Grade level (γ_{011})	0.06		0.07	0.08
	(.10)		(.11)	(.11)
Competence × Grade level (γ_{021})		0.02	0.03	0.02
		(.06)	(.06)	(.06)
Gender × Competence × Grade level (γ_{031})				0.01
				(.06)
Achievement × Grade level (γ_{041})	–0.05	–0.04	–0.04	–0.04
	(.06)	(.06)	(.06)	(.06)
Cross-level interactions L1-L2-L3				
Trait/State × Gender × Grade level (γ_{111})	0.00		–0.12	–0.13
	(.18)		(.17)	(.17)
Trait/State × Competence × Grade level (γ_{121})		–0.07	–0.08	–0.08
		(.07)	(.07)	(.07)
Trait/State × Gender × Competence × Grade level (γ_{131})				0.12
				(.08)
Variance components				
Within-student (L1) variance (σ^2)	0.307	0.307	0.307	0.307
Intercept (L2) variance (τ_{00})	0.177	0.177	0.176	0.176
Slope (L2) variance (τ_{11})	0.541	0.437	0.426	0.423
Intercept-slope (L2) covariance (τ_{01})	–0.034	–0.043	–0.043	–0.043
Intercept (L3) variance	0.009	0.006	0.007	0.008
Explanatory power	0.316	0.448	0.461	0.465

Table 4.3 continued

	Study 2			
	Model 1	Model 2	Model 3	Model 4
Level 1				
Intercept (γ_{000})	1.60***	1.58***	1.70***	1.71***
	(.11)	(.08)	(.13)	(.15)
Trait/State (γ_{100})	0.25	0.71***	0.54***	0.37
	(.19)	(.16)	(.20)	(.26)
Level 2				
Gender (γ_{010})	–0.05		–0.20	–0.21
	(.14)		(.18)	(.20)
Competence (γ_{020})		–0.10	–0.19*	–0.19*
		(.07)	(.09)	(.09)
Gender × Competence (γ_{030})				0.02
				(.08)
Achievement (γ_{040})	–0.04	0.03	0.09	0.09
	(.07)	(.07)	(.09)	(.10)
Cross-level interactions L1-L2				
Trait/State × Gender (γ_{110})	0.69*		0.29	0.42
	(.32)		(.31)	(.35)
Trait/State × Competence (γ_{120})		–0.71**	–0.59**	–0.54**
		(.19)	(.19)	(.15)
Trait/State × Gender × Competence (γ_{130})				–0.23
				(.14)
Trait/State × Achievement (γ_{140})	–0.17*	0.22	0.15	0.13
	(.08)	(.15)	(.15)	(.15)
Level 3				
Grade level (γ_{001})	–0.12	–0.13	–0.20	–0.25
	(.21)	(.13)	(.22)	(.23)
Cross-level interactions L1-L3				
Trait/State × Grade level (γ_{101})	–0.25	–0.33	–0.53	–.33
	(.32)	(.22)	(.32)	(.36)
Cross-level interactions L2-L3				
Gender × Grade level (γ_{011})	0.04		0.14	0.13
	(.24)		(.27)	(.28)
Competence × Grade level (γ_{021})		0.02	0.10	0.09
		(.20)	(.21)	(.20)
Gender × Competence × Grade level (γ_{031})				–0.15
				(.11)
Achievement × Grade level (γ_{041})	–0.09	–0.13	–0.17	–0.17
	(.09)	(.17)	(.18)	(.18)
Cross-level interactions L1-L2-L3				
Trait/State × Gender × Grade level (γ_{111})	0.06		0.41	0.22
	(.47)		(.46)	(.47)
Trait/State × Competence × Grade level (γ_{121})		0.30	0.31*	0.25
		(.18)	(.15)	(.14)
Trait/State × Gender × Competence × Grade level (γ_{131})				0.16
				(.19)
Variance components				
Within-student (L1) variance (σ^2)	1.035	1.045	1.034	1.030
Intercept (L2) variance (τ_{00})	0.045	0.035	0.041	0.034
Slope (L2) variance (τ_{11})	0.128	0.096	0.072	0.062
Intercept-slope (L2) covariance (τ_{01})	0.073	0.057	0.053	0.044
Intercept (L3) variance	0.043	0.036	0.036	0.040
Explanatory power	0.566	0.675	0.756	0.790

Note. Description of variables: Trait/State (0 = state, 1 = trait); Gender (0 = male, 1 = female); Competence and Achievement: *z*-standardized scale/score; Grade level: 0 = grade 5/6/7, 1 = grade 8/9/10 for Study 1, 0 = grade 8, 1 = grade 11 for Study 2; L1 = Level 1, L2 = Level 2, L3 = Level3. Explanatory power refers to the proportion of slope variance explained by the L2 and L3 predictors. The slope variance of the model in which no cross-level interactions are included was $\tau_{11} = 0.791$ for Study 1 and $\tau_{11} = 0.295$ for Study 2.

* $p < .05$, ** $p < .01$, *** $p < .001$.

5 General Discussion

Emotions are credited with increasing importance in the academic context. On the one hand, emotions are important antecedents of learning outcomes (e.g., the relation between anxiety and academic performance; Zeidner, 2007). On the other hand, they are also important outcomes of the learning process such that positive affect and the reduction of negative emotions such as anxiety with regard to learning is important for encouraging students, for example, with regard to life-long learning (Goetz et al., 2003) and fostering their well-being (Lipnevich & Roberts, 2012). Based on their importance, the aim of the present dissertation was to get granular on two different conceptualizations of emotions, namely trait and state emotions, in the learning and achievement context. Both conceptualizations are used frequently and may lead to different results. Furthermore, as assessing state emotions via the experience-sampling method increases in popularity (Linnenbrink-Garcia & Pekrun, 2011), it is important to compare the two conceptualizations in order to gain a better understanding of their similarities and differences. The aim of the present dissertation was to gain insight into the structural similarities and differences of trait and state emotions with regard to the relation between cognitive appraisal antecedents and academic emotions (Study 1). Another aim was to test whether the previously found mean-level differences of trait and state emotional assessments also emerge in the educational context with students' academic emotions. The results from comparing the two assessment methods of trait and state emotions provides initial evidence regarding whether findings from trait assessments can be generalized to draw conclusions about students' actual state emotions or in other words, whether trait assessments via generalized self-reports are valid for capturing students' emotions (Studies 2 and 3). The present dissertation also went one step further and attempted to identify variables that may be related to the assumed discrepancy between trait and state emotions, namely control or competence beliefs (Study 2 and Study 3) and gender (Study 3). This final chapter will summarize and discuss the main findings of the three studies. Furthermore, strengths and weaknesses will be depicted, and implications for research and practice will be provided.

5.1 Overall Summary of Main Study Findings

In Study 1, the relations between cognitive appraisal antecedents and students' emotions were investigated against the background of the control-value theory of achievement emotions (Pekrun, 2006). The relations between trait and state control and value appraisals of the emotions of pride, anxiety, and boredom, were analyzed using an intraindividual approach.

Control appraisal was found to positively predict positive emotions and negatively predict negative emotions. As expected, value appraisals were found to be positively related to pride and anxiety but negatively related to boredom. Study 1 also investigated the interactive effect of control and value on achievement emotions, which was largely neglected in past research (Nagengast et al., 2011). The interaction between control and value was found to significantly predict emotions over and above the main effects of control and value, meaning that the strength of the relation between one appraisal and the emotion is dependent on the level of the other appraisal. Furthermore, it was found that the structural relations were similar for trait and state assessments, meaning that the effect of control, value, and their interaction is similar in the actual situation (state assessment) and when investigating the relationships through the use of trait assessments. The study findings strengthen the validity of the control-value theory such that the expected relations between antecedents and emotions were found even when investigating the assumptions of the theory in multiple state *and* trait assessments per person while using an intraindividual analysis. The first study therefore highlights the necessity of assessing and analyzing data intraindividually when intraindividual functioning is to be examined (Molenaar & Campbell, 2009).

Study 2 investigated the discrepancy between mean-levels of students' trait and state emotions against the background of the accessibility model of emotional self-report (Robinson & Clore, 2002). Furthermore, according to this theory, and based on assumptions from control-value theory (Pekrun, 2006), self-concept was examined as a moderating variable in explaining the discrepancy between trait and state emotions. First, as hypothesized, a discrepancy between trait and state emotions of enjoyment, pride, anger, and anxiety was found with trait emotions generally being rated higher than state emotions. Second, the trait-state discrepancy could be explained by self-concept beliefs such that self-concept was positively related to the discrepancy in positive emotions and negatively related to the discrepancy in negative emotions. Thus, beyond finding a discrepancy between trait and state emotional assessments, results from the present study were able to account for this discrepancy. Study 2 revealed that trait and state emotions are not the same as trait emotions seem to reflect *beliefs* about emotions rather than the actual emotions themselves. Future researchers are therefore advised to make clear distinctions between trait and state emotional assessments according to the research question at hand when investigating emotions.

Study 3 was aimed at providing a more detailed examination of the highly publicized gender gap in math anxiety. As in Study 2, the accessibility model of emotional self-report

(Robinson & Clore, 2002) provided the theoretical background for the study. As hypothesized, results from the two studies showed gender differences to only emerge in trait anxiety but not in state math test and math classroom anxiety. Thus, gender is a moderator of the trait-state discrepancy. Further, competence beliefs predicted a higher trait-state discrepancy in girls as compared to boys. This finding is in some way the most thought-provoking as it challenges previous findings regarding gender differences based on trait assessments as there may be no gender differences in state assessments. Rather, gender differences emerged in previous trait-based investigations because gender stereotypic beliefs seem to have a much stronger influence on trait assessments compared to state assessment (see Porter et al., 2000; Robinson, Johnson, & Shields, 1998), and the results are once again an example of the importance of peoples' subjective beliefs when assessing trait emotions as opposed to their actual experiences.

5.2 Overall Discussion of Results

5.2.1 Similarities and Differences Between Trait and State Emotions

Generally, structural relations as well as mean-level differences can be examined when investigating different conceptualizations of a construct: structural relations may be similar but mean-levels can differ or vice versa; alternatively, all relations could be identical or all could diverge (e.g., Frenzel et al., 2007). As was shown in the present dissertation, there were no profound structural differences between trait and state academic emotions with appraisal antecedents of control, value, and their interaction forming similar relationships to the investigated emotions in trait and state data when analyzed intraindividually (Study 1). Thus, when conceptualizing emotions as habitual traits and assessing them with their appraisal antecedents from a memory-based perspective, the same relationships emerge as when conceptualizing emotions as states and assess them and their appraisals via the experience-sampling method.

At the same time, clear discrepancies in the perceived intensities of trait and state emotions (i.e., mean-level differences) were found in the present dissertation (Study 2 and Study 3). When reviewing the literature on emotional self-reports it became clear that subjective beliefs are assumed to play a vital role in answering global trait questionnaires while being less important when completing state-based assessments (Robinson & Clore, 2002). It is assumed that individuals are unable to recall actual emotional experiences so they have to rely on semantic memory when asked about their trait emotions, whereas they can

retrieve emotions from episodic memory while answering state questionnaires. This is one reason why ecological validity is assumed to be higher in real-time or state assessments (Shiffman et al., 2008).

Given that both trait assessments and state assessments attempt to capture emotions, the results of Study 2 and Study 3 in the present dissertation bring the validity of trait assessments into question. One proposition from classical test theory is that there should be no systematic bias in the data assessment (Gulliksen, 1950; Kempf, 2003). Only random error in test scores is admissible. However, in trait assessments, some form of systematic bias seems to distort the data, given that in trait emotional self-reports 'actual' emotions and not beliefs about emotions are to be assessed. Looking back on the long tradition of trait emotional assessments this seems critical. As was found herein, control beliefs and gender bias trait emotional assessments systematically. Thus, one proposition from classical test theory is not met. When doing research on academic emotions, one central aim should be to assess actual emotions rather than other constructs (e.g., subjective beliefs) if one is interested in students' actual emotional experiences. Otherwise, it should become clear that trait emotional assessments do not only assess emotions independently but incorporate other constructs such as control beliefs.

One solution to this problem may be to adjust the intentions with which trait assessments are used. Trait assessments do not appear to capture actual emotions, thus they do not seem to be suitable when researchers are interested in operationalizing actual emotions. State data are sometimes assumed to 'capturing life as it is lived' (Bolger, Davis, & Rafaeli, 2003). Perhaps trait assessments can be labeled as 'capturing life as it is believed to be lived' or, regarding emotional assessments, 'capturing emotions as they are believed to be experienced.' This should become clear when researchers attempt to investigate emotions through the use of trait reports. Thus, trait assessments seem to be an operationalization of *beliefs* about emotions rather than actual emotions (see Robinson & Clore, 2002). Other researchers have proposed additional reasons for a possible discrepancy between trait and state emotional assessments that could be taken into account in future studies. For example, in the present dissertation, peak or end effects were not investigated (Fredrickson & Kahneman, 1993; Levine et al., 2009). For example, students may focus mainly on tests when asked about their anxiety in mathematics, although compared to the number of regular lessons few tests are actually written during a school year.

The idea that memory can influence trait assessments is not new, however, it likely was not expressed explicitly enough until now. Trait assessments and state assessments are both valuable but they indeed capture different facets of a construct (Conner & Barrett, 2012). Given the importance of trait assessments with regard to future choices (Wirtz et al., 2003), it is clear that they do have predictive value. When considering the critical role that trait assessments play in relation to the assumptions of classical test theory, they seem to be best utilized when predicting future behavior, which was the original role that trait constructs were intended to fill. Although this was not explicitly investigated in the present studies, we assume that students would base their future domain and career choices on what they think they feel (Eccles, 1985) and this could only be assessed by the use of trait emotional assessments.

Generally, the question arises why people seem unable to report their trait emotions accurately (when taking actual state emotions as the reference point). Furthermore, people do not appear to learn from their errors in estimating their emotions (Meyvis, Ratner, & Levav, 2010). Is it adaptive to misremember emotions as they actually have been experienced or usually are experienced? On the one hand, one would expect that memory should not betray us as we have to rely on it when making decisions, and "memory is what we get to keep from our experience" (Kahneman & Riis, 2005, p. 286). On the other hand, it can also be advantageous to misremember emotions as this can protect motivational resources (e.g., by facilitating goal-striving following negative emotional events) and can also helps individuals handle difficult situations in which an overly optimistic view about one's emotions can buffer against potential setbacks such as failing a test (Levine et al., 2009).

5.2.2 The Importance of Subjective Control

A common theme across all three studies was the importance of subjective control with regard to emotions but also with regard to the discrepancy between trait and state emotional assessments. Control is assumed to positively relate to positive emotions and negatively relate to negative emotions[11] according to Pekrun's control-value theory of achievement emotions (Pekrun, 2006). As was shown in Study 1, state control relates to state emotions in the expected direction. Also, trait control (operationalized as self-concept or self-efficacy) and trait emotions are related. Furthermore, it was shown that, at least to some extent, trait control is able to predict state emotions (see main effect of control on state

[11] As a cautionary note it should be mentioned that this relation might be different with boredom; see the discussion section of Study 1.

emotions in Study 2). This relation can likely be explained by trait control relating to state control and state control (although not explicitly assessed in Study 2) once again relating to emotions. In sum, this implies that control appraisals are very important with regard to academic emotions[12].

The fact that trait and state control predict trait and state emotions is only one conclusion that can be drawn from the results of the studies that were conducted herein. Further, and somewhat more surprising, we found control beliefs to be a very important predictor of not only the emotions per se but also of the *discrepancy* between trait and state emotional assessments. In Study 2 control beliefs (measured as self-concept) and in Study 3 competence beliefs (operationalized as academic self-concept and self-efficacy) were able to predict the trait-state discrepancy such that the higher the control beliefs, the higher the discrepancy between trait and state emotional assessments for positive emotions. Conversely, control or competence beliefs negatively predicted the discrepancy in negative emotions. To summarize, control and control beliefs were important predictors of trait and state emotions but were also related to the extent to which individuals overestimate their trait emotions compared to their actual state emotions. The findings of the present studies contribute to ongoing research on self-concept not only as an important expected outcome and antecedent in the learning process but also as a moderator of the discrepancy between trait and state emotional assessments.

5.3 Strengths and Limitations

The strengths and limitations of this dissertation are to be discussed in the following section. More specifically, the advantages and disadvantages of the theoretical rationale behind the studies and the instruments, study design, and data analysis techniques that were used will be illuminated. Also, some considerations concerning the generalizability of the results will be presented.

5.3.1 Theoretical Rationale

The main theoretical groundings for the present dissertation were the control-value theory of achievement emotions (Pekrun, 2006) and the accessibility model of emotional self-

[12] Although the present studies focused largely on control, it should be noted that it interacts with subjective value appraisal in the prediction of academic emotions. Subjective value appraisals are also very important predictors of academic emotions. As was implied by the current findings, self-concept beliefs had lower explanatory power for enjoyment compared to all the other emotions in Study 2. As such, in the case of enjoyment for example, value appraisals may be even more predictive than control.

report (Robinson & Clore, 2002). The control-value theory provided a backdrop for the study of trait and state appraisal antecedents and their relations to academic emotions (Study 1). Both theories proved helpful in the comparison of trait and state emotions and their assessment (Study 2 and Study 3). The strengths of basing the research on the control-value theory (Pekrun, 2006) was that it was not only valuable for the identification of important antecedents of academic emotions but also provided a good starting point from which to search for possible subjective beliefs that may specifically influence students' emotional trait self-reports as proposed in Robinson and Clore's model (2002). As such, in Study 2, self-concept as an important control belief could be utilized to explain the trait-state discrepancy. Further, in Study 3, competence beliefs (operationalized as self-efficacy or self-concept) were able to partly account for gender differences in the trait-state discrepancy. Although this research was based on a sound theoretical background, more basic theories of related processes during the assessment of psychological constructs would have provided additional insight into the discrepancies that emerge when answering trait and state self-reports (e.g., Bornstein, 2011; Jobe, 2003; Schwarz, 2012). Several cognitive processing steps are necessary when participants are asked to rate their emotions, however, we did not start from the very beginning (i.e., a student has to comprehend the question which is asked, then process and interpret it, etc.). This was assumed to be a precondition but may have also contributed to a possible discrepancy between trait and state assessments and was not controlled for, for example, participants may have made different interpretations of the wording 'in general' and therefore made different estimates of their trait emotions.

5.3.2 Instruments and Study Design

An intraindividual perspective was taken in all of the studies included in the present dissertation. As it is not justifiable to draw conclusions from interindividual analyses on intraindividual functioning (Molenaar & Campbell, 2009), a long-standing gap in the literature on appraisal-emotion relations was closed with intraindividual analyses of trait and state data in Study 1.

In order to assess students' state emotions, the experience-sampling method was used in the present dissertation (Csikszentmihalyi & Larson, 1987). This rather new procedure seems suitable to capture students' emotional experiences in class (e.g., Nett et al., 2011) and has clear advantages over trait assessments including higher ecological validity and the possibility of capturing intraindividual fluctuations of constructs in daily life, such as emotions or appraisals. However, this method has the disadvantage that the state

questionnaire needs to be short in order to avoid inadvertently influencing the emotions experienced by participants as a result of having to complete the state assessment in class. As such, single items stemming from existing scales instead of the complete scales were utilized for the state assessment. If the state questionnaires had contained more items, and therefore taken longer to complete, there would have been a greater risk of unintentionally influencing the emotional state of the students (Goetz et al., 2010). Therefore, single-items to assess emotions and state appraisals were used, for which there is an issue of reliability. However, there is research showing that single-item measures can be useful under certain circumstances while maintaining a satisfying level of reliability (Gogol, Brunner, Preckel, & Götz, 2013; Wanous et al., 1997). Furthermore, in our study we decided to adopt trait and state items that used parallel wording in order to make trait and state assessments comparable, thus, trait emotions were assessed by the use of single-items as well. The trait items that were used to assess control and value, which were originally assessed using the entire scales, had very high item-scale correlations (each $r > .88$; in Study 1, for example). Nevertheless, it is one potential weakness of the studies conducted in the present dissertation.

Given that several state assessments should be representative of habitual emotional reactions, it would be of interest to have a raw estimate of how many state assessments are necessary to gain a reliable and valid insight into actual behavior. Depending on the emotion in question, this number can be quite different (see Augustine & Larsen, 2012 for a calculation with personality facets). In general, the literature is not explicit in identifying concrete numbers. According to Snijders and Bosker (2012), no definitive improvement of reliability of an aggregated variable is to be expected for more than ten measurement points for a variable with a medium intraclass correlation (see Lüdtke, Trautwein, Kunter, & Baumert, 2006 for a detailed discussion of reliability issues with aggregated measures). Assessing a small number of state measurement points may therefore be a weakness of Studies 2 and 3.

With the endeavor to compare trait and state emotions, it may be criticized that trait emotions were assessed with the wording ‘in general’ rather than trait emotional reports which referred to the same time frame for which state data was assessed (e.g., retrospective assessment or prospective assessment of emotions of the same time period in which state emotions were assessed). However, the wording is commonly used in the literature, and further, although the period of time during which state emotions were assessed was relatively short, trait assessments should be related to the state assessments regardless.

In the studies included in the present dissertation, discrete emotions were assessed. This is a clear strength of the studies as there is empirical evidence that only assessing the two dimensions of positive and negative valenced emotions or even just arousal on its own is not sufficient (Levine & Pizarro, 2004). For example, pride and enjoyment – although both positive emotions – can function quite differently (see, for example, the different explanatory power of self-concept in enjoyment and pride in Study 2). However, the differentiation of discrete emotions in the present studies was left unchecked and it was therefore the responsibility of the study participants to distinguish between their emotions according to the questions they were asked. This required a certain amount of reflection that may not be possible for every participant.

5.3.3 Statistical Methods

In all three studies, hierarchical linear regression models were used. As all data sets were structured with multiple measurement points per student (and students being nested in classes), accounting for the nested data structure seemed vital. Advantages of this analytic approach is the unbiased estimate of standard errors, accounting for different numbers of measurement points, and a sophisticated way to account for the discrepancy between trait and state emotions (slope-as-outcome models, Studies 2 and 3). Modeling the discrepancy between trait and state assessments by means of the slope-as-outcome model is preferable to former methods such as aggregating state assessments. It is also an elaborate way of testing moderators of the trait-state discrepancy rather than measuring how one variable related to trait assessments compared to state assessments, meaning the state assessment were taken into account as reference in the present studies.

Using the experience-sampling method entails having to cope with missing data to a greater degree than when analyzing trait assessments. In the present study, no special method to impute missing data was used, however, at least two state measurement points had to be available in order for a participant to be considered for data analysis. A more elaborate way of handling missing data would have likely generated more robust findings.

Although not employed in the current paper, modeling latent effects of the influences of control and value appraisals on emotions in addition to modeling their interaction on a latent basis (Klein & Moosbrugger, 2000; Trautwein et al., 2012) may have been a more elaborate way of handling the data, especially in the case of Study 1.

5.3.4 Generalizability

In order to answer the research questions, a large variety of student samples were used. Students from different age ranges (especially Study 3) and from different countries (Study 2) participated in the studies. However, the student samples that were used only consisted of students in the upper track of the German and Swiss state school system (Gymnasium; about one third of the total student population in secondary school). Future research needs to investigate whether the findings are replicable in other student populations.

In Study 1, different academic domains were analyzed as it was not assumed that the structural relations between antecedents and emotions would differ across domains. In Studies 2 and 3, however, math was the only domain in which the relations were investigated. This seems necessary because of the domain-specificity of emotions when talking about mean-levels (Goetz et al., 2007a), however, the trait-state relation might have been investigated in other domains as well. Regardless, it is not possible to draw conclusions about mean-level relations in other domains, which may be interesting with regard to generalizability issues.

Another question of generalizability would be whether the results of the comparison of trait and state emotional assessments can be generalized to constructs other than emotions. The general idea of trait assessments being biased by subjective beliefs to a greater degree than state assessments has been found for several other constructs as well, such as personality facets (Augustine & Larsen, 2012; Weber & Wiedig-Allison, 2007). For example, motivation is another construct for which a conceptualization of trait (cf. motives, goal orientation) and state is assumed. Thus, similar conclusions can likely be drawn about the relations between trait and state conceptualizations of motivation and their assessment. A similar memory influence may contribute to a different estimation of trait motivation as compared to several assessments of state motivation. Moderating variables could once again be self-concept or perhaps interest (which itself can be conceptualized in terms of trait and state; Hidi, 2000; Krapp, 2002). This is another potential avenue for future investigation.

5.4 Implications

5.4.1 Implications for Future Research

The necessity and value of intraindividual analyses seem to be underestimated in research thus far. In addition to the fact that it is critical to draw conclusions from interindividual analyses on intraindividual functioning (Molenaar & Campbell, 2009), intraindividual designs seem promising in enabling researchers to investigate functioning

within persons (e.g., appraisal-emotion relationships, see Study 1). Thus, researchers should be encouraged to use intraindividual designs and intraindividual analytical approaches.

Generally, although it can be quite labour-intensive, findings from the present studies supports the continued use of state data in emotion research when researchers are interested in actual emotions and not in the subjective *beliefs* about emotions (Robinson & Clore, 2002). The experience-sampling method provides one suitable method to assess students' actual emotional experiences. Researchers should be encouraged to use this method more frequently in order to address actual events in the school context – not only to assess the construct of emotions, but also motivation and interest, for example. With regard to the idea that several states should represent a trait, future research may identify a guideline of how many state assessments are necessary to represent a trait. Zuckerman (1976) found ratings over the time span of one week (7 assessments) to be highly correlated with 77 assessments (11 weeks), but further investigation is needed to find more robust recommendations.

Another implication for future research is to more clearly differentiate between trait and state emotional assessments. Until now, the differentiation between actual emotions and beliefs about emotions does not seem to be explicit enough. However, as has been shown in Study 2 and Study 3, there are several differences between trait and state assessments with trait assessments being more strongly influenced by subjective beliefs (see Robinson & Clore, 2002). More studies that use state assessments are necessary to further examine how valid previous finding from trait assessments are when it comes to students' actual emotions in learning and achievement settings. Interestingly, the view on traditional trait self-reports has changed from trait assessments being perceived as biased and invalid to the opinion that trait and state assessments attempt to capture different constructs (Conner & Barrett, 2012). In fact, global trait estimations do not seem to really reflect students' actual emotions, however, they do have high predictive validity regarding future choices (Wirtz et al., 2003). Thus, the research question at hand should guide the decision about which assessment method to utilize. Assessing trait and state emotions in one sample seems to be optimal as this allows stable influences and beliefs about emotions as well as transient emotional states to be assessed and therefore a complete picture about students' emotional lives may be reached. Future investigation is needed in order to complete the picture about trait and state emotions and their assessment.

The results of the present study concerning moderators of the trait-state discrepancy allow researchers to roughly estimate how far trait and state emotional assessments resemble

each other when a students' self-concept is known. Further identifying moderating variables of the trait-state discrepancy such as value beliefs, gender stereotypes, interest, etc. may contribute to our understanding of how reliable a trait assessment is for actual state assessments when additionally moderating variables are also available.

There are various operationalizations that can be employed when assessing a trait (see Table 1.1 in the General Introduction section and description in paragraph 1.2.2). The results of the present studies might have looked different had the research been based on another operationalization. The focus of the present studies was on the comparison of two different assessment methods that are used frequently (or at least becoming more popular) in educational psychology research. According to latent state-trait models (Hagemann & Meyerhoff, 2008; Steyer et al., 1999) the underlying latent trait can be extracted from states. This may be done in future studies in order to compare trait data obtained from trait questionnaires and latent trait data extracted from state measures.

An interesting and new procedure that was suggested recently with regard to the assessment of emotions or affective experiences is Affective Averaging (Comerford, 2011). The main idea of this procedure is to ask participants to recall one episode (e.g., a school lesson) as precisely as possible and answer the trait questions afterwards. Participants also have to rate how prototypical this episode was. This procedure can enable more accurate trait assessments. Although the method has only been tested thus far with participants rating their commuting experience, it may also be possible to transfer it to emotional assessments in the academic context. This may be helpful to focus attention on actual affective components of the classroom situation and less so on subjective beliefs. For example, students could be asked to remember a previous lesson from a specific subject and then rate how prototypical it was. This may be more economic than the effortful gathering of state assessments while better reflecting actual emotions compared to previous trait self-reports. Future studies will need to determine whether Affective Averaging is a promising method that can be applied in research on academic emotions.

A completely different way of explaining the discrepancy between trait and state emotional assessments focuses on the process of how state questions are answered. In a recently published study it was found that state intensity ratings were diminished if there was cognitive load during state assessments (Kron, Schul, Cohen, & Hassin, 2010). The authors argue that this resulted in lower intensities in state ratings and therefore discrepancies between trait and state ratings emerged. Although this explanation may not account for the systematic

bias between trait and state emotional ratings, it is an approach that seems worth pursuing in future studies.

5.4.2 Implications for Practice

As was found in Study 1, control and value appraisal antecedents and their interaction were related to students' trait and state academic emotions. There are many suggestions for how to influence students' control and value appraisals, and this should contribute to shaping a positive learning environment that is conducive to learning. Appraisal antecedents can be influenced by transparency, autonomy support, or feedback on achievement to name just a few (Pekrun, 2006). Control and value appraisals are of importance in concrete situations (i.e., states) as are the beliefs about control and value (i.e., traits). Thus, teaching practices that influence state appraisals as well as programs to influence long-term subjective beliefs (e.g., self-concept) can be valuable in the classroom (Dresel & Haugwitz, 2008; Dresel & Ziegler, 2006; Hall, Perry, Chipperfield, Clifton, & Haynes, 2006). As one aim of the school system is to not only motivate students to learn for the test but to also encourage them to learn their whole lives, it is helpful for teachers to have background knowledge about the importance of students' academic emotions and possibilities for influencing them.

Our study revealed that students with high self-concepts tended to overestimate their positive emotions. Literature on academic self-concept suggests that unjustifiable high self-concept is not conducive to academic achievement (Blanton, Buunk, Gibbons, & Kuyper, 1999). Hence, helping students to gather realistic insight about their self-concept (e.g., by giving them realistic feedback about their abilities) is important. Furthermore, as emotions also contribute to domain and career choices, it is especially critical to have realistic estimates of emotions, and acquiring a realistic self-concept may serve as a reasonable starting point. However, it seems to be a double-edged sword to 'disenchant' students with high self-concepts and be insistent that they should not be overly optimistic about their abilities. As there have been positive effects found for 'biased' memories of emotions, with memory of positive emotions facilitating goal striving (Lench & Levine, 2010; Levine et al., 2009), having a rosy view of one's self-concept, and subsequently of trait emotions, may be adaptive and helpful with regard to academic achievement and for pursuing a career in a specific domain.

In the same way that positive emotions contribute to pursuing a career, negative emotions may contribute to students refraining from career aspirations in specific fields in

which they experience negative emotions (Eccles, 1985). However, as it is not possible to directly retrieve emotional experiences, students have to rely on memories about emotions that are biased by subjective beliefs and do not necessarily reflect actual emotional experiences. Therefore, one very important point seems to be to encourage teachers and students alike to be aware of the importance of subjective beliefs with regard to their attitudes, emotional beliefs, and future career choices (Wirtz et al., 2003). This may be done by simply pointing out to students and teachers that a discrepancy exists between what students (and perhaps also teachers) think they feel and what they really feel. This information is not only necessary for female students in mathematics as discussed in Study 2 but for all students who have lower self-concepts (as is the case with the majority of girls in the domain of mathematics). At the moment, empirical evidence is lacking concerning whether simply pointing out this discrepancy to students with lower self-concepts is sufficient to induce change with regard to their beliefs about emotions and subsequent pursuance of careers in a certain domain. At the very least it appears to be a worthwhile pursuit and when paired with programs to enhance subjective control, it may prove helpful in decreasing student attrition from fields in which they are desperately needed.

Generally, it is important to encourage students to be mindful about their emotions. As emotions and the beliefs about emotions seem to differ, it can be helpful to be explicitly mindful of one's emotions, as is suggested in psychotherapy (Hofmann, Sawyer, Witt, & Oh, 2010). Although not reported herein, preliminary analyses suggested that after the experience-sampling procedure, students provided lower ratings of their trait emotions compared to their trait ratings prior to the experience-sampling intervention. Unfortunately, as no control group was available we could not control for possible confounding influences but it appears as if a simple prompt to report about one's emotions can lead to changes in rated intensities of trait emotions and therefore diminish the trait-state discrepancy. Thus, explicitly using mindfulness-based approaches may help students, teachers, and researchers to gain further insight into affective lives of students.

5.5 Conclusion

Beginning with the importance of emotions in the academic context and their different conceptualizations as traits and states, the present dissertation contributes to current developments in the field of psychology – especially educational psychology – by endeavoring to systematically compare trait and state emotions and their assessment.

Trait and state emotions were found to be similar with regard to relations between appraisal antecedents and emotions, but a clear discrepancy with regard to the rated intensities of trait and state emotions was found. As a conclusion, the present dissertation highlights the fact that the conceptualization and assessment of emotions in the educational context is of vast importance. It will hopefully encourage researchers to clearly distinguish between trait and state emotions and their assessments in future studies, and choose appropriate methodological approaches according to the research question under investigation. The present results confirm that trait and state conceptualizations are different and that the assessment method utilized is highly important with regard to the conclusions one wants to make. Although it seems trivial at first, the differentiation of trait and state conceptualizations of emotions and their assessments has not been explicit enough in the past.

Furthermore, the results of the present dissertation emphasize that emotions may be open to influence. Appraisal-emotion relations suggest that state emotions are malleable, but altering trait emotions is also possible, for example, by influencing students' control beliefs. Evaluations of one's self, such as the academic self-concept, have far-reaching effects on the ways we evaluate our emotions but also on the extent to which students remember or forecast their emotions (trait-state discrepancy). In summary, researchers and practitioners must pay attention to 'life as it is lived' (Bolger et al., 2003) but also to 'life as it is believed to be lived'. As "it is a basic fact of the human condition that memories are what we get to keep from our experience […]" (Kahneman & Riis, 2005 p. 286), both trait and state emotions matter.

6 References

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, 1-39. doi: 10.1177/0149206313478188

Ahmed, W., Minnaert, A., van der Werf, G., & Kuyper, H. (2010a). The Role of Competence and Value Beliefs in Students' Daily Emotional Experiences: A Multilevel Test of a Transactional Model. *Learning and Individual Differences, 20*, 507-511.

Ahmed, W., van der Werf, G., Minnaert, A., & Kuyper, H. (2010b). Students' daily emotions in the classroom: Intra-individual variability and appraisal correlates. *British Journal of Educational Psychology, 80*, 583-597.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks: Sage Publications.

Allen, J., Robbins, S. B., & Sawyer, R. (2010). Can measuring psychosocial factors promote college success? *Applied Measurement in Education, 23*, 1-22.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*, i-171. doi: 10.1037/h0093360

Amelang, M., Bartussek, D., Stemmler, G., & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6 ed.). Stuttgart: Kohlhammer.

Augustine, A. A., & Larsen, R. J. (2012). Is a trait really the mean of states? Similarities and differences between traditional and aggregate assessments of personality. *Journal of Individual Differences, 33*, 131-137. doi: 10.1027/1614-0001/a000083

Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin, 23*, 1100-1110. doi: 10.1177/01461672972310010

Ben-Zeev, D., McHugo, G. J., Xie, H., Dobbins, K., & Young, M. A. (2012). Comparing retrospective reports to real-time/real-place mobile assessment in individuals with schizophrenia and a nonclinical comparison group. *Schizophrenia Bulletin, 38*, 396-404. doi: 10.1093/schbul/sbr171

Blanton, H., Buunk, B. P., Gibbons, F. X., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology, 76*, 420-430. doi: 10.1037/0022-3514.76.3.420

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

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment, 23*, 532-544. doi: 10.1037/a0022402

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

Cattell, R. B., & Scheier, I. H. (1961). *The meaning and measurement of neuroticism and anxiety*. Oxford England: Ronald.

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

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Erlbaum.

Comerford, D. A. (2011). Attenuating focalism in affective forecasts of the commuting experience: Implications for economic decisions and policy making. *Journal of Economic Psychology, 32*, 691-699. doi: 10.1016/j.joep.2011.06.005

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

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease, 175*, 526-536. doi: 10.1097/00005053-198709000-00004

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist, 55*, 34-43. doi: 10.1037/0003-066x.55.1.34

Dresel, M., & Haugwitz, M. (2008). A computer-based approach to fostering motivation and self-regulated learning. *Journal of Experimental Education, 77*, 3-18. doi: 10.3200/jexe.77.1.3-20

Dresel, M., & Ziegler, A. (2006). Langfristige Förderung von Fähigkeitsselbstkonzept und impliziter Fähigkeitstheorie durch computerbasiertes attributionales Feedback. *Zeitschrift für Pädagogische Psychologie / German Journal of Educational Psychology, 20*, 49-63. doi: 10.1024/1010-0652.20.12.49

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405-432. doi: 10.1111/j.1467-8624.2010.01564.x

Eccles, J. S. (1985). Why doesn't Jane run? Sex differences in educational and occupational patterns. In F. D. Horowitz & M. O'Brien (Eds.), *The gifted and talented: Developmental perspectives.* (pp. 251-295). Washington, DC US: American Psychological Association.

Eccles, J. S. (2012, April). *Expectancy-value theory and gendered academic and occupational choices*. Paper presented at the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada.

Efklides, A., & Volet, S. (2005). Emotional experiences during learning: Multiple, situated and dynamic. *Learning and Instruction, 15*, 377-380. doi: 10.1016/j.learninstruc.2005.07.006

Eid, M., Schneider, C., & Schwenkmezger, P. (1999). Do you feel better or worse? The validity of perceived deviations of mood states from mood traits. *European Journal of Personality, 13*, 283-306. doi: 10.1002/(sici)1099-0984(199907/08)13:4<283::aid-per341>3.0.co;2-0

Ellis, A. (1999). Early theories and practices of rational emotive behavior therapy and how they have been augmented and revised during the last three decades. *Journal of Rational-Emotive & Cognitive Behavior Therapy, 17*, 69-93. doi: 10.1023/a:1023048830350

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*, 103-127. doi: 10.1037/a0018053

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138. doi: 10.1037/1082-989x.12.2.121

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51*, 360-392. doi: 10.1111/j.1467-6494.1983.tb00338.x

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*, 1011-1027. doi: 10.1037/0022-3514.80.6.1011

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65*, 45-55. doi: 10.1037/0022-3514.65.1.45

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics--A 'hopeless' issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 22*, 497-514. doi: 10.1007/bf03173468

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science, 317*, 1351-1354. doi: 10.1126/science.1144161

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

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology, 33*, 9-33.

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007a). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology, 99*, 715-733. doi: 10.1037/0022-0663.99.4.715

Goetz, T., Frenzel, A. C., Stoeger, H., & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion, 34*, 49-62. doi: 10.1007/s11031-009-9152-2

Goetz, T., & Hall, N. C. (2013). Emotion and achievement in the classroom. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 192-195). New York: Routledge.

Goetz, T., Pekrun, R., Hall, N. C., & Haag, L. (2006). Academic emotions from a social-cognitive perspective: Antecedents and domain specificity of students' affect in the context of Latin instruction. *British Journal of Educational Psychology, 76*, 289-308.

Goetz, T., Preckel, F., Pekrun, R., & Hall, N. C. (2007b). Emotional experiences during test taking: Does cognitive ability make a difference? *Learning and Individual Differences, 17*, 3-16.

Goetz, T., Zirngibl, A., Pekrun, R., & Hall, N. C. (2003). Emotions, learning and achievement from an educational-psychological perspective. In P. Mayring & C. v. Rhoeneck (Eds.), *Learning emotions. The influence of affective factors on classroom learning* (pp. 9–28). Frankfurt am Main: Peter Lang.

Gogol, K., Brunner, M., Preckel, F., & Götz, T. (2013). *"Mein Fragebogen ist zu lang!" Messung von affektiv-motivationalen Konstrukten mit drei Items und Einzelitems ["My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures]*. Paper presented at the 1st meeting of young academics of the Gesellschaft für Empirische Bildungsforschung (GEBF), Kiel.

Gross, J. J., & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3*, 8-16. doi: 10.1177/1754073910380974

Guindon, G. E., & Boyle, M. H. (2012). Using anchoring vignettes to assess the comparability of self-rated feelings of sadness, lowness or depression in France and Vietnam. *International Journal of Methods in Psychiatric Research, 21*, 29-40. doi: 10.1002/mpr.1346

Gulliksen, H. (1950). *Theory of mental tests*. Hoboken, NJ US: John Wiley & Sons Inc.

Hagemann, D., & Meyerhoff, D. (2008). A simplified estimation of latent state-trait parameters. *Structural Equation Modeling, 15*, 627-650. doi: 10.1080/10705510802339049

Hall, N. C., Chipperfield, J. G., Perry, R. P., Ruthig, J. C., & Goetz, T. (2006). Primary and secondary control in academic development: Gender-specific implications for stress and health in college students. *Anxiety, Stress & Coping: An International Journal, 19*, 189-210. doi: 10.1080/10615800600581168

Hall, N. C., Perry, R. P., Chipperfield, J. G., Clifton, R. A., & Haynes, T. L. (2006). Enhancing primary and secondary control in achievement settings through writing-based attributional retraining. *Journal of Social and Clinical Psychology, 25*, 361-391. doi: 10.1521/jscp.2006.25.4.361

Hall, N. C., Perry, R. P., Goetz, T., Ruthig, J. C., Stupnisky, R. H., & Newall, N. E. (2007). Attributional retraining and elaborative learning: Improving academic development through writing-based interventions. *Learning and Individual Differences, 17*, 280-290. doi: 10.1016/j.lindif.2007.04.002

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*, 1-51.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage Publications.

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the trait-state distinction for the structural modeling of developmental change. *Child Development, 58*, 93-109. doi: 10.2307/1130294

Hidi, S. (2000). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance.* (pp. 309-339). San Diego, CA US: Academic Press.

Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 169-183. doi: 10.1037/a0018555

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.

Hsee, C. K., & Hastie, R. (2006). Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Sciences, 10*, 31-37. doi: 10.1016/j.tics.2005.11.007

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14*, 299-324. doi: 10.1111/j.1471-6402.1990.tb00022.x

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*, 494-495.

Jobe, J. B. (2003). Cognitive psychology and self-reports: Models and methods. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 12*, 219-227. doi: 10.1023/a:1023279029852

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY US: Farrar, Straus and Giroux.

Kahneman, D., & Riis, J. (Eds.). (2005). *Living, and thinking about it: Two perspectives on life*. Oxford: Oxford University Press.

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47*, 193-198.

Kempf, W. (2003). *Forschungsmethoden der Psychologie: Zwischen naturwissenschaftlichem Experiment und sozialwissenschaftlicher Hermeneutik*. Berlin: Verlag Irena Regener.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*, 457-474. doi: 10.1007/bf02296338

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12*, 383-409. doi: 10.1016/s0959-4752(01)00011-1

Kron, A., Schul, Y., Cohen, A., & Hassin, R. R. (2010). Feelings don't come easy: Studies on the effortful nature of feelings. *Journal of Experimental Psychology: General, 139*, 520-534. doi: 10.1037/a0020008

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., & Neubrand, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente [PISA 2000: Documentation of scales]*. Berlin: Heenemann GmbH & Co.

Larson, R. W., & Richards, M. H. (1991). Boredom in the middle school years: Blaming schools versus blaming students. *American Journal of Education, 99*, 418-443. doi: 10.1086/443992

Lench, H. C., & Levine, L. J. (2010). Motivational biases in memory for emotions. *Cognition and Emotion, 24*, 401-418. doi: 10.1080/02699930802650788

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

Levine, L. J., & Pizarro, D. A. (2004). Emotion and memory research: A grumpy overview. *Social Cognition, 22*, 530-554. doi: 10.1521/soco.22.5.530.50767

Linnenbrink-Garcia, L., & Pekrun, R. (2011). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology, 36*, 1-3. doi: DOI: 10.1016/j.cedpsych.2010.11.004

Linnenbrink, E. (2006). Emotion Research in Education: Theoretical and Methodological Perspectives on the Integration of Affect, Motivation, and Cognition. [Article]. *Educational Psychology Review, 18*, 307-314. doi: 10.1007/s10648-006-9028-x

Lipnevich, A. A., & Roberts, R. D. (2012). Noncognitive skills in education: Emerging research and applications in a variety of international contexts. *Learning and Individual Differences, 22*, 173-177. doi: 10.1016/j.lindif.2011.11.016

Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Reliability and agreement of student ratings in the classroom environment: A reanalysis of TIMSS data. *Learning Environment Research, 9*, 215-230. doi: 10.1007/s10984-006-9014-8

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for research in mathematics education*, 520-540.

MacCann, C., Lipnevich, A. A., & Roberts, R. D. (2012). New directions in assessing emotional competencies from kindergarten to college. *Journal of Psychoeducational Assessment, 30*, 315-319. doi: 10.1177/0734282912449438

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*, 129-149.

Marsh, H. W. (1990). *Self-Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph.* Macarthur, Australia: University of Western Sydney.

Matthews, G., Deary, I. J., & Whiteman, M. C. (2009). *Personality Traits* (3 ed.). Cambridge: University Press.

McCrae, R. R., & Costa, P. T. (1995). Trait explanations in personality psychology. *European Journal of Personality, 9*, 231-252. doi: 10.1002/per.2410090402

McFarland, C., Ross, M., & DeCourville, N. (1989). Women's theories of menstruation and biases in recall of menstrual symptoms. *Journal of Personality and Social Psychology, 57*, 522-531. doi: 10.1037/0022-3514.57.3.522

Meyvis, T., Ratner, R. K., & Levav, J. (2010). Why don't we learn to accurately forecast feelings? How misremembering our predictions blinds us to past forecasting errors. *Journal of Experimental Psychology: General, 139*, 579-589. doi: 10.1037/a0020285

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

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

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267-316.

Naemi, B., Gonzalez, E., Bertling, J., Betancourt, A., Burrus, J., Kyllonen, P., et al. (2013). Large-scale group score assessments: Past, present, and future. In D. H. Saklofske, C. R. Reynolds & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment*. Cambridge, MA: Oxford University Press.

Nagengast, B., & Marsh, H. W. (2012). Big Fish in Little Ponds Aspire More: Mediation and Cross-Cultural Generalizability of School-Average Ability Effects on Self-Concept

and Career Aspirations in Science. *Journal of Educational Psychology*. doi: 10.1037/a0027697; 10.1037/a0027697.supp (Supplemental)

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the "×" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science, 22*, 1058-1066. doi: 10.1177/0956797611415540

National Academy of Sciences. (2006). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. Washington, DC: National Academies Press.

Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with Boredom in School: An Experience Sampling Perspective. *Contemporary Educational Psychology, 36*, 49-59.

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do Self-Concept Interventions Make a Difference? A Synergistic Blend of Construct Validation and Meta-Analysis. *Educational Psychologist, 41*, 181-206. doi: 10.1207/s15326985ep4103_4

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2007). Students' emotions: A key component of self-regulated learning? In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education.* (pp. 185-204). San Diego, CA US: Elsevier Academic Press.

Organisation for Economic Co-operation and Development. (2003). *The PISA 2003 assessment framework*. Paris, France: Author.

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris, France: Author.

Partnership for 21st Century Skills. (2006). Results that matter: 21st century skills and high school reform. Retrieved from http://www.p21.org/storage/documents/RTM2006.pdf

Pekrun, R. (2000). A social-cognitive, control-value theory of achievement emotions. In J. Heckhausen (Ed.), *Motivational psychology of human development: Developing motivation and motivating development.* (pp. 143-163). New York, NY US: Elsevier Science.

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

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007a). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A.

Schutz & R. Pekrun (Eds.), *Emotion in education.* (pp. 13-36). San Diego, CA US: Elsevier Academic Press.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102*, 531-549. doi: 10.1037/a0019243

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*, 36-48. doi: DOI: 10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the test emotions questionnaire (TEQ). *Anxiety, Stress & Coping: An International Journal, 17*, 287-316. doi: 10.1080/10615800412331303847

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*, 91-105.

Pekrun, R., & Stephens, E. J. (2010). Achievement emotions: A control-value approach. *Social and Personality Psychology Compass, 4*, 238-255. doi: 10.1111/j.1751-9004.2010.00259.x

Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A. C., Götz, T., & Wartha, S. (2007b). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report of the DFG Priority Programme.* (pp. 17-37). Münster, Germany: Waxmann.

Peyton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., et al. (2008). *The positive impact of social and emotional learning for kindergarten to eighth grade students: Findings from three scientific reviews*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Porter, L. S., Marco, C. A., Schwartz, J. E., Neale, J. M., Shiffman, S., & Stone, A. A. (2000). Gender differences in coping: A comparison of trait and momentary assessments. *Journal of Social and Clinical Psychology, 19*, 480-498.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185-227.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2 ed.). Thousand Oaks: Sage.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2009). *HLM 6.08. Hierarchical Linear and Nonlinear Modeling*. Lincolnwood: Scientific Software International.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*, 261-288.

Robbins, S. B., Oh, I.-S., Le, H., & Button, C. (2009). Intervention effects on college performance and retention as mediated by motivational, emotional, and social control factors: Integrated meta-analytic path analyses. *Journal of Applied Psychology, 94*, 1163-1184. doi: 10.1037/a0015738

Robinson, M. D., & Barrett, L. F. (2010). Belief and feeling in self-reports of emotion: Evidence for semantic infusion based on self-esteem. *Self and Identity, 9*, 87-111. doi: 10.1080/15298860902728274

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

Robinson, M. D., Johnson, J. T., & Shields, S. A. (1998). The gender heuristic and the database: Factors affecting the perception of gender-related differences in the experience and display of emotions. *Basic and Applied Social Psychology, 20*, 206-219. doi: 10.1207/s15324834basp2003_3

Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review, 15*, 351-357. doi: 10.2307/2087176

Safer, M. A., Levine, L. J., & Drapalski, A. L. (2002). Distortion in memory for emotions: The contributions of personality and post-event knowledge. *Personality and Social Psychology Bulletin, 28*, 1495-1507. doi: 10.1177/014616702237577

Schutz, P. A., & Lanehart, S. L. (2002). Introduction: Emotions in Education. *Educational Psychologist, 37*, 67-68.

Schutz, P. A., & Pekrun, R. (Eds.). (2007). *Emotion in education*. San Diego: Academic Press.

Schwarz, N. (2012). Why researchers should think 'real-time': A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life.* (pp. 22-42). New York, NY US: Guilford Press.

Schwenkmezger, P. (1985). *Modelle der Eingeschafts- und Zustandsangst*. Göttingen: Hogrefe.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4*, 5-34. doi: 10.1023/a:1023605205115

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1-32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Sibley, C. G. (2008). Utilities for examinig interactions in multiple regression [comupter software]. University of Auckland.

Skinner, E. A. (1996). A guide to constructs of control. *Journal of personality and social psychology, 71*, 549-570.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analyisis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage Publications Ltd.

Spielberger, C., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology, 69*, 797.

Steptoe, A., Gibson, E. L., Hamer, M., & Wardle, J. (2007). Neuroendocrine and cardiovascular correlates of positive affect measured by ecological momentary assessment and by questionnaire. *Psychoneuroendocrinology, 32*, 56-64. doi: 10.1016/j.psyneuen.2006.10.001

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8*, 79-98.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality, 13*, 389-408. doi: 10.1002/(sici)1099-0984(199909/10)13:5<389::aid-per361>3.0.co;2-a

Strenze, T. (2007). Intelligence and Socioeconomic Success: A Meta-Analytic Review of Longitudinal Research. *Intelligence, 35*, 401-426.

Titz, W. (2001). *Emotionen von Studierenden in Lernsituationen: Explorative Analysen und Entwicklung von Selbstberichtsskalen*. München: Waxmann.

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology, 104*, 763-777. doi: 10.1037/a0027470

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*, 111-133.

Valsiner, J. (1986). Between groups and individuals: Psycholgists' and laypersons' interpretations of correlational findings. In J. Valsiner (Ed.), *The individual subject and scientific psychology* (pp. 113-151). New York: Plenum Press.

van den Brink, M., Bandell-Hoekstra, E. N. G., & Huijer Abu-Saad, H. (2001). The occurrence of recall bias in pediatric headache: A comparison of questionnaire and diary data. *Headache: The Journal of Head and Face Pain, 41*, 11-20. doi: 10.1046/j.1526-4610.2001.111006011.x

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82*, 247-252. doi: 10.1037/0021-9010.82.2.247

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

Weber, H., & Wiedig-Allison, M. (2007). Sex differences in anger-related behaviour: Comparing expectancies to actual behaviour. *Cognition and Emotion, 21*, 1669-1698. doi: 10.1080/02699930701202111

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*, 548-573. doi: 10.1037/0033-295x.92.4.548

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797-826. doi: 10.1037/0033-2909.127.6.797

Wigfield, A., Battle, A., Keller, L. B., & Eccles, J. S. (2002). Sex differences in motivation, self-concept, career aspiration, and career choice: Implications for cognitive development. In A. McGillicuddy-De Lisi & R. de Lisi (Eds.), *Biology, society, and behaviour. The development of sex differences in cognition* (pp. 93-124). Westport, CT: Ablex.

Wiggins, J. S., & Pincus, A. L. (1992). Personality: Structure and assessment. *Annual Review of Psychology, 43*, 473-504. doi: 10.1146/annurev.ps.43.020192.002353

Wilson, T. D., & Gilbert, D. T. (2005). Affective Forecasting: Knowing What to Want. *Current Directions in Psychological Science, 14*, 131-134. doi: 10.1111/j.0963-7214.2005.00355.x

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78*, 821-836. doi: 10.1037/0022-3514.78.5.821

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

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York, NY US: Plenum Press.

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 165-183). Burlington, MA: Elsevier Academic Press.

Zuckerman, M. (1960). The development of an affect adjective check list for the measurement of anxiety. *Journal of Consulting Psychology, 24*, 457-462. doi: 10.1037/h0042713

Zuckerman, M. (1976). General and situation-specific traits and states: New approaches to assessment of anxiety and other constructs. In M. Zuckerman & C. D. Spielberger (Eds.), *Emotions and anxiety: New concepts, methods, and applications.* (pp. 133-174). Oxford England: Lawrence Erlbaum.

Zuckerman, M., Persky, H., & Link, K. (1967). Relation of mood and hypnotizability: An illustration of the importance of the state versus trait distinction. *Journal of Consulting Psychology, 31*, 464-470. doi: 10.1037/h0024966

7 Index of Figures

8 Index of Tables