One Measure Can't Capture It All –

Comparing Different Assessment Methods of Anxiety in Learning and Achievement Situations

Dissertation

zur Erlangung des akademischen Grades

des Doktors der Naturwissenschaften

(Dr.rer.nat.)

vorgelegt von

Anna-Lena Roos

an der

Universität Konstanz

Mathematisch-Naturwissenschaftliche Sektion

Fachbereich Psychologie

Tag der mündlichen Prüfung: 15.07.2019

1. Referent: Prof. Dr. Thomas Götz

2. Referentin: Prof. Dr. Anne Frenzel

Danksagung

Somewhere,
something incredible
is waiting
to be known.

Carl Sagan

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich auf meinem Weg während meines Dissertationsprojektes an der Universität Konstanz begleitet und unterstützt haben.

Besonders gilt mein Dank Professor Dr. Thomas Götz, für die stets sehr angenehme, wertschätzende und vor allem humorvolle Zusammenarbeit, die vielen wertvollen Rückmeldungen, die Ermöglichung zahlreicher inspirierender Forschungs- und Konferenzreisen, sowie für die Ermutigungen mit dem Einsatz von physiologischen Maßen etwas Neues zu wagen und für die damit verbundenen Lern- und Weiterentwicklungsmöglichkeiten. Außerdem danke ich meiner Bürokollegin Maike, mit der ich diesen Weg gemeinsam gehen durfte und die mir einfach immer zur Seite stand.

Des Weiteren gilt mein Dank all meinen Kolleginnen und Kollegen der Arbeitsgruppe Empirische Bildungsforschung, den Hilfskräften, meinen Koautorinnen und Koautoren, sowie meinen nationalen und internationalen Kooperationspartnerinnen und Kooperationspartnern, die ich durch dieses Projekt kennenlernen durfte und durch die gemeinsame Zusammenarbeit sehr schätze. Besonders danken möchte ich in diesem Rahmen Professorin Dr. Anne Frenzel, für ihr bereicherndes Feedback zum ersten Artikel dieser Arbeit, ihren Enthusiasmus und die Bereitschaft meine Dissertation zu begutachten.

Vor allem aber danke ich meiner Familie und meinen Freunden, für ihre unglaubliche Unterstützung in allen Phasen dieses Projektes. Merci Regina, Klaus, Johannes, Christina, Timo, Carola, Julie und Vanessa – ohne euch gäbe es diese Arbeit nicht!

Veröffentlichungen der Dissertation

Teilergebnisse dieser Dissertation wurden bereits in folgenden Beiträgen vorgestellt:

Zeitschriftenartikel

Roos, A.-L., Bieg, M., Goetz, T., Frenzel, A., Taxer, J., & Zeidner, M. (2015). Experiencing more mathematics anxiety than expected? Contrasting trait and state anxiety in high achieving students. *High Ability Studies, 26*(02), 245-258. doi: 10.1080/13598139.2015.1095078

Roos, A.-L., Goetz, T., Krannich, M., Jarrell, A., Donker, M., & Mainhard, T. (submitted). Test anxiety components: An intra-individual approach testing their control antecedents and effects on performance. *Manuscript submitted for publication.*

Roos, A.-L., Goetz, T., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (submitted). Test anxiety and physiological arousal: A systematic review. *Manuscript submitted for publication.*

Konferenzbeiträge

Roos, A.-L., Goetz, T., Krannich, M., & Jarrell, A. (2019, April). *Antecedents and effects of test anxiety components: An intra-individual real-time approach including physiological measures.* Paper presented at the annual meeting of the American Educational Research Association (AERA 2019), Toronto, Canada.

Roos, A.-L., Goetz, T., Krannich, M., Gogol, K., & Jarrell, A. (2018, April). *Optimizing emotion regulation: Zooming in on the antecedents and effects of anxiety components.* Paper presented at the annual meeting of the American Educational Research Association (AERA 2018), New York, NY, USA.

Roos, A.-L., Bieg, M., Goetz, T., Frenzel, A., Taxer, J., & Zeidner, M. (2017, September). *Experiencing more anxiety than expected? Trait and state mathematics anxiety in high achieving students.* Paper presented at the conference of the European Society for Cognitive Psychology (ESCoP), Potsdam, Germany.

Roos, A.-L., Goetz, T., Bieg, M., Krannich, M., & Jarrell, A. (2017, September). *Test anxiety and physiological arousal: A systematic review*. Paper presented at the conference of the European Association for Research on Learning and Instruction (EARLI), Tampere, Finland.

Roos, A.-L., Goetz, T., Bieg, M., Krannich, M., Donker, M., & Taxer, J. (2017, April). *Test anxiety and physiological arousal: A systematic review*. Paper presented at the annual meeting of the American Educational Research Association (AERA), San Antonio, TX, USA.

Roos, A.-L., Bieg, M., Goetz, T., Frenzel, A., Taxer, J., & Zeidner, M. (2016, April). *Experiencing more mathematics anxiety than expected? – Contrasting trait and state anxiety in high achieving students*. Paper presented at the annual meeting of the American Educational Research Association (AERA), Washington, DC, USA.

Roos, A.-L., Bieg, M., Goetz, T., Frenzel, A., Taxer, J., & Zeidner, M. (2015, September). *Mehr Angst als gedacht? Trait- und State-Mathematikangst bei hochleistenden Schülerinnen und Schülern.* Posterbeitrag bei der Fachgruppentagung Pädagogische Psychologie, Kassel.

Eigenabgrenzung

Die der Arbeit zugrundeliegenden Manuskripte (siehe Veröffentlichungen der Dissertation) sind unter der Mithilfe von Koautorinnen und Koautoren hervorgegangen. Für ihre darin investierte Arbeit, ihre Ideen und ihr Wissen möchte ich mich an dieser Stelle noch einmal ganz herzlich bedanken. Alle drei Forschungsartikel sind unter meiner Erstautorenschaft entstanden. Im Folgenden möchte ich genauer darlegen, welche Beiträge ich zu den jeweiligen Artikeln geleistet habe.

Im ersten Artikel habe ich noch als wissenschaftliche Hilfskraft den Prozess der Studienplanung und die Datenerhebung begleiten dürfen und unterstützt. Danach war ich als wissenschaftliche Mitarbeiterin hauptverantwortlich für die Entwicklung der Fragestellung, Datenanalyse und die Erstellung und Einreichung des Manuskripts. Im zweiten und dritten Artikel war ich Hauptverantwortliche für den kompletten Forschungsprozess. Dies beinhaltete bei Artikel II alle Schritte von der Entwicklung der Forschungsidee und des Studiendesigns, der Planung und Durchführung der Studie (Entwicklung der Studienmaterialien, Korrespondenz mit den teilnehmenden Schulen, Datenerhebung, Dateneingabe und -management), der Durchführung der statistischen Analysen sowie das Verfassen und Einreichen des Manuskripts. Im dritten Artikel war ich ebenfalls hauptverantwortlich für den gesamten Forschungsprozess, angefangen bei der Entwicklung der Fragestellung bis hin zum Verfassen und Einreichen des Manuskripts.

Contents

Summary

The current dissertation is based on three related Research Papers, in which different dimensions of assessing anxiety in learning and achievement situations, namely trait and state self-report assessments of anxiety (Paper I), the different anxiety components (Paper II) and physiological measures of anxiety (Paper III) are examined and compared with each other concerning their results. The first study investigated the discrepancy between generalized trait and situation-specific state self-report measures of anxiety in mathematics in a sample of students with high and low achievement levels. Previous studies have revealed that trait anxiety measures are typically rated higher than state measures. In addition, the academic self-concept has been identified to play a moderating role in the so-called trait-state discrepancy, with a higher academic self-concept leading to a lower discrepancy (i.e., less overestimation of trait anxiety when state assessments reflect actual experience). Therefore, it was predicted in this study that high achievers, due to their higher academic self-concept in mathematics, exhibit a smaller trait-state discrepancy in anxiety than students with low achievement levels. The results were in line with this hypothesis and demonstrated that high achievers even underestimated their anxiety when judging it from a trait as compared to a state perspective. Taken together, this study shows that state and trait self-report measures of anxiety can yield different results and that even high achievers experience some anxiety in class.

The second study investigated the different components of test anxiety (i.e., cognitive, affective, physiological, and motivational components) using an intra-individual approach. The first aim of this study was to examine the relative impact of the different components of anxiety in the mediating mechanism that connects control, anxiety and performance - as proposed by the control-value theory. Thereby it sought to identify which component should be primarily addressed by anxiety interventions and instructional techniques in order to improve students' performance. A second aim of this study was to uncover which anxiety component underlies the commonly used and very convenient single-item anxiety measures. In line with the expectations, the anxiety components indeed differed in their relative impact on test performance, with the cognitive component being central for this outcome. Furthermore, results indicated that behind a single-item anxiety measure we most likely find the affective anxiety component and thus, not the component most relevant for performance. Therefore, it is important to consider and differentiate between the anxiety components and

select the assessment method based on the research question at hand.

The third study used the methodological approach of a systematic review to investigate the relationship between traditional self-report measures of test anxiety with another type of measure, namely physiological measurement. This assessment method has gained increased attention in pedagogical psychological research in recent years. Results of the systematic review showed that in line with theoretical considerations, self-report measures of test anxiety and physiological measures (e.g., cardiovascular measures, electrodermal measures, cortisol sampling) were significantly correlated in the majority of the reviewed studies. The correlations were moderate, suggesting that although the two measures overlap, they should not be used interchangeably. In conclusion, each of these measures provides information about a person's anxiety and applying both of the two measures might contribute to a better measurement and thus also a better understanding of this often detrimental emotion. Furthermore, variables that can influence the relation between self-report and physiological measures were identified (e.g., sampling rate, choice of physiological measure and the way the data is analyzed) and are discussed with regard to a successful assessment of physiological data in future research.

Taken together, from the results of these three Research Papers it can be concluded that it is important to distinguish between different ways of assessing anxiety and to select the assessment method according to the research focus. Besides its implications for research, the present thesis provides novel perspectives on how to successfully influence anxiety in learning and achievement situations.

Zusammenfassung

Die vorliegende Dissertation basiert auf drei verwandten Forschungsarbeiten, in denen verschiedene Dimensionen der Erfassung von Angst in Lern- und Leistungssituationen, nämlich Trait- und State-Selbstberichtsmaße von Angst (Paper I), die verschiedenen Angstkomponenten (Paper II) und physiologische Messungen von Angst (Paper III), differenziert betrachtet und hinsichtlich ihrer Forschungsergebnisse miteinander verglichen werden. Die erste Studie untersuchte die Diskrepanz zwischen generalisierten Trait- und situationsspezifischen State Selbstberichtsmaßen von Angst im Fach Mathematik in einer Stichprobe von Schülerinnen und Schülern mit hohem und niedrigem Leistungsniveau. Frühere Studien haben gezeigt, dass Trait-Maße zur Erfassung von Angst typischerweise zu höheren Angstwerten führen als State Maße. Darüber hinaus wurde festgestellt, dass das akademische Selbstkonzept eine moderierende Rolle in der sogenannten Trait-State-Diskrepanz spielt, wobei ein höheres akademisches Selbstkonzept zu einer geringeren Diskrepanz führte (d. h. weniger Überschätzung der Trait-Angst, wenn State Maße das tatsächliche Erleben widerspiegeln). Daher wurde in dieser Studie angenommen, dass Schülerinnen und Schüler mit hohem Leistungsniveau aufgrund ihres höheren akademischen Selbstkonzepts im Fach Mathematik eine geringere Trait-State-Diskrepanz der Angst aufweisen als Schülerinnen und Schüler mit niedrigerem Leistungsniveau. Die Ergebnisse entsprachen dieser Hypothese und zeigten, dass Hochleistende ihre Angst sogar unterschätzten, wenn sie sie aus einer Trait- im Vergleich zu einer State-Perspektive beurteilten. Zusammenfassend zeigte diese Studie, dass Trait- und State-Selbstberichtsmaße von Angst unterschiedliche Ergebnisse liefern können und dass selbst Hochleistende im Unterricht ein gewisses Maß an Angst erleben.

Die zweite Studie untersuchte die verschiedenen Komponenten der Prüfungsangst (d.h. kognitive, affektive, physiologische und motivationale Komponenten) mit einem intraindividuellen Ansatz. Das erste Ziel dieser Studie war es, den relativen Einfluss der verschiedenen Angstkomponenten zu untersuchen in dem von der Kontroll-Wert-Theorie postulierten Vermittlungsmechanismus, der Kontrolle, Angst und Leistung miteinander verbindet. Dadurch sollte die Komponente identifiziert werden, die in erster Linie durch Angstinterventionen und Unterrichtstechniken angesprochen werden sollte, um die Leistung der Schülerinnen und Schüler zu verbessern. Ein zweites Ziel dieser Studie war es, herauszufinden, welche Angstkomponente den häufig verwendeten und sehr praktischen

Einzelitem-Angstmessungen zugrunde liegt. Wie erwartet unterschieden sich die Angstkomponenten tatsächlich in ihrem relativen Einfluss auf die Prüfungsleistung, wobei die kognitive Komponente hierfür von zentraler Bedeutung war. Darüber hinaus deuteten die Ergebnisse darauf hin, dass wir hinter einem Einzelitem höchstwahrscheinlich die affektive Komponente und damit nicht die für die Leistung zentrale Komponente finden. Daher ist es wichtig, die Komponenten entsprechend zu berücksichtigen und zwischen ihnen zu differenzieren und die Erhebungsmethode anhand der jeweils vorliegenden Forschungsfrage auszuwählen.

Die dritte Studie verwendet den methodischen Ansatz eines systematischen Reviews, um den Zusammenhang zwischen traditionellen Selbstberichtsmaßen von Prüfungsangst und einer weiteren Erhebungsmethode, nämlich physiologischen Messungen, zu untersuchen. Diese Maße haben in den letzten Jahren in der pädagogisch psychologischen Forschung zunehmend an Aufmerksamkeit gewonnen. In Übereinstimmung mit theoretischen Überlegungen, zeigten die Ergebnisse des systematischen Reviews, dass Selbstberichtsmaße von Prüfungsangst und physiologische Maße (z.B. kardiovaskuläre Messungen, elektrodermale Messungen, und Cortisolproben) in der Mehrheit der untersuchten Studien signifikant korreliert waren. Die Korrelationen waren moderat, was darauf hindeutet, dass sich die beiden Messungen zwar überschneiden, aber nicht austauschbar verwendet werden sollten. Zusammenfassend lässt sich sagen, dass jede dieser Erhebungsmethoden Informationen über die Angst einer Person liefert und die Anwendung beider Methoden zu einer besseren Messung und damit auch zu einem besseren Verständnis dieser oftmals nachteiligen Emotion beitragen könnte. Darüber hinaus wurden Aspekte identifiziert, die den Zusammenhang zwischen Selbstberichtsmaßen und physiologischen Messungen beeinflussen können (z.B. die Samplingrate, die Wahl des physiologischen Maßes und die Art und Weise der Datenanalyse); diese werden im Hinblick auf eine erfolgreiche Erhebung physiologischer Parameter in zukünftigen Studien diskutiert.

Zusammenfassend lässt sich aus den Ergebnissen dieser drei Forschungsarbeiten schlussfolgern, dass es wichtig ist, zwischen verschiedenen Methoden zur Erhebung von Angst zu differenzieren und die Methode entsprechend dem Forschungsschwerpunkt auszuwählen. Neben den Implikationen für die Forschung bietet die vorliegende Arbeit auch neue Perspektiven, wie man Angst in Lern- und Leistungssituationen erfolgreich beeinflussen kann.

1 Synopsis

1.1. Introduction and Overview

In recent decades, there has been an increase in the use of various methods to investigate the complex emotion of anxiety in learning and achievement situations (Harley, 2015; Pekrun & Linnenbrink-Garcia, 2014). However, many fundamental questions either have remained unanswered or need to be taken up again in view of new empirical studies and an increasing interest in new methods of measuring anxiety, such as physiological measurements (D'Mello & Calvo, 2013). These questions include*: Are there differences between different self-report measures of anxiety? Is a differentiated component-specific assessment of anxiety relevant? Is more perceived anxiety also reflected in higher physiological arousal, for example a faster heartbeat?*

Across three related Research Papers, the present dissertation addresses these questions and compares different dimensions of assessing anxiety in educational settings, namely trait and state self-report assessments of anxiety (Paper I), the different anxiety components (Paper II), and physiological measures of anxiety (Paper III). The aim of this synopsis is to give a short outline of the three included studies and to discuss the three papers in a common framework, showing how each of them contributes to the overarching topic of the present thesis. I will therefore proceed as follows:

In the first section, I will give an overview of *anxiety in learning and achievement situations* and introduce two different self-report measures of anxiety, namely *trait and state measures* along with outlining their advantages and disadvantages. Thereby I provide the theoretical framework for the first Research Paper that examined the *discrepancy between trait and state self-report measures of anxiety in a sample of students with high vs. low achievement levels*.

In the second section, I will focus on another dimension that can be considered when assessing anxiety, especially when examining its antecedents and effects on performance - namely taking the *different components of anxiety* (i.e., cognitive, affective, physiological, and motivational components) into account. This was the focus of the second study, which used an *intra-individual* approach to examine the different test anxiety components in order to determine which of the anxiety components is most important in the *mediating mechanism that connects control antecedents and performance*. Thereby we aimed at identifying the component that should be primarily addressed by anxiety interventions and instructional techniques that focus on improving students' performance. The second goal of this study was to explore which component of anxiety underlies the convenient and often used single-item measures of anxiety.

In the third section, I will turn to the third Research Paper, which addresses a further dimension of anxiety assessment, namely physiological measurement – which has become increasingly popular in educational research in recent year (D'Mello & Calvo, 2013). In detail, this Paper examined the *relation between self-report measures of test anxiety and physiological arousal* using the methodological approach of a systematic review to provide a research basis on the relationship between self-report and physiological measures and get more grip on how to apply them successfully in educational settings.

Figure 1 provides a graphical overview of the relations between the study variables of the three Research Papers. The synopsis closes with a summary of the results of the three studies and a general discussion that points out implications for research and practice, as well as strengths and limitations of the present research and sketches promising directions for future research. The three introduced studies are then presented separately in the style of self-contained manuscripts with the common repeating structure (introduction, theoretical framework, methods section, results and discussion) in the following chapters 2-4.

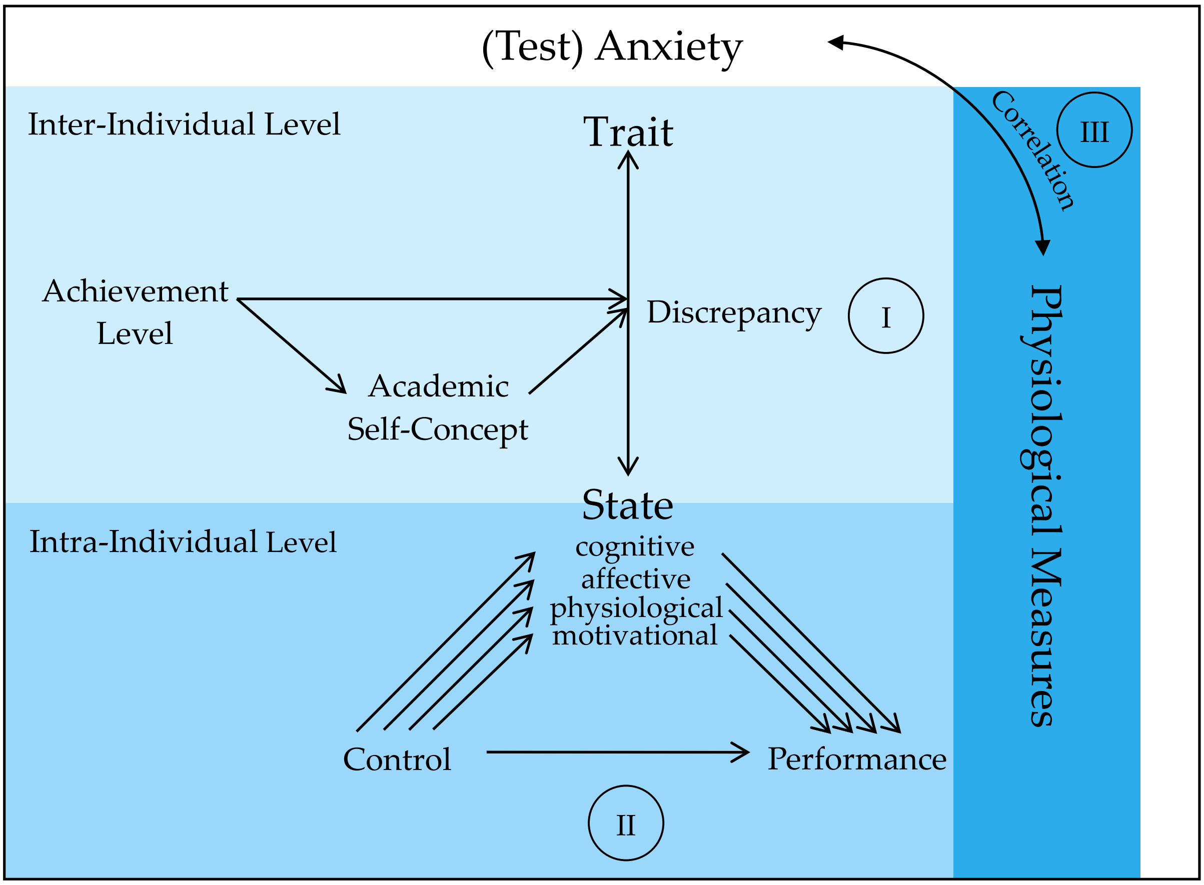

Figure 1. Overview of relations between variables of the three research projects.

1.2. Anxiety in Learning and Achievement Situations

Anxiety is an emotion that is familiar to all of us and that can be seen as the most common source of emotional distress in the school context (Cassady, 2010; Cassady & Johnson, 2002; Goetz & Hall, 2013). Two very prevalent forms of anxiety are math anxiety and test anxiety (Jain & Dowson, 2009; Zeidner, 2014). *Math anxiety* refers to the aversive feelings of tension, concern, mental disorganization, and bodily symptoms that are evoked in situations involving mathematics such as mathematical computations, problem-solving, and assessment (Ashcraft, 2002; Ashcraft, Krause, & Hopko, 2007). It is conceptually related to test anxiety through a common theme of concerns about evaluation (e.g., Rosen & Maguire, 1990). *Test anxiety* is evoked when a person believes that her or his intellectual, motivational, and social capabilities do not meet the demands of the test or evaluative situation (Reeve, Bonaccio, & Charles, 2008). It can be experienced before, during and after an examination or some other test situation (Zeidner, 2007, 2014).

Because of its prevalence (test anxiety: 25 to 40% of the population; Zeidner, 2014) and various debilitating effects (e.g., deficits in information processing and retrieval, low performance, low self-esteem and lower well-being; Cassady, 2004; Enright, Baldo, & Wykes, 2000; Eysenck, Derakshan, Santos, & Calvo, 2007; Hembree, 1990; OECD, 2016; Seipp, 1991; Van Yperen, 2007), knowledge about the detrimental and complex emotion of anxiety is crucial - not only for researchers, but also for teachers and educators (Zeidner & Matthews, 2011). For example, for planning an anxiety free lesson or designing tests, and for developing, planning and evaluating interventions (Astleitner, 2000; Zeidner, 2007, 2014). In order to gain this knowledge, a very important aspect for researchers is the measurement of anxiety (Cipora, Artemenko, & Nuerk, 2019).

1.3. Assessing Anxiety via Trait and State Self-Report Measures

There are different ways to assess anxiety. The most prominent and widely used ways of measurement are self-report measures (Larsen & Prizmic-Larsen, 2006). They can be further divided into so-called trait and state assessment methods (Zuckerman & Spielberger, 2015).

1.3.1. Trait Self-Report Measures of Anxiety

Trait self-report assessments of anxiety have a long-standing history in the research literature (Wilhelm, Perrez, & Pawlik, 2012). They are regarded to assess habitual anxiety without referring to a specific situation and require a recall of this emotion via generalized self-reports (Kaplan, Dalal, & Luchman, 2013; Schutz & Davis, 2000)[1]. Trait self-report assessments typically involve paper-pencil questionnaires in which subjects respond to items like *"How much anxiety do you usually experience in mathematics class?"* and are usually of cross-sectional nature (i.e., they are only assessed at one point of time). Thus, the researcher's focus is generally on comparisons across people (Kaplan et al., 2013). The advantages of trait measures of anxiety are that they offer quick and cost-effective assessments of self-observed thoughts and emotional experiences (Wilhelm & Grossman, 2010).

1.3.2. State Self-Report Measures of Anxiety

In contrast to trait self-report measures, *state self-report measures* of anxiety examine momentary anxiety as it is experienced in a specific situation (e.g., Eid, Schneider, & Schwenkmezger, 1999; Kärner & Kögler, 2016). In this case, participants are asked to indicate their anxiety over a short time interval (e.g., during the past lesson, during the past test, or at the current moment). An example item would be: *"How much anxiety are you experiencing right now?"*. Usually researchers are interested in multiple repeated measures of state anxiety, where the focus is on within-person comparisons across time that allow for intra-individual analyses (Molenaar & Campbell, 2009). This can be achieved with so-called experience sampling methods (Hektner, Schmidt, & Csikszentmihalyi, 2007; Shiffman, Stone, & Hufford, 2008; Trull & Ebner-Priemer, 2009), in which participants complete repeated surveys in near real-time in a natural setting (e.g., at multiple time points during a school day, a lesson, or a test). Such assessments can either be conducted by using electronic devices that send multiple random reminders to fill in a short electronic questionnaire during a lesson (e.g., ambulatory psychological assessment; Trull & Ebner-Priemer, 2009) or by filling in a short paper-pencil questionnaire after each lesson (i.e., kind of a diary as used in the study byBecker, Keller, Goetz, Frenzel, & Taxer, 2015).

The idea of differentiating traits and states by asking students about their anxiety "in general" versus "at the moment" is closely related to the approach of Spielberger's State-Trait-Anxiety-Inventory (Spielberger, Gorsuch, & Lushene, 1970), but actually goes back to Zuckerman (1960). Existing studies on anxiety have predominately used the less costly and more convenient trait assessment method (Wilhelm et al., 2012). In more recent research, momentary assessments of anxiety in ecologically valid achievement and learning situations (i.e., state anxiety) are becoming increasingly popular (e.g., Ahmed, van der Werf, Minnaert, & Kuyper, 2010; Goetz et al., 2012; Pekrun & Bühner, 2014).

[1] It has to be clarified that the concept of "trait anxiety" or "trait anxiousness" could suggest a general and almost pathological proneness or vulnerability to anxiety, regardless of the nature of threat (Endler & Kocovski, 2001). However, the present dissertation uses a methodologically centered definition of "trait anxiety". When referring to "trait anxiety" or "trait measures of anxiety" in the following sections, what is meant is that anxiety is assessed via generalized self-reports (Amelang, Bartussek, Stemmler, & Hagemann, 2006). When discussing state anxiety or state assessments, these terms refer to anxiety as assessed directly in a specific context (i.e., perceived anxiety in mathematics class at the time the questionnaire is completed).

1.3.3. The Discrepancy Between Trait and State Self-Report Measures

Previous studies found that there is a discrepancy in the intensity rating of the two above mentioned measures, in such a way that trait self-report measures in comparison to scores obtained by multiple state assessments, are generally rated higher (e.g., Barrett, 1997; Ben-Zeev, McHugo, Xie, Dobbins, & Young, 2012; Wilson & Gilbert, 2005). This has also been termed *intensity bias* by Buehler and McFarland (2001).

Robinson and Clore (2002) explain this discrepancy between trait and state measures in their *accessibility model of emotional self-report.* Their model proposes that when using momentary or online self-reports to assess anxiety (i.e., like in state self-report measures of anxiety), the direct experiential information is accessible to the student. Thus, state self-report measures of anxiety only require the student to recall experiences that just occurred. When self-reports become more distant from the actual events they are intended to cover, such as in trait self-report measures of anxiety, students have to recall or summarize memories about their anxiety that are not readily available. Thus, they may not attempt to recall single episodes or events in which they experienced anxiety. Instead, they answer the questionnaire based on their memory and their beliefs about how they or others typically feel in such circumstances (i.e., using semantic knowledge; Kahneman, 2011). Robinson and Clore (2002) postulate that this in turn causes trait measures to be strongly influenced by subjective beliefs. In contrast, due to their temporal proximity to the situation, state measures are less susceptible to such a bias (Kahneman, 2011; Wilhelm & Grossman, 2010). Therefore, state measures of momentary emotions that are more dependent on contextual and environmental factors are considered to evaluate actual emotions, while trait measures are regarded to rather reflect and individual’s beliefs about emotions (Robinson & Clore, 2002). With respect to the so-called *trait-state discrepancy*, this would mean that the discrepancy (i.e., higher intensity rating of trait self-reports) arises due to the overestimation of anxiety as assessed with trait self-report measures. This raises the question of whether it is appropriate to draw conclusions about anxiety as assessed with state measures from anxiety assessed with trait measures and vice versa, and calls for comparing the two self-report measures by applying both a trait as well as a state approach to assess students’ anxiety.

There are limited studies that have already embarked on this path (e.g., Bieg, Goetz, & Lipnevich, 2014; Bieg, Goetz, Wolter, & Hall, 2015; Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013; Nett, Bieg, & Keller, 2017). Studies investigating the discrepancy between trait and state self-report measures of anxiety in girls and boys found that girls only reported higher trait but not state anxiety than boys (Goetz et al., 2013). In addition, they found that the trait-state discrepancy in anxiety was smaller in students with higher academic self-concepts (Bieg et al., 2014; Goetz et al., 2013).

The *academic self-concept* is an identity related subjective belief that can be defined as a mental representation of abilities and competencies a person has (Nagengast & Marsh, 2012). It is seen as a crucial antecedent of emotions according to the control-value theory of achievement emotions (Pekrun, 2006). With regard to the trait-state discrepancy a smaller discrepancy (i.e., due to a higher academic self-concept) would mean less overestimation of trait anxiety and thus can be viewed as a more realistic evaluation of one's anxiety. Consequently, students with a high academic self-concept should evaluate their anxiety more realistically than students with a low academic self-concept.

High achievers (i.e., students with high achievement levels) can in general be expected to have high academic self-concepts (Bailey, 1971). Yet, until now, there is no knowledge about how the trait-state discrepancy in anxiety looks like in such students with very high achievement levels. A meta-analysis by Ma (1999) found that for high achieving students' anxiety levels seem to be relatively unproblematic, because high achievers tend to be less anxious about mathematics than their low achieving peers. However, the majority of the studies on mathematics anxiety have been based on the assessment of trait anxiety with trait self-report measures, and little is known whether this also holds true for actual state anxiety measures, or in other words whether trait measures of anxiety are able to validly gauge what high achieving students actually feel in the classroom.

1.3.4. Research Questions of Research Paper I

In response to this lack of research, the first study of the present dissertation examined the discrepancy between trait and state assessments of anxiety in a group of high and low achieving students and further investigated academic self-concept as a predictor of the trait-state discrepancy. The study was conducted with $N = 237$ high school students (grades 9 and 10) from Germany and was using a multi-method approach with traditional questionnaires (i.e., to assess trait mathematics anxiety) and experience sampling (i.e., diaries) during multiple real-life mathematics lessons to assess students' state (i.e., momentary) anxiety in mathematics. This rendered it possible to evaluate whether findings from previous studies assessing high achieving students' trait anxiety hold true when actual state anxiety measures are also assessed.

The study addressed the following research questions:

(1) Do high achievers experience less trait as well as state anxiety in mathematics than low achievers?

(2) Do high achievers exhibit a smaller trait-state discrepancy?

(3) Given that high achievers exhibit a smaller trait-state discrepancy, is this due to their higher academic self-concept in mathematics?

Based on previous findings that have shown that academic achievement (Ma, 1999) is negatively related to trait mathematics anxiety, along with the assumption that high achievers have higher control experience in class, which according to the control-value theory (Pekrun, 2006) should result in high achievers also experiencing less state anxiety than low achievers, it was assumed that high achievers report lower levels of both trait and state anxiety. Moreover, since high achievers can be expected to have very high academic self-concepts, it was assumed that they would less strongly overestimate their trait anxiety (i.e., smaller trait-state discrepancy) in comparison to low achievers. In more technical terms, we expected achievement group (i.e., high vs. low achievers) to moderate the trait-state discrepancy in mathematics anxiety and assumed that the academic self-concept partially mediates this relationship (i.e., mediated moderation, see Figure 1).

1.4. Assessing the Different Components of Anxiety

Another aspect of measuring anxiety is taking the different components of anxiety into account. Although emotions, and more specifically anxiety, in the academic context are generally defined as multidimensional constructs in which different components are distinguished (e.g., cognitive, affective, motivational, physiological, behavioral-expressive components; Pekrun et al., 2004; Pekrun, Goetz, Titz, & Perry, 2002; Pekrun & Linnenbrink-Garcia, 2014; Reeve, 2014; Shuman & Scherer, 2014), in empirical studies this multi-componential nature of anxiety is often neglected. As a consequence, anxiety is usually assessed and examined relatively undifferentiated as a one-dimensional construct, for example by building sum scores or by using single-item measures (e.g., *"How much anxiety are you experiencing right now?"*), which do not clearly indicate whether one, some, or all components of anxiety are perceived (Shuman, Clark-Polner, Meuleman, Sander, & Scherer, 2017). Regarding test anxiety, researchers often make only a rather rough differentiation between "worry" and "emotionality" components (Zeidner, 2007). Worry can be described as cognitive concerns especially about the consequences of failure and emotionality refers to the affective arousal experience (Liebert & Morris, 1967). However, the dichotomous distinction between worry and emotionality is nowadays replaced by a finer distinction between the components (e.g., Hodapp & Benson, 1997; Matthews, Hillyard, & Campbell, 1999; Pekrun & Bühner, 2014; Sarason, 1984; Scherer, 1984, 2009).

1.4.1. Four Components of Anxiety

A common distinction of the emotional components, which is also reflected in emotion questionnaires such as the Achievement Emotions Questionnaire (AEQ; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011), is the differentiation between four components[2] (Pekrun & Bühner, 2014; Pekrun et al., 2004; Scherer, 2009). Regarding the emotion of anxiety they can be described by intrusive thoughts and worries (i.e., *cognitive component*), a feeling

[2] Sometimes a fifth component, namely the behavioral-expressive component, is included (Pekrun & Bühner, 2014). 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. However, this component is not part of the typical anxiety questionnaires such as the Achievement Emotions Questionnaire (AEQ; Pekrun et al., 2011), as it is not easily accessible with self-report measures. Often, this component is indirectly assessed, for example via observation (Harley, 2015)).

of nervousness (i.e., *affective component* – often also referred to as the component that describes the core feeling of an emotion), increased physiological arousal such as higher heart rate or sweating (i.e., *physiological component*), and an urge to withdraw from the situation (i.e., *motivational component*; Pekrun et al., 2011; Pekrun et al., 2004; Zeidner, 2014; Zeidner & Matthews, 2005). These four components are theoretically distinct, yet related (Zeidner, 2014).

1.4.2. Control Antecedents and Achievement Effects of the Anxiety Components

Especially in theories that explain antecedents and effects of academic emotions, this multi-component perspective of anxiety is often neglected and the components are not considered separately. Knowledge on the causes (i.e., antecedents) and effects of the individual components may proof helpful, for example to develop and apply component-focused strategies to reduce anxiety according to their effect intensities (Zeidner, 2014). One of the most prominent theories in the educational context that deals with the antecedents and effects of emotions is the *control-value theory* (Pekrun, 2006), which has already been briefly touched upon before. This theory refers to control as a central antecedent of anxiety and lower performance as a major effect of anxiety.[3]

Control can be described as the cognitive appraisal of the possibility to personally influence activities and outcomes and may include perceptions such as competence beliefs and causal attributions (Goetz, Pekrun, Hall, & Haag, 2006; Pekrun & Stephens, 2010). It is assumed that perceived control is positively associated with positive emotions and negatively related to negative emotions (Pekrun, 2006). The control-value theory states that increased anxiety especially arises in achievement situations that a student perceives to be important, such as a test (i.e., high value) and when the student experiences a loss of control or insufficient control over the achievement activities or outcomes (Pekrun, 2006; Scherer, Schorr, & Johnstone, 2001). This increased anxiety is in turn associated with decreased *performance*. Thus, this relationship can be described as a *mediation: anxiety mediates the relationship between control and performance*. In detail, anxiety resulting from perceived lack of control reduces cognitive resources and the use of flexible learning and test-taking strategies and self-regulation (Pekrun, 2006). Subsequently, increased anxiety has negative effects on performance (Pekrun, Elliot, & Maier, 2009).

A large number of empirical findings consistently confirm Pekrun's assumptions about the relationship between control, anxiety and performance (Frenzel, Pekrun, & Goetz, 2007; Goetz et al., 2006; Hembree, 1988; Lee, 2009; Perry, Hladkyj, Pekrun, & Pelletier, 2001).

[3] *Value* appraisals, as already reflected in the name of the control-value theory, are also important antecedents of anxiety (Pekrun, 2006). The focus of the present dissertation, however, was on control appraisals.

These studies were mainly based on trait assessments of anxiety, but there is also some supporting evidence for state assessments (e.g., Ahmed et al., 2010).

However, although some previous findings indicate that the anxiety components might differ regarding their antecedents and effects (e.g., Cassady & Johnson, 2002; Keogh, Bond, French, Richards, & Davis, 2004; Liebert & Morris, 1967; McIlroy, 2000; Morris & Liebert, 1970; Putwain, 2007; Putwain, Connors, & Symes, 2010; Seipp, 1991), it remains unclear whether all of the anxiety components are equally important in this proposed mediating mechanism connecting control, anxiety and performance at school, since no study that examined this relation has differentiated between the components thus far (see Figure 1). Therefore, it seems reasonable to explicitly investigate the different anxiety components in the framework of the control-value theory.

1.4.3. Research Questions of Research Paper II

Against this background, the aim of the second study in the current dissertation was to investigate the relative impact the anxiety components have in the above described mediating mechanism that connects control and performance, and to identify the component that is central in this respect (i.e., has the strongest impact on students' performance). Furthermore, as the control-value theory describes intra-individual processes (i.e., within-person associations; Pekrun, 2006), we aimed at complementing findings from traditional inter-individual studies (i.e., studies that are only looking at between-person associations) by applying an *intra-individual* real-time approach (Murayama et al., 2017). In this way, it was possible to gain insight into the intra-individual functioning of the control-anxiety components-performance relation-ship. Particularly, we were examining this relationship in a sample of N = 137 German 8th graders, when students were writing a math exam using multiple within-person measurements (i.e., state assessments) of the study variables. Since anxiety, as previously stated, is usually examined relatively undifferentiated and in more recent studies for practical reasons, is even assessed only with *single-item measures* (Ashcraft, 2002; Csikszentmihalyi & Larson, 2014; Goetz et al., 2013; Schimmack, 2003; Tong et al., 2007), the second goal of the study was to investigate which anxiety component is hidden behind such an undifferentiated measure of anxiety. Such measures have clear advantages because they are less intrusive and easier to administer (especially in large-scale assessments), and previous research suggested that they could be sufficient for adequately measuring subjective experiences such as anxiety (Ainley & Patrick, 2006; Ashcraft, 2002; Gogol et al., 2014; Núñez-Peña, Guilera, & Suárez-Pellicioni, 2014). In the case of test anxiety however, these single-item measures are only able to gauge anxiety as an overall construct (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012) and little is known what anxiety component such measures actually tap into. In combination with our first research question, answering this question may help future research in deciding whether and how differentiated anxiety should be assessed depending on the research question and the conclusions they want to draw.

Our study was guided by the following research questions:

(1) a) Is the effect of control and performance mediated by the test anxiety components (see Figure 1)?
b) Does control affect all anxiety components in similar ways?
c) Do the different anxiety components affect performance in the same way?

(2) What anxiety component is mainly assessed with single-item self-report measures of test anxiety?

It was hypothesized that anxiety mediates the relationship between control and performance and that the anxiety components vary in the strength of their mediating effects. Since according to the control-value theory control is a cognitive appraisal, we assumed that it has the strongest negative effect on the cognitive anxiety component (Pekrun, 2006). Based on information-processing models that state that thinking about negative consequences of one's performance takes up a lot of cognitive resources, and building on previous research that found strong associations between performance deficits and high levels of worry and cognitive interference, it was assumed that the cognitive anxiety component has the strongest effect on performance as well as the strongest mediating effect (Naveh-Benjamin, 1991; Van Yperen, 2007; Zeidner & Matthews, 2011). Furthermore, it was assumed that when students are asked about their anxiety with a single-item, their answer will most likely be associated with the affective anxiety component – as this is according to most models the key anxiety component (Scherer, 1984, 2005).

1.5. Assessing Anxiety Using Physiological Measures

Based on theoretical considerations in Paper II of the current dissertation, for a more complete picture of anxiety and its antecedents and effects it might be important to take the different components into account. However, as reflected in the theoretical assumptions of Paper I, trait self-report measures can be biased (e.g., by memory biases, or subjective beliefs; Robinson & Clore, 2002) and even state measures of anxiety can be influenced by certain factors, such as the willingness and capability to report about feelings (Kaplan et al., 2013). For example, a state self-report measure could surprise students if they are engaged in deep concentration and are not expecting the questionnaire to pop up. Alternatively, students may feel bored by having to repeatedly complete the questionnaire if they are administered too closely together or they may be frustrated because of the repeated interruption (Harley, 2015). In addition, self-reports can only capture the subjective experience of a student. These limitations of self-report measures call for applying measures that are less susceptible to such biases and cannot be masked or controlled by the individual, such as for example *physiological measures* of arousal, in order to obtain a more objective and therefore also more complete assessment of anxiety (Podsakoff, 2003). This is especially promising when examining the physiological component of test anxiety, as increased levels of physiological arousal can be continuously assessed in real-life testing situations through physiological measures such as heart rate or skin conductance levels (Betancourt, Dethorne, Karahalios, & Kim, 2017; Harley, 2015; Zeidner & Matthews, 2011).

1.5.1. The Physiological Response to Anxiety

Strong emotional arousal, such as that associated with stress, and more specifically, test anxiety, activates two primary biological systems. The first is the autonomic nervous system which is responsible for modulating peripheral function and consists of sympathetic and parasympathetic branches that are generally associated with activation and relaxation, respectively (Mauss & Robinson, 2009). The second system is the adrenocortical system, which is activated via the hypothalamic-pituitary-adrenal (HPA) axis. This system is responsible for regulating the hormone cortisol, often referred to as the classical stress hormone (Lundberg, 2005).

Since the body's responses to stress are numerous (e.g., increased pulse and heart rate, heightened blood pressure, scanty secretion of salivary glands, increased sweating, cold and clammy hands, …), there are several different physiological measures to assess physiological arousal (Gatchel, Baum, & Krantz, 1989; Zeidner & Matthews, 2011). The focus in the third Paper of the current dissertation was therefore only on measures that are typically and currently used to examine anxiety (Hodges, 2015). These are cardiovascular measures such as heart rate (HR), heart rate variability (HRV) and blood pressure (BP); electrodermal activity measures (EDA; Boucsein, 2012) such as skin conductance (SC) or skin resistance (SR); endocrine measures such as cortisol sampling, and pH level in saliva which recently has gained increased attention (Marques, Silverman, & Sternberg, 2010).

Interest in physiological measures to complement traditional self-report measures of emotions has grown immensely over the last five years (D'Mello & Calvo, 2013). However, interpreting findings obtain from these measures in relation to existing educational research and theories remains a challenge in the field of educational psychology (Kreibig, 2010; Kreibig & Gendolla, 2014; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005; Mauss & Robinson, 2009). In particular, when attempting to measure psychological constructs like (test) anxiety, one needs not only to show that the measurement is reliable, but also that the measurement of the construct is valid (e.g., by showing that different measures of the same construct are related; Cipora et al., 2019). Since heightened physiological arousal is one component of test anxiety (Zeidner & Matthews, 2005), it is reasonable to assume that students' objectively measurable level of physiological arousal during test taking (i.e., as assessed with physiological measures) and their level of self-reported test anxiety (i.e., often including the subjective experience of all components) would converge. However, empirical findings investigating this relationship are mixed and ambiguous (e.g., Mauss et al., 2005; Tryon, 1980; Zeidner, 1998; Zeidner & Matthews, 2011).

1.5.2. Research Questions of Research Paper III

To yield a more coherent picture of the relationship between self-report and physiological measures, and to gain further insight of their similarities and differences, the systematic review of the present dissertation investigated whether higher self-reported test anxiety is associated with objectively measurable increased physiological arousal. To our knowledge, no review has systematically examined or quantified this relation. As such, this review aims to systematically integrate the existing evidence.

Furthermore, another goal of the review was to identify aspects that potentially influence the relationship between self-report measures of test anxiety and physiological measures (e.g., study design and setting, sample characteristics or the assessment and analysis of physiological data). This may help future research to apply them successfully in educational contexts.

Finally, some researchers (e.g., Morris, Davis, & Hutchings, 1981) state that physiological indices provide information concerning the emotionality component of anxiety but little information about the worry component. They assumed that the inconsistent findings on the correlations between self-report and physiological measures may be partly due to combining emotionality and worry in the same self-report measures. Therefore, we were additionally aiming to find out whether the relationship between physiological arousal and the emotionality components of test anxiety is different from the relationship between physiological arousal and the worry component.

Thus, the following research questions drove our systematic review:

(1) Is self-reported test anxiety positively associated with objectively measurable physiological arousal (and can we quantify this relation)?

(2) Are there methodological variables that moderate this relation (e.g., study design, setting, sample characteristics, assessment and analysis of physiological data)?

(3) Are there differential relations between worry and emotionality components of test anxiety, on the one hand and physiological arousal, on the other?

To answer these questions, a systematic review of the literature in PsycINFO, PubMed, and ERIC databases was conducted to identify all studies addressing the relationship between self-reported test anxiety and physiological arousal applying cardiovascular measures (e.g., heart rate), electrodermal measures (e.g., skin conductance), cortisol sampling, or saliva pH.

1.6. General Discussion

The aim of the present dissertation was to gain insight into the structural similarities and differences of the described assessment methods of anxiety in learning and achievement situations. The following general discussion will answer the general questions of the introduction by combining the results and the implications of the three included studies. First, the *main findings* of each research project of my dissertation are briefly summarized and discussed, and several *implications for research* and *practice* are drawn, many of them already detailed in the three Research Papers. In the subsequent section, general *strengths* and *limitations* of the three Research Papers are addressed, and I will sketch potential *directions for future research*. The last section closes with an overall *conclusion* that can be drawn from the presented findings.

1.6.1. Summary of Main Findings and Implications for Research

Paper I: Trait and state self-report measures of anxiety

Research Paper I examined the discrepancy between trait and state self-report measures of anxiety in mathematics in high versus low achieving students while also taking into account their academic self-concept in mathematics – a subjective belief previously shown to play a role in explaining this discrepancy. The goal of this study was to investigate whether findings from trait measures hold true when also applying state anxiety measures. The findings showed that high achievers experienced less trait as well as less state anxiety in mathematics class than low achievers, which was consistent with the theoretical assumptions and our hypotheses. Thus, overall high achievers seem to be rather on the low end of the anxiety continuum. However, results also showed that multiple state assessments of self-reported anxiety in mathematics yield different results than single trait assessments, especially in students with different achievement levels and different levels of academic self-concept. In more detail, low achievers rated their trait anxiety significantly higher than their state anxiety (i.e., they overestimated their trait anxiety), which is consistent with the intensity bias in emotional self-reports (Buehler & McFarland, 2001). Yet, even though the trait-state discrepancy in anxiety was lower in high achievers, surprisingly unlike low achievers high achieving students due to their higher academic self-concepts experienced more anxiety (i.e., as assessed with state measures) than we would have expected from their trait self-reports (i.e., they underestimated their trait anxiety). Given that state assessments are regarded to reflect actual emotional experience more validly than trait assessments of anxiety (Robinson & Clore, 2002), our findings imply that, even though high achievers appear well adapted to the classroom environment due to lower levels of trait and state anxiety than low achievers, they actually experience more anxiety in mathematics class than they are aware of (i.e., as reported in trait self-report measures). Thus, by applying fine-grained state assessments, we get a more realistic picture of the anxiety students experience in class. Additionally, this indicates that when using trait anxiety measures as a proxy for students' actual anxiety, one has to always bear in mind the effects of other variables, such as academic achievement and academic self-concept, which can influence self-reported trait anxiety in either direction (i.e., over- or underestimation of trait anxiety).

While our study shows that trait assessments, in contrast to state assessments, are unable to measure the students' actual momentary anxiety and cannot capture situational fluctuations, they were found to be relevant in other studies: If one is interested in more distal outcomes, empirical evidence suggests that trait measures of emotions might be a better choice, because they are more predictive for future behavior and decision making, such as career choice and presumably other outcomes which are based on students' expectations and beliefs (Amelang et al., 2006; Conner & Barrett, 2012; Kihlstrom, Beer, & Klein, 2003; Krannich et al., 2019; Levine, Lench, & Safer, 2009; Robinson & Sedikides, 2009; Wirtz, Kruger, Scollon, & Diener, 2004). In conclusion, and in response to the first question posed in the introductory section, *trait assessments and state assessments are both valuable but they indeed capture different facets of a construct* and should be selected based on the outcome variable researchers want to examine.

Paper II: Component-specific assessment of self-reported anxiety

Research Paper II went one step beyond and examined the relative importance of the different components of state anxiety (i.e., cognitive, affective, physiological, and motivational components) during a mathematics test in the framework of the control-value theory by using an intra-individual approach (i.e., multiple state measurements of the anxiety components, control and performance). The second goal of this study was to investigate which anxiety component underlies the frequently used single-item measures of test anxiety.

As expected, the results of the study suggest that the different anxiety components indeed differed with regard to their control antecedents and effects on performance. In more detail, control was negatively related to all anxiety components but the associations varied in strength. In addition, the components differed in their relative impact on performance, with the cognitive component being central for this outcome. In line with the hypotheses, we found a relationship between control and performance mediated via the cognitive test anxiety component (i.e., significant indirect effect): lower control went along with an increase on the cognitive anxiety component, which in turn resulted in lower performance. These findings support the assumptions of the control-value theory and thereby additionally show that this theory on intra-individual relations indeed also holds true when examining it on an intra-individual level (Goetz et al., 2006). Since the cognitive component is the only component for which we found such a mediating effect (i.e., significant indirect effect), our results suggest that the cognitive anxiety component is most important in this mediating mechanism.

Moreover, consistent with the expectations, results revealed that with a relatively undifferentiated measure of anxiety, namely with a single-item, we most likely assess the affective anxiety component. As stated earlier, the affective component is usually regarded as the core anxiety component in multi-componential models of anxiety (Scherer, 1984). Our results thus confirm that single-item measures are capable of gauging this key component of the anxiety construct. Since this component seems to be so saliently represented in a student's mind when asked about her or his subjective anxiety with single-items, one could further speculate that this component might be important for factors other than performance (e.g., well-being, school attendance or self-esteem; Diener, 2000; Wigfield, Battle, Keller, & Eccles, 2002). This underlines that the focus should not only be on the cognitive anxiety component and that it is required to examine the affective component more closely with regard to its effects on other variables than performance in future studies. The same could apply to the physiological and motivational anxiety components, which have been understudied thus far and for which we also did not find a significant indirect effect on performance in the current study. However, this does not mean that they are not important for other factors.

Taken together, it is crucial for researchers to realize that when using single-item self-report anxiety measures such as *"How much anxiety are you experiencing right now?"*, they are most likely working with a predominantly affective measure. Therefore, when examining performance outcomes in the context of the control-value theory, assessing anxiety with a single-item measures that is focusing predominantly on the affective anxiety component might not be sufficient. In this case, items that largely tap into the cognitive component appear to be more appropriate as they are a better predictor for performance. However, for investigating more distal outcomes, single-item measures that capture the affective component, or other multiple-item measures that capture the physiological and motivational anxiety components could ultimately be a good choice. Therefore, researchers should carefully choose which measure of anxiety they are using, depending on their particular research question at hand and the implications they want to draw. In conclusion, to answer the second question posed in the introduction of this synopsis, *a differentiated component-specific assessment of anxiety is not only relevant but highly important and inevitable for certain research questions*.

Paper III: Self-report and physiological measures of test anxiety

Research Paper III synthesized research on the relationship between self-report measures of test anxiety and physiological measures - another dimension which can be considered when conducting research on anxiety - by systematically reviewing studies that applied both measures. Furthermore, the systematic review aimed at finding out whether the relationship between physiological arousal and the emotionality component of test anxiety is different from the relationship between physiological arousal and the worry component.

The systematic literature search yielded 193 articles and a structured study selection process resulted in 27 eligible articles with 29 studies. The majority of the reviewed studies (i.e., 21 studies, 72 %) found significant positive correlations between self-reported test anxiety levels and at least one measure of physiological arousal (mean significant correlation across all measures: $r_{\text{Fisher's } z} = .32$). These results are promising as they are generally consistent with theoretical concepts and support the assumption that higher self-reported test anxiety is associated with objectively measurable higher physiological arousal. Thus, it can be said that *more perceived anxiety is often also reflected in higher physiological arousal, for example a faster heartbeat* – which answers the third question from the introduction.

Furthermore, results from nine studies that reported differential results for the two test anxiety components, showed that, for the most part, correlations were stronger between the physiological measure and the emotionality component. This is in line with theoretical assumptions as the items for the emotionality component more strongly tap into aspects of physiological arousal than the items of the worry component (Deffenbacher & Hazaleus, 1985; King, Ollendick, & Prins, 2000; Spielberger & Vagg, 1995).

The overall only medium significant correlation (ranging from $r = .19$ to $r = .51$) indicates that the two measures do not completely overlap. Consequently, although the two measures are related, they should not be used interchangeably. This supports theoretical considerations, since physiological measures are assumed to only tap into the physiological component of test anxiety whereas self-report assessments usually represent all test anxiety components as subjectively experienced (i.e., also cognitive, affective, motivational, and behavioral components; Pekrun et al., 2011; Wilhelm & Roth, 2001). However, knowing that self-report measures can be strongly biased by subjective beliefs (as seen by the results from Paper I) and that they are more likely to capture the affective component of anxiety because they assess subjective experience that are represented in the conscious mind (Pekrun & Bühner, 2014), one could conclude that using physiological in addition to self-report measures is highly relevant and a promising enterprise. Together, they might provide additional information, especially about the physiological test anxiety component, that cannot be assessed with self-reports alone (Wilhelm & Roth, 2001). This implies for educational research that in order to capture the complex construct of test anxiety as objective and detailed as possible it is important to use both measures.

Additionally, we identified methodological aspects that could provide an explanation for the apparent divergence in the studies from the current review that did not find significant results ($N = 8$ studies) as well as for the relatively low convergence in the studies that found significant correlations. This leads to several implications and recommendations for future research. For instance, when applying physiological measures in future research, the study setting and sample characteristics should be carefully selected and the choice of physiological indices should be well grounded. Additionally, researchers should statistically consider influencing variables such as previous food intake or drug use (such as alcohol or smoking; e.g., Conley & Lehman, 2012) as moderators in their research designs. Moreover, as correlations seemed to emerge more often in the interaction of subject and task (i.e., during a test), it seems to be important to look at within-person associations between self-report and physiological measures across time (i.e., multiple measurement points), rather than only looking at between-person associations at a single time point.

Summary of research implications

Taken together, the findings of the present dissertation in general show that a more granular assessment of anxiety is vital. Furthermore, across all three Papers, this dissertation highlights, that it is not only important that the assessment of anxiety is fine-grained, but it also indicates that anxiety should be assessed with a combination of different measures (e.g., self-report in combination with physiological measures) and that these measures (or their combination) should be selected thoroughly based on theoretical considerations and the research question at hand. In this way, we might get the most comprehensive assessment of anxiety.

1.6.2. Practical Implications

A particular focus in two of the three research papers (i.e., Paper I and II) was on applied research questions. Thus, several implications for educational practice can also be derived.

Anxiety - An underestimated emotion in high achievers?

A first implication can be drawn from research Paper I that showed that high achievers underestimated their trait anxiety and experienced more anxiety in class than we would have expected from their trait anxiety ratings. Since high levels of trait anxiety are negatively related to well-being, motivation and learning behavior (OECD, 2016; Pekrun et al., 2002) on the one hand, it could be advantageous for them to believe that they experience less anxiety in mathematics class than they really felt (i.e., when taking state anxiety as the reference point). Moreover, an optimistic view about one's own emotions and a positive self-image could also help students handle difficult situations, as it can buffer against potential setbacks such as failing a test (Levine et al., 2009). Hence, this underestimation of trait anxiety in comparison to actual state anxiety levels due to their high academic self-concepts could be adaptive and favorable regarding high achievers' future academic choices, resiliency or general mental health (Levine et al., 2009; Taylor & Brown, 1994; Wirtz et al., 2004). Another explanation for the negative trait-state discrepancy would be that high achievers have better coping skills to deal with their state anxiety and do not interpret it as threatening in the situation, but rather as motivating. This is also consistent with the direct positive effect of the affective anxiety component on performance we found in Paper II. In detail, the results from Paper II suggest that higher state anxiety (i.e., on the affective anxiety component) during a mathematics test could be beneficial for performance. This could have resulted from a positive interpretation of the experienced nervousness during the test that might have motivated the students to work hard on the test. Consequently, it is questionable if this negative trait-state discrepancy in high achievers should be diminished as it could be considered a healthy coping mechanism.

On the other hand, this misperception of their actual anxiety (i.e., underestimating how much anxiety they really experience in mathematics class) could also have detrimental effects for high achievers in the long run. Particularly, this overconfidence may lead high achievers to overestimate their abilities, which is related to risky decision-making and could negatively impact their academic performance (Johnson & Fowler, 2011; Kahneman & Riis, 2005). Especially, in very threatening situations, high achievers might be overwhelmed by their sudden feelings of anxiety and then could lack the coping skills to adequately deal with them.

Whichever interpretation is correct, as classrooms become more and more heterogeneous (e.g., Gröhlich, Scharenberg, & Bos, 2009; Krannich et al., 2019), especially with respect to different achievement levels, it is crucial for teachers to remember that not only low achieving students, but also high achievers experience a certain amount of anxiety in class. A vital responsibility of teachers and educational practitioners is to address and reduce this anxiety. In this respect, it is necessary that teachers have excellent diagnostic competencies to enable them to recognize the individual needs of their students and provide both low and high achievers with adequate coping skills (Ohle & McElvany, 2015).

Selecting appropriate anxiety interventions

The results of Paper II showed that such coping skills and anxiety interventions should be carefully selected and that it is important to distinguish between the different components of anxiety because they differed in their relative impact on performance. As previously stated, the cognitive component had the strongest effect in the control-anxiety-performance relationship, which is consistent with the control-value theory and information processing models of anxiety (e.g., Naveh-Benjamin, 1991; Schwarzer & Jerusalem, 1992). In order to increase students' performance, it is therefore important to apply strategies that can especially target the cognitive component of anxiety. Likewise, also research on challenge and threat suggests that the interpretation of a situation as threatening and uncontrollable goes along with increased cognitive anxiety and is detrimental for performance (Moore, Vine, Wilson, & Freeman, 2012), while reinterpreting it as a challenge that can be met and overcome is related to more positive emotions and ultimately better performance. Such a positive interpretation can be achieved, for example, with cognitive reappraisal strategies that aim at changing how one thinks about the situation or about one's capacity to manage the demands it poses (Gross & Thompson, 2007). Consequently, this can increase students' perceptions of control and decreases the cognitive component of anxiety.

With regard to the positive direct effect of the affective component on performance (i.e., higher affective anxiety was related to increased performance in our mediation model), which has been briefly discussed previously, the findings imply that it could also be helpful to tell students that nervousness during a test is not always harmful for performance but rather the thoughts (i.e., cognitive anxiety component) about the consequences of failing can lead to performance decrements (Jamieson, Mendes, & Nock, 2013). However, this effect has to be further investigated in future studies, as eventually also other mechanisms could operate behind this relationship.

Furthermore, the findings of Paper II indicated that, since the affective component seems to be most saliently represented in a student's mind when asked about her or his subjective anxiety with single-items, it is reasonable to assume that it may be connected to factors other than performance, such as psychological well-being, school attendance, self-esteem, sleeping problems, or other stress related physical ailments (Austin, Partridge, Bitner, & Wadlington, 1995; Damer & Melendres, 2011; Huberty, 2010; Jamieson, Nock, & Mendes, 2013; OECD, 2016; Salend, 2012). All these factors can have consequences especially in the long run, for example for future career choice or mental and physical health (Diener, 2000; Wigfield et al., 2002). Therefore, the affective anxiety component should not be neglected by interventions and it should be more closely examined with regard to its effects on other variables than performance in future studies.

Increasing subjective control

Another practical implication that can be drawn is based on the findings about the experience of control in both empirical Papers. Control appraisals are very important as subjective beliefs about control (i.e., operationalized as academic self-concept) as shown by Paper I, or in concrete situations as shown by Paper II (i.e., control appraisals while taking a test). Paper I found that the academic self-concept influences the way students evaluate their anxiety (i.e., lower academic self-concept was related to higher anxiety) but also influences the extent to which students remember or forecast their anxiety (i.e., higher academic self-concept was associated with a lower or even negative trait-state discrepancy). Accordingly, also Paper II found that control affects all of the four anxiety components in the same direction (i.e., lower control was associated with increased anxiety; negative direct effect) and only the sizes of these effects differed. Thus, teaching practices that influence control appraisals in concrete situations as well as programs to influence long-term subjective beliefs, such as the academic self-concept, can be valuable in the classroom and will also contribute to shaping a positive test-taking environment that leads to performance facilitation and more favorable responses (Dresel & Haugwitz, 2006, 2008; Hall, Perry, Chipperfield, Clifton, & Haynes, 2006; Moore et al., 2012; Moore, Vine, Wilson, & Freeman, 2014). There are many suggestions on how to influence students' control appraisals. For example, right at the beginning of a test, the task could be framed in a manner consistent with challenge, as this has the potential to increase perceived control (Feinberg & Aiello, 2010). Furthermore, control appraisals can be influenced by providing transparency, autonomy support, and feedback during learning and test-taking (Pekrun, 2006), or by teaching students strategies such as cognitive-reappraisal or attributional retaining which they can apply in a test situation (e.g., Davies, 2010; Denny & Ochsner, 2014; Jamieson, Mendes, et al., 2013; Jamieson, Nock, & Mendes, 2012; Ruthig, Perry, Hall, & Hladkyj, 2004).

A general overall implication that can be drawn from the two empirical studies is that it may proof very helpful for teachers and educators to have background knowledge about the importance of students' anxiety and its antecedents and effects as well as about possibilities for successfully influencing them. Especially instructions and teaching techniques that foster a state of control seem to be of high relevance.

1.6.3. General Strengths, Limitations, and Considerations

The strengths and limitations and the consequent considerations of this dissertation are to be discussed in the following. A special focus is thereby placed on the advantages and disadvantages of the study designs, instruments, data analyses techniques, data abstraction, and the generalizability of the results.

Study design

An important strength of this dissertation pertains to the study designs of the two empirical studies. Anxiety and its antecedents and effects were assessed on a trait level via traditional trait-based questionnaires (Paper I), and on a state level via experience-sampling with diaries filled in at the end of each lesson (Paper I) or via state questionnaires during a mathematics test (Paper II). As both studies employed the experience-sampling method with multiple assessments on momentary experiences in realistic learning and assessment environments (i.e., in class and during a test) the study findings have high ecological validity (e.g., Csikszentmihalyi & Larson, 1992, 2014; Goetz, Bieg, & Hall, 2016; Hektner et al., 2007). Since understanding psychological processes in the real world is a key driver to answering applied research questions and most importantly to constructing intervention programs (Jones & Johnston, 2011), researchers should be encouraged to use this method more frequently - not only to assess the construct of anxiety, but also other constructs such as motivation or well-being.

Instruments

Furthermore, it should be highlighted that in both empirical Papers we used well-established instruments that contained multiple items to assess-trait and state mathematics anxiety and academic self-concept (Paper I) and the different test anxiety components (Paper II), which is a clear strength of these studies. In order to keep the state assessments during the mathematics test less intrusive and as short as possible and because we did not assume that there are different components of control, we used a single-item to assess state control appraisals in Paper II. Moreover, research has shown that it is reasonable to use single-item measures when long scales are not applicable (Gogol et al., 2014; Nagy, 2002; Wanous & Hudy, 2001), and we knew from previous studies (e.g., Bieg, Goetz, & Hubbard, 2013) that single state items that were used to assess control, had very high item-scale correlations. Nevertheless, this may be viewed as a potential shortcoming and future research could consider also assessing state control experience with multiple items.

Data analyses techniques

Another strength of the two empirical studies pertains to the analyses. In both Papers an intra-individual perspective was taken and multilevel regression models that are accounting for the nested data structure were used. This seemed vital as all data sets were structured with multiple measurement points per student (and students being nested in classes) and especially in the data of Paper II there was high inter-individual variance, as well as variance within students from task to task (see Figure 7 in Paper II).

The advantage of the hierarchical linear analytic approach we used in Paper I is the unbiased estimation of standard errors, accounting for different numbers of measurement points. Furthermore, it is a sophisticated way to account for the discrepancy between trait and state measures of anxiety (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011; Raudenbush & Bryk, 2002) and we could also include potential mediators and moderators of the trait-state discrepancy, such as achievement level and academic self-concept (for a similar approach see Bieg et al., 2014).

An important benefit of the intra-individual approach we used in Paper II is that, as one of the first studies, we could investigate the control-anxiety-performance relationship on an intra-individual level. The control-value theory actually refers to within-person functioning, but the majority of emotion research is based on between-person comparisons instead of manipulating or assessing variations within an individual (Hamaker, 2012; Voelkle, Brose, Schmiedek, & Lindenberger, 2014). In this way, with our approach we could contribute to closing a long-standing research gap and proof that the control-value theory holds true on an intra-individual level. Since it can sometimes be critical to draw conclusions from inter-individual analyses on intra-individual functioning, researchers should be encouraged to use intra-individual analytical approaches more often (Molenaar, 2004; Molenaar & Campbell, 2009; Murayama et al., 2017). Despite these encouraging findings, a point open to critique is that from a methodological point of view, although the sample size in the second study was sufficient for conducting intra-individual analyses, yet, is was still comparably small, which may have influenced the statistical power of the study (Maas & Hox, 2005). Future investigators could ensure a high number of participants in their studies to increase the power. A further point worth mentioning in this respect is that although we used multiple measurement points to assess state anxiety in Paper I and II, especially in Paper II, the analyses would have benefited from more measurement points. However, conducting more state assessments can be intrusive (Goetz, Frenzel, Stoeger, & Hall, 2010) and the number of assessments was restricted by the timeframe the schools provided for the data collection. Future studies could try to optimize the research design and thereby increase the number of state assessments, which would allow for running even more complex intra-individual models, like for example cross-lagged models (Bringmann, Ferrer, Hamaker, Borsboom, & Tuerlinckx, 2018; Hamaker, Ceulemans, Grasman, & Tuerlinckx, 2015; Hamaker & Wichers, 2017; Schuurman, Ferrer, de Boer-Sonnenschein, & Hamaker, 2016).

Data abstraction

In the last Research Paper, the systematic review, we not only qualitatively summarized the selected research articles, but also systematically investigated the relationship between self-report and physiological measures and determined a correlation between the two assessment methods. This is a clear strength of this Paper, as such a correlation index is important for future research in order to know the size of the correlation coefficients they can expect between the two measures. In addition, the systematic review addressed potential influencing variables of the relationship between the self-report and physiological measures. In particular, the studies that found non-significant correlations proved very informative, since future research can learn from them and can consider and control the factors thus identified when applying physiological measures in their studies. Moreover, the current review may be a pioneering step towards the development of guidelines for the application of physiological measures in educational settings.

Generalizability

Our sample of high school students in both empirical studies (Paper I and II) was relatively selective as we examined students from grades 8 to 10 in the top track of the German school system (i.e., "Gymnasium"; about one-third of a student cohort attend this school track; Federal Statistical Office, 2015). However, for Paper I we needed to be selective to ensure that we received a sample of students with high achievement levels. Regarding Paper II, although we do not assume that the selection of the sample may have limited the generalizability of the findings, since from a theoretical perspective, the proposed structural relations should be universal across different samples and different age groups, it would be interesting to investigate these effects in students from different school types and with different age groups. Furthermore, although the focus on mathematics in both empirical studies allowed us to draw conclusions related to this field as a very relevant STEM (science, technology, engineering, and math) domain, the exploration of a possible trait-state discrepancy as well as the examination of the control-anxiety-performance relationship in other subject domains than mathematics is indicated as well.

Summary

Taken together, the two empirical studies conducted in the present dissertation benefit from novel research designs that utilized repeated state assessments in ecologically valid settings, which allow for intra-individual analyses. Likewise, the systematic review has a clear strength as it is the first Paper that provided a correlation between self-report and physiological measures and thereby can help to advance physiological research in education and provide a deeper understanding of the complex and mostly detrimental emotion of test anxiety. Yet, the research Papers in the current dissertation are not without their limitations.

1.6.4. Future Directions

In the following section I want to continue with some potential routes for future research based on the findings as well as the identified strengths and limitations of the current dissertation.

Beyond academic self-concept and control

The results of the present study concerning achievement level as a moderator of the trait-state discrepancy and the academic self-concept as a mediator of this relationship allow researchers to roughly estimate the extent to which trait and state anxiety assessments resemble each other when a students' performance level and self-concept is known. In this respect, the academic self-concept in mathematics, due to its high explanatory power in Paper I (65.3%), seems to play a major role. However, further identifying moderating and mediating variables of the trait-state discrepancy beyond academic self-concept in future research, such as perceived value, gender stereotypes, or interest in mathematics may contribute to our understanding of how reliable a trait assessment is for actual state assessments when additional variables are also available.

Likewise, identifying such additional variables seems to be important for the relationship between control, the anxiety components and performance in Paper II. Although in this Paper, we found a partial mediation of the relationship between control and performance via the cognitive anxiety component, there was a direct positive effect of control on performance. This effect could be explained in part by other variables that might play an additional mediating role (e.g., other emotions or variables such as self-efficacy or value) and thus corresponds with assumptions of the control-value theory. Especially subjective value appraisals might be very important, because they are a central variable in the control-value theory, in such a way that they interact with control antecedents (Pekrun, 2006). Therefore, our results suggest that it is not only necessary to regulate anxiety, but also to consider additional variables during test taking, as they might affect performance as well (Goetz, Preckel, Pekrun, & Hall, 2007). Such variables should be incorporated in future research designs.

Complimentary methods to self-report measures

The two empirical studies conducted in the present dissertation used multiple methods to assess students' anxiety and its antecedents and effects in the classroom and during a mathematics exam. Yet, all methods relied on self-reports, which can be rather intrusive, especially in situations such as a test (Pekrun & Bühner, 2014). As indicated by the results of Paper III, future research could benefit from recent advancements in technology, allowing for ambulatory assessment of physiological indices. This is also a reoccurring issue on many conferences.

By combining self-reports with continuous physiological measurement, for example during an exam, we might be able to expand our knowledge of the mechanisms underlying test anxiety in ways not possible using self-report measures alone. In the long run, this more ecologically valid, more objective and detailed assessment might also aid the design of effective theory-led interventions (Jones & Johnston, 2011).

Furthermore, by additionally enriching this data with data from observational techniques (e.g., Donker, Van Gog, & Mainhard, 2018; Trull & Ebner-Priemer, 2014), or data from automated facial coding with software such as "Facereader" or "FACET" (D'Mello, Dieterle, & Duckworth, 2017; D'Mello & Graesser, 2010; Frenzel, Marx, Pekrun, Reck, & Müller, 2019; Harley, 2015; Lewinski, den Uyl, & Butler, 2014; Littlewort et al., 2011; Taub, 2018), we could objectively assess an additional component of anxiety, namely the behavioral-expressive component, which is otherwise not readily accessible via self-report (Harley, 2015). Furthermore, such methodological triangulations can increase the validity and reliability of the findings (D'Mello & Calvo, 2013; Kantor, Endler, Heslegrave, & Kocovski, 2001; Onwuegbuzie, Johnson, & Collins, 2010).

Although beyond the scope of the systematic review in the current dissertation, considerable progress has also been made in elucidating the various brain systems that control the physiological response to anxiety. These include the limbic system in the medial temporal lobe, in particular the amygdala, as well as higher-level cortical brain activity (Barrett & Wager, 2006; Caprara & Cervone, 2000; Immordino-Yang, Christodoulou, Pekrun, & Linnenbrink-Garcia, 2014). Examining these brain systems could also be a route for future research and further enrich our understanding of the underlying mechanisms of anxiety.

Another interesting approach that goes beyond mere self-reports is to investigate and measure anxiety and other emotions by mapping them in the body. The studies by Nummenmaa, Hari, Hietanen, and Glerean (2018) and Nummenmaa, Glerean, Hari, and Hietanen (2014) used such a topographic approach in which participants reported where they experienced a certain emotion in the body and found that different emotions were associated with culturally universal, topographically distinct bodily sensations.

Cross-cultural research

Another avenue for future research would be to examine cultural differences regarding the trait-state discrepancy in anxiety (Paper I). Eventually, other cultures exhibit an even higher or lower trait-state discrepancy in mathematics anxiety because in their culture it is not common to admit that one is generally anxious in mathematics. This could apply for instance to Eastern societies (i.e., Asian countries), as they were found to suppress their emotions more often because this is more socially and positively evaluated by others, as compared to people from Western cultures (Butler, Lee, & Gross, 2007; De Leersnyder, Boiger, & Mesquita, 2013; Ma, Tamir, & Miyamoto, 2018; Qu & Telzer, 2017). Cultural differences would also be worth investigating with regard to test anxiety (Bodas & Ollendick, 2005) and especially the role of the anxiety components in the mediating mechanism that connects control and performance (Paper II). Perhaps for students with other cultural backgrounds, another component of anxiety is more relevant for performance, or there may also be differences between cultures when it comes to the intensity at which the different components are rated.

Gender differences

An important aspect that unfortunately is still an issue in educational contexts, is gender differences (Else-Quest, Hyde, & Linn, 2010; Maloney, Waechter, Risko, & Fugelsang, 2012). Although, we did not find any gender differences in state anxiety in mathematics class in Paper I, in Paper II we found differences between female students in state test anxiety and control experiences, but not in performance. This is in line with earlier findings reporting gender discrepancies in emotional experiences, even though girls and boys had achieved at similar levels in mathematics (e.g., Frenzel et al., 2007; Goetz et al., 2013; OECD, 2004; meta-analysis by Seipp, 1991). Hence, our results demonstrate that there still remains a gap between female and male students in terms of competence beliefs and emotions in mathematics tests. Future work could more closely examine these effects, in particular with regard to interventions that aim at closing this long-standing gender gap. Since regrettably, women are severely underrepresented in STEM fields and one factor that has been implicated in female students' avoidance of careers in these fields is math anxiety (Chipman, Krantz, & Silver, 1992), this is especially important with regard to the subjects of mathematics and science.

Emotional contagion of anxiety

Finally, as anxiety does not only develop within a student, but it can also be "transmitted" to other students or it can be influenced by other emotions of the surrounding persons (Hartel & Page, 2009), another route for future research would be to investigate emotional contagion of anxiety. This is especially interesting in a classroom setting where, in addition to the classmates' emotions, the teachers' emotions might also contribute to the anxiety students experience in class (Becker, Goetz, Morger, & Ranellucci, 2014; Frenzel, Becker-Kurz, Pekrun, Goetz, & Lüdtke, 2018; Frenzel, Goetz, Lüdtke, Pekrun, & Sutton, 2009).

1.6.5. Conclusion

Starting with a general (i.e., trait) assessment of anxiety, followed by a more situation specific (i.e., state) and more differential (i.e., component-specific) assessment, up to a more objective and often continuous (i.e., physiological) measurement, the present dissertation compared different methods to assess anxiety in educational settings and highlighted their similarity and differences. This procedure resulted in a very fine grained and complete picture of anxiety as experienced in educational settings.

The results from the two empirical studies in the present dissertation confirm that trait and state conceptualizations of anxiety are different regarding their intensity ratings (Paper I) and that the assessment method employed is highly important with regard to the conclusions one wants to make. The same applies to the different components of anxiety, which differ with regard to their control antecedents and effects on performance (Paper II). Although it may seem trivial at first, these differentiations have not been explicit enough in the past. Furthermore, the results of the present dissertation emphasize that anxiety can be influenced. With regard to performance, for state anxiety this could be done by altering students' control during a test or in class and for trait anxiety this could be done by influencing students' control beliefs (i.e., academic self-concept). Concerning variables other than performance, the results point to the fact, that it might be prolific to more closely examining the other anxiety components. The systematic review indicates that there is a substantial relationship between self-reported test anxiety and physiological arousal as assessed with physiological measures (which might also depend on methodological aspects), but the moderate correlations suggest that applying both measures may add to a better understanding of this complex emotion. Thus, overall, although it can be quite labor-intensive, the findings from the present dissertation support the continued use of self-report measures while combining them with physiological measurement to gain a more complete view of anxiety in educational settings.

As a conclusion, the present dissertation underpins the notion that "*The person's subjective sense of being anxious depends on a complex integration of the different components of anxiety, depending on the individual*" (Zeidner & Matthews, 2011, p. 19) and emphasizes the fact that the assessment method adopted to capture this emotion is of vast importance. It will ideally stimulate researchers to consider using physiological measures in combination with self-report measures in classroom-based settings. Most importantly, it will hopefully also encourage them to clearly distinguish between trait and state self-report measures of anxiety, as well as the different components and their assessments in future studies, and to choose appropriate methodological approaches guided by the research question under investigation - Since *One Measure Can't Capture It All*.

2 Research Paper I

Experiencing More Mathematics Anxiety Than Expected? Contrasting Trait and State Anxiety in High Achieving Students

Anna-Lena Roos [1,2], Madeleine Bieg [1], Thomas Goetz [1,2], Anne Frenzel [3], Jamie Taxer [3], & Moshe Zeidner [4]

(2015)

High Ability Studies, 26(02), 245-258

doi: 10.1080/13598139.2015.1095078

[1] Empirical Educational Research, University of Konstanz, Konstanz, Germany

[2] Thurgau University of Teacher Education, Kreuzlingen, Switzerland

[3] Ludwig Maximilians University, Munich, Germany

[4] University of Haifa, Haifa, Israel

Experiencing More Mathematics Anxiety Than Expected? Contrasting Trait and State Anxiety in High Achieving Students

2.1. Abstract

This study examined mathematics anxiety among high and low achieving students (N = 237, grades 9 and 10) by contrasting trait (habitual) and state (momentary) assessments of anxiety. Previous studies have found that trait anxiety measures are typically rated higher than state measures. Furthermore, the academic self-concept has been identified to play a moderating role in the trait-state discrepancy, with higher academic self-concept leading to a lower discrepancy (i.e., less overestimation of trait anxiety if state assessments reflect actual experience). Therefore, we assumed that high achievers who were expected to have high academic self-concepts would exhibit a smaller trait-state discrepancy than low achievers. Results confirmed these assumptions and revealed that high achievers even underestimated their trait anxiety. Implications are discussed.

Keywords: state; trait; intensity bias; anxiety; academic self-concept; high achievers

2.2. Introduction

In comparison to low achievers, high achievers generally face fewer problems at school. One widespread problem in the school context is anxiety, which is a detrimental emotion in the classroom particularly in the domain of mathematics (Goetz & Hall, 2013). A meta-analysis by Ma (1999) on math anxiety and its scholastic outcomes found that high achievers tend to be less anxious about mathematics than their low achieving peers. However, when comparing different levels of anxiety, it is important to distinguish between different assessment methods.

Students' class-related anxiety can be assessed using either trait or state assessment methods. Trait measures examine habitual anxiety, whereas state measures examine momentarily experienced anxiety in real-life situations (Spielberger et al., 1970). Existing studies on mathematics anxiety have predominately used the less costly and more convenient trait assessment method (Wilhelm et al., 2012). However, in comparison to scores obtained by multiple state assessments, which are regarded as reflecting actual emotions as they are experienced (Robinson & Clore, 2002), trait assessments of emotions are generally overestimated (Buehler & McFarland, 2001). It has been argued that the discrepancy between trait and state measures of emotions (i.e., overestimation of trait emotions) results from trait measures being recall based, which causes them to be strongly influenced by subjective beliefs (Robinson & Clore, 2002). Thus, trait measures are often believed to not reflect actual emotions (Robinson & Clore, 2002). This raises the question of whether trait measures of anxiety are able to validly gauge what students actually feel in the classroom and calls for applying a state as well as a trait approach to assess students' emotions.

Therefore, the aim of the present study was to compare the discrepancy between trait and state assessments of mathematics anxiety among the particular groups of high and low achievers. Thus, we assessed students' momentary mathematics anxiety during real-life mathematics class with a state-based measure and compared these ratings with the students' trait mathematics anxiety ratings. This rendered it possible to evaluate whether findings from previous studies assessing high achieving students' trait anxiety hold true when actual state anxiety measures are also assessed.

2.3. The Trait-State Discrepancy in Anxiety and the Role of Academic Self-Concept

The discrepancy between trait and state self-reports of emotion (referred to from here on as the trait-state discrepancy) reflects the observation that trait emotions are usually overestimated in comparison to state emotions (also termed sometimes as intensity bias; Buehler & McFarland, 2001). In their *accessibility model of emotional self-report*, Robinson and Clore (2002) postulated that subjective beliefs play an important role in the trait-state discrepancy. Their model proposes that trait measures, which examine habitual emotions, are strongly influenced by subjective beliefs because they are recall based. Whereas subjective beliefs are less impactful for state measures of emotions because they assess emotions as they are experienced. Therefore, state measures of momentary emotions are considered to evaluate actual emotions, while trait measures reflect individual's *beliefs* about emotions (Robinson & Clore, 2002).

Concerning the trait-state discrepancy in anxiety, previous studies have found that academic self-concept, which is an identity related-subjective belief, influences the extent trait anxiety measures are overestimated in comparison to state anxiety measures (Bieg et al., 2014; Goetz et al., 2013). The academic self-concept is a mental representation of the abilities and competencies a person has (Nagengast & Marsh, 2012) and can be defined as an evaluative self-perception that is formed through experience with and interpretation of one's school environment (Marsh & Craven, 1997). Specifically, studies have found that the trait-state discrepancy for anxiety (i.e. the overestimation of trait anxiety) is *smaller* for students with *higher* academic self-concepts (Bieg et al., 2014; Goetz et al., 2013). This smaller trait-state discrepancy in students with higher academic self-concepts implies that these students show a lower overestimation of trait in comparison to state anxiety levels. Since state measures are assumed to assess actual emotions, a smaller trait-state discrepancy in emotional self-reports can be viewed as a more realistic evaluation of trait anxiety because it is less overestimated. Thus, students with a high academic self-concept should evaluate their anxiety more realistically than students with a low academic self-concept. Research comparing the academic self-concepts of high and low achieving students has shown that high achieving students report higher levels of academic self-concept (Bailey, 1971). Therefore, we assumed that high achievers, who are also expected to have high academic self-concepts, would have a smaller trait-state discrepancy in mathematics anxiety (i.e., less overestimation of trait anxiety), meaning that they would evaluate their anxiety more realistically.

2.4. The Present Research

The aim of the present study was to examine the trait-state discrepancy in mathematics anxiety among high and low achievers, and to further investigate academic self-concept as a predictor of the trait-state discrepancy. Prior studies have shown that academic achievement (Ma, 1999) is negatively related to trait mathematics anxiety. Furthermore, high achievers can be assumed to have higher subjective control appraisals in classroom situations than low achievers, which as assumed by the control-value theory (Pekrun, 2006) should result in high achievers also experiencing less state anxiety than low achievers. Therefore, we assumed that high achievers would experience less trait as well as state mathematics anxiety than low achievers (Hypothesis 1). Furthermore, since previous studies have found a discrepancy between ratings of trait and state mathematics anxiety, which was smaller in students with a higher academic self-concept in mathematics (Bieg et al., 2014; Goetz et al., 2013), we expected high achievers to have a smaller trait-state discrepancy in mathematics anxiety than low achievers. This smaller discrepancy was expected to be due to a large extent to high achievers having a higher mathematics self-concept than low achievers. In more technical terms, we expected achievement group (high vs. low achievers) to moderate the trait-state discrepancy in mathematics anxiety (Hypothesis 2). Furthermore, we expected mathematics self-concept to partially mediate the relationship between the achievement groups (high vs. low achievers) and the trait-state discrepancy (Hypothesis 3; i.e., mediated moderation; Preacher, Rucker, & Hayes, 2007). Figure 2 is a graphical depiction of our expectations for the high achieving students.

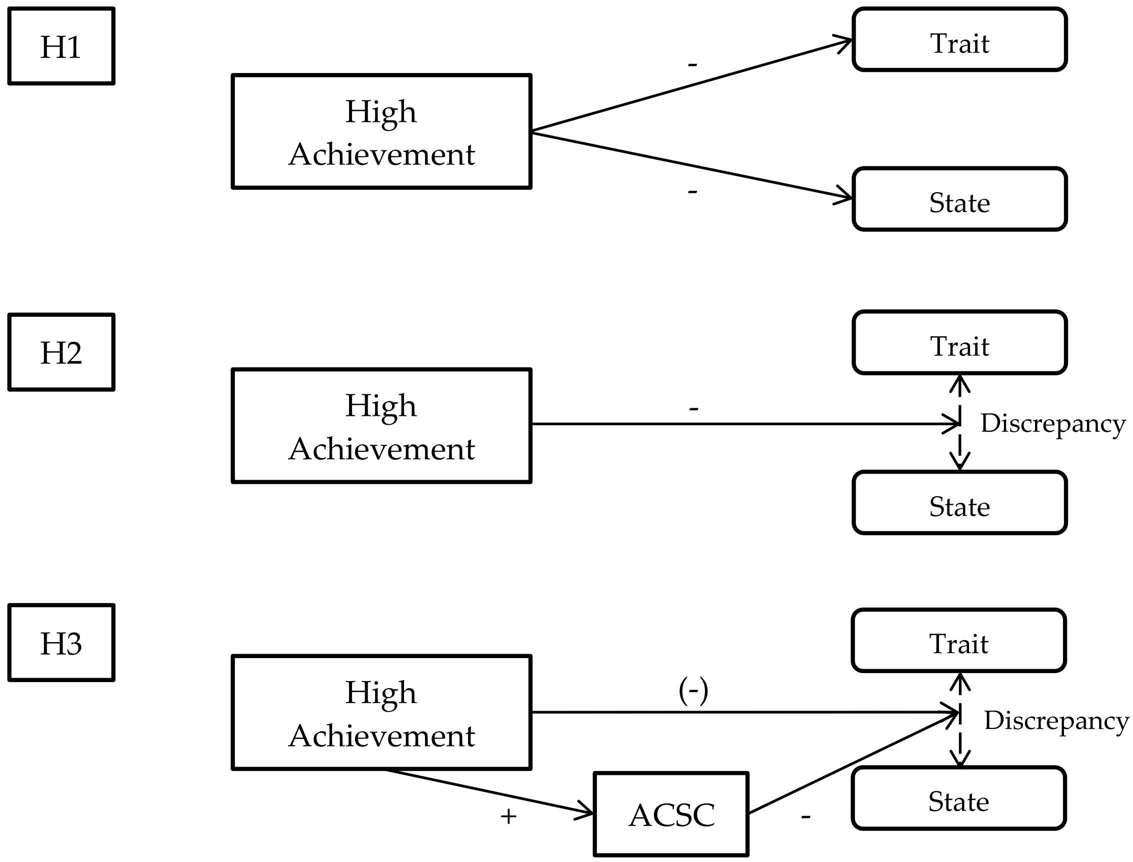

Figure 2. Graphical depiction of the expected results for the high achieving students (hypotheses 1 to 3).

Note. H = hypothesis, ACSC = academic self-concept.

2.5. Methods

2.5.1. Sample

The present study was part of a large research project that examined students' emotions in mathematics using both trait and state assessment methods. For the current study, a subsample of *N* = 237 high and low achieving students (from 43 classes at 21 schools;

47% female, M_{age} = 15.71 years, *SD* = 0.78) was retained from an initial sample consisting of *N* = 828 students (50.6% female). The students were in the ninth (57.4 %) and tenth (42.6 %) grades and attended the top track of the German education system (i.e., Gymnasium; approximately one third of the total student cohort in Germany attend this track; Federal Statistical Office [Statistisches Bundesamt], 2015).

For the present study the *high* and *low* achievement groups were defined using students' self-reported midterm mathematics grades. Grades in the German school system range from 1, *very good*, to 6, *insufficient*, with higher numbers representing poorer performance. Midterm grades additionally differentiate within grades by applying 0.25 gradations (e.g., grade 1 is subdivided in grades 1, 1.25, 1.5 and 1.75). In this study high achievers were defined as students whose mathematics grades were better than *good* (i.e., < 2). Low achievers were defined as students whose mathematics grades were *sufficient* or worse (grades 4 to 6). Based on this definition, *N* = 116 students (14 % of the entire sample, 55 % female) were labeled as high achievers and *N* = 121 students (14.6 % of the entire sample, 49 % female) were labeled as low achievers. Since we were only interested in the two extreme groups of high and low achievers, students with a *good* or *satisfactory* academic performance (grades 2 to 3.75) were excluded from the present analyses.

2.5.2. Procedure

After students consented to participate in the study, trait mathematics anxiety, mathematics self-concept and demographic data were assessed using a paper-and-pencil questionnaire that was administered by trained researchers. The state assessment phase began directly after the trait assessment of anxiety and lasted for 3 weeks. State mathematics anxiety was assessed via a personal diary that was designed as a short questionnaire to be filled out by each student after each mathematics lesson. Students who made fewer than three diary entries were excluded from the analyses. In total, this resulted in *N* = 1,687 state assessments (diary entries) with an average of 7.12 (*SD* = 2.40; range: 3 to 14) state assessments per student.

2.5.3. Study Measures

Trait and state mathematics anxiety

Since we wanted to compare levels of mathematics anxiety as assessed by trait versus state measures, we used parallel item wording and an identical answer format for both assessment types. Trait and state mathematics anxiety were each assessed with two items. Trait anxiety items were developed for a study that analyzed learning and achievement in mathematics (PALMA; Goetz, 2004; Pekrun, Goetz, & Frenzel, 2006). The *trait anxiety* items included: "When thinking of mathematics classes, I am worried" and "I usually experience anxiety in mathematics classes". The sample specific internal consistency of this 2-item scale was α = .63. The parallel item wordings for the *state anxiety* items were as follows: "In this mathematics class I was worried" and "In this mathematics class I experienced anxiety" (the reliability of this 2-item scale was α = .75). Responses were made on a 5-point Likert-type scale ranging from 1, *strongly disagree*, to 5, *strongly agree.*

Assessment of mathematics self-concept

The academic self-concept in the domain of mathematics was assessed via three items (the reliability of this 3-item scale was α = .94) adapted from Marsh's Self-Description Questionnaire (German version, Kunter et al., 2002; sample item: "I have always done well in mathematics"; SDQ; Marsh, 1990). Responses were made on a 5-point Likert-type scale ranging from 1, *strongly disagree*, to 5, *strongly agree.*

2.5.4. Data Analysis

Comparison of mean levels

As a first step, we conducted *t*-tests to compare high and low achievers' trait anxiety, state anxiety and academic self-concepts. Effect sizes are reported as Cohen's *d* (Cohen, 1988).

Hierarchical Linear Regression

Similar to previous research (e.g., Bieg et al., 2014; Goetz et al., 2013), we used a multilevel, intra-individual modeling approach to evaluate our hypotheses. This analyses method was chosen to account for the nested structure of the data (multiple measurement points per student and students from different classes) and because there were different numbers of state assessment measurement points per student. Our three-level multilevel analyses (measurement points nested within students and students nested within classrooms) were conducted using Hierarchical Linear Modeling software (Version 7.01; Raudenbush et al., 2011).

Level 1 (measures within students). Since we were interested in the discrepancy between students' trait and state mathematics anxiety ratings (i.e., trait-state discrepancy), students' anxiety scores served as the outcome variable in all the hierarchical linear regression models. We formed this outcome variable by including the one trait anxiety score and several state anxiety scores of each student into one variable. Thereby, the outcome variable comprised of two types of measures (one trait and several state measures) for each person. Including the combined variable in the analyses is only meaningful when a dichotomous independent variable, in the present analyses labeled as "State/Trait" (uncentered), is simultaneously introduced on Level 1. In this dichotomous variable the state measures were coded as 0 and the trait measures were coded as 1. This coding scheme allowed the state and trait measures of anxiety to be separated in the analyses and the trait-state discrepancy in anxiety to be examined. 1. Due to the coding of this variable, the intercept value (γ_{000}) in all of our models represents the overall state anxiety mean when other linear terms are also coded as 0. The effect of the State/Trait variable (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 (trait-state discrepancy).

Level 2 (student level). Two Level 2 variables were included in the respective models, each as a predictor of the intercept and slope of the State/Trait variable. These Level 2 variables were "High/Low Achievers" (0 = high, 1 = low, uncentered) and "Mathematics Self-concept" (z score standardized across students).

Cross-level interactions Level 1–Level 2. To analyze the effect of achievement group (High/Low Achievers) on the discrepancy between trait and state anxiety scores, a cross-level multiplicative interaction term was included in the respective models to predict the slope, namely State/Trait × High/Low Achievers (γ_{110}). To analyze the effect of academic self-concept in mathematics on the trait-state discrepancy, an additional interaction term was included, namely State/Trait × Mathematics Self-Concept (γ_{120}).

Level 3 (class level). The different classes (N = 43) in which students were nested constituted the third level. This level took the clustering of students within classes when estimating standard errors into account.

Calculated Models

In order to test the study hypotheses as illustrated in Figure 1, the models described below were calculated, each of which was constructed as a slopes-as-outcome model (Raudenbush & Bryk, 2002). The mixed model regression equations for the three models were as follows:

Model 1

Y_{ijk}[Anxiety measures i of student j in class k] = γ_{000} + γ_{100}(State/Trait) + γ_{010}(High/Low Achievers) + γ_{110}(State/Trait*High/Low Achievers) + r_{0jk} + r_{1jk} (State/Trait) + u_{00k} + e_{ijk}

In *Model 1* we investigated whether high achievers have lower state anxiety than low achievers (*Hypothesis 1*) by examining the effect of achievement group (High/Low Achievers; γ_{010}) on state anxiety. In this model a significant positive effect of achievement group indicates that low achievers have significantly higher state anxiety ratings than high achievers. This model also investigated whether high achievers showed a smaller trait-state discrepancy than low achievers (*Hypothesis 2*) by examining the effect of the State/Trait × High/Low Achievers interaction (γ_{110}). A significant positive effect of this interaction indicates that low achievers have a significantly larger trait-state discrepancy than high achievers.

Model 2

Y_{ijk}[Anxiety measures i of student j in class k] = γ_{000} + γ_{100}(State/Trait) + γ_{020}(Mathematics Self-concept) + γ_{120}(State/Trait*Mathematics Self-concept) + r_{0jk} + r_{1jk} (State/Trait) +u_{00k} + e_{ijk}

In *Model 2* we investigated whether mathematics self-concept could predict the discrepancy between trait and state anxiety scores by examining the effect of the State/Trait × Mathematics Self-concept interaction (γ_{120}). A significant negative effect of this interaction indicates that higher levels of mathematics self-concept correspond with a smaller trait-state discrepancy in mathematics anxiety.

Model 3

Y_{ijk}[Anxiety measures i of student j in class k] = γ_{000} + γ_{100}(State/Trait) + γ_{010}(High/Low Achievers) + γ_{020}(Mathematics Self-concept) + γ_{110}(State/Trait*High/Low Achievers) + γ_{120}(State/Trait*Mathematics Self-concept) + r_{0jk} + r_{1jk} (State/Trait) +u_{00k} + e_{ijk}

Model 3 examined mathematics self-concept as a possible mediator of the trait-state discrepancy in mathematics anxiety among high versus low achievers (*Hypothesis 3*). In this model, achievement group (High/Low Achievers) and mathematics self-concept were both included as predictors of the trait-state discrepancy (γ_{110}, γ_{120}). With this model we were able to see whether the effect of achievement group (High/Low Achievers) on the trait-state discrepancy (γ_{110}), as tested by *Model 1,* was reduced when mathematics self-concept was included in the model (mediated moderation).

To fully meet the criteria of mediated moderation as postulated by Preacher, Rucker, and Hayes (2007), we additionally tested whether achievement level predicted mathematics self-concept (mediator). Since these variables are both within person variables we examined the Spearman correlation between mathematics self-concept and achievement because achievement was measured on an ordinal scale.

2.6. Results

2.6.1. Comparison of Mean Levels

Table 1 provides an overview of the mean level comparisons. *T* tests revealed that, as anticipated, high and low achievers significantly differed in their trait and state anxiety levels and mathematics self-concepts. Specifically, high achievers reported less trait and state anxiety and higher mathematics self-concepts than low achievers. The effect sizes were medium to large. These results partly confirmed *Hypothesis 1,* in which we assumed that high achievers would report lower levels of trait anxiety than low achievers. Since trait anxiety was assessed with a single measurement point per student, we were able to use a *t* test to examine the prediction made in *Hypothesis 1* regarding trait anxiety. However, the prediction that high achievers would experience less state anxiety than low achievers (the second part of *Hypothesis 1*) could not be tested with a t test due to the differing number of state anxiety assessments per student. So, we used *Model 1* to further test the predictions made in *Hypothesis 1.*

Table 1. Descriptive statistics, mean-level differences, and related effect sizes of key study variables

	High Achievers (N = 116)		Low Achievers (N = 121)			
Measure	M	SD	M	SD	$T(df)$	Cohen's d
Trait Anxiety	1.36	0.61	2.38	1.07	$t(191.96) = -9.12$***	-0.79
State Anxiety	1.53	0.55	1.92	0.58	$t(235) = -5.31$***	-0.37
Mathematics Self-concept	4.51	0.68	1.79	0.67	$t(235) = 30.97$***	2.34

Note. Positive t values reflect higher scores for high achievers than for low achievers. For multi-item measures, item means were used to calculate scale values. * $p < .05$, ** $p < .01$, *** $p < .001$.

2.6.2. Hierarchical Linear Regression

The results of the hierarchical linear regressions, including the regression coefficients and the variance components, as well as the explanatory power, are depicted in Table 2.

Model 1

In Model 1 the coefficient for the intercept (γ_{000}) is interpreted as the mean state anxiety score for high achievers (γ_{000} = 1.53). Results showed that the effect of achievement group (High/Low Achievers; γ_{010}) on anxiety scores was significant (γ_{010} = 0.36, $p < .001$). This effect corresponds to the t tests results and can be interpreted as state anxiety scores being significantly lower in high achievers than in low achievers, which supports our first hypothesis. Analyses further revealed that the coefficient for the State/Trait variable (γ_{100}) was significant (γ_{100} = -0.17, $p < .05$) and revealed that trait anxiety was lower than state anxiety in high achievers. This means that high achievers even underestimated their anxiety when judging it from a trait rather than a state perspective. Since this result was unexpected we examined it further with post-hoc analyses, which are described in the last paragraph of the results section.

The effect of the State/Trait x High/Low Achievers interaction (γ_{110}) was significant (γ_{110} = 0.64, $p < .001$) and revealed that the discrepancy between trait and state mathematics anxiety was higher for low achievers, as was proposed in *Hypothesis 2*. Hence, achievement level moderated the trait-state discrepancy in anxiety.

Model 2

In this model the intercept coefficient (γ_{000}) is interpreted as the mean state anxiety score for a student with a mean mathematics self-concept score (mathematics self-concept was z-standardized). The effect of the State/Trait x Mathematics Self-concept (γ_{120}) interaction was significant ($\gamma_{120} = -0.34$, $p < .001$). This effect showed that higher levels of mathematics self-concept correspond with smaller trait-state discrepancies in mathematics anxiety.

Model 3

In Model 3 the effect of the State/Trait x Mathematics Self-concept (γ_{120}) interaction was significant ($\gamma_{120} = -0.30$, $p < .001$) and the effect of achievement group (High/Low Achievers) on the trait-state discrepancy (γ_{110}) was no longer significant. In this model the moderating effect of achievement group (High/Low Achievers) on the trait-state discrepancy (γ_{110}) found in *Model 1* was reduced and no longer significant when the State/Trait x Mathematics Self-concept interaction term (γ_{120}) was included.
The results of the Spearman correlation analyses showed that achievement and academic self-concept were significantly correlated, with higher achievement accompanied by higher levels of self-concept ($r = -.82$, $p < .01$; negative correlation due to higher grades indicating lower achievement in the German grading system).[4] Therefore, academic self-concept met the criteria to be a mediator between achievement group and the trait-state discrepancy.

Along with the reduced effect of the State/Trait x High/Low Achievers (γ_{110}) coefficient when the State/Trait x Mathematics Self-concept (γ_{120}) interaction was added in *Model 3*, this shows that the moderating effect of achievement group (High/Low Achievers) on the trait-state discrepancy in mathematics anxiety (as found in *Model 1*) was mediated by academic self-concept. Thus, the described pattern is a mediated moderation (Preacher et al., 2007). Our results therefore support the assumption that achievement group moderates the trait-state discrepancy in mathematics anxiety (*Hypothesis 2*) and that the academic self-concept in the domain of mathematics mediates the relationship between achievement group and the trait state-discrepancy in mathematics anxiety (*Hypothesis 3*).

[4] The correlation between achievement and academic self-concept was quite high ($r = -.82$) which was mainly due to only comparing the two extreme groups (high vs. low achievers). As assumed, the Pearson correlation between achievement level and academic self-concept in the whole sample (with the whole range of grades and levels of self-concept) was lower ($r = -.75$, $p < .01$). The size of this correlation as related to the whole sample was in line with findings from numerous previous studies and meta-analyses (e.g., Goetz et al., 2013; Möller, Pohlmann, Köller, & Marsh, 2009; Valentine, DuBois, & Cooper, 2004).

Table 2. Results of models predicting mathematics anxiety

Level and predictor	Model 1	Model 2	Model 3
Level 1			
Intercept (γ_{000})	1.53***	1.70***	1.85***
	(.05)	(.04)	(.10)
State/Trait Anxiety (γ_{100})	-0.17*	0.13*	0.08
	(.07)	(.06)	(.11)
Level 2			
High/Low Achievers (γ_{010})	0.36***		-0.31
	(.09)		(.19)
Mathematics Self-concept (γ_{020})		-0.23***	-0.36***
		(.04)	(.09)
Cross-level interactions L1-L2			
State/Trait Anxiety × High/Low Achievers (γ_{110})	0.64***		0.10
	(.09)		(.15)
State/Trait Anxiety × Mathematics Self-concept (γ_{120})		-0.34***	-0.30***
		(.05)	(.09)
Variance components			
Within-student (L1) variance (σ^2)	0.437	0.437	0.437
Intercept (L2) variance (τ_{00})	0.233	0.210	0.203
Slope (L2) variance (τ_{11})	0.182	0.165	0.165
Intercept-slope (L2) covariance (τ_{01})	-0.063	-0.089	-0.088
Intercept (L3) variance	0.019	0.023	0.027
Explanatory power	0.617	0.653	0.653

Note. Description of variables: State/Trait (0 = state, 1 = trait); High/Low Achievers (0 = high achievers, 1 = low achievers); Mathematics Self-concept: *z*-standardized scale; L = Level; N_{L1} = 1924 (measures within students), N_{L2} = 237 (measures between students), N_{L3} = 43 (measures between classes). 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} = 0.475. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.[5]

[5] In the above analyses gender was not analyzed as a possible moderator because gender differences in anxiety were not the main concern of the present study. However, as gender has been found to play a role in trait mathematic anxiety, but not in state mathematic anxiety (Goetz et al., 2013), we also calculated all models with gender as an additional variable. Results in these models revealed no significant effects of gender above and beyond the effects of self-concept, and therefore, left the conclusions of the study unaffected.

2.6.3. Post-Hoc Analyses

Model 1 revealed that high achievers unlike low achievers did not overestimate their anxiety, but rather underestimated their trait anxiety. Since these results were surprising and unanticipated, we examined the distribution of the values for the trait-state discrepancy in high and low achievers (see Figure 3). To calculate these values, we subtracted students' mean trait anxiety scores from their mean state anxiety scores. Therefore, positive trait-state discrepancy values reflected an overestimation of trait anxiety, whereas negative values reflected an underestimation of trait anxiety. As depicted in Figure 3, the majority of high achievers underestimated their anxiety in the trait approach, whereas the majority of low achievers overestimated their trait anxiety.

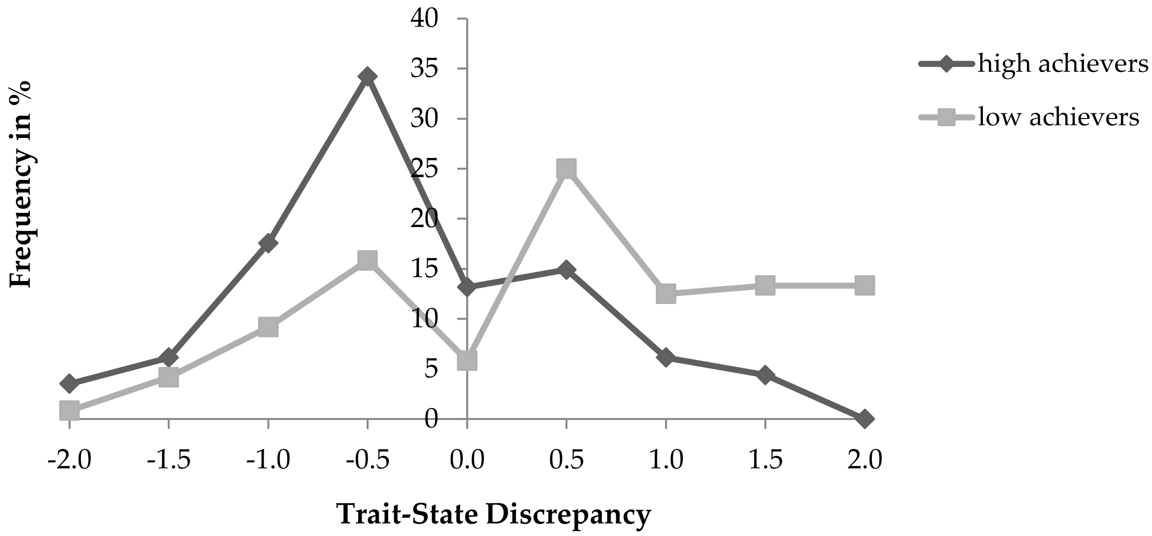

Figure 3. Distribution of high and low achievers' trait-state discrepancy in mathematics anxiety.

Note. The figure shows the distribution of the trait-state discrepancy in mathematics anxiety for high and low achievers. The x-axis depicts the difference between trait and state anxiety measures with higher numbers indicating a stronger trait-state discrepancy. To calculate these values, we subtracted students' mean trait anxiety scores from their aggregated state anxiety scores. For simplification we applied 0.5 gradations to group the trait-state discrepancy values. Negative values on the trait-state discrepancy (i.e., trait anxiety was rated lower than state anxiety) are depicted on the left side of the x-axis (negative numbers), whereas positive values on the trait-state discrepancy (i.e., trait anxiety was rated higher than state anxiety) are depicted on the right side of the x-axis. The y-axis depicts the frequencies (i.e., the percentage (%) of students) in the respective groups (high vs. low achievers) for the different trait-state discrepancies.

2.7. Discussion

The aim of the present study was to examine the trait-state discrepancy in mathematics anxiety among high and low achieving students while also taking into account their academic self-concepts in mathematics – an identity-related subjective belief previously shown to play a role in explaining the trait-state discrepancy in students' emotions (Goetz et al., 2013). We examined whether high achievers show a smaller trait-state discrepancy in mathematics anxiety in comparison to low achievers because as expected they have higher academic self-concepts in the domain of mathematics. If state anxiety reflects actual anxiety (Robinson & Clore, 2002), this would imply that high achievers estimate their anxiety more realistically in comparison to low achievers.

Consistent with previous research (Ma, 1999) and supporting our first hypothesis, we found that high achievers reported lower levels of trait anxiety than low achievers. Extending previous research on trait anxiety, we also found that high achievers reported lower levels of state mathematics anxiety than low achievers. Furthermore, we found a discrepancy between trait and state self-reports of anxiety. The majority of low achievers rated their trait anxiety significantly higher than their state anxiety, which means that they overestimated their trait anxiety (e.g., Bieg et al., 2014). This finding supports the *intensity bias* in emotional self-reports (Buehler & McFarland, 2001). However, contrary to previous findings and our initial expectations, we did not find an overestimation of trait anxiety among high achievers. Even though the trait-state discrepancy in mathematics anxiety was lower in high achievers than in low achievers, as assumed in our second hypothesis, we actually found that the majority of high achievers rated their trait mathematics anxiety significantly *lower* than their actual state mathematics anxiety. This means that high achievers underestimated their trait anxiety in comparison to their actual state anxiety levels. Furthermore, in support of our assumption (Hypothesis 3) we found that the relationship between the trait-state discrepancy in mathematics anxiety and achievement group is mediated by academic self-concept. Higher levels of academic self-concept in mathematics corresponded with a smaller trait-state discrepancy in mathematics anxiety. Specifically, we found that high achievers with high mathematics self-concepts not only showed a smaller trait-state discrepancy in mathematics anxiety than low achievers, but also had negative trait-state discrepancy values. This finding demonstrates the influence that academic self-concept has on the trait-state discrepancy – it can even lead to an underestimation of trait anxiety in high achievers (i.e., negative values on the trait-state discrepancy).[6]

---

[6] Based on these findings we were interested in the value of the mathematics self-concept in which the intensity of trait and state anxiety was estimated equally or in other words at which point the values of the trait-state discrepancy turned from positive (overestimation of trait anxiety) to negative (underestimation of trait anxiety). We called this turning point the "flip threshold". To find this turning point we conducted post-hoc analyses which revealed that the majority of students in our sample with a mathematics self-concept of 3.81 (range 1 to 5) or higher showed negative values on the trait-state discrepancy in anxiety, meaning higher values for state than for trait anxiety. Therefore, in our sample we expected the so-called "flip threshold" to be at a mathematics self-concept level of 3.81. Results further revealed that only high achievers were in this group (85 % of the high achievers), which confirms past findings that a high academic self-concept seems to be closely related to high academic achievement (Bailey, 1971). All the high achievers with a mathematics self-concept of 3.81 or lower (15 %) had positive values on the trait-state discrepancy, which indicates higher values for trait as compared to state anxiety.

If state anxiety indeed reflects actual experiences of anxiety more validly than trait anxiety (Robinson & Clore, 2002), our findings imply that even though high achievers appear well adapted to the classroom environment due to lower levels of trait and state anxiety than low achievers, they actually experience more anxiety in mathematics class than they are aware of (i.e., as reported in trait self-report measures). Therefore, when using trait anxiety measures as a proxy for student's actual anxiety, one has to always bear in mind the effects of other variables, such as academic achievement and academic self-concept, which can influence self-reported trait anxiety in either direction (i.e., overestimation or underestimation of trait anxiety).

The underestimation of trait anxiety could impact high achievers in multiple ways. On the one hand, it could be advantageous for high achievers to believe they experience less anxiety in mathematics class than they really felt because high levels of trait anxiety is negatively related to well-being, motivation and learning behavior (Pekrun et al., 2002). So this misperception of their actual anxiety (i.e., underestimation of trait anxiety in comparison to actual state anxiety levels) due to their high academic self-concepts could therefore result in conducive effects for high achiever's future academic choices, resiliency or general mental health (Taylor & Brown, 1994; Wirtz et al., 2004).

On the other hand, underestimating how much anxiety they really experience in mathematics class could in the long run also have detrimental effects for high achievers. Specifically, this overconfidence (i.e., believing they are less anxious than they actually are) may lead high achievers to overestimate their abilities. Overestimating one's abilities is related to risky decision-making (Johnson & Fowler, 2011), which could in turn negatively impact students' academic performance – specifically if the situation turns out to be highly threatening. This could result in high achievers being overwhelmed by their sudden feelings of anxiety and lacking the coping skills to deal with them. Therefore, it is important for teachers to remember that high achievers also experience anxiety while in class and they should address this anxiety by providing both low and high achievers with adequate coping skills.

Consistent with prior research (e.g., Goetz et al., 2013), our findings demonstrate that academic self-concept in the domain of mathematics, due to its high explanatory power (65.3 %), plays an important role in the discrepancy between trait and state mathematics anxiety. However, additional variables such as the perceived value of mathematics and interest in mathematics might also influence the trait-state discrepancy in mathematics anxiety, and should therefore also be investigated.

In his meta-analysis on math anxiety and performance, Hembree (1990) found that math anxiety peaks in grades 9 to 10. Therefore, our sample only included 9th and 10th grade students in mathematics classes. Future research should extend our study and explore whether our findings generalize to different age groups. Furthermore, the exploration of a possible trait-state discrepancy in other subject domains than mathematics is indicated as well.

In summary, our results were in line with previous studies concerning the role of academic self-concept in explaining the trait-state discrepancy. We found that achievement group moderates the trait-state discrepancy in mathematics anxiety and that this relationship is mediated by academic self-concept. High achievers exhibited a smaller trait-state discrepancy than low achievers because of their higher academic self-concepts. However, unlike low achievers, high achievers underestimated their trait anxiety and actually experienced more state anxiety than would have been expected based on their self-reported trait anxiety measures. Since state measures are believed to reflect actual anxiety, this study shows that math anxiety is a widespread emotion in classrooms – even among high achieving students – that deserves our attention.

3 Research Paper II

Test Anxiety Components:

An Intra-Individual Approach Testing their Control Antecedents and Effects on Performance

Anna-Lena Roos [1,2], Thomas Goetz [1,2], Maike Krannich [1,2], Amanda Jarrell [3], Monika Donker [4], & Tim Mainhard[4]

[1] Empirical Educational Research, University of Konstanz, Konstanz, Germany

[2] Thurgau University of Teacher Education, Kreuzlingen, Switzerland

[3] Educational and Counselling Psychology, McGill University, Montreal, Canada

[4] Department of Education, Utrecht University, Utrecht, The Netherlands

Test Anxiety Components: An Intra-Individual Approach Testing their Control Antecedents and Effects on Performance

3.1. Abstract

Although anxiety constitutes of multiple components, including cognitive, affective, motivational, and physiological, and some findings suggest that there might be differences regarding their control antecedents and effects on performance, previous studies have largely neglected to examine these components separately and for reasons of convenience often assessed test anxiety as a unified construct using a single-item.

Therefore, this study investigated the different test anxiety components with the goal to: (1) examine the relative impact of the anxiety components in the mediating mechanism that connects control and performance - as proposed by Pekrun's control-value theory, and (2) determine which specific anxiety component is underlying common single-item anxiety measures.

The research questions were investigated using an intra-individual approach in a sample of *N*=137 German 8th graders during a mathematics exam.

As expected, control was negatively related to all anxiety components, but associations varied in strength. Additionally, the components differed in their relative impact on performance, with the cognitive component being central for this outcome. Furthermore, common single-item measures seem to assess specifically the affective component, and thus not the component most relevant for test performance. Consequently, our study strongly recommends to distinguish between the anxiety components depending on the research question at hand.

Keywords: test anxiety; components; performance; antecedents; control-value theory; intra-individual

3.2. Introduction

Anxiety, and especially test anxiety is a prevalent emotion in school (prevalence 25 to 40 %; Zeidner, 2014) and one of the most powerful obstacles to learning in an educational setting (Enright et al., 2000). Therefore, a deeper understanding of test anxiety's nature is important, especially in educational psychology, to develop effective anxiety interventions or instructional techniques that can help students deal with this detrimental emotion (Astleitner, 2000; Zeidner, 2007, 2014). The control-value theory (CVT) refers to control (appraisals) as a central antecedent and lower performance as a major consequence of anxiety (Goetz et al., 2006). Although in the research literature anxiety is typically conceptualized as constituting of multiple components (i.e., cognitive, affective, motivational, and physiological components), empirical knowledge about the antecedents and effects of individual anxiety components is relatively scarce (Scherer, 2009; Zeidner, 2007). This may result from anxiety being usually assessed and examined relatively undifferentiated as an overall construct (i.e., by building a sum score, or by using single-item measures) or by only roughly differentiating between worry and emotionality components (Zeidner, 2007). Thus, it remains unclear whether all of the anxiety components are equally important in the mediating mechanism proposed by the CVT that connects control, anxiety (components) and test performance. A more detailed understanding about the relative importance of the different components in this model may prove helpful to identify which anxiety components should be addressed to increase students' performance – a very important outcome at school (Hembree, 1990) and to gain a better understanding about how strongly the different components depend on control appraisal antecedents.

The goal of the present study was twofold. First, we investigated the individual anxiety components based on assumptions of the CVT to examine the relative impact that antecedent appraisals of control have on the individual anxiety components and which of the anxiety components has the strongest impact on student's performance (i.e., mediation). As the CVT describes intra-individual processes (i.e., processes within a person), we applied an intra-individual real time approach (Murayama et al., 2017). Particularly, we examined the within-person association of control appraisal, anxiety components (cognitive, affective physiological, and motivational) and performance during a real-life math exam.

Because anxiety is, despite its componential nature, usually examined as a relatively undifferentiated construct and in more recent studies, mostly for practical and economic reasons, it is assessed using single-item measures (Ashcraft, 2002; Csikszentmihalyi & Larson, 2014; Goetz et al., 2013; Schimmack, 2003; Tong et al., 2007), the second goal of the current study was to investigate which anxiety component is hidden behind such undifferentiated measures of test anxiety. In other words, we wanted to explore which anxiety component affects student reports when we ask them about their anxiety with single-item measures of anxiety. In combination with our first research questions, this may help future research in deciding whether and how differentiated anxiety should be assessed depending on the conclusions the researchers want to draw.

3.3. Anxiety and its Components

Early research conceptualized test-related anxiety as a single, unidimensional construct (Mandler & Sarason, 1952). However, nowadays it is usually conceived as consisting of multiple components (Pekrun et al., 2004; Scherer, 2009; Zeidner, 2014). Liebert and Morris (1967) were the first to introduce the distinction between the "worry" and "emotionality" components of anxiety. The worry component has been defined as a cognitive concern about one's performance (e.g., thinking about the potential consequences of failure during examinations and tests), as well as doubts about one's ability to perform adequately. Emotionality refers to negative experiences, such as negative arousal and unease (e.g., feeling nervous, sweaty palms, or generally upset feelings) that may occur during a stressful academic event (Morris & Liebert, 1970). Recently, researchers have adopted a multiple component rather than dichotomous view of anxiety (e.g., Hodapp & Benson, 1997; Sarason, 1984; Scherer, 2009). Contemporary definitions conceptualize anxiety as an unpleasant emotional state or feeling with cognitive, affective, physiological, and motivational components (Matthews et al., 1999; Pekrun et al., 2004; Scherer, 1984; Zeidner, 2014). These components are theoretically distinct, yet related and can be described by intrusive thoughts (i.e., cognitive component), a feeling of nervousness (i.e., affective component), increased physiological arousal (such as higher heart rate or sweating; i.e., physiological component), and an urge to withdraw from the situation (i.e., motivational component; e.g., Pekrun et al., 2011; Pekrun et al., 2004; Zeidner, 2014; Zeidner & Matthews, 2005).

3.4. Control Antecedents and Achievement Effects of Anxiety and its Components

Pekrun's control-value theory (2006) is a prominent and well acknowledged appraisal theory in the educational context that deals with the antecedents and effects of emotions. It refers to control (appraisals) as a central antecedent and lower performance as a major effect of anxiety in achievement situations (Pekrun, 2006). Control can be described as the cognitive appraisal of the possibility to personally influence activities and outcomes and may include perceptions such as competency beliefs and causal attributions (Goetz et al., 2006; Pekrun & Stephens, 2010). The CVT suggests that an increase in anxiety occurs for achievement situations that a student perceives to be important, such as a test (i.e., high value) and when the student experiences a loss of control or insufficient control over the achievement activities or outcomes (i.e., negative control appraisals; Pekrun, 2006; Scherer et al., 2001). This increased anxiety is in turn is associated with decreased performance (i.e., anxiety mediates the relationship between control and performance). In detail, anxiety resulting from perceived lack of control reduces cognitive resources and the use of flexible strategies and self-regulation (Pekrun, 2006). Subsequently, increased anxiety has negative effects on performance (Pekrun et al., 2009). In accordance with Pekrun's assumptions, a large number of empirical findings consistently demonstrate that a perceived lack of control (i.e., low control appraisals such as low self-concept of ability, self-efficacy expectations, and academic control beliefs) is closely related to increased (test) anxiety (e.g., Frenzel et al., 2007; Goetz et al., 2006; Hembree, 1988; Pekrun et al., 2004; Perry et al., 2001) and that anxiety and more specifically test anxiety is associated with lower performance (Hembree, 1988, 1990; Lee, 2009; Ma, 1999; Seipp, 1991).

So far, it remains unclear whether all of anxiety components are equally important in this proposed mediating mechanism that connects control and performance at school, as the multi-component view of anxiety is still largely neglected in many empirical studies. Regarding control antecedents, until now, to our knowledge there is no study that has examined control appraisals and their relation to the individual anxiety components. However, as control is a cognitive appraisal, it is reasonable to expect that it has the strongest effect on the cognitive anxiety component (Goetz et al., 2006). Investigations of the effect of different anxiety components on performance, with the exception of findings from research that differentiated between worry and emotionality (e.g., Liebert & Morris, 1967), are also scarce. In this research, worry showed stronger negative correlations with performance than emotionality (Cassady & Johnson, 2002; Deffenbacher, 1977; meta-analysis: Seipp, 1991). In research by McIlroy (2000) that used a four-component model differentiating between worry, test-irrelevant thinking, tension and bodily symptoms it was found that in addition to worry, the test irrelevant thinking and tension components were also negatively related to performance. Keogh et al. (2004), however, reported that only the worry component was significant. Putwain et al. (2010) found that bodily symptoms (i.e., the physiological component), in addition to worry, were significantly negatively related to achievement. In contrast to this finding, some evidence points to a positive relation between physiological arousal and performance in athletes, where increased physiological arousal can sometimes positively impact performance (e.g., Burton, 1988; Parfitt, Hardy, & Pates, 1995). Given these inconsistencies, the current study aimed to take a closer look at the anxiety components in two ways: by investigating their relation with control appraisals – one of its most important antecedents, and by examining their effects on performance – a major outcome at school (e.g., Pekrun, 2006; Scherer, 2001).

3.5. Assessing Test Anxiety

Single-items versus multiple items

An important aspect when examining test anxiety is the differentiation between multiple item measures and general, single-item measures of anxiety. With the emergence of authentic studies that make use of repeated measurements of the same construct (i.e., experience sampling) single-item measures (e.g., "How much anxiety are you experiencing right now?") have become more popular (Ashcraft, 2002; Csikszentmihalyi & Larson, 2014; Goetz et al., 2013; Nett, Goetz, & Hall, 2011; Schimmack, 2003; Tong et al., 2007). Such measures have practical and economic advantages because they are less intrusive and easier to administer (especially in large scale assessments), and previous research suggested that they can be sufficient for adequately measuring subjective experiences such as anxiety (e.g., Ainley & Patrick, 2006; Ashcraft, 2002; Gogol et al., 2014; Nagy, 2002; Núñez-Peña et al., 2014). In the case of test anxiety, however, these single-item measures are only able to gauge anxiety as a global and undifferentiated construct (Diamantopoulos et al., 2012) and little is known which anxiety components such measures actually tap into (i.e., what component of anxiety do students consider when we ask them about their anxiety). As in most theoretical models, the affective component is seen as the key anxiety component (Scherer, 1984) it can therefore be assumed that single-item anxiety measures are closely associated with the affective component. More knowledge about undifferentiated single-item measures can be helpful when interpreting results and drawing implications from studies that are applying single-item anxiety measures. Moreover, we can determine which anxiety component is most likely in students' minds when they are asked about their test anxiety.

Trait versus state assessments

A further distinction that can be made when assessing anxiety are so called trait and state assessments of anxiety. In most studies on test anxiety the measure is not directly connected to specific test situations (i.e., participants have to indicate how they typically or generally feel in exams) and represents therefore a *trait test anxiety measure* (e.g., TAQ, Mandler & Sarason, 1952; TAS, Sarason, 1978; TAI, Spielberger & Gonzalez, 1980; a typical sample item "How much anxiety are you generally experiencing in exams?"). Because of low ecological validity, this is not an ideal design when interested in the connection between anxiety and performance. Furthermore, such trait measures can be considered biased, for instance by memory effects and generalized subjective beliefs (Buehler & McFarland, 2001; Fredrickson & Kahneman, 1993; Goetz et al., 2013; Robinson & Clore, 2002).

So called *state test anxiety* scores obtained by multiple (state) assessments examine students momentarily experienced anxiety in a specific situation and reduce these effects because of the better proximity in terms of time (i.e., real time) and space of the assessment (e.g., in the school context during a test; Barrett, 1997; Wilhelm & Grossman, 2010). A typical sample item for these state anxiety measures would be "How much anxiety are you experiencing right now?". Furthermore, repeated state assessments allow for analyzing the control, anxiety, performance relationship on an intra-individual level (Murayama et al., 2017). This is more in line with CVT, which in fact describes intra-individual functioning, that is, within-person relationships between appraisals, emotions and performance (e.g., Goetz et al., 2006; Pekrun et al., 2009). Understanding the intra-individual functioning is also highly relevant when designing interventions for anxiety (Murayama et al., 2017). However, only a few studies have taken an intra-individual approach (Ahmed et al., 2010; Bieg et al., 2013). As findings indicate that between-person relations (i.e., inter-individual approaches) can sometimes be different from within-person relations and do not necessarily reflect internal psychological processes (i.e., intra-individual functioning; e.g., Hox, Moerbeek, & Van de Schoot, 2017; Kenny, Bolger, & Kashy, 2002; Molenaar & Campbell, 2009), in the current study we investigated the relationship between control experience, anxiety components, and performance in a specific situation (i.e., when students are writing a math exam) on an intra-individual level (i.e., how control experience within a person is related to the anxiety experienced and how this relates to performance) using state measures at multiple points in time.

3.6. The Present Study

An overall goal of the present intra-individual study was to examine the relative importance of four different components of test related anxiety in mathematics (i.e., cognitive, affective, physiological, and motivational) in the mediating mechanism that connects control and test performance. Furthermore, we were investigating what anxiety component is hidden behind more frequently used single-item measures of test anxiety. Our study was guided by the following research questions:

(1) a) Is the effect of control on performance mediated by the test anxiety components (Figure 4, ab)?
b) Does control affect the anxiety components in similar ways (Figure 4, a)?
c) Do the different anxiety components affect performance in similar ways (Figure 4, b)?

(2) What anxiety component is mainly assessed with single-item self-report measures of test anxiety?

Hypothesis 1

Based on the described theoretical assumptions we hypothesized that low levels of control are associated with increased test anxiety and that increased anxiety is associated with decreased performance. However, it was also expected that the different anxiety components vary with regard to their antecedents and effects. In more detail, we hypothesized that anxiety mediates the relationship between control and performance and that the anxiety components vary in the strength of their mediating effects (i.e., indirect effects; Figure 4). According to the CVT, control is a cognitive appraisal and one would expect that it has the strongest effect on the cognitive anxiety component (i.e., negative relationship; Goetz et al., 2006; Pekrun, 2006). From studies using inter-individual approaches and the worry and emotionality distinction, we know that the worry component seems to play a bigger role for performance than the emotionality component (Cassady & Johnson, 2002; Liebert & Morris, 1967; Morris & Liebert, 1970). Therefore, we assumed that the cognitive component has the strongest effect on performance as well as the strongest mediating effect in our intra-individual approach. In all instances, mediation is expected to be partial rather than complete, because other mediational processes, such as test emotions other than anxiety (e.g., joy, pride, anger) or study or test-taking strategies, are likely to be operative as well (Elliot, McGregor, & Gable, 1999; Goetz et al., 2006; Pekrun et al., 2004).

Hypothesis 2

Regarding research question two, we assumed that when students are asked about their anxiety with a single-item, their answer will most likely be related to the affective anxiety component because this component is according to most models the key anxiety component (Scherer, 1984, 2005).

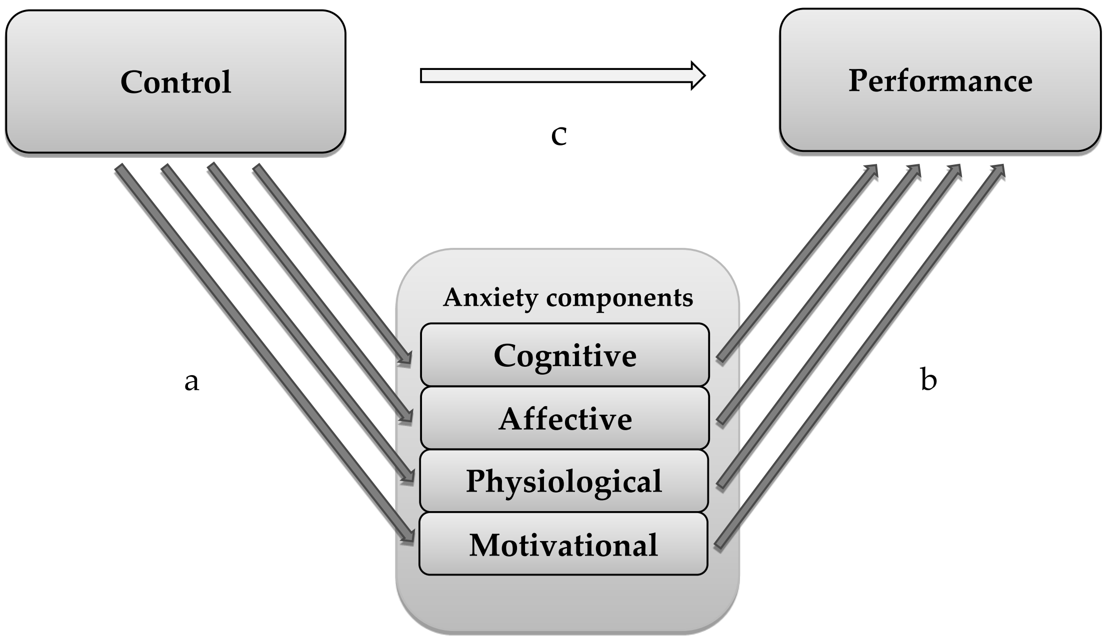

Figure 4. Hypothesized mediation.

Note. It was hypothesized that lower control is associated with lower performance (c) as well as with higher anxiety (i.e., all four components) (a). Furthermore, it was expected that the anxiety components mediate the relationship between control and performance (ab) and the components vary in the strength of their mediating effects.

3.7. Methods

3.7.1. Sample

The sample of the present study consisted of N = 137 high school students (from 6 classes at 3 schools, M_{age} = 13.82, SD = 0.55, 44.5 % female). The students went to eighth grade and attended the highest track of the German school system (i.e., *Gymnasium*; approximately one third of the total student cohort in Germany attend this track; Federal Statistical Office, 2015).

3.7.2. Ethics Statement

Data collection, data protection and ethical issues of the present study were handled according to the guidelines of the German Association for Psychology (Deutsche Gesellschaft für Psychologie, (DGPs), 2016) and the American Psychological Association (2010). Participation was voluntarily and informed consent was received by both students and parents. All participants, the students' parents, the principals, and teachers were informed about the study's purpose, duration, and procedure. Written confirmation of participation was also provided by the parents and principals. All data was completely anonymized.

3.7.3. Procedure

After students and parents consented to participate in the study, the study started with a general assessment of students' cognitions, motivation, and emotions in mathematics and demographic data using a paper-and-pencil questionnaire that was administered by trained researchers. Approximately one week later, this general assessment was followed by the state assessments during a mathematics exam specifically developed for the current study.

Procedure during the mathematics exam

The exam was a written mathematics test designed in cooperation with a mathematics teacher. It included tasks adapted from the VERA8 exam – a standardized exam all students from 8th grade all over Germany have to take at the same time of the school year (developed by the Institute for Educational Quality Improvement; IQB, Berlin, Germany). We chose an exam that is very similar to an authentic exam to ensure ecological validity. An actual exam could not be used for ethical considerations and to ensure consistency between exams across all participating classes. This was essential to allow for reliably comparisons between the tasks within the exam and the emotional reactions to these tasks. The mathematics exam consisted of six task-blocks and task difficulty varied within the exam to create variation in students' appraisals of control (i.e., we induced the feeling of losing control by difficult tasks, to examine whether loss of control goes along with increased anxiety). Each of the six task-blocks had to be solved within five minutes. After these five minutes, students had to hand in the task sheet and fill in the short state-questionnaire (i.e., paper-pencil questionnaire). Following the completion of the questions, students started the next task-block. Thus, there were six measurement points of all state study variables per student. In total, the exam lasted approximately 60 minutes. To incentivize students to take the exam seriously, a monetary prize was awarded to the class with the highest performance (most points). Therefore, the performance of each student, not only the high performing students, contributed to the overall class performance and therefore to winning the prize.

3.7.4. Measures

Assessment of performance

The tasks in the math exam were corrected following the recommendations from the Institute for Educational Quality Improvement (IQB). Student's performance in each task-block served as performance measure in our analysis. We used the percentage of correctly solved tasks per task-block in our analysis. The performance measure was highly correlated with students' grades in mathematics (for exact values see *Descriptive statistics and correlations*).

Assessment of state control and state anxiety components via questionnaire

All items in the questionnaires always referred to the student's experience during the previous task (i.e., *"During this task I…."*).

Assessment of control. Control was assessed with one item adapted from the Perceived Academic Control Scale (PAS; Perry & Hamm, 2017; Perry et al., 2001). A five-point Likert scale was used ranging from 1 (*strongly disagree*) to 5 (*strongly agree; "During this task I had a great deal of control"*).

Assessment of the anxiety components. The anxiety components (cognitive, affective, physiological, and motivational component) were assessed using three items for each component adapted from the Achievement Emotions Questionnaire (AEQ; Pekrun et al., 2011). As for subjective control, answers were given on a five-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Example items: "*During this task*: *…I worried whether I will perform poorly* (cognitive); *…I was very nervous* (affective); *…my heart was beating faster* (physiological); *…I was so anxious that I'd rather be anywhere else* (motivational)." The sample-specific internal consistency across all six measurement points of these three item scales were α = .84 for the cognitive, α = .89 for the affective, α = .83 for the physiological, and α = .89 for the motivational component. The sample specific internal consistencies per task were all higher than .80.

Single-item measure to assess students' anxiety over the course of the exam

The single-item anxiety measure was assessed at the end of the exam in such a way that students were asked to draw a profile of the course of their anxiety during the exam. In this profile they indicated their anxiety for each task as well as their anxiety 30 minutes before the exam, at the beginning of the exam (i.e., when receiving the instructions), at the end of the exam (i.e., after handing in the final task), and right now (i.e., when filling in the last questionnaire). The scale was ranging from 0 (*no anxiety at all*) to 5 (*very strong anxiety*). For an example, see Figure 5. In the present study we were only interested in the anxiety values during the *six tasks*, as we wanted to compare them with what the students reported on the different anxiety components after each task (see *Assessment of the anxiety components*).

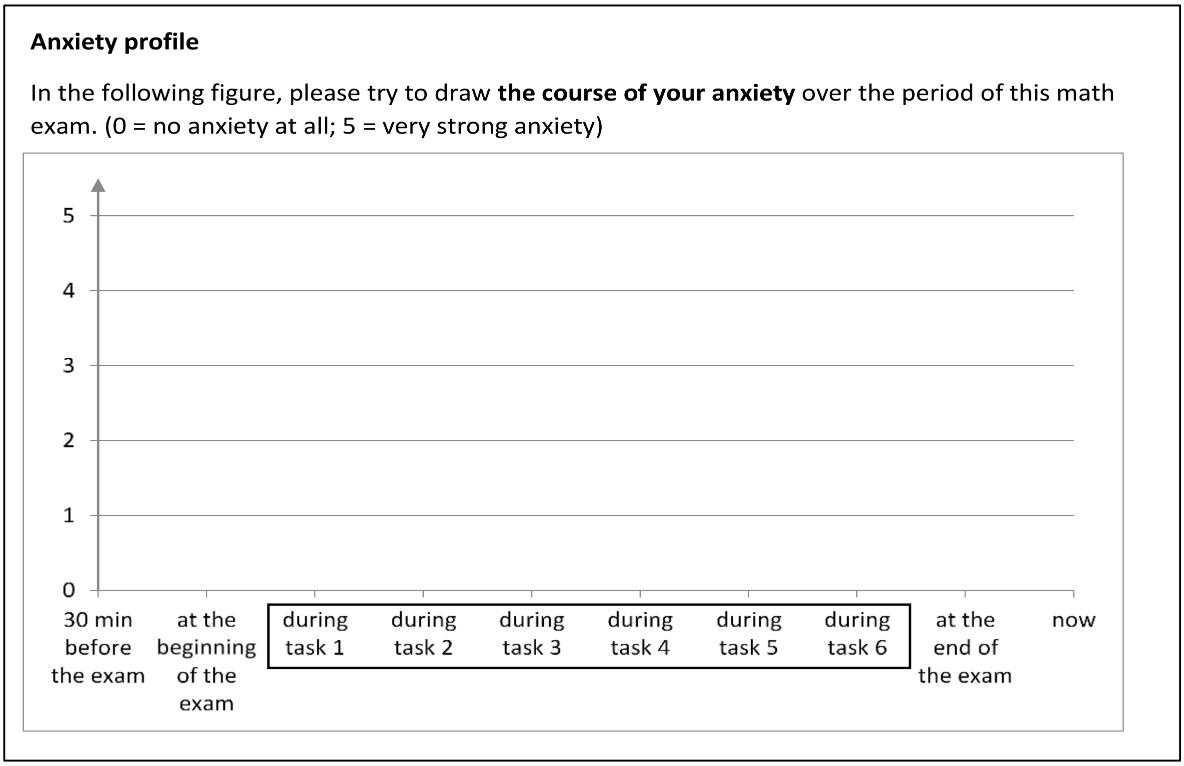

Figure 5. Anxiety profile.

Note. Single-item measure to assess students' anxiety over the course of the exam. The six tasks we are interested in in the current study are highlighted with a box.

3.7.5. Data Analyses

Our analyses were conducted with M*plus* 7.11 (Muthén & Muthén, 1998-2017) and missing data was handled with full information maximum likelihood procedures (Rubin, 1976).

Hypothesis 1

Mediation analysis with "type = complex" was conducted to examine the hypothesized mediation on the intra-individual level and to account for the three-level nested data structure (multiple measurement points (6) of state anxiety, subjective control and task performance nested within student; students nested in different classes). In our model we tested the following effects: The direct effect of control on performance (c, see Figure 4) and the indirect effects of control on performance via the different anxiety components (i.e., cognitive, affective, physiological, and motivational component). Moreover, we report the direct effects of control on the anxiety components (a) as well as the direct effects of the components on performance (b). Additionally, we controlled for gender in our analysis as some studies indicated that there might be gender differences in test anxiety (i.e., female students were found to be more test anxious but there was no performance difference between girls and boys; Hembree, 1988; Zeidner, 1990) and in control experience in mathematics (i.e., female students were found to have lower control in mathematics; Frenzel et al., 2007; Hyde, Fennema, Ryan, Frost, & Hopp, 1990). For all indirect effects, we used the "model indirect" option of M*plus* to calculate unstandardized and standardized effects (Muthén & Muthén, 1998-2017).

Hypothesis 2

To test the second hypothesis, we built an intra-individual regression model with "type = complex" in which we included the anxiety components as predictors for single-item anxiety measure (i.e., outcome) as indicated in the anxiety profiles the students were drawing. This renders it possible to find out which of the anxiety component is the strongest predictor for the single-item anxiety measure (i.e., which component is more closely related to the single-item as indicated in the anxiety profiles).

3.8. Results

3.8.1. Descriptive Statistics and Correlations

Mean scores and standard deviations of the test anxiety components, control, and performance for each task of the exam (i.e., between-person) are depicted in Table 3. Figure 6 shows the means of the test anxiety components and control per task and across persons. It can be seen that more difficult tasks (i.e., tasks 2, 3, 5, 6 in comparison to tasks 1 and 4) are associated with less control and more anxiety (especially the cognitive anxiety component).

Table 4 shows the correlations between the anxiety components, control and performance across tasks (i.e., multiple measurement points) along with the respective means and standard deviations. All anxiety components were significantly and positively correlated with each other (all $ps < .01$) with the affective and physiological components showing the strongest correlation ($r = .74$). All anxiety components significantly and negatively correlated with subjective control (all $ps < .01$). Furthermore, the anxiety components significantly negatively correlated with performance, with the cognitive component showing the strongest negative correlation ($r = -.25$, $p < .01$). Subjective control also significantly and positively correlated with performance ($r = .30$, $p < .01$).

The correlations between the performance measure (i.e., total performance in the math exam) and student's grades in their last math exam, in their last mid-term school report and in their school report were $r = .58$, $r = .52$, and $r = .61$, respectively. All correlations were significant (all $ps < .01$) and in the expected directions (i.e., better grades were associated with better performance). There were no significant gender differences in any of the performance measures.

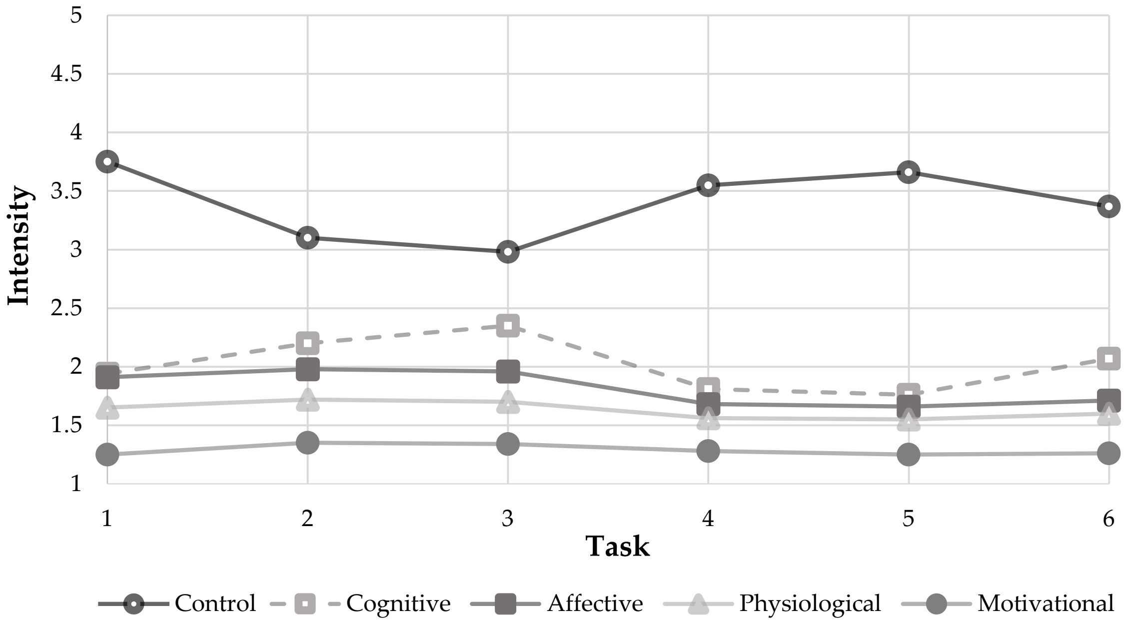

Figure 6. Control and anxiety components per task.

Note. Means of anxiety components (cognitive, affective, physiological, motivational), and control per task.

3.8.2. Hypothesis 1: Control-Anxiety Components-Performance Mediation

To test our first hypothesis, we looked at the indirect effects of control on performance via the different anxiety components (i.e., mediation). For an overview of all relations of this mediation model, as well as direct, indirect, and total effects and the respective regression coefficients see Table 5. The results of our model show that control was significantly negatively associated with all anxiety components, with the cognitive component showing the strongest effect (b = -0.37, SE = 0.05, p < .001). This effect was significantly different from the effects of the other components as indicated by the results of a scaled difference chi-square test (p < .001; Bryant & Satorra, 2013). Mediation analyses further revealed a significant positive direct effect (c) of control on performance (b = 0.24, SE = 0.04, p < .001) and a significant indirect effect via the cognitive anxiety component (b = 0.09, SE = 0.02, p < .001). Specifically, control had a negative effect on the cognitive component (a), which in turn had negative effects on performance (b) (i.e., positive indirect effect). In sum, the cognitive component partially mediated the relationship between control and performance.

There was a significant negative effect of the cognitive component on performance (b = -0.25, SE = 0.04, p < .001; see Table 5). This means that higher anxiety measured by the cognitive component is associated with poorer performance. However, there was a significant positive direct effect of the affective anxiety component on performance when controlling for the effects of control and the effects of the other anxiety components (b = 0.12, SE = 0.06, p < .05; i.e., higher affective anxiety is associated with better performance). The effects of the motivational and physiological did not reach significance.

Effects of gender

Results revealed a significant effect of the gender covariate on control (b = -0.15, SE = 0.06, p <. 05), that is female students experienced significantly less control than male students. Furthermore, there was an effect of gender on the cognitive component (b = 0.21, SE = 0.06, p < .001). Specifically, with the same levels of control female students reported significantly higher values on the cognitive anxiety component than male students.

3.8.3. Hypothesis 2: Single-Item Anxiety Measure

The results from the regression model with the anxiety components predicting the single-item anxiety measure are depicted in Table 6. Results show that the affective anxiety component was, as expected, the strongest predictor for the single-item anxiety measure (b = 0.43, SE = 0.08, p < .001; correlation with single-item anxiety measure: r = .47; when controlling for the effects of the other components) followed by the physiological component (b = 0.21, SE = 0.09, p < .05; correlation with single-item anxiety measure: r = .23). The cognitive and motivational components were not significant predictors for the single-item measure. Thus, in line with hypothesis 2, results show that when students are asked to report their anxiety experienced during a test with a single-item measure (i.e., when drawing a profile of the course of their anxiety) their answer is mostly associated with the affective anxiety component.

Table 3. Mean scores and standard deviations of the study variables per task

	Task 1		Task 2		Task 3		Task 4		Task 5		Task 6	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Cognitive	1.94	0.98	2.20	1.05	2.35	1.02	1.81	.83	1.76	0.91	2.07	1.07
Affective	1.91	0.96	1.98	0.98	1.96	1.05	1.68	.93	1.66	0.87	1.71	0.98
Physiological	1.65	0.86	1.72	0.90	1.70	0.93	1.56	.87	1.55	0.83	1.60	0.86
Motivational	1.25	0.53	1.35	0.76	1.34	0.81	1.28	.74	1.25	0.67	1.26	0.67
Control	3.75	1.08	3.10	0.97	2.98	1.08	3.55	1.08	3.66	1.05	3.37	1.16
Performance	80.47	19.45	28.52	15.16	10.68	13.28	77.74	13.18	68.75	16.86	47.01	22.90
Anxiety (SIP)	1.24	1.32	1.60	1.41	1.59	1.30	1.51	1.32	1.43	1.26	1.36	1.41

Note. N = 137. SIP = Single-item from the anxiety profiles.

Table 4. Mean scores, standard deviations, and correlations between the study variables across all tasks and persons

	1	2	3	4	5	6
1. Cognitive Component	--					
2. Affective Component	.65**	--				
3. Physiological Component	.53**	.74**	--			
4. Motivational Component	.50**	.69**	.65**	--		
5. Control	-.40**	-.29**	-.29**	-.31**	--	
6. Performance (in %)	-.25**	-.08*	-.07*	-.08*	.30**	--
Mean	2.02	1.81	1.63	1.29	3.40	52.19
SD	1.00	0.97	0.88	0.70	1.10	31.09

Note. N = between 819 and 822 (resulting from 6 measurement points per student and N = 137 students). Pure descriptive values without control for nesting of data within persons (i.e., no type is complex). * $p < .05$, ** $p < .01$.

Table 5. Total, direct, and indirect effects of the proposed multilevel mediation model

	Standardized coefficients	Standard errors
Control		
Total Performance on Control	.29***	.03
Total Indirect Performance on Control	.05***	.01
Indirect Performance on Control via Cognitive	.09***	.02
Indirect Performance on Control via Affective	-.03	.02
Indirect Performance on Control via Physiological	-.01	.01
Indirect Performance on Control via Motivational	.003	.01
Direct Cognitive on Control	-.37***	.05
on Gender	.21***	.06
Direct Affective on Control	-.28***	.06
on Gender	.10	.07
Direct Physiological on Control	-.28***	.06
on Gender	.07	.08
Direct Motivational on Control	-.26***	.05
on Gender	.02	.07
Direct Performance on Control	.24***	.04
Control on Gender	-.15*	.06
Anxiety components		
Direct Performance on Cognitive	-.25***	.04
Direct Performance on Affective	.12*	.06
Direct Performance on Physiological	.05	.05
Direct Performance on Motivational	-.01	.04

Note. All regression coefficients are standardized; Gender: male = 0, female = 1; standard errors are displayed in brackets; *p < .05; **p < .01; ***p < .001.

Table 6. Results from the regression model with the anxiety components predicting the single-item anxiety measure (i.e., as indicated in the anxiety profiles)

	Single-item measure of anxiety		
Components	Standardized coefficients	Standard errors	Standardized correlation with single-item
Cognitive	.01	.06	.01
Affective	.43***	.08	.47***
Physiological	.21*	.09	.23*
Motivational	.05	.07	.05

Note. All regression coefficients and correlations are standardized. Standard errors are displayed in brackets; *p < .05; **p < .01; ***p < .001.

3.9. Discussion

3.9.1. Summary of the Study

The present study examined four test anxiety components (i.e., cognitive, affective, physiological, and motivational components) in a sample of 8th graders and had two major goals. Based on assumptions of the CVT, the first goal was to investigate which of the anxiety components is the most important mediating mechanism that explains the relationship between control appraisals and students' performance (i.e., has the strongest impact on students' performance). To draw upon previous findings from inter-individual studies and ensure high ecological validity, we examined this relationship during an authentic exam situation using multiple within person state measurements of these variables and analyzing this relationship on an intra-individual level (Murayama et al., 2017). In order to inform future research, decide whether and how to differentiate anxiety, the second goal of the study was to investigate which anxiety component is hidden behind a single-item measures of test anxiety.

3.9.2. Discussion and Implications of the Findings

Which anxiety component is most central in the mediating mechanism that connects control and performance (Hypothesis 1)?

In order to determine which component is the most important predictor of performance, we examined the relationship between control, anxiety components and performance. As expected, we found a relationship between control and performance mediated via the cognitive test anxiety component (i.e., significant indirect effect). More specifically, lower levels of control were associated with an increase in the cognitive anxiety component, which in turn resulted in lower performance. These findings support the assumptions of the CVT and show that this theory on intra-individual relations indeed also holds true when examining it on an intra-individual level (Goetz et al., 2006). As the cognitive component is the only component for which we found such a mediating effect, our results further suggest that this component appears to be the most important for explaining the relationship between control and performance. These findings are consistent with the CVT and with information processing models of anxiety (e.g., Naveh-Benjamin, 1991; Schwarzer & Jerusalem, 1992) which state that thinking about negative consequences of one's performance (cognitive component) consumes cognitive resources, which may interfere with optimal performance. Therefore, our findings suggest that this component should be the main focus of instructional techniques and anxiety interventions that are aimed at increasing student performance.

Furthermore, we found that appraisals of control affect all four anxiety components in the same direction (i.e., lower control was associated with increased anxiety; negative direct effect). However, the sizes of the effects differed. Control showed the strongest association with the cognitive anxiety component. The strengths of the size of the effects of control on the affective, physiological, and motivational components were similar. This finding is in line with the assumptions of our study that as control is a cognitive appraisal, therefore it is probably more closely related to the cognitive anxiety component (Goetz et al., 2006). In this respect, it may prove helpful that practitioners implement appropriate strategies during learning and test-taking that serve to enhance perceptions of control and thereby indirectly target the cognitive anxiety component.

Moreover, our results showed that the individual anxiety components differed in their relative impact on performance. As previously stated, the cognitive component had the strongest effect. Although there was no significant indirect effect of control on performance via the affective anxiety component, we found a direct positive effect of the affective component on performance (i.e., higher affective anxiety was associated with increased performance). This effect requires further investigation as mechanisms not included in our statistical model may explain this relationship. One possibility is that there could be a moderating effect of extrinsic motivation on the relationship between the affective anxiety component and performance, which would explain the positive relationship. For example, students that work hard and know a lot but are nervous about tests, might work harder because they are nervous. In other words, their experience of anxiety motivates them to work hard on the exam, which leads to better exam performance. Thus, our finding implies that it can be helpful to tell students that nervousness during a test is not always harmful for performance but rather the thoughts (i.e., cognitive anxiety component) about the consequences of failing can lead to poorer performance. Finally, we found a direct positive effect of control on performance and thus, as expected, we found a partial mediation. This effect could in part be explained by other variables (e.g., other emotions, or constructs like self-efficacy) that might play an additional mediating role between control and performance and also corresponds with assumptions of the CVT. Therefore, our results suggest that it is not only necessary to regulate anxiety but also to look at other variables during test taking as they might have effects on performance as well (Goetz et al., 2007).

In sum, as expected and also in line with hypothesis 1, the results of the study suggest that the components of anxiety indeed differ with regard to their control antecedents and achievement effects. For anxiety interventions, these findings imply that interventions must be carefully selected and that it is important to distinguish between the components of anxiety. In addition, the results suggest that to increase students' performance strategies that target the cognitive anxiety component might be the most effective. This can be achieved for example with cognitive reappraisal strategies that aim at changing how one thinks about the situation or about one's capacity to manage the demands of the situation (Gross & Thompson, 2007) and consequently, increase student's perceptions of control and decrease the cognitive component of anxiety.

In addition to providing evidence that supported our first hypotheses, our findings demonstrated *further interesting results regarding the effects of gender* that although they were not the main focus of the current study will be briefly discussed here. Female students experienced significantly less control than male students which is in line with findings from Arch (1987) who found that in test situations females were more likely to rate themselves as less capable and confident in their abilities and felt more uncomfortable (i.e., less in control) than male students. Furthermore, with the same levels of control, female students reported significantly higher levels of the cognitive component of anxiety than male students, but we could not find differences in exam performance between females and males. This is in line with earlier findings reporting gender discrepancies in emotional experiences even though girls and boys had achieved at similar levels in mathematics (e.g., OECD, 2004; meta-analysis: Seipp, 1991). Hence, our results show gender equity in terms of performance. However, there still remains a gap in terms of competence beliefs and emotions in mathematics tests which could be closed by promoting positive affect (e.g., by interventions that address students' competence beliefs) and preventing anxiety in girls (Goetz et al., 2013).

What anxiety component affects undifferentiated single-item measures (Hypothesis 2)?

Also consistent with our expectations, our findings concerning the single-item measure of anxiety showed that students will most likely report their affective anxiety when assessing anxiety with a single-item. This is in line with the multi-componential model of anxiety (Scherer, 1984) in which the affective component is seen as the core anxiety component and therefore confirms that single-item measures are able to gauge this key component of the anxiety construct. Furthermore, the findings indicate that this component might be very important, as is seems to be most saliently represented in a student's mind when asked about her or his subjective anxiety with single-items. Although, as shown by our first research question, the affective component did not show a significant indirect effect on achievement in our mediation model and therefore does not seem to be the most relevant component for performance. It may, however, be closely related to other factors such as psychological well-being, school attendance, self-esteem, or other stress related physical ailments (Austin et al., 1995; Damer & Melendres, 2011; Huberty, 2010; OECD, 2016; Salend, 2012). All these factors can also be negatively affected by test anxiety and should be taken seriously as they might have long term consequences, for example for future career choice or mental and physical health (Diener, 2000; Wigfield et al., 2002). Therefore, the affective anxiety component should not be neglected in interventions or by researchers. Rather it should be more closely examined with regard to its effects on variables other than performance in future studies. The same is true for the physiological and motivational anxiety components for which we also did not find a significant indirect effect on performance in the current study – which however does not mean that they are not important for other factors.

In sum, it is crucial for researchers to realize that when using single-item self-report anxiety measures they are most likely working with a predominantly affective measure. Therefore, they should carefully choose which measure of anxiety they are using depending on their research question and the conclusions they want to draw. When examining performance outcomes in the context of the CVT, assessing anxiety with a single-item measure that is focusing predominantly on the affective component might not be sufficient. In this case, items that largely tap into the cognitive component might be more adequate as they are a better predictor for performance. However, for investigations of more distal outcomes, single-item measures that examine the affective component, or other multiple item measures that examine the physiological and motivational test anxiety components could eventually be a good choice.

3.9.3. Strengths, Limitations, and Future Directions

This study was one of the first studies that looked at the antecedents and effects of the anxiety components on an intra-individual level. Our approach offered a higher level of ecological validity than in most test anxiety studies, as the measures were taken directly in the evaluative situation at multiple time points. Second, the investigation of the relationship in a realistic assessment environment allows for greater confidence in the approximation of the relative impact of the anxiety components on high school students performance. Thus, the results of this investigation may be more credible to the reality of test anxiety than in experimental settings in a lab environment. Third, in our data there was high inter-individual variance as well as variance within students from task to task (see Figure 7), which suggests that analyzing the relationship between control, anxiety components and performance on an intra-individual level (i.e., as done in the present study) might be more adequate than only looking at between-person associations. Thus, this once again underpins the importance of conducting intra-individual analyses when looking at intra-individual relationships such as those postulated by the CVT.

Our study also has some limitations, which suggests a number of directions for future research. First, the data of this study are based on student's self-reports. It would be interesting to investigate if, what these students think they feel about their physiological arousal, corresponds to their actual physiological arousal (e.g., heart rate or skin conductance) and how this relates to their performance. Furthermore, our sample of high school students was relatively selective as we examined students in the top track of the German school system. Hence, it would be interesting to investigate these mediational effects in students from lower school tracks or in students with another cultural background to confirm the generalizability of these findings. Furthermore, although we assessed students' anxiety during a test situation, for ethical reasons students did not receive grades for their test performance. Therefore, one might argue that our test situation was artificial too, which could have impacted the results. Finally, we investigated students' test anxiety in one specific domain, namely mathematics. Although our theoretical approach implies that the relationship between control, anxiety components and performance is the same in other subject areas as well, the results of the present study cannot be generalized beyond the domain of mathematics. Future research should also focus on other domains, for example, by investigating the relationship when students are working on an achievement test in language classes (i.e., English or French).

Despite these limitations, our study provides additional evidence that emotions are complex constructs by using an elaborate intra-individual real-time approach. Additionally, it proposes that extending research on the antecedents and effects of different test anxiety components on an intra-individual level may contribute to the development of more effective anxiety interventions and thereby facilitate not only learning and academic achievement, but also critical developmental outcomes including career choice, health and psychological well-being.

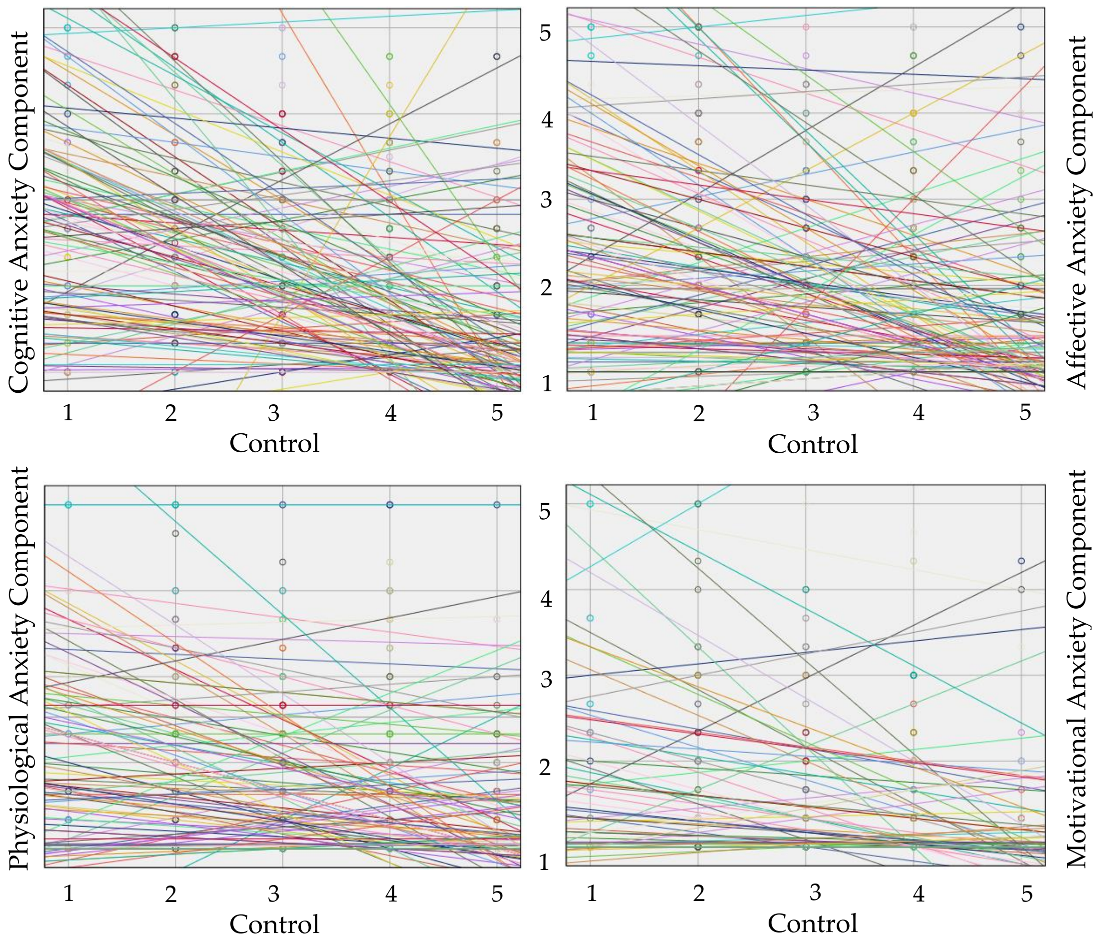

Figure 7. State control (x-axis) and state anxiety cognitive, affective, physiological, and motivational components (y-axis).

Note. Within-person relationship between state anxiety (i.e., separate for each component) and state control (i.e., six measurement points per person). Each colored line represents one student. It can be seen that there is high inter- and intra-individual variance.

4 Research Paper III

Test Anxiety and Physiological Arousal: A Systematic Review

Anna-Lena Roos [1,2], Thomas Goetz [1,2], Maike Krannich [1,2], Madeleine Bieg [1], Amanda Jarrell [3], & Reinhard Pekrun [4,5]

[1] Empirical Educational Research, University of Konstanz, Konstanz, Germany

[2] Thurgau University of Teacher Education, Kreuzlingen, Switzerland

[3] Educational and Counselling Psychology, McGill University, Montreal, Canada

[4] Department of Psychology, University of Munich, Munich, Germany

[5] Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia

Test Anxiety and Physiological Arousal:
A Systematic Review

4.1. Abstract

Test anxiety is a widespread and mostly detrimental emotion in learning and achievement settings. Thus, it is a construct of high interest for researchers and its measurement is an important issue. So far, test anxiety has typically been assessed using self-report measures. However, physiological measures (e.g., heart rate or skin conductance level) have gained increasing attention in educational research as in contrast to frequently used self-report measures, they allow for an objective and often also continuous assessment of students' physiological arousal (i.e., the physiological component of test anxiety) in real-life situations, such as a test. Although, theoretically one would assume convergence between traditional self-report measures of test anxiety and objective physiological measures, empirical evidence is scarce and findings are mixed. To gain a more coherent picture of the relationship between these measures, the present systematic review investigates whether higher self-reported test anxiety is associated with objectively measurable increased physiological arousal. A systematic literature search yielded 193 articles and a structured selection process resulted in 27 eligible articles with 29 studies that met the specified criteria and provided sufficient information about the relationship under investigation. In line with theoretical models, in 21 out of the final 29 studies there was a significant positive relationship between self-reported test anxiety and physiological arousal. The strengths of these correlations were small to large (mean correlation: $r_{\textit{Fisher's z}} = .32$; range of mean correlations: .19 to .51). Influencing aspects of the relation between these two measures are discussed along with implications for the assessment of physiological data in future classroom-based research.

Keywords: physiological measures; assessment; test anxiety; self-report; systematic review

4.2. Introduction

In educational research, there is a growing interest in using physiological measures of affect (e.g., cardiovascular measures such as heart rate measurement, electrodermal measures or saliva samples) to complement traditional self-report measures of emotion (D'Mello & Calvo, 2013). However, the reliability of these measures, which assess physiological arousal as a proxy for emotional processes, are still questioned, and interpreting findings obtained from these measures in relation to existing educational research and theory remains a challenge in educational psychology (Kreibig, 2010; Mauss et al., 2005; Mauss & Robinson, 2009). Therefore, the purpose of the present review is to address these issues by reviewing and synthesizing research findings on the relationship between physiological measures of test anxiety - one of the most prominent emotions in the academic context - and self-report measures, which traditionally have been considered as benchmarks for assessing emotions (Larsen & Prizmic-Larsen, 2006).

Test anxiety has been extensively investigated since the early 1950s (Mandler & Sarason, 1952). To date there are more than 2,000 published studies on this often detrimental emotion (e.g., Hembree, 1988; McDonald, 2001; Zeidner, 2014). A primary indicator of test anxiety is heightened levels of physiological arousal. It can be unobtrusively assessed through physiological measures, such as cardiovascular measures (e.g., heart rate) or electrodermal measures (e.g., skin conductance response). Physiological responses associated with anxiety are also difficult to suppress or control. Therefore, physiological measures represent an objective way of assessing students' arousal in testing situations (Harley, 2015; Houtveen & de Geus, 2009; Wilhelm & Grossman, 2010).

Nevertheless, efforts to measure test anxiety has typically relied on self-report measures, which can be influenced by factors such as social desirability or subjective beliefs (Goetz et al., 2013; Kaplan et al., 2013; Pekrun & Bühner, 2014; Robinson & Clore, 2002). Therefore, it is possible that self-report assessments of test anxiety might not accurately reflect the "true" level of test anxiety.

Since heightened physiological arousal is one component of test anxiety (Zeidner, 1998), it is reasonable to assume that students' objectively measurable level of physiological arousal during test taking and level of self-reported test anxiety (i.e., often including the subjective experience of all components) would converge (i.e., a certain degree of overlap). For instance, when test anxious students are asked about their feelings while taking a test, they should report both high levels of anxiety and perceptions of high physiological arousal (e.g., a faster heart beat). However, empirical findings investigating relations between self-report measures of anxiety and physiological measures are mixed (e.g., Mauss et al., 2005; Mauss & Robinson, 2009). Knowledge about the relationship between these measures might contribute a deeper understanding of test anxiety's nature, which is an important research theme, especially in educational psychology (Campbell, 1986; Van Yperen, 2007; Zeidner, 2014). In the long run, this understanding might also help guiding the development and implementation of test anxiety interventions.

The aim of the current review is to yield a more coherent picture of the relationship between these two measures of test anxiety in school and academic contexts and to identify possible moderating variables between them by systematically investigating whether self-reported test anxiety is associated with objectively measurable physiological arousal (e.g., heart rate, blood pressure, or cortisol level).

We start with defining test anxiety and its components. We then describe the two assessment methods discussed in this review in detail and point out advantages and disadvantages of each method followed by a description of the research questions of the current review. Next, we provide a systematic review of empirical findings on the relationship between these self-report and physiological measures of anxiety. The last section summarizes the findings and discusses possible moderator variables (i.e., in terms of study design and setting or sampling rate of physiological measure) influencing the relation between the two measures as well as methodological and practical implications for the assessment of physiological data in future research in educational psychology.

4.3. Concept and Assessment of Test Anxiety

4.3.1. Test Anxiety and its Components

Test anxiety is a specific type of anxiety that can be experienced before, during and after an examination or other test situations (Beidel & Turner, 1988; Zeidner, 2014). It is usually conceptualized as comprising multiple theoretically distinct, yet correlated components (Pekrun et al., 2004; Zeidner, 2014). Liebert and Morris (1967) first introduced the distinction between the “worry” (cognitive) and “emotionality” (affective) components of test anxiety. The worry component has been defined as a cognitive concern about one’s performance (e.g., thinking about the potential consequences of failure in examinations and tests, as well as doubts about one’s ability to perform adequately). Emotionality refers to negative affective and physiological experiences, such as arousal and unease (e.g., nervous stomach, muscle tension, sweaty palms, or generally upset feelings) that may occur during the stressful testing situation (King et al., 2000; Morris & Liebert, 1970). In more recent studies, researchers have moved from a dichotomous view to a multi-component view of emotions and more specifically test anxiety (Pekrun et al., 2002; Shuman & Scherer, 2014; Zeidner & Matthews, 2005) often including affective, cognitive, physiological, motivational, and behavioral components (Matthews et al. 1999; Pekrun et al., 2004; Scherer, 1984).

4.3.2. Assessing Test Anxiety

Test anxiety can be assessed through multiple measures (Zeidner & Matthews, 2005) including self-report measures, behavioral measures (e.g., detection and classification of facial expressions), neuroscientific measures (e.g., fMRI), and physiological measures (e.g., heart rate or skin conductance). This review focuses on two types of test anxiety measures, self-report as the traditionally used measures, and physiological measures as emerging measures in the field of educational psychology, which are described in detail in the following paragraphs.

Self-report measures

Self-report measures are the most common measures to assess achievement emotions such as test anxiety (Larsen & Prizmic-Larsen, 2006). These measures ask individuals to report their typical or habitual level of test anxiety (i.e., trait test anxiety) or their current level of test anxiety (i.e., state test anxiety). Trait test anxiety measures are the most popular type of self-report measures. One of the most frequently used self-report instruments is the Test Anxiety Inventory (TAI, Spielberger et al., 1980; Szafranski, Barrera, & Norton, 2012). Trait test anxiety measures typically involves paper-pencil questionnaires (e.g., a typical sample item would be *"Even when I am well prepared for a test I feel very anxious about it"*; TAS, Sarason 1978). It is common for these measures to assess both the cognitive component (also referred to as worry) and the affective component of test anxiety (also referred to as emotionality; Morris et al., 1981). Worry is assessed with items about being concerned about one's achievement or the consequences of failure, whereas the emotionality component is assessed with items about the emotional distress experience (e.g., nervousness or tension) or self-perceived arousal in terms of reactions of one's autonomic nervous system (Spielberger et al., 1980). The advantages of trait measures of anxiety are that they offer quick and cost-effective assessments of self-observed thoughts and emotional experiences (Wilhelm & Grossman, 2010). However, these measures can be biased, for instance by memory effects and by subjective beliefs (Buehler & McFarland, 2001; Fredrickson & Kahneman, 1993; Goetz et al., 2013; Robinson & Clore, 2002).

State test anxiety measures examine students momentarily experienced anxiety in a test situation (e.g., via ambulatory psychological assessment; a typical sample item could be *"How much anxiety are you experiencing right now?"*). These state assessments of anxiety are less susceptible to bias due to closer temporal proximity to the test (i.e., real time or "online self-reports") and because they are situated in the testing situation (i.e., real life e.g., in the school context during a test; Wilhelm & Grossman, 2010). However, during tests it can be difficult for students to report their current emotional state (e.g., because of high cognitive load during a test; Putwain, 2007; Wilhelm & Grossman, 2010). Moreover, there are concerns that even these "online self-reports" are also susceptible to bias. For example, individual differences in the awareness of and willingness (i.e., response tendencies or social desirability) to report on emotional states can influence responding (Mauss & Robinson, 2009). Therefore, researchers often emphasize the need to use measures that are less susceptible to such biases (Podsakoff, 2003).

Physiological measures

The above mentioned limitations of self-report measures call for applying physiological measures of emotions (Houtveen & de Geus, 2009; Wilhelm & Roth, 2001).

Physiological channels (e.g., heart rate, cortisol sampling or skin conductance) are difficult for individuals to mask or control which creates the possibility of more objectively gauging individuals' arousal in testing situations (Harley, 2015; Houtveen & de Geus, 2009; Scollon, Prieto, & Diener, 2009). There are two primary biological systems that become active in a coordinated effort to respond to stress and, by extension, test anxiety. The first is the autonomic nervous system which is responsible for modulating peripheral function and consists of sympathetic and parasympathetic branches, which are generally associated with activation and relaxation, respectively (Mauss & Robinson, 2009). In response to stress, the autonomic nervous system reacts by making the body ready for action, including increasing heart rate and vasodilation. However, it should be noted that different emotions (e.g., anxiety and excitement) can have the same physiological signatures and that as the autonomic nervous system serves a general-purpose, its activity is not exclusively a function of emotional responding, but rather encompasses a wide variety of other functions such as homeostasis (i.e., keeping the body's internal environment in balance) and digestion (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000). The second system is the adrenocortical system which is activated via the hypothalamic-pituitary-adrenal (HPA) axis. This system is responsible for regulating the hormone cortisol, which is often referred to as the classical stress hormone (Lundberg, 2005). Physiological measures of arousal measure the body's above-mentioned response to stress.

As the body's responses to stress are numerous, there are several physiological measures to examine test anxiety. For the present review, we selected the measures that are typically and currently used (Hodges, 2015; Zeidner & Matthews, 2011). These are cardiovascular measures, such as *heart rate (HR), heart rate variability (HRV)* and *blood pressure (BP);* electrodermal activity measures (i.e., often also referred to as electrodermal activity, EDA; Boucsein, 2012), such as *skin conductance (SC)* or *skin resistance (SR);* and endocrine measures, such as *cortisol sampling* (Kantor et al., 2001; Martin, 1961). Recently, measures of *pH level in saliva* has gained increased attention (Marques et al., 2010) as there is evidence of an association between saliva pH and anxiety (Morse, Schacterle, Furst, Esposito, & Zaydenburg, 1982). We therefore decided to additionally include this measure in our review. In the next paragraphs we describe each of these measures in more detail.

Heart rate (HR) is a cardiovascular measure (i.e., response of the blood circulatory system) and refers to the number of heart beats per time interval (e.g., beat per minute, bpm; Hugdahl, 1995). The normal resting adult heart rate is between 60 and 100 bpm (Avram et al., 2018; Spodick, 1993). Heart rate provides non-invasive information about modulation of heart rate by the autonomic nervous system (McCraty, Atkinson, Tiller, Rein, & Watkins, 1995) and is sometimes assumed to directly reflect anxious arousal (i.e., the higher the heart rate, the more anxiety; Calvo & Miguel-Tobal, 1998). However, measuring heart rate also has some limitations as it does not only increase due to emotional states but also with increased mental work load at the beginning of a task or with increased physical activity (Myrtek, Dieterle, & Brügner, 1990; Saito & Nakamura, 1995). Therefore, in emotion research heart rate change scores or so called additional heart rate is often used, which is defined as heart rate acceleration without a corresponding increase of physical activity (Turner, Carroll, Hanson, & Sims, 1988).

Heart rate variability (HRV) is commonly described by the normally occurring variation of the time interval between heartbeats and is expressed by the interbeat interval (IBI; Shaffer, McCraty, & Zerr, 2014). IBI are fluctuations in heart rate that result from complex interactions among a number of different physiological systems. Interest in the relationship between heart rate variability and negative emotional reactions has increased after findings that suggest that heart rate variability is associated with perceived stress (Sloan et al., 1994). An optimal level of heart rate variability within a person is presumed to reflect healthy function, adaptability, and resilience (McCraty & Zayas, 2014). While very high heart rate variability (i.e., too much instability; e.g., arrhythmias) can be detrimental to efficient physiological functioning, low heart rate variability (i.e., too little variation) may be a risk factor for cardiac events (e.g., heart attack, stroke) and can predict the development of hypertension (i.e., high blood pressure), because it reflects reduced regulatory capacity to respond to physiological challenges such as stress and exercise (Berntson, Norman, Hawkley, & Cacioppo, 2008; Bigger et al., 1992; Matthews, Salomon, Brady, & Allen, 2003; Singh et al., 1998).

Blood pressure (BP) is measured in millimeters of mercury (mmHg) and is multiply determined: by the amount of blood pumped by the heart and by how easily blood flows through arteries (Obrist, 1976). *Systolic blood pressure (SBP)* refers to the amount of pressure exerted against arterial walls when the heart is contracting (i.e., maximum during one heart beat). *Diastolic blood pressure (DBP)* describes the amount of pressure exerted against arterial walls between heart beats (i.e., minimum between two heart beats; Hugdahl, 1995). Temporary increases in blood pressure and heart rate provide an indication of the activation of the sympathetic nervous system, which helps supply the body with resources to fight an immediate stressor. With age, blood pressure reactivity to stressful events typically increases (Uchino, Berg, Smith, Pearce, & Skinner, 2006; Uchino, Holt-Lunstad, Bloor, & Campo, 2005). Therefore, it can be important to consider age as a factor that can potentially influences the association between anxiety and physiological response.

Skin conductance (SC) and *skin resistance (SR)* are electrodermal parameters (i.e., sweat gland responses). They are often also referred to as electrodermal activity (EDA; Boucsein, 2012). Electrodermal activity varies with the state of sweat glands in the skin. If the sympathetic branch of the autonomic nervous system is highly aroused, then sweat gland activity also increases, which in turn increases skin conductance. In this way, increases in skin conductance (measured in microsiemens, μS) have been found to be correlated with emotional states (Carlson, 2013). It is useful to differentiate between tonic skin conductance level (SCL) and short-duration (i.e., phasic) skin conductance responses (SCRs). Skin resistance (SR) is an inverse index of skin conductance (i.e., for test anxiety one would expect that the higher the SR the less test anxiety). Electrodermal activity can be most reliably and validly recorded on the palm of the non-dominant hand, however temperature and humidity can affect EDA measurements, which can lead to inconsistent results (Boucsein, 2012). In comparison to heart rate responses electrodermal responses are relatively slow. They appear after a stimulus with a delay of approximately one to three seconds. This shows the complexity of determining a relationship between electrodermal activity and other physiological measures (Martin, 1961).

Cortisol can be sampled from the blood, the urine and the saliva and is expressed in micrograms per deciliter (i.e., mcg/dL). Additionally, in the case of chronic stress it can be found in the hair (Meyer & Novak, 2012). Salivary cortisol levels are highly correlated with free serum cortisol levels (Dorn, Lucke, Loucks, & Berga, 2007). Levels of cortisol follow a diurnal rhythm, which means they are usually high in the morning and then decrease throughout the day. An increased release of cortisol activated via the hypothalamus pituitary adrenal axis is often associated with the experience of academic stress during examinations (Dickerson & Kemeny, 2004). In addition, caffeine consumption, eating patterns, physical activity or sleep patterns can also affect the release of cortisol (Bouma, Riese, Ormel, Verhulst, & Oldehinkel, 2009; Liu et al., 2017; Pollard, 1995; Schwartz, Granger, Susman, Gunnar, & Laird, 1998).

PH level in saliva has gained increased interest and has recently been suggested as a possible useful biomarker indicating psychological stress levels (Marques et al., 2010). The regulation of saliva volume and its composition is regulated by the sympathetic and parasympathetic nervous system. Under stress this may lead to a lower rate of secretion of the salivary glands in the mouth (often expressed by dry mouth) in reaction to stress, which in turn leads to increased acidity and a decrease in oral pH (Humphrey & Williamson, 2001). The advantage of this measure is that it is inexpensive, noninvasive and relatively easy to collect.

4.4. Research Questions

As noted, from a theoretical perspective, one would assume convergence between self-reported test anxiety (i.e., including the subjective experience of all components) and the above described physiological measures since heightened physiological arousal is one component of test anxiety (Zeidner & Matthews, 2005). Furthermore, prominent models of stress (Selye, 1976) would also suggest a relationship between emotional self-reports and physiological reactions to exams. To our knowledge, no review has systematically looked at this relation or has quantified the relation. As such, this review aims to systematically integrate the existing evidence. Furthermore, we want to identify aspects that potentially influence the relationship between self-report measures of test anxiety and physiological measures (e.g., study design and setting, sample characteristics or the assessment and analysis of physiological data). Finally, some researchers (e.g., Morris et al., 1981) state that physiological indices provide information concerning the emotionality component of anxiety but little information about the worry component. They assumed that the inconsistent findings on the relations between physiological and self-report measures may be partly due to combining emotionality and worry in the same self-report measures. Therefore, we additionally address the question whether the relationship between physiological arousal and the emotionality component of self-reported test anxiety is different from the relationship between physiological arousal and the worry component.

Thus, we aim to answer the following research questions:

(1) Is self-reported test anxiety positively related to objectively measurable physiological arousal (and can we quantify this relation)?

(2) Are there methodological variables that moderate this relation (e.g., study design, setting, sample characteristics, assessment and analysis of physiological data)?

(3) Are there differential relations between the worry and emotionality components of test anxiety, on the one hand, and physiological arousal, on the other?

4.5. Methods

4.5.1. Search Strategy

We conducted a comprehensive literature search in PsycINFO, PubMed, and ERIC databases to identify studies addressing the relationship between trait or state self-reported test anxiety and measures of physiological arousal. The physiological measures included cardiovascular measures, electrodermal measures, cortisol sampling, and saliva pH. No date restrictions were placed on any searches.

4.5.2. Inclusion Criteria, Screening, and Coding

Eligible studies were those that were published in peer-reviewed journals and reported findings on the relationship between test anxiety in school or academic contexts (e.g., elementary school, high school, university) and the above-mentioned physiological measures. In the first phase of the literature search, studies were preselected using the search terms *"test anxiety"* OR *"exam anxiety"* OR *"examination anxiety"* AND "[PHYSIOLOGICAL MEASURE, e.g., *heart rate*]". The search terms had to appear either in the title, the keywords or the abstract of the study. The exact search terms are listed in Table 7. Next, the results from the initial search results were screened for peer-reviewed empirical studies (e.g., no theoretical work) published in English (see Figure 8).

In the second phase, the full texts of the remaining articles were retrieved. Then the abstracts and method sections of these studies were screened and only studies that used at least one self-report measure of test anxiety and one kind of physiological measure (i.e., as defined by the search terms) were included for a detailed review. To be eligible for inclusion, the self-report measure needed to be either a test anxiety questionnaire (i.e., mostly trait test anxiety questionnaires like the Test Anxiety Scale; TAS, Sarason 1978; or the Test Anxiety Inventory; TAI, Spielberger et al., 1980) or state anxiety items that asked about test anxiety immediately in the context of an exam (i.e., exam-related anxiety before, during or after an exam) or other testing situation, for example in a lab setting (e.g., intelligence test or problem/anagram solving test). Social-evaluative situations (e.g., public speaking) that did not involve a test instruction were not considered for the current review. The same applied for trait measures of general anxiety that were not administered within the context of an exam or testing situation.

In the last step of the second phase, after detailed review of the results, only studies were retained that fulfilled the following criteria: (1) they reported about non-significant or significant relationships between self-reported test anxiety and physiological measures, or (2) they explicitly reported correlation coefficients, or (3) they gave sufficient statistical information to enable us to calculate these correlations. Intervention studies were only included when they reported results for the pretreatment period or for the control group. Results of the treatment group were not included in this review because the effects of the intervention may have confounded the relations. Similarly, in studies that used experimental designs (e.g., designs that involved classical conditioning to induce increased arousal) where the relationship between self-reported test anxiety and physiological measures could not be disentangled from the effects of the experimental paradigm, were not included. Moreover, studies including participants with a clinical diagnosis, for example an anxiety disorder, were excluded from the current review because test anxiety could not be disentangled from the clinical diagnosis.

In the third phase reference lists of the identified publications were screened for further studies meeting the specified criteria. The procedure of the literature search is depicted in Figure 8.

Table 7. Search terms

		PHYSIOLOGICAL MEASURE
		Physiology
		Physiological arousal
		Autonomic nervous system
		Cardiac
		Cardiovascular
"TEST ANXIETY" OR "EXAM ANXIETY" OR "EXAMINATION ANXIETY"	**AND**	Heart rate
		Pulse
		Blood pressure
		Electrodermal
		Skin conductance
		Skin resistance
		Cortisol
		Saliva pH

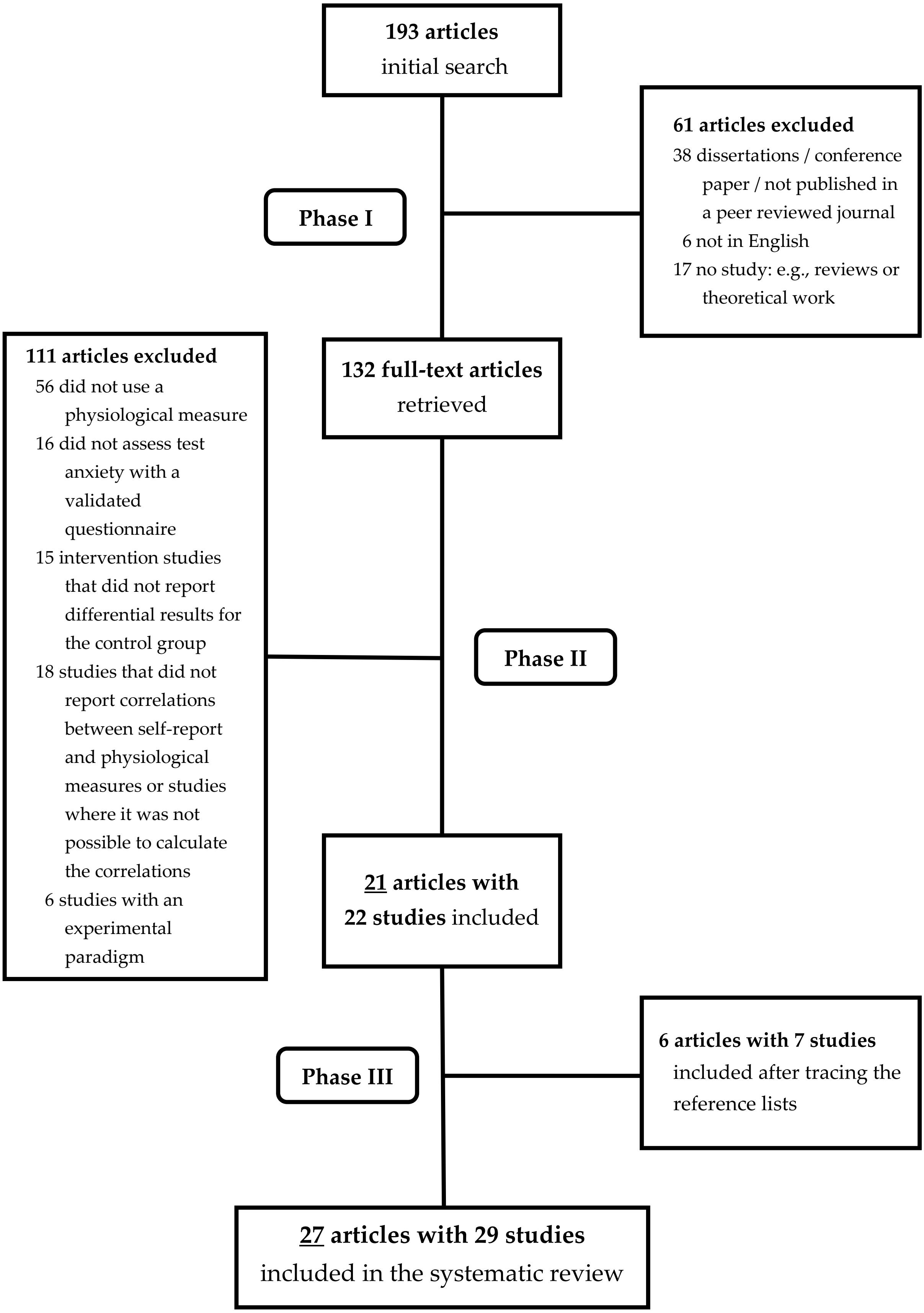

Figure 8. Procedure and results of the literature search.

4.5.3. Data Coding and Synthesis

Studies included for review were coded for the variables identified a priori (for a full description of the variables, see A1 and A2 in the Appendix). Any inconsistencies in the study selection process were resolved by discussion between the first author and the coauthors.

Due to the limited comparability of the data and studies (e.g., heterogeneity among results and methodologies in the studies because of variations in sample size or use of different physiological and self-report measures), it was difficult to conduct a formal meta-analysis. Therefore, we decided to qualitatively summarize the studies. However, we also provide a quantification of the relationship between self-reported test anxiety and physiological measures. Statistically, correlations allow for a meaningful integration of findings across studies with correlational as well as (quasi-)experimental designs in which groups are compared (Schmidt & Hunter, 2015). Therefore, for significant effects, correlation coefficients and effect sizes d were determined. They were either extracted from the primary studies (e.g., Pearson's r in studies with a correlational design) or calculated from descriptive and inferential statistics (e.g., from a t or F value in group-based studies). Where studies yielded multiple effect sizes for a given relation, all were extracted.

A mean standardized correlation coefficient across all studies that found significant correlations was calculated separately for each physiological measure. This was done by a Fisher's z transformation of the raw correlation coefficients accounting for the different sample sizes of the studies. Furthermore, to provide a correlation across all measures, a mean correlation coefficient across all studies and across all physiological measures was also calculated using the method mentioned above. Additionally, in the same way we calculated a mean correlation coefficient that also included the results of the studies that found non-significant correlations whenever these results were available. If studies reported results from more than one physiological measure, mean correlations and effect sizes from these measures were calculated. The magnitude of the correlations was interpreted according to Cohen (1988). Correlation coefficients of .10 are considered as small, those of .30 are medium, and those of .50 as large.

4.6. Results

4.6.1. Results from the Systematic Literature Search

The initial literature search was concluded on February 27, 2017 and resulted in 193 studies with publication dates ranging from 1957 to 2016. The majority of studies (188) were found with the search word "test anxiety". The synonyms "exam anxiety" and "examination anxiety" only yielded five additional relevant studies. We then screened these studies and excluded 61 studies not meeting the eligibility requirements of the first phase (e.g., 38 studies were not published in a peer-reviewed journal; see Figure 8 for details about these exclusions).

As depicted in Figure 8, we evaluated the remaining 132 studies in the second phase and excluded 111 studies for not meeting the eligibility requirements of this phase. For example, we excluded 56 studies that did not use a physiological measure and 16 studies that did not assess test anxiety but rather other constructs related to test anxiety such as stress during public speaking or math anxiety. There was a total of 22 studies from 21 articles with publication dates ranging from 1957 to 2014 that met the specified criteria and were included for review.

In the third phase, we screened the reference lists of these studies which led to the inclusion of six additional articles reporting seven studies. Thus, our systematic literature search resulted in 27 articles with 29 studies. The selected studies and the study characteristics are depicted in Tables A1 and A2 in the Appendix.

4.6.2. Characteristics of the Selected Studies

Samples

Participants in the selected studies were primarily college or university students (23 out of 29 studies) among them mostly were psychology undergraduates (16 studies) who were for the most part from the United States. Only six studies included younger age groups (i.e., under 16 years old, e.g., elementary school or high school students). The mean age of participants across studies was M = 20.86 years (SD = 6.73, median = 21.5, range from 7 to 43 years). However, in more than half of the studies (55 %) the exact age of the participants was not reported. The average sample size was N = 58.38 (SD = 35.76, median = 56). However, it should be noted that there was broad range between N = 6 and N = 171 in terms of sample size across studies.

Study setting and design

Approximately half of the studies (i.e., 15 out of 29 studies) were laboratory studies and used an experimental test anxiety elicitation paradigm. These experimental test situations involved tasks like anagrams or other problem-solving tasks (six studies), general aptitude tests (six studies), vocabulary tests (two studies) or psychology tests (one study). Most of the real-life studies (i.e., 10 out of 14 studies) were conducted within the context of a psychology course exam for undergraduate students. Two studies were conducted in the context of a medical exam. Other real-life settings were a mock French exam and a nursing exam. None of the naturalistic studies assessed self-reported test anxiety during the test, instead it was assessed before or after the test or with a trait test anxiety inventory. However, one study used a video-based interview to retrospectively assess self-reported test anxiety that participants experienced during the test (Spangler, Pekrun, Kramer, & Hofmann, 2002; study 2). For more details about the study setting and design, see Tables A3 and A4 in the Appendix. Most of the reviewed studies focused on test anxiety and performance outcomes (15 studies), other common variables of interest were coping styles (four studies) or general anxiety (five studies). One study looked at treatment effects (McGlynn, Bichajian, Giesen, Rullan, & Pulver, 1981).

Assessment of self-reported test anxiety

Most of the reviewed studies used trait self-report measures of test anxiety (i.e., 21 out of 29 studies). Among these measures, the Test Anxiety Scale (TAS; Sarason 1978) or the child version of this scale (TASC; Sarason et al. 1958) were the most common measures (i.e., 11 out of 29). The second most frequently employed measure was the Test Anxiety Questionnaire (TAQ; Mandler & Sarason, 1952), which was used in five studies. The Test Anxiety Inventory (TAI; Spielberger et al., 1980) was used in three studies. Two studies used other trait test anxiety questionnaires. From the studies that applied trait measures there was one study that used a combination of two trait measures to assess test anxiety, and there were four studies in which additionally to the trait test anxiety measure the researchers applied a state measure. The remaining studies used state anxiety items before a test or exam (eight studies). More details can be retrieved from Table A5 in the Appendix.

Assessment of physiological parameters

Multiple physiological measures were used in 14 out of 29 studies but only 10 of these studies applied a combination of two measures that were considered for the current review. The other four studies used a combination of additional measures (not considered for the review) as described below. Cardiovascular measures were the most commonly used measures: 21 studies assessed heart rate or heart rate variability and four studies assessed blood pressure. Seven studies applied electrodermal measures, five studies used cortisol sampling and two studies assessed saliva pH. Additional measures that were employed but not considered for the current review (e.g., as they are not regarded as a typical indicator of test anxiety or because they are no longer in use or were not successfully assessed in these studies) were respiration rate (three studies) or FPV (finger pulse volume), testosterone, prolactin, LH, skin and urinary pH, and sIgA (secretory immunoglobulin; one study each). The majority of the studies used well-established devices to assess physiological arousal; however, in four of the reviewed studies participants counted their own heart beats.

Ways of comparing self-report and physiological measures

In 14 studies samples were split into high and low test anxious groups based on a general index of self-reported test anxiety as indicated by their sum score on a test anxiety scale. Thus, these studies only compared physiological arousal between high and low test anxiety groups. Eleven studies used a continuous assessment of test anxiety and correlated the scores on the test anxiety scales with physiological arousal without grouping participants. Four studies used both categorical and continuous measures.

Twelve studies differentiated between the worry and emotionality components of test anxiety, but only nine of these studies reported separate results on the components' correlation with physiological measures. The other three studies only compared total scores of test anxiety with physiological measures.

4.6.3. Correlations between Self-Report and Physiological Measures

Correlation coefficients for different physiological measures

Table 8 shows the correlation coefficients (with and without being Fisher's *z*-transformed) for the significant relations between self-reported test anxiety and the different kinds of measures. In the second column we noted how many studies found significant correlations out of the total number of studies that used the respective measure. Note that multiple measures were used in 14 out of 29 studies.

Table 8. Correlation coefficients for different physiological measures

Physiological measure	Number of studies with significant results	Total number of studies that used this measure	r	$r_{\text{Fisher's } z}$
Electrodermal measures	4	7	0.37	0.34
Heart rate/ heart rate variability	13	21	0.34	0.31
Blood pressure	2	4	0.32	0.27
Cortisol	2	5	0.38	0.38
Saliva pH	2	2	0.39	0.37

Note. Multiple measures were used in 14 out of 29 studies. The mean correlations were calculated based on the significant correlations.

Correlation coefficients across all physiological measures

Results revealed that 21 studies found significant positive correlations between self-reported test anxiety and at least one measure of physiological arousal. The mean correlation across these studies with significant correlations was $r_{\text{Fisher's } z} = 0.32$ (range: 0.19 to 0.51), and the mean effect size was $d = 0.76$ (range: 0.39 to 1.01). Seven of the reviewed studies did not find significant correlations, and in one study low anxious students were found to be more aroused than high anxious students. The mean correlation across all studies including these non-significant correlations was $r_{\text{Fisher's } z} = 0.29$ (range: -0.12 to 0.51), and the mean effect size was $d = 0.65$ (range: -0.24 to 1.01). However, it has to be noted that four from the eight studies that found non-significant correlations did not report their results (for details see Table A2 in the Appendix) and therefore could not be included into this mean correlation coefficient. We noted that the different strengths of correlations can be explained by different aspects of the reviewed studies like the study setting or the sample size. We will address these aspects and their proposed effect on the convergence between the two measures in the discussion section.

Correlation coefficients for different test anxiety components

In total, nine of the 29 reviewed studies differentiated between the worry and emotionality components of test anxiety and reported differential results for the two components. Only one study (Herbert, Moore, De La Riva, & Watts, 1986) did not find significant results between either worry and serum cortisol or emotionality and serum cortisol. From the remaining eight studies (Cohen and Khalaila 2014; study 1 and 2 by Conley & Lehman, 2012; Deffenbacher, 1986; Deffenbacher & Hazaleus, 1985; Endler & Magnusson, 1977; Kantor et al., 2001; study 1 and 2 by Morris & Liebert, 1970), six studies found higher correlations between the emotionality component of test anxiety and physiological measures. Only two studies found higher correlations between the worry component and the physiological measure. The exact values of the correlations can be seen in Table 9. In most of these studies it was not reported if the correlations for worry and emotionality were significantly different from each other.

Table 9. Correlations between test anxiety components and physiological measures

Author	Physiological measure	Worry		Emotionality	
		r	*d*	*r*	*d*
Conley and Lehman (2012)	SBP	**0.25***	0.53	0.14	0.28
Deffenbacher (1986)	HR	0.14	0.28	**0.17***	0.34
Deffenbacher and Hazaleus (1985)	HR	0.20	0.41	**0.26***	0.54
Morris and Liebert (1970)	HR	0.29	0.60	**0.34***	0.71
	HR change[a]	0.15	0.30	**0.27***	0.56
Morris and Liebert (1970)	HR	n.s.[b]	-	**0.24***	0.49
Cohen and Khalaila (2014)	SpH	0.11	0.22	**0.33**	0.49
Endler and Magnusson (1977)[c]	HR	**0.56****	1.33	0.38**	0.81
Kantor et al. (2001)	HR	0.41*	0.84	**0.53****	1.18

Note. Abbreviations and full names: Physiological parameters: HR = Heart rate; SBP = Systolic blood pressure; SpH = Saliva pH. n.s. = in this study physiological measure did not correlate significantly with the component. **Bold** = this correlation was higher. **p* < .05; ***p* < .01.

[a] worry and emotionality were not significantly different for HR. However, HR change scores significantly correlated with emotionality (r = 0.27) but not worry.

[b] correlation coefficient was not mentioned in this study.

[c] in this study the worry component was called cognitive component and the emotionality component was called physiological component.

4.7. Discussion

The present investigation was intended to synthesize past research findings on the relationship between self-report measures of test anxiety and physiological measures in order to help advance physiological research in education and provide a deeper understanding of this widespread emotion. In the following sections we discuss the results of our review and implications for future research.

4.7.1. Relationship between Self-Reported Test Anxiety and Objectively Measurable Physiological Arousal

The majority of the reviewed studies (i.e., 21 studies, 72 %) found significant positive correlations between self-reported test anxiety and at least one measure of physiological arousal (mean significant correlation across all measures: $r_{\textit{Fisher's z}}$ = .32; mean correlation across all measures including also the non-significant correlations $r_{\textit{Fisher's z}}$ = .29). Thus, the results of the present review are generally consistent with theoretical concepts and support the assumption that higher self-reported test anxiety is associated with higher physiological arousal. The size and range of the observed significant correlations (from r = .19 to r = .51) indicate that the two types of measures do not completely overlap. This is in line with theoretical assumptions, as physiological measures are assumed to only tap into the physiological component of test anxiety, whereas self-report assessments usually represent all test anxiety components as subjectively experienced (i.e., also cognitive, affective, motivational, and behavioral components; Pekrun et al., 2011; Wilhelm & Roth, 2001). From the significant but moderate relationship between self-report and physiological measures, one could further conclude that using physiological in addition to self-report measures is highly relevant because they might add information about the physiological component of test anxiety that is not accessible with self-report alone (Wilhelm & Roth, 2001). Thus, this result also provides support for the view that in order to get the most comprehensive assessment of test anxiety, different kinds of test anxiety measures should be used (e.g., Allen, Elias, & Zlotlow, 1980; Kantor et al., 2001; Zeidner & Matthews, 2011).

4.7.2. Aspects That Are Potential Moderators of the Relationship

As our results suggest, a substantial relationship between test anxiety and physiological arousal in the majority of the reviewed studies, we now want to more closely look at aspects that might influence this relationship. We first start discussing aspects of the studies with non-significant or contradictory results that may explain the apparent divergence in findings and draw implications and considerations for future research. This will be followed by a detailed discussion of the studies with significant results.

Studies with non-significant or contradictory results

Seven of the reviewed studies (24 %) did not find significant correlations between self-reported test anxiety and objectively measurable physiological arousal, and in one study low anxious students (i.e., as classified by self-report measures) were found to be more physiologically aroused than high anxious students. These contradictory and non-significant findings of the reviewed studies are very informative.

Study setting. We noted that more than half of the studies with non-significant or contradictory results were conducted within a lab setting (five from eight studies). These laboratory experimental approaches have distinct advantages. For example, because of the constrained lab environment physiological measurements contain relatively few artifacts from participant's movement or technical problems (Houtveen & de Geus, 2009). Moreover, conducting a study in a lab avoids difficulties associated with the lack of portable, sophisticated equipment and additionally allows for precise control over the task (Wilhelm & Grossman, 2010). However, a major disadvantage is that laboratory studies are often conducted under artificial conditions and lack ecological validity and therefore often personal relevance for the participants (Houtveen & de Geus, 2009). Furthermore, it might be the case that subjects are simply excited about the lab or the experiment and not necessarily test anxious. In this case, increased physiological arousal in laboratory settings could also be interpreted as an indicator of activation due to positive interest (i.e., approach motivation) in the task or a positive interest or excitement cued off by the complex experimental setup that participants face in a lab (Schlosberg, 1954).

Thus, elevated physiological arousal may merely be an artifact of the laboratory situation and not necessarily an indicator of anxiety. This is also supported by a review of Johnston, Anastasiades, and Wood (1990) who compared cardiovascular responses in the laboratory to those under natural conditions and found that approximately half of the reviewed studies reported contradictory findings. Zanstra and Johnston (2011) published a review on the extent to which responses to laboratory and natural realistic stressors are comparable. They found that responses obtained in real life are often larger than those obtained in the laboratory and that subjective ratings of stress, emotion and cognitive determinants in real life also relate to real- life cardiovascular responses. This calls for determining the relation between self-reported test anxiety and physiological arousal under *more natural conditions* (i.e., real-life assessment).

Sample characteristics. The three naturalistic studies with non-significant results also showed methodological problems that could in part explain the non-significant findings. Two of the three studies had problems of sample size. One study used a mock exam and had a relatively *small sample size* (N = 23) and a mostly female sample. The authors (Daly, Chamberlain, & Spalding, 2011) stated that their work can be seen more like a pilot study, that the results need to be tested in a larger sample under more naturalistic conditions, and that their *sample (high performers) might have been generally less anxious* about taking a test. The second real-life study

(Herbert et al., 1986) looked at the cortisol response of male medicine students before a major medicine exam. This is also a very *special sample* as medical students from the UK are most likely used to taking such exams and are probably also high-achieving students. Moreover, because the participants *were all male*, the findings of this study (i.e., non-significant relationship) are not necessarily generalizable to other samples (e.g., female participants).

The lab study had the smallest sample size ($N = 6$). This study found the contradictory result of low test anxious students being more physiologically aroused (i.e., higher HR) than high anxious students (Hollandsworth, Glazeski, Kirkland, Jones, & Van Norman, 1979). The small size of the sample reduces the generalizability of these results.

Assessment and analysis of physiological data. The third naturalistic study with non-significant results (Huwe, Hennig, & Netter, 1998) had a larger sample size ($N = 58$) but also showed some major methodological problems as heart rate was counted manually by the experimenter (i.e., not with a *sophisticated device*). However, this study found that heart rate and cortisol response at least tended to be higher in high test anxious students than in low test anxious students – although this difference did not reach significance.

Another methodological issue of one of the studies with non-significant results (Glazeski, Hollandsworth, & Jones, 1986) was the use of a *single arousal index* calculated from three different measures (heart rate, skin conductance and respiration) averaged across the whole test situation. In detail, Glazeski et al. (1986) used a single score of arousal for each participant which they computed from mean scores on three different measures administered during test taking. Later, based on these single scores, they additionally subdivided their sample into groups with high, moderate and low arousal. They did not find any differences between high and low test anxiety participants in arousal group membership ($N = 15$ for each group). By averaging and merging the different physiological measures, the authors might have simply overseen the relations between different physiological parameters and test anxiety. As mentioned, different physiological measures have *different temporal resolutions*, and there may be systematic differences between parameters representing different physiological systems. Therefore, it is problematic to combine them (Martin, 1961). This shows that the selection of physiological measures should be well grounded and that the different sampling rates (i.e., temporal resolutions) and links to different physiological systems should be taken into account when using more than one measure.

Studies with significant results

From the studies with significant results we also found some interesting aspects that might have had an influence on the correlations and that might explain the relatively high variance of the correlations coefficients and thereby help to guide the development of future studies. These aspects concern mostly the *assessment and analysis of physiological data*.

Sophisticated devices. The manner physiological arousal was assessed seemed to play an important role for the magnitude of convergence in the significant studies as well. Correlations were lower or as mentioned above even nonexistent like in the study by Huwe et al. (1998) when physiological assessment was poor (e.g., when heart rate was measured manually). For example, in four studies (Deffenbacher, 1986; Deffenbacher & Hazaleus, 1985; study 1 and 2 by Morris & Liebert, 1970) participants had to count their own heart beats, which yielded low correlations (mean $r_{\text{Fisher's } z} = 0.23$; range: 0.19 to 0.30) as compared with the results from sophisticated devices like heart rate monitors (mean $r_{\text{Fisher's } z} = 0.38$; range: 0.26 to 0.51). As also mentioned by the authors of one of these studies (Morris & Liebert, 1970), a main reason is that participants or even experimenters often fail to obtain a satisfactory count, which leads to missing or unreliable data. This implies that it is highly important to use technologically adequate devices rather than manual counting.

Detailed assessment of physiological data. Another explanation for the high variance of the correlation coefficients might be that some studies included two assessments, one during baseline (i.e., as a control condition) and a second during stress conditions (e.g., Deffenbacher, 1986; Deffenbacher & Hazaleus, 1985; Morris & Liebert, 1970), or reported a single mean score to reflect the level of arousal over the entire session. Other studies included *multiple assessments* (i.e., went more into detail) and investigated parameters under baseline, anticipation, stress, and recovery or even looked at variability of changes in physiological arousal within these phases. Correlations were generally higher with these more fine-grained assessments. For example, Montgomery (1977) focused on the anticipatory cardiac response and took a detailed look at the waveforms. He found relatively high correlations between heart rate and test anxiety as well as a greater heart rate slowing in low test anxious subjects. Raphelson (1957) first found that low test anxious subjects initially (in their null period, i.e., baseline) had higher skin conductance levels than high test anxious (other factors than anxiety might have played a role here). However, when looking at variation during the actual task, high test anxious participants' skin conductance increased and low test anxious participants' skin conductance decreased during the task. Beidel (1988) looked at such variation during a test using heart rate and Spangler et al. (2002) used cortisol sampling and looked at response patterns and found similar response patterns in their studies.

Harleston, Smith, and Arey (1965) also found that the low test anxious group showed a slight drop in heart rate during a test while the high test anxious group sustained the highest increase in heart rate during the test. Furthermore, they found that the relation between anxiety and arousal was significant only during the test when working on the task. If these researchers had only looked at group differences of absolute values at a single point of time (e.g., at baseline) without inspecting change over time, they would have missed important information and may not have found any substantial association between self-report and physiological measures. Significant correlations seemed to emerge more often when participants interacted with the task (i.e., during the test). Furthermore, it seems to be important to consider within-person associations between self-report and physiological measures across time (e.g., Spangler et al. (2002) rather than only considering between-person associations at a single time point (Mauss & Robinson, 2009).

Influencing factors that should be controlled. Additionally, the correlations can be affected by variables like previous food intake or drug use (such as alcohol or smoking). Conley and Lehman (2012) included several covariates (i.e., smoking, activity level, food, alcohol and caffeine) in their data analysis. They found that these factors influenced their results in different ways. Specifically, cigarette use and more strenuous physical activity during ambulatory assessment led to elevated heart rate, diastolic blood pressure and systolic blood pressure; food consumption led to elevated systolic blood pressure and heart rate; alcohol consumption led to elevated heart rate; whereas caffeine had no influence. Controlling these variables is especially relevant for studies that use cortisol or other saliva samples as the results of these measures can be strongly biased by such factors (Hansen, Garde, & Persson, 2008; Pollard, 1995). This can be seen clearly in the study by Cohen and Khalaila (2014) which only found correlations in the predicted directions for emotionality and pH when they controlled for the degree of physical activity per week and habitual smoking.

Specific characteristics of different physiological measures. We also found differences between the different physiological measures (see Table 4). Saliva pH, cortisol and electrodermal measures generally yielded higher correlations than heart rate and blood pressure measurements (i.e., cardiovascular measures). This can be explained by the fact that some of these measures may be more specific to emotional arousal, like electrodermal activity which is known to be solely controlled by the sympathetic nervous system (e.g., Setz et al., 2010). On the other hand, some measures such as heart rate scores do not only increase during sympathetic activation but also show increased values when cognitive load is high or during general activation or physical activity (Myrtek et al., 1990; Saito & Nakamura, 1995). For example, McGlynn et al. (1981) found a correlation between test anxiety and skin conductance but no correlation with heart rate. This shows that it is important to keep in mind that the different measures might provide different indices of arousal and are not necessarily correlated with each other. Moreover, they may also vary in the strength of their association with emotions (Kreibig, 2010).

Appropriate data analysis. A final important point is the way the data is analyzed. As mentioned earlier, there might be an intra-individual (i.e., within-person) relationship between physiological and self-report measures that cannot be accounted for when using a between-persons approach and only considering inter-individual differences. Conley and Lehman (2012) accounted for these intra-individual processes by using hierarchical linear regression modelling, based on multiple blood pressure readings for each participant (i.e., high sampling rate) along with self-reports at each blood pressure reading. In future work it could be worthwhile to apply more of such intra-individual approaches.

4.7.3. Worry and Emotionality

The third goal of the current review was to find out whether the relationship between the emotionality component of test anxiety and physiological arousal is different from the relationship for the worry component. Although valid instruments (such as the TAI) that include items for both components were used in most of the reviewed studies, only nine studies differentiated between the two components and reported differential results. Only one of these studies (Herbert et al., 1986) did not find significant relations between either worry and serum cortisol nor emotionality and serum cortisol. From the remaining eight studies, six studies found higher correlations between the emotional component and physiological measures. Of the remaining two studies, one study (Conley & Lehman, 2012) found that only worry predicted elevations in SBP. The other study (Endler & Magnusson, 1977) also found higher correlations for worry and HR, but the correlations for the emotionality component also reached significance.

However, most of these studies did not report whether the differences between the correlations were significant. Liebert and Morris (1967) tested the differences in correlations for significance and found that the correlations between emotionality and heart rate and worry and heart rate did not significantly differ. They explain this due to the fact that the correlations they found were much smaller than expected, which might have resulted from poor physiological assessment (i.e., participants had to count their own pulse rates).

In sum, in comparison to the worry component, the emotionality component seems to be more closely related to physiological arousal. The stronger association between emotionality and physiological arousal is in line with theoretical assumptions, as the self-report items for the emotionality component more strongly tap into aspects of physiological arousal than the items for the worry component (Deffenbacher & Hazaleus, 1985; King et al., 2000; Spielberger & Vagg, 1995). The medium size of the correlation coefficients in these studies also is not very surprising as perceptions of physiological arousal (i.e., as partly reflected in emotionality items) are not necessarily strongly correlated with actual autonomic reactivity as assed with objective physiological measures (Hodges, 2015). By implication, physiological measures can provide additional information (especially about the physiological test anxiety component) that cannot be assessed with self-reports alone.

Surprisingly all eight studies with significant results that were differentiating between the worry and emotionality components only used cardiovascular measures (six studies used heart rate and one study used blood pressure measurements) and saliva pH (one study). Moreover, we identified a lack of research on electrodermal activity and cortisol and their relation to the worry and emotionality components. Future work could close this research gap.

4.7.4. Strengths and Limitations of the Current Review

Due to the heterogeneity of the included studies, we chose to perform a systematic review rather than a meta-analysis. The advantage of a systematic literature review is that a broader and more complete picture can be gained because the review can be more inclusive than in a formal meta-analysis. Thus, we could include samples with different demographic backgrounds (e.g., different age groups) and physiological and self-report measures without a restrictive focus on only a few comparable dimensions. For example, only one specific sample or physiological parameter. This was done to provide ideas for generating hypotheses and future directions for applying physiological measures in the field of educational psychology.

There are several limitations of the present review that should be considered when interpreting the findings and can be used to derive directions for future research. One limitation is that the research discussed in this review was obtained from articles that were published in English and in peer-reviewed journals. Therefore, this review does not include findings and information provided by unpublished studies (e.g., unpublished dissertations) or articles published in other languages. Future work could take this limitation into account. Another point is that the number of studies in our review was limited because we did not include studies that looked at constructs that appear similar to test anxiety like for example, exam stress or social evaluative stress. This was done because stress is a broader construct that can include anxiety but does not always do so (e.g., positive stress; Jamieson et al., 2012; Putwain, 2007; Selye, 1976, 1976). Therefore, we only included studies of which we were sure that they did indeed assess test anxiety (i.e., with a test anxiety questionnaire).

4.7.5. Conclusion, Educational Significance, and Future Directions

In sum, we found mostly positive associations between self-reported and physiological measures of test anxiety. The results of this review are promising and in line with theoretical assumptions. Knowing that self-report measures can be strongly biased by subjective beliefs (Robinson & Clore, 2002) and are more likely to capture the affective or cognitive components of anxiety as they assess subjective experience that are represented in the conscious mind (Pekrun & Bühner, 2014), the results of our review suggest that physiological measures can provide objective information about test anxiety, especially about the physiological component of test anxiety that is difficult to capture through self-reports. For research in educational psychology, this implies that is important to use both types of measures in order to capture the complex construct of test anxiety as objective and detailed as possible.

In most of the reviewed studies, the samples included students (i.e., 23 from 29 studies), especially students in psychology (i.e., 16 studies). This is a general problem of studies in this field. The selectiveness of samples limits the conclusions and inferences for educational research that can be drawn (Gordon, Slade, & Schmitt, 1986; Shen et al., 2011). This calls for applying these measures in classroom-based research to get results that are not only specific to a psychology students sample and additionally have high ecological validity. By continuously assessing students' physiological arousal (for example during an exam), we might be able to expand our knowledge of the mechanisms underlying test anxiety in ways not possible using self-report measures alone. This more ecologically valid, more objective and detailed assessment might in the long run also help to improve interventions targeting test anxiety (e.g., cognitive test anxiety treatment and relaxation techniques).

Moreover, as there are studies that question the debilitative aspects of physiological arousal in testing situations (Glazeski et al., 1986), by using both types of measures it would be possible to identify whether it is the interpretation of physiological arousal (i.e., self-reported physiological test anxiety component) or the arousal as assessed by objective physiological measures that influences student's performance, and to investigate which direction this influence takes (Mandler & Kremen, 1958). As this review also shows that psychophysiological results seem to be influenced by different aspects of the methodology used, future research could statistically consider these influencing variables as moderators in their research designs. Furthermore, when applying physiological measures, the study design, setting, sample characteristics, data assessment and data analysis should be carefully selected.

Finally, as our results suggest that each of the two measures (i.e., self-report and physiological measures) provides information about a person's test anxiety and much of the information provided by each measure might be specific to a particular component of test anxiety, in future research, besides physiological measures, studies should also include additional measures of test anxiety like behavioral measures to get a more comprehensive assessment and a more complete picture of the different components of this widespread and mostly detrimental emotion (Larsen & Prizmic-Larsen, 2006; Zeidner & Matthews, 2011).

4.8. Appendix Tables

Table A 1. Description of the study variables

Variables	Description
Authors, Year, Country, Journal	
Sample size,	Sample size of the study (*N*),
Gender ratio (f/m),	Proportion female/male subjects (N),
Age,	Age in years: Mean (M_{age}); Standard deviation (*SD*); Range,
Characteristics,	Sample characteristics: i.e., UPSY, HS, ES, USO,
Cont. or G	Assessment: Continuous (Cont.) i.e., without grouping or Grouping (G) i.e., into HTA/LTA
Setting (L/RL),	Setting (i.e., lab vs real life): L / RL; T: study that involves a treatment. Only results from the pretreatment are reported.
Design,	Design: e.g., experimental anxiety induction (Exp.); exam, test,
Area	Area: e.g., language, problem solving
Self-report measure of test anxiety,	Inventory to assess test anxiety: Note: studies sometimes used other measures to examine other constructs as well (e.g., depression, general anxiety). Only the self-report measure for test anxiety are mentioned here.
Test anxiety components (Y/N)	Distinction between the test anxiety components (Y/N)
Physiological measure(s),	Physiological Measure(s): if in ***(i.e., + respiration)*** this measure was not considered for the current review, no results are reported
Assessment,	Assessment of Physiology: device and site,
Measurement schedule	*Physiology measurement schedule.* Anticipation: After subjects received the instructions but did not start with the test yet
Significant measures,	Measures that yielded significant results, correlation with self-report measure and effect size.
Correlation,	When it is not further mentioned: HTA had higher arousal (as expected): For SR this means lower values for HTA. For HRV this
Effect size	means greater variability for LTA. For pH this means pH sores tend to acidity in HTA.
Measures not significant	Measures where no significant correlation was found.
Baseline (Y/N/NN)	Baseline: Was there a BL assessment (Y/N/NN)? It is also noted when BL was used to calculate change/range corrected scores,
Increase (Y/N/NN)	Increase (y/n/NN): was there an increase from BL,
Significant (sign. / not sign. / Sign. NN)	Significant (sign. / not sign. / Sign. NN): Information about the significance of the increase from BL.
(Mean) significant Corr.,	(Mean) significant correlation: when there was more than one significant correlation a mean score was calculated,
Effect size	Effect size.

Note. All abbreviations (in alphabetical order) can be retrieved from the Note in Table A2.

Table A 2. Studies in systematic review

No.	Authors, Year, Country, Journal	Sample size, Gender ratio (f/m), Age, Characteristics, Cont. or G	Setting (L/RL) Design, Area	Self-reported test anxiety measure, Test anxiety components (Y/N)	Physiological measure(s), Assessment, *Measurement schedule*	Measures significant, Correlation, Effect size	Measures not significant	Baseline (Y/N/NN), Increase (Y/N/NN), Significant (Sign./ not sign./Sign. NN)	(Mean) sign. Corr., Effect size
1	Beidel 1988 USA Journal of Abnormal Psychology	*N* = 50 (23/27) *M*M_{age} = 9.1; 8 - 12 **ES** **G:** 25/25 based on TASC	**L** Exp. 2 tasks: Vocabulary test + oral reading session	**TASC** (grouping) Before exp. **N**	**BP + HR** BP: automatic electrosphygmomanometer HR: pulse-rate monitor Non-dominant arm *(measured during adaptation, BL, vocabulary test, oral reading)*	**HR change** higher in HTA in both tasks r = 0.36 d = 0.79	**SBP** **DBP**	**Y** BL to calculate change scores **NN**	**r = 0.36** **d = 0.79**
2	Calvo & Miguel-Tobal 1998 Spain Motivation and Emotion	*N* = 56 (36/20) M_{age} = 21.0; *SD* = 2.2 **UPSY** **G:** 28/28 based on TAI	**L** Exp. 2 aptitude tasks (motor skills + linguis-tic ability) in front of a camera with ego-threat instructions in writing	**TAI** (grouping) Before exp. **N**	**HR + SR** HR: photoelectric finger pulse transducer (middle finger of non-dominant hand) SR: electrodes (thenar and hypothenar) *(measured during adaptation, BL, anticipation, tasks, recovery)*	**HR** higher in HTA in anticipation phase r = 0.37 d = 0.80 **SR** lower in HTA in anticipation r = 0.22 d = 0.45		**Y** **Y** **Sign.** for HR + SC	**mean** **r = 0.30** **d = 0.62**
3	Cohen & Khalaila 2014 Israel Journal of Psychosoma-tic Research	*N* = 68 (51/32) M_{age}= 21.9, *SD* = 3.0, 19 - 36 **USO** **Cont.**	**RL** Nursing studies Exam (T1) vs non exam period (T2)	**TAI** one hour before exam (T1) and 3 months post-exam (T2) **Y** worry emotionality (from TAI)	**Salivary pH** Saliva samples (Measured at T1 and T2, i.e., BL)	*T1* ***saliva PH*** *and* ***T2 emotionality*** *r = - 0.33* *d = 0.69*	*T1* ***saliva PH*** *and* ***T2 worry*** **corr. with-out control variables:** no corr. between pH and TAI	**Y** + control variables: physical activity, smoking **Y** pH decrease from BL **Sign.**	**r = 0.33** **d = 0.69**

4	Conley & Lehman 2012 USA Stress and Health	*N* = 99 (65/34) M_{age} = 21 **UPSY** **Cont.** and **G:** low, mean, high based on TAI	**RL** Most stressful events of the day (e.g., writing an exam) = acute academic stressor vs. anticipated academic stressors	**TAI short** **Y** Worry Emotionality (8 items each) from TAI short	**BP + HR** ambulatory blood pressure monitors *(Ambulatory assessment over 4 days both at times with and without academic stressor, reports at each time of BP reading, online survey at the end of the day)*	**SBP** higher in HTA r = 0.20 d = 0.41 during acute academic stressors ***SBP** and higher **worry**:* *r = 0.25* *d = 0.53*	**DBP** **HR**	**N** but **intra-individual analysis** and covariates (alcohol, smoking, alcohol, food, caffeine)	**mean** **r = 0.23** **d = 0.47**
5	Deffenbacher 1986 USA Cognitive Therapy and Research	*N* = 171 (114/57) M_{age} = NN **UPSY** **Cont.** and **G:** 42/52 based on TAS	**RL** Psychology exam	**TAS** 5 weeks before exam **State test anxiety**: **STAR** after exam **Y** State worry and emotionality (from STAR)	**HR** Participants count own pulse rate Non-dominant wrist *(measured by participants themselves immediately upon completion of the exam for a 15-second period)*	**HR** higher in HTA **Group** r = 0.26 d = 0.54 **Cont. (TAS)** r = 0.20 d = 0.40 ***Worry*** *r = 0.14* *d = 0.28* ***Emotionality*** *r = 0.17* *d = 0.34*		**N**	**Mean** **r = 0.19** **d = 0.39**
6	Deffenbacher & Hazaleus 1985 USA Cognitive Therapy and Research	*N* = 129 (65/64) M_{age} = NN **UPSY** **Cont**. and **G:** 66/63 based on TAS	**L** Exp. carried out in a classroom General aptitude test	**TAS** before exp. **State test anxiety**: **STAR** After exp. **Y** State worry and emotionality (from STAR)	**HR** Participants count own pulse rate wrist *(measured by participants themselves for a 15-second period before the exp. (i.e., BL) and immediately after the exp.)*	**HR** higher in HTA **Cont. (TAS)** r = 0.26 d = 0.54 ***Worry*** *r = 0.20* *d = 0.41* ***Emotionality*** *r = 0.27* *d = 0.56*	**HR** when **grouping**	**Y** Prepulse rate as BL **NN**	**Mean** **r = 0.24** **d = 0.5**

7	Endler & Magnusson 1977 Canada Canadian Journal of Behavioural Science	*N* = 56 (NN) M_{age} = NN **UPSY** **Cont.**	**RL** Psychology exam	**State exam anxiety**: **BRQ** before exam **Y** Cognitive Physiological (from BRQ)	**HR** device not mentioned *(measured before exam, and one week later (i.e., BL))*	**HR** and **BRQ total score** r = 0.59 d = 1.44 ***BRQ Cognitive*** *r = 0.56* *d = 1.33* ***BRQ Physiological*** *r = 0.38* *d = 0.81*	-	**Y** HR one week after exam as BL **Y** **Sign.**	**Mean** **r = 0.51** **d = 1.19**
8	Harleston, Smith, & Arey 1965 USA Journal of Personality and Social Psychology	*N* = 42 (42/0) M_{age} = NN **UPSY** **G:** 13 HTA/ 14 MTA/ 15 LTA based on TAQ	**L** Exp. Anagram solving	**TAQ** modified 2 months before the exp. **N**	**HR** Photocell amplifier-recorder unit right hand (thumb) *(measured during pretest, anagram-solving and posttest)*	**HR change** higher in HTA whole task r = 0.31 d = 0.65 first 5 min of task r = 0.30 d = 0.63		**Y** Pretest as BL, BL to calculate change scores **NN**	**Mean** **r = 0.31** **d = 0.64**
9	Holroyd, Westbrook, Wolf, & Badhorn 1978 USA Journal of Abnormal Psychology	*N* =72 (72/0) M_{age} = NN **UPSY** **G:** 36/36 based on TAS	**L** Exp. Stroop task, Anagram task	**TAS** (grouping) Before exp. **State test anxiety**: **STAI-State** **N**	**HR, HRV + SC/SR** SC/SR: active electrode non-dominant hand (thenar) HR/HRV: electrodes non-dominant (forearm and opposite leg) *(measured during pretest, stroop task and anagram-solving)*	**HRV** greater variability in LTA during all assessment periods r = 0.26 d = 0.545	**HR** **SC/SR**	**Y** Pretest as BL **Y** **Sign. NN**	**r = 0.26** **d = 0.55**

10	Kantor, Endler, Heslegrave, & Kokovski 2001 Canada Current Psychology	*N* = 17 (13/4) *Median*$_{age}$ = 26, 24-43 **UPSY** **Cont.**	**RL** Class seminar presentation (students received a grade)	**State exam anxiety**: **EMAS-State** prior to presenting in a seminar **Y** Worry Emotionality (from EMAS-State)	**HR** Ambulatory heart rate recorder; electrodes attached to upper sternum and fifth left rib *(measured cont. during BL and seminar presentation)*	**HR** **state exam anxiety total score** r = 0.46 d = 0.98 ***Emotionality*** *r = 0.53* *d= 1.18* ***Worry*** *r = 0.41* *d = 0.84*	-	**Y** **Y** **Sign. NN**	**Mean** **r = 0.47** **d = 1.00**
11	Kissel & Littig 1962 USA Journal of Abnormal and Social Psychology	*N* = 96 (NN) *M*$_{age}$ = NN **UPSY** **G:** 48/48 based on TAQ	**L** Exp. Perceptual reasoning test: Line drawing problems (unsolvable: induced failure)	**TAQ** (grouping) specific to testing situation **N**	**SC** electrodes non-dominant hand (1. and 3. finger) *(measured during rest period, 1-min. BL, tasks)*	**SC** higher in HTA under induced failure r = 0.28 d = 0.59	-	**Y** BL used as a control variable **NN**	**r = 0.28** **d = 0.59**
12	McGlynn, Bichajian, Giesen, Rullan, & Pulver 1981 USA Psychological Reports	*N* = 47 (20/27) *M*$_{age}$ = NN **UPSY** **G:** 38/9 based on TAS and palmar sweating item on the PSRSQ	**L + T** Exp. Psychology test (HTA: potential therapy offered LTA: study of test-taking behavior)	**TAS +** **PSRSQ** (grouping) **N**	**SC + HR** SC: electrodes non-dominant hand (index and 3. finger) HR: photoelectric pulse-pickup non-dominant hand (2. finger) *(pretreatment: measured during habituation, anticipation, and test)*	**SC** higher in HTA during pretreatment Anticipation: r = 0.49 d = 1.142 Pretreatment test-taking: r = 0.48 d = 1.086	**HR**	**Y** HR: habituation as BL SC: range corrected scores **NN**	**Mean** **r = 0.49** **d = 1.12**

13	Montgomery 1977 USA Psycho- physiology	*N* = 48 (0/48) M_{age} = NN **UPSY** **G:** 24/24 based on TAS: 12/12 (evaluative stress condition) 12/12 (no stress)	**L** Exp. (evaluative stress vs. no stress) Anagram solving	**TAS** (grouping) **N**	**HR *(+Respiration)*** HR: silver electrodes Left and right anterolateral lower ribs Respiration: not reliably obtained *(measured during BL and anticipation)*	**HR change** higher in HTA in anticipatory phase of the evaluative stress condition r = 0.43 d = 0.97		**Y** BL to calculate change scores **NN**	**r = 0.43** **d = 0.97**
14	Morris & Liebert 1970 USA Journal of Consulting and Clinical Psychology	*N* = 95 (NN) M_{age} = NN **UPSY** **Cont.**	**RL** Psychology exam	**TAQ** before the exam (i.e., after HR assessment) **Y** Worry Emotionality (from TAQ)	**HR** Participants count own pulse rate wrist *(measured during 4 times for 15 sec. before normal class and before a regular course examination)*	***HR** and emotionality* *r = 0.34* *d = 0.71* *worry* *r = 0.29* *d = 0.60* ***HR change** and **emotionality*** *r = 0.27* *d = 0.56*	***HR change** and worry*	**Y** Normal class HR as BL (to calculate HR change scores) **NN**	**Mean** **r = 0.30** **d = 0.63**
15	Morris & Liebert 1970 USA Journal of Consulting and Clinical Psychology	*N* = 91 (NN) M_{age} = NN **HS** **Cont.**	**RL** Sociology-psychology final exam	**TAQ** Before the exam Groups: before HR assessment vs. after HR assessment **Y** Worry Emotionality (from TAQ)	**HR** Participants count own pulse rate wrist *(Measured before the exam: either before TAQ administration or after TAQ administration; BL collected one month after the exam)*	***HR change** and **emotionality** when TAQ was administered first* *r = 0.24* *d = 0.49*	**HR** **HR change** and worry and emotionalit y when HR was assessed first	**Y** BL to calculate change scores **NN**	**r = 0.24** **d = 0.49**

16	Raphelson, 1957 USA The Journal of Abnormal and Social Psychology	*N* = 25 (0/25) M_{age} = NN, 18 – 25 **USO** **G:** 8/8 + (8 MTA) based on TAQ	**L** Exp. Test instruction: Intellectual and motor abilities	**TAQ** After the test (grouping) **N**	**SC *(+Respiration)*** SC: electrode left hand (palm and wrist) Respiration: pneumograph, no data reported *(measured during null period (i.e., BL), anticipation, test, and posttest)*	**SC** r = 0.47 d = 1.07 Difference in the direction of the trend of SC: HTA increased, LTA decreased		**Y** BL to calculate SC **NN**	**r = 0.47** **d = 1.07**
17	Sandin & Chorot 1985 Spain Psycho-physiology	*N* = 32 (32/0) M_{age} = 22.8, 20 – 25 **UPSY** **Cont.**	**RL** Psychology Comprehensive oral exam	**State anxiety**: **ECAE** **8SQ-a** 15 days before exam (BL1), immediately before exam, 15 days after exam (BL2) **Y but no results**	**Saliva pH** ***(+ Skin and urinary pH)*** Saliva samples pH meter (measured during BL1, immediately before exam and BL2)	**Saliva pH change** and **ECAE (exam)** r = 0.56 d = 1.31 **8SQ-a (exam)** r = 0.34 d = 0.70	-	**Y** Mean of BL1+2 to calculate change scores **Y** pH decrease from BL **Sign.**	**r = 0.45** **d = 1.01**
18	Smith, Houston, Zurawski 1984 USA Psycho-physiology	*N* = 62 (31/31) M_{age} = NN **USO** **Cont.**	**L** (Exp.) Instruction: Intelligence interview (stress condition) vs. listening and speaking (no-stress condition)	**State exam anxiety** 3 times: baseline, after instruction (i.e., anticipation), after interview **N**	**HR *(+ FPV)*** Physiograph Non-dominant middle finger *(measured during BL, anticipation period and interview)*	**HR** in interview and **state exam anxiety** in anticipation: r = 0.34 d = 0.71	-	**Y** **Y** higher in stress condition **Sign.**	**r = 0.34** **d = 0.71**

19	Spangler, Pekrun, Kramer, & Hofmann 2002 Germany Anxiety, Stress, & Coping	*N* = 40 (24/16) M_{age} = 25.3, *SD* = 2.5, 22 – 32 **USO** **Cont.** and **G:** by median of trait exam anxiety to look at the cortisol response	**RL** Final oral exam in psychology	**Trait exam anxiety** (12 items; see **TEQ**) 4 months before the exam (grouping) **State test anxiety** (Schmitz and Skinner, 1993)) immediately before and after the exam (asked for anxiety during and after the exam) **Y in the trait** **measure**; **but no** **results reported**	**Cortisol** Saliva samples salivettes *(measured in the morning of* *the exam, immediately before* *the exam, as well as 5 and 15* *min after the exam and in the* *morning of an exam free day* *(i.e., BL))*	**Cortisol (5 and 15** **min.) after the** **exam** and trait exam anxiety r = 0.49 d = 1.1 r = 0.45 d = 0.99 **Cortisol after the** **exam** and state anxiety before the exam: r = 0.28 d = 0.57 during the exam: r = 0.30 d = 0.62 **Cortisol response** and **trait** exam anxiety: During: HTA increase, LTA decrease After: HTA higher	-	**Y** + controlled for circadian differences in cortisol level **Y?** anticipatory cortisol response to the exam **NN**	**Mean** **r = 0.38** **d = 0.82**
20	Spangler, Kramer, Pekrun, & Hofmann 2002 Germany Anxiety, Stress, & Coping	*N* = 26 (15/11) M_{age} = 25.2, 22 – 31 **USO** **G:** Low, medium, high frequency of state test anxiety	**RL** Oral psychology exam videotaped	**State test anxiety** **during the exam:** Event sampling with a video-based semi-structured interview (grouping) **N**	**Cortisol** Saliva samples *(measured in the morning of* *the exam, immediately before* *the exam, as well as 15 min* *after the exam and in the* *morning of an exam free day)*	**Cortisol** in the morning and test anxiety: r = 0.34 d = 0.7 before exam and test anxiety: r = 0.40 d = 0.84	-	**Y** morning cortisol of an exam free day **Y?** anticipatory response in HTA **NN**	**Mean** **r = 0.37** **d = 0.77**

21	Zhang, Su, Peng, Yang, & Cheng 2011 China Clinical and Experimental Hypertension	*N* = 64 (33/31) M_{age} = 20.0, *SD* = 0.1 **UMS** **G:** 20/44 based on SAS (scores over 50)	**RL** Medicine Exam (Pre-review, review, exam period)	**State anxiety**: **SAS** on the first day of the exam period **N**	**HR + BP** Auscultatory blood pressure readings, Cuff around right upper arm *(measured during pre-review (i.e., BL), review and exam period)*	**HR** IA higher in HTA r = 0.44 d = 0.97 **SBP** IA higher in HTA r = 0.36 d = 0.76 **DBP** IA higher in HTA r = 0.45 d = 0.99 in exam period	-	**Y** Pre-review period as BL **Y** Increase from BL **Sign.**	**Mean** **r = 0.42** **d = 0.91**
22	Beidel, Turner, & Trager 1994 USA Journal of Anxiety Disorders	*N* = 62 (NN) M_{age} = 10 **ES** **G:** 36/26 based on TASC	**L** (Exp.) vocabulary test + reading aloud before an audience	**TASC** (grouping) Before exp. to identify children with HTA and LTA (initial sample N = 195) **N**	**HR + BP** automated BP/HR monitor Non-dominant arm BP: controlled for weight. *(measured at 2min intervals during BL, test-taking, reading aloud)*	-	**HR + SBP +** **DBP**: no difference between HTA and LTA	**Y** **NN**	**NS** **r = NN** **d = NN**
23	Daly, Chamberlain, & Spalding 2011 UK Educational Research	N = 23 (21/2) M_{age} = 17.31; SD = 0.47; 16 – 19 **SEC (UK college)** **Cont.**	**RL** **Pilot study** Mock French A-level exam	**TAS** and **CTA** one week before exam **Y but no results** Worry Emotionality (from CTA and TAS)	**HR** Suunto heart rate Memory Belt Chest *(measured in 10-second intervals during BL, immediately before and during the exam)*	**-**	**HR** r = 0.08 NS	**Y** **Y** **Sign.**	**NS** **r = 0.08** **d = 0.16**

24	Darley & Katz 1973 USA Child Development	N = 20 (0/20) M_{age} = NN **HS** 5th grade **Cont.**	**L** Exp. Mental task test group (N= 10) vs. game instruction group (N= 10)	**TASC** after experiments **N**	**HR** EKG electrodes Wrist and ankle *(measured cont. during BL and during and after the experimental manipulations)*	-	**HR change** across both groups did not correlate with TASC scores	**Y** Before experiment **Y** **Sign.** Increase from BL in test group	**NS** **r = NN** **d = NN**
25	Glazeski, Hollands- worth, & Jones 1986 USA Educational & Psychological Research	*N* = 30 (28/2) M_{age} = NN **UPSY + USO** (business) **G:** 15/15 based on upper and lower 10% of AAT difference score scores	**L** Exp. Problem solving: Mental ability test	**Trait exam anxiety** **(AAT)** (grouping) Before exp. **N**	**HR + SR + Respiration** **(= only a single combined arousal index was used)** HR: Silver electrodes Standard Lead II placement SR: Finger-tip electrodes (non-dominant hand) Respiration: Velcro chest band *(measured during BL and test)*	-	Single arousal index (combi- nation of HR, SR and respiration): No differ- ence be- tween HTA and LTA groups	**Y** **Y** **Sign.**	**NS** **r = NN** **d = NN**
26	Herbert, Moore, De La Riva, & Watts 1986 UK Biological Psychology	N = 38 (0/38) M_{age} = NN **UMS** **Cont.**	**RL** Medicine before major medicine exam	**State anxiety**: **STAI X, worry, emotionality** 3 weeks before (= BL) and immediately before exam **Y** Worry Emotionality	**Cortisol** ***(+ Testosterone + Prolactin + LH)*** Serum from venous blood *(measured 3 weeks before the exam (BL) and immediately before exam)*		**Cortisol** **(serum)** and state anxiety r = -0.05 NS, ***worry*** r = -0.03 NS, ***emotion- ality*** r = -0.13 NS	**Y** BL to calculate change **Y** Increase from BL to exam **Sign.**	**NS** **Mean** **r = -0.07** **d = -0.14**

27	Hollands- worth, Glazeski, Kirkland, Jones, & Van Norman 1979 USA Cognitive Therapy and Research	*N* = 6 (6/0) M_{age} = NN **UPSY** **G:** 3/3 based on TAS scores	**L** (Exp.) Mental ability/ scholastic aptitude test (videotaped)	**TAS** (grouping) Before exp., N = 6 selected from N = 239 students based on extreme TAS and average GAS scores. + interview to validate TAS scores **N**	**HR + SR *(+ Respiration)*** HR: Silver electrodes, Standard Lead II placement SR: Electrodes, thenar and hypothenar, non-dominant hand *(measured at 1 min intervals during BL, test instruction, anticipatory period and test)*	-	**HR *(and respiration):*** LTA higher arousal than HTA during BL, test & as test began; **SRR:** incon- sistent : HTA highly aroused but 2 of the 3 LTA as well	**Y** **NN/ not sign.** inconsistent	**NS** **r = NN** **d = NN**
28	Huwe, Hennig, & Netter 1998 Germany Anxiety, Stress & Coping	N = 58 (36/22) M_{age} = 24.81, SD = 3.82 **UPSY** **G:** 29/29 based on STAI	**RL** Psychology oral examination (video-taped)	**State exam anxiety:** **STAI – State** (grouping) immediately before the exam **N**	**HR + Salivary Cortisol** ***(+slgA secretory immunoglobulin)*** HR: manually by experimenter: Palpitation of right radial artery Salivary cortisol (+slgA): collected by salivettes *(measured immediately before and after the exam)*	-	**Cortisol** r = 0.23 NS, **HR** r = 0.23 NS**,** correlations go in the expected directions	**Y** Four weeks after exam **Y** **Sign**.	**NS** **Mean** **r = 0.23** **d = 0.47**
29	Tennes & Kreye 1985 USA Psycho- somatic Medicine	*N* = 70 (32/38) M_{age} = 7.7, 7 – 9 **ES** 2nd grade **Cont. +** **G:** median split based on TASC scores	**L** standardized achievement test + classroom observation on normal school days (**RL**)	**TASC** adapted version (grouping) one week after achievement test **N**	**Cortisol** urine samples, *(2hr collections at the same time each day, once a month on normal school days and on days of achievement tests)* Controlled for age, gender, and weight.	-	**Cortisol and TASC** scores **(cont.)** r = -0.12 NS, **Cortisol and HTA/LTA groups**	**Y** **Y** **Sign.** increase from normal days to test day	**NS** **r = -0.12** **d = -0.24**

Table A 2. Studies in systematic review

Note. Results for the worry and emotionality components are printed in *italics.*

Abbreviations (in alphabetical order): **AAT** = Achievement Anxiety Test (Alpert & Haber, 1960); **BL** =baseline; **BP** = Blood pressure; **BRQ** = Behavioral Reactions Questionnaire (Endler & Okada, 1975); **CTA** = Cognitive Test Anxiety Scale (Cassady & Johnson, 2002); **DBP** = diastolic blood pressure; **ECAE** = Anxiety State Behavioral Scale (Chorot & Sandin, 1985); **EMAS** = Endler Multidimensional Anxiety Scales (Endler, Edwards, & Vitelli, 1991); **ES** = Elementary school; **EWL** = Adjective checklist on emotional states (Janke, Debus, & 1978); **Exp.** = Experiment; **f** = female; **FPV** = finger pulse volume**; GAS** = General Anxiety Scale (Sarason, 1972); **HR** = Heart rate; **HRV** = Heart rate variability**; HS** = High school; **HTA** = High test anxious; **IA** = increase amplitudes; **L** = Laboratory; **LTA** = Low test anxious; **m** = male; **MTA** = Medium test anxious; **N** = No; **NN** = not mentioned; **PSRSQ** = Perceived Somatic Reactions to Stress Questionnaire (Stern & Higgins, 1969); **RL** = real life; **SAS** = Zung Self-Rating Anxiety Scale (Zung, 1976); **SBP** = systolic blood pressure; **SC** = Skin conductance; **SEC** = Secondary school students; **SR** = skin resistance; **STAI** = State-Trait Anxiety Inventory (Spielberger et al., 1970); **STAR** = State Test Anxiety Report (Deffenbacher & Hazaleus, 1985); **T** = treatment; **TAI** = Test Anxiety Inventory (Spielberger et al., 1980); **TAQ** = Test Anxiety questionnaire (Mandler & Sarason, 1952) modified by Harleston (1962); **TAS** = Test Anxiety Scale (Sarason, 1972); **TASC** = Test Anxiety Scale for Children (Sarason et al., 1958); **TAI short** = Test Anxiety Inventory short form (Taylor & Deane, 2002); **TEQ** = Test Emotions Questionnaire (Pekrun et al., 2004); **UMS** = University Medicine students; **UPSY** = University Psychology students; **USO** = University students (other than Psychology) **Y** = Yes; **8SQ-a** = Eight State Questionnaire Anxiety (Curran & Cattell, 1976).

Table A 3. Sample characteristics

Characteristics		Number of studies	Study No.
College or university students:	Psychology students	16	2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 25, 27, 28
	Medicine students	3	18, 21, 26
	Other students	4	3, 16, 19, 20
Elementary school or high school students		6	1, 15, 22, 23, 29

Note. Study No. can be retrieved from Table A2.

Table A 4. Study design and setting

Design		Number of studies	Study No.
Lab studies with experimental paradigms:	Problem solving task	6	8, 9, 11, 13, 24, 25
	General aptitude test	6	2, 5, 6, 16, 18, 29
	Vocabulary test	2	1, 22
	Psychology test	1	12
Real-life studies:	Psychology exam	10	4, 5, 7, 10, 14, 15, 17, 19, 20, 28
	Medicine exam	2	21, 26
	French exam	1	23
	Nursing exam	1	3

Note. Study No. can be retrieved from Table 3.

Table A 5. Self-report test anxiety measures

Measure	Number of studies		Study No.
TAS	7		12, 13, 27
		+ trait measure (1)	23
		+ state measure (3)	5, 6, 9
TASC	4		1, 22, 24, 29
TAQ	5		8, 11, 14, 15, 16
TAI	3		2, 3, 4
Other trait test anxiety measures	2		25
		+ state measure (1)	19
State anxiety before a test	8		7, 10, 17, 18, 20, 21, 26, 28

Note. Study No. can be retrieved from Table 3.

5 References

Ahmed, W., van der Werf, G., Minnaert, A., & Kuyper, H. (2010). Students' daily emotions in the classroom: Intra-individual variability and appraisal correlates. *British Journal of Educational Psychology, 80*, 583-597. doi:10.1348/000709910X498544

Ainley, M., & Patrick, L. (2006). Measuring self-regulated learning processes through tracking patterns of student interaction with achievement activities. *Educational Psychology Review, 18*, 267-286. doi:10.1007/s10648-006-9018-z

Allen, G., Elias, M., & Zlotlow, S. (1980). Behavioral interventions for alleviating test anxiety: A methodological overview of current therapeutic practices. *Test anxiety: Theory, research, and applications*, 155-185.

Alpert, R., & Haber, R. N. (1960). Anxiety in academic achievement situations. *The Journal of Abnormal and Social Psychology, 61*(2), 207. doi:10.1037/h0045464

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

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*, 181-185. doi:10.1111/1467-8721.00196

Ashcraft, M. H., Krause, J. A., & Hopko, D. R. (2007). Is math anxiety a mathematical learning disability? In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 329-348). Baltimore, MD, US: Paul H Brookes Publishing.

Astleitner, H. (2000). Designing emotionally sound instruction: The FEASP-approach. *Instructional Science, 28*, 169-198. doi:10.1023/A:1003893915778

Austin, J. S., Partridge, E., Bitner, J., & Wadlington, E. (1995). Prevent school failure: Treat test anxiety. *Preventing School Failure: Alternative Education for Children and Youth, 40*, 10-13. doi:10.1080/1045988X.1995.9944644

Avram, R., Kuhar, P., Vittinghoff, E., Aschbacher, K., Tison, G., Pletcher, M., . . . Olgin, J. (2018). Redefining Normal Resting Heart Rate Values Using Big Data. *Circulation, 138*(Suppl_1), A15098-A15098. doi:10.1161/circ.138.suppl_1.15098

Bailey, R. C. (1971). Self-concept differences in low and high achieving students. *Journal of Clinical Psychology, 27*(2), 188-191.

Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin, 23*(10), 1100-1110. doi:10.1177/01461672972310010

Barrett, L. F., & Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science, 15*(2), 79-83.

Becker, E. S., Goetz, T., Morger, V., & Ranellucci, J. (2014). The importance of teachers' emotions and instructional behavior for their students' emotions: An experience sampling analysis. *Teaching and Teacher Education, 43*, 15-26. doi:10.1016/j.tate.2014.05.002

Becker, E. S., Keller, M. M., Goetz, T., Frenzel, A. C., & Taxer, J. L. (2015). Antecedents of teachers' emotions in the classroom: an intraindividual approach. *Frontiers in Psychology, 6*, 635. doi:10.3389/fpsyg.2015.00635

Beidel, D. C. (1988). Psychophysiological assessment of anxious emotional states in children. *Journal of Abnormal Psychology, 97*(1), 80-82. doi:10.1037/0021-843X.97.1.80

Beidel, D. C., & Turner, S. M. (1988). Comorbidity of test anxiety and other anxiety disorders in children. *Journal of Abnormal Child Psychology, 16*(3), 275-287. doi:10.1007/BF00913800

Ben-Zeev, D., McHugo, G. J., Xie, H., Dobbins, K., & Young, M. A. (2012). Comparing Retrospective Reports to Real-Time/Real-Place Mobile Assessments in Individuals With Schizophrenia and a Nonclinical Comparison Group. *Schizophrenia Bulletin, 38*(3), 396-404. doi:10.1093/schbul/sbr171

Berntson, G. G., Norman, G. J., Hawkley, L. C., & Cacioppo, J. T. (2008). Cardiac autonomic balance versus cardiac regulatory capacity. *Psychophysiology, 45*(4), 643-652. doi:10.1111/j.1469-8986.2008.00652.x

Betancourt, M. A., Dethorne, L. S., Karahalios, K., & Kim, J. G. (2017). Skin conductance as an in situ marker for emotional arousal in children with neurodevelopmental communication impairments: Methodological considerations and clinical implications. *ACM Transactions on Accessible Computing (TACCESS), 9*(8), 1-29. doi:10.1145/3035536

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. (2014). What students think they feel differs from what they really feel–academic self-concept moderates the discrepancy between students' trait and state emotional self-reports. *PLoS ONE, 9*(3), 1-9. doi:10.1371/journal.pone.0092563

Bieg, M., Goetz, T., Wolter, I., & Hall, N. C. (2015). Gender stereotype endorsement differentially predicts girls' and boys' trait-state discrepancy in math anxiety. *Frontiers in Psychology, 6*, 1404. doi:doi: 10.3389/fpsyg.2015.01404

Bigger, J. T., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Kleiger, R. E., & Rottman, J. N. (1992). Frequency domain measures of heart period variability and mortality after myocardial infarction. *Circulation, 85*(1), 164-171. doi:10.1161/circ.85.1.1728446

Bodas, J., & Ollendick, T. H. (2005). Test anxiety: A cross-cultural perspective. *Clinical Child and Family Psychology Review, 8*(1), 65-88. doi:10.1007/s10567-005-2342-x

Boucsein, W. (2012). *Electrodermal activity* (2 ed.). New York: Springer Science & Business Media.

Bouma, E. M., Riese, H., Ormel, J., Verhulst, F. C., & Oldehinkel, A. J. (2009). Adolescents' cortisol responses to awakening and social stress; effects of gender, menstrual phase and oral contraceptives. The TRAILS study. *Psychoneuroendocrinology, 34*(6), 884-893.

Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D., & Tuerlinckx, F. (2018). Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model. *Multivariate Behavioral Research, 53*(3), 293-314. doi:10.1080/00273171.2018.1439722

Bryant, F. B., & Satorra, A. (2013). EXCEL macro file for conducting scaled difference chi-square tests via LISREL 8, LISREL 9, EQS, and Mplus. *Available from the authors*.

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

Burton, D. (1988). Do anxious swimmers swim slower? Reexamining the elusive anxiety-performance relationship. *Journal of Sport and Exercise Psychology, 10*(1), 45-61.

Butler, E. A., Lee, T. L., & Gross, J. J. (2007). Emotion regulation and culture: Are the social consequences of emotion suppression culture-specific? *Emotion, 7*(1), 30-48. doi:10.1037/1528-3542.7.1.30

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. Haviland-Jones (Eds.), *The handbook of emotions* (pp. 173-191). New York: Guildford Press.

Calvo, M. G., & Miguel-Tobal, J. J. (1998). The anxiety response: Concordance among components. *Motivation and Emotion, 22*(3), 211-230. doi:10.1023/A:1022384022641

Campbell, S. B. (1986). Developmental issues in childhood anxiety. In R. Gittelman (Ed.), *Anxiety Disorders of Childhood* (pp. 24-57). New York: Guilford.

Caprara, G. V., & Cervone, D. (2000). *Personality: Determinants, dynamics, and potentials*. New York, NY, US: Cambridge University Press.

Carlson, N. (2013). *Physiology of behavior*. New Jersey: Pearson Education, Inc.

Cassady, J. (2010). Test anxiety: Contemporary theories and implications for learning. In J. Cassady (Ed.), *Anxiety in schools: The causes, consequences, and solutions for academic anxieties* (pp. 7-26). New York, NY: Peter-Lang.

Cassady, J. C. (2004). The impact of cognitive test anxiety on text comprehension and recall in the absence of external evaluative pressure. *Applied Cognitive Psychology, 18*, 311-325. doi:10.1002/acp.968

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*(2), 270-295. doi:10.1006/ceps.2001.1094

Chipman, S. F., Krantz, D. H., & Silver, R. (1992). Mathematics Anxiety and Science Careers among Able College Women. *Psychological Science, 3*(5), 292-296. doi:10.1111/j.1467-9280.1992.tb00675.x

Chorot, P., & Sandin, B. (1985). The Anxiety State Bahavioral-Scale *Psiquis, 6*(3), 60-65.

Cipora, K., Artemenko, C., & Nuerk, H.-C. (2019). Different ways to measure math anxiety. In I. C. Mamarella, S. Caviola, & A. Dowker (Eds.), *Mathematics Anxiety: What Is Known, and What is Still to be Understood* (pp. 20). New York, NY: Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2 ed.). Hillsdale, NJ: Erlbaum.

Cohen, M., & Khalaila, R. (2014). Saliva pH as a biomarker of exam stress and a predictor of exam performance. *Journal of Psychosomatic Research, 77*(5), 420-425. doi:10.1016/j.jpsychores.2014.07.003

Conley, K. M., & Lehman, B. J. (2012). Test anxiety and cardiovascular responses to daily academic stressors. *Stress and Health, 28*(1), 41-50. doi:10.1002/smi.1399

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine, 74*(4), 327-337. doi:10.1097/PSY.0b013e3182546f18

Csikszentmihalyi, M., & Larson, R. (1992). Validity and reliability of the Experience Sampling Method. In M. W. deVries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings.* (pp. 43-57). New York, NY: Cambridge University Press.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In M. Csikszentmihalyi & R. Larson (Eds.), *Flow and the foundations of positive psychology* (pp. 35-54). Dordrecht, Netherlands: Springer.

Curran, J. P., & Cattell, R. B. (1976). *Manual for the eight state questionnaire: 8SQ*. Champaign: Institute for Personality and Ability Testing.

D'Mello, S., & Calvo, R. A. (2013). *Beyond the basic emotions: what should affective computing compute?* Paper presented at the CHI'13 Extended Abstracts on Human Factors in Computing Systems.

D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist, 52*(2), 104-123. doi:10.1080/00461520.2017.1281747

D'Mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction, 20*(2), 147-187. doi:10.1007/s11257-010-9074-4

Daly, A. L., Chamberlain, S., & Spalding, V. (2011). Test anxiety, heart rate and performance in A-level French speaking mock exams: An exploratory study. *Educational Research, 53*(3), 321-330. doi:10.1080/00131881.2011.598660

Damer, D. E., & Melendres, L. T. (2011). "Tackling test anxiety": A group for college students. *The Journal for Specialists in Group Work, 36*, 163-177. doi:10.1080/01933922.2011.586016

Davies, E. W. (2010). Tools for overcoming students' emotional obstacles to learning psychology. In S. A. Meyers & J. R. Stowell (Eds.), *Essays from excellence in teaching* (Vol. 9, pp. 7-10). Washington, DC: Society for the Teaching of Psychology.

De Leersnyder, J., Boiger, M., & Mesquita, B. (2013). Cultural regulation of emotion: individual, relational, and structural sources. *Frontiers in Psychology, 4*, 55-55. doi:10.3389/fpsyg.2013.00055

Deffenbacher, J. L. (1977). Relationship of worry and emotionality to performance on the Miller Analogies Test. *Journal of Educational Psychology, 69*, 191-195. doi:10.1037/0022-0663.69.2.191

Deffenbacher, J. L. (1986). Cognitive and physiological components of test anxiety in real-life exams. *Cognitive Therapy and Research, 10*(6), 635-644. doi:10.1007/BF01173751

Deffenbacher, J. L., & Hazaleus, S. L. (1985). Cognitive, emotional, and physiological components of test anxiety. *Cognitive Therapy and Research, 9*(2), 169-180. doi:10.1007/BF01204848

Denny, B. T., & Ochsner, K. N. (2014). Behavioral effects of longitudinal training in cognitive reappraisal. *Emotion, 14*(2), 425-433. doi:10.1037/a0035276

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science, 40*, 434-449. doi:10.1007/s11747-011-0300-3

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355-391. doi:10.1037/0033-2909.130.3.355

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist, 55*(1), 34-43. doi:10.1037/0003-066X.55.1.34

Donker, M., Van Gog, T., & Mainhard, T. (2018). A Quantitative Exploration of Two Teachers with Contrasting Emotions: Intra-Individual Process Analyses of Physiology and Interpersonal Behavior. *Frontline Learning Research, 6*, 162-185.

Dorn, L. D., Lucke, J. F., Loucks, T. L., & Berga, S. L. (2007). Salivary cortisol reflects serum cortisol: Analysis of circadian profiles. *Annals of Clinical Biochemistry, 44*(3), 281-284. doi:10.1258/000456307780480954

Dresel, M., & Haugwitz, M. (2006). The relationship between cognitive abilities and self-regulated learning: evidence for interactions with academic self-concept and gender. *High Ability Studies, 16*(2), 201-218.

Dresel, M., & Haugwitz, M. (2008). A computer-based approach to fostering motivation and self-regulated learning. *The Journal of Experimental Education, 77*(1), 3-20.

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*(4), 283-306.

Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology, 91*, 549-563. doi:10.1037/0022-0663.91.3.549

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*(1), 103-127. doi:10.1037/a0018053

Endler, N. S., Edwards, J. M., & Vitelli, R. (1991). *Endler multidimensional anxiety scales (EMAS)*. CA, Los Angeles: Western Psychological Services

Endler, N. S., & Kocovski, N. L. (2001). State and trait anxiety revisited. *Journal of Anxiety Disorders, 15*(3), 231-245. doi:10.1016/S0887-6185(01)00060-3

Endler, N. S., & Magnusson, D. (1977). The interaction model of anxiety: An empirical test in an examination situation. *Canadian Journal of Behavioural Science, 9*(2), 101-107. doi:10.1037/h0081612

Endler, N. S., & Okada, M. (1975). A multidimensional measure of trait anxiety: The SR Inventory of general trait anxiousness. *Journal of Consulting and Clinical Psychology, 43*(3), 319-329.

Enright, M., Baldo, T. D., & Wykes, S. D. (2000). The efficacy of eye movement desensitization and reprocessing therapy technique in the treatment of test anxiety of college students. *Journal of College Counseling, 3*, 36-48. doi:10.1002/j.2161-1882.2000.tb00162.x

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. J. E. (2007). Anxiety and cognitive performance: attentional control theory. *7*(2), 336.

Federal Statistical Office [Statistisches Bundesamt]. (2015). *Schnellmeldungsergebnisse zu Schülerinnen und Schülern der allgemeinbildenden und beruflichen Schulen - Schuljahr 2014-15 [Preliminary results of general and vocational school students: academic year 2014-15].* Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/SchnellmeldungSchueler5211003158004.pdf

Feinberg, J. M., & Aiello, J. R. (2010). The effect of challenge and threat appraisals under evaluative presence. *Journal of Applied Social Psychology, 40*(8), 2071-2104.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65*(1), 45. doi:10.1037/0022-3514.65.1.45

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

Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology, 101*(3), 705-716. doi:10.1037/a0014695

Frenzel, A. C., Marx, A. K. G., Pekrun, R., Reck, C., & Müller, M. (2019). *Teachers' and Learners' Emotional Experiences in Class: A Field-based Video Study*. Paper presented at the American Educational Research Association (AERA) annual conference, Toronto, Canada.

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

Gatchel, R. J., Baum, A., & Krantz, D. S. (1989). *An Introduction to Health Psychology*. New York, NY: McGraw Hill.

Glazeski, R. C., Hollandsworth, J. G., & Jones, G. E. (1986). An investigation of the role of physiological arousal in test anxiety. *Educational & Psychological Research, 6*(2).

Goetz, T. (2004). *Emotionales Erleben und selbstreguliertes Lernen bei Schülern im Fach Mathematik [Emotions and self-regulated learning of students in mathematics]*. Munich, Germany: Utz.

Goetz, T., Bieg, M., & Hall, N. (2016). Assessing Academic Emotions via the Experience Sampling Method (pp. 245-258).

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

Goetz, T., Frenzel, A. C., Stoeger, H., & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion, 34*(1), 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, NY: Routledge.

Goetz, T., Nett, U. E., Martiny, S. E., Hall, N. C., Pekrun, R., Dettmers, S., & Trautwein, U. (2012). Students' emotions during homework: Structures, self-concept antecedents, and achievement outcomes. *Learning and Individual Differences, 22*(2), 225-234. doi:10.1016/j.lindif.2011.04.006

Goetz, T., Pekrun, R., Hall, N., & 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. doi:10.1348/000709905X42860

Goetz, T., Preckel, F., Pekrun, R., & Hall, N. C. (2007). Emotional experiences during test taking: Does cognitive ability make a difference? *Learning and Individual Differences, 17*, 3-16. doi:10.1016/j.lindif.2006.12.002

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., . . . Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*, 188-205. doi:10.1016/j.cedpsych.2014.04.002

Gordon, M. E., Slade, L. A., & Schmitt, N. (1986). The "science of the sophomore" revisited: From conjecture to empiricism. *Academy of Management Review, 11*(1), 191-207. doi:10.5465/amr.1986.4282666

Gröhlich, C., Scharenberg, K., & Bos, W. (2009). Does classroom heterogeneity in secondary schools affect student achievement? *Journal for Educational Research Online/Journal für Bildungsforschung Online, 1*(1), 86-105.

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of Emotion Regulation* (pp. 3–24). New York, NY: Guilford Press.

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*(4), 361-391. doi:10.1521/jscp.2006.25.4.361

Hamaker, E. L. (2012). Why researchers should think 'within-person': A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life.* (pp. 43-61). New York, NY: The Guilford Press.

Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review, 7*(4), 316-322. doi:10.1177/1754073915590619

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science, 26*(1), 10-15. doi:10.1177/0963721416666518

Hansen, Å. M., Garde, A. H., & Persson, R. (2008). Sources of biological and methodological variation in salivary cortisol and their impact on measurement among healthy adults: A review. *Scandinavian Journal of Clinical and Laboratory Investigation, 68*(6), 448-458. doi:10.1080/00365510701819127

Harleston, B. W. (1962). Test anxiety and performance in problem-solving situations. *Journal of Personality, 30*(4), 557-573. doi:10.1111/j.1467-6494.1962.tb01689.x

Harleston, B. W., Smith, M. G., & Arey, D. (1965). Test-anxiety level, heart rate, and anagram problem solving. *Journal of Personality and Social Psychology, 1*(6), 551. doi:10.1037/h0021991

Harley, J. M. (2015). Measuring emotions: A survey of cutting-edge methodologies used in computer-based learning environment research. In S. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 89 - 114). London, UK: Academic Press, Elsevier.

Hartel, C., & Page, K. (2009). *Discrete emotional crossover in the workplace: The role of affect intensity* (Vol. 24).

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

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47-77. doi:10.2307/1170348

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*, 33-46. doi:10.2307/749455

Herbert, J., Moore, G., De La Riva, C., & Watts, F. (1986). Endocrine responses and examination anxiety. *Biological Psychology, 22*(3), 215-226. doi:10.1016/0301-0511(86)90027-X

Hodapp, V., & Benson, J. (1997). The multidimensionality of test anxiety: A test of different models. *Anxiety, Stress, and Coping, 10*, 219-244. doi:10.1080/10615809708249302

Hodges, W. (2015). The psychophysiology of anxiety. In M. Zuckerman & C. D. Spielberger (Eds.), *Emotions and Anxiety (PLE: Emotion): New Concepts, Methods, and Applications* (pp. 175-192): Psychology Press.

Hollandsworth, J. G., Glazeski, R. C., Kirkland, K., Jones, G. E., & Van Norman, L. R. (1979). An analysis of the nature and effects of test anxiety: Cognitive, behavioral, and physiological components. *Cognitive Therapy and Research, 3*(2), 165-180. doi:10.1007/BF01172603

Houtveen, J. H., & de Geus, E. J. (2009). Noninvasive psychophysiological ambulatory recordings: Study design and data analysis strategies. *European Psychologist, 14*(2), 132-141. doi:10.1027/1016-9040.14.2.132

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Huberty, T. J. (2010). Test and performance anxiety. *The Education Digest, 75*(9), 34-38.

Hugdahl, K. (1995). *Psychophysiology: The mind-body perspective*: Cambridge, MA: Harvard University Press.

Humphrey, S. P., & Williamson, R. T. (2001). A review of saliva: Normal composition, flow, and function. *The Journal of Prosthetic Dentistry, 85*(2), 162-169. doi:10.1067/mpr.2001.113778

Huwe, S., Hennig, J., & Netter, P. (1998). Biological, emotional, behavioral, and coping reactions to examination stress in high and low state anxious subjects. *Anxiety, Stress & Coping, 11*(1), 47-65. doi:10.1080/10615809808249313

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

Immordino-Yang, M., Christodoulou, J., Pekrun, R., & Linnenbrink-Garcia, L. (2014). *Neuroscientific contributions to emotion measurement in educational contexts*. New York, NY: Routledge.

Jain, S., & Dowson, M. (2009). Mathematics anxiety as a function of multidimensional self-regulation and self-efficacy. *Contemporary Educational Psychology, 34*(3), 240-249.

Jamieson, J. P., Mendes, W. B., & Nock, M. K. (2013). Improving acute stress responses: The power of reappraisal. *Current Directions in Psychological Science, 22*(1), 51-56. doi:10.1177/0963721412461500

Jamieson, J. P., Nock, M. K., & Mendes, W. B. (2012). Mind over matter: Reappraising arousal improves cardiovascular and cognitive responses to stress. *Journal of Experimental Psychology. General, 141*(3), 417-422. doi:10.1037/a0025719

Jamieson, J. P., Nock, M. K., & Mendes, W. B. (2013). Changing the Conceptualization of Stress in Social Anxiety Disorder: Affective and Physiological Consequences. *Clinical Psychological Science, 1*(4), 363-374. doi:10.1177/2167702613482119

Janke, W., Debus, G., & (1978). *Die Eigenschaftswörterliste: EWL; Eine mehrdimensionale Methode zur Beschreibung von Aspekten des Befindens.* Goettingen: Hogrefe.

Johnson, D. D., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature, 477*(7364), 317-320.

Johnston, D. W., Anastasiades, P., & Wood, C. (1990). The relationship between cardiovascular responses in the laboratory and in the field. *Psychophysiology, 27*(1), 34-44. doi:10.1111/j.1469-8986.1990.tb02175.x

Jones, M., & Johnston, D. (2011). Understanding Phenomena in the Real World: The Case for Real Time Data Collection in Health Services Research. *Journal of Health Services Research & Policy, 16*(3), 172-176. doi:10.1258/jhsrp.2010.010016

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

Kahneman, D., & Riis, J. (2005). Living, and thinking about it: Two perspectives on life. In F. A. Huppert, N. Baylis, & B. Keverne (Eds.), *The science of well-being* (pp. 285–304). Oxford, England: Oxford University Press.

Kantor, L., Endler, N. S., Heslegrave, R. J., & Kocovski, N. L. (2001). Validating self-report measures of state and trait anxiety against a physiological measure. *Current Psychology, 20*(3), 207-215. doi:10.1007/s12144-001-1007-2

Kaplan, S., Dalal, R. S., & Luchman, J. N. (2013). Measurement of emotions *Research methods in occupational health psychology: Measurement, design, and data analysis* (pp. 61-75). New York, NY: Routledge.

Kärner, T., & Kögler, K. (2016). Emotional states during learning situations and students’ self-regulation: process-oriented analysis of person-situation interactions in the vocational classroom. *Empirical Research in Vocational Education Training, 8*(1), 12. doi:10.1186/s40461-016-0038-8

Kenny, D. A., Bolger, N., & Kashy, D. A. (2002). Traditional methods for estimating multilevel models. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 1-24). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Keogh, E., Bond, F. W., French, C. C., Richards, A., & Davis, R. E. (2004). Test anxiety, susceptibility to distraction and examination performance. *Anxiety, Stress & Coping, 17*, 241-252. doi:10.1080/10615300410001703472

Kihlstrom, J. F., Beer, J. S., & Klein, S. B. (2003). Self and identity as memory *Handbook of self and identity* (pp. 68-90). New York, NY, US: The Guilford Press.

King, N. J., Ollendick, T. H., & Prins, P. J. (2000). Test-anxious children and adolescents: Psychopathology, cognition, and psychophysiological reactivity. *Behaviour Change, 17*(03), 134-142. doi:10.1375/bech.17.3.134

Krannich, M., Goetz, T., Lipnevich, A. A., Bieg, M., Roos, A.-L., Becker, E. S., & Morger, V. (2019). Being over-or underchallenged in class: Effects on students' career aspirations via academic self-concept and boredom. *Learning and Individual Differences, 69*, 206-218.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394-421. doi:10.1016/j.biopsycho.2010.03.010

Kreibig, S. D., & Gendolla, G. H. (2014). Autonomic nervous system measurement of emotion in education and achievement settings. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 625–642). New York, NY: Routledge/Taylor & Francis Group.

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.

Larsen, R. J., & Prizmic-Larsen, Z. (2006). Measuring emotions: Implications of a multimethod perspective. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 337-351). Washington, DC: American Psychological Association.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*, 355-365. doi:10.1016/j.lindif.2008.10.009

Levine, L. J., Lench, H. C., & Safer, M. A. (2009). Functions of remembering and misremembering emotion. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 23*(8), 1059-1075.

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics, 7*(4), 227-236. doi:10.1037/npe0000028

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*(3), 975-978. doi:10.2466/pr0.1967.20.3.975

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). *The computer expression recognition toolbox (CERT).* Paper presented at the IEEE.

Liu, J. J. W., Ein, N., Peck, K., Huang, V., Pruessner, J. C., & Vickers, K. (2017). Sex differences in salivary cortisol reactivity to the Trier Social Stress Test (TSST): A meta-analysis. *Psychoneuroendocrinology, 82*, 26-37. doi:10.1016/j.psyneuen.2017.04.007

Lundberg, U. (2005). Stress hormones in health and illness: The roles of work and gender. *Psychoneuroendocrinology, 30*(10), 1017-1021. doi:10.1016/j.psyneuen.2005.03.014

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*, 520-540. doi:10.2307/749772

Ma, X., Tamir, M., & Miyamoto, Y. (2018). A socio-cultural instrumental approach to emotion regulation: Culture and the regulation of positive emotions. *Emotion, 18*(1), 138-152. doi:10.1037/emo0000315

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86-92. doi:10.1027/1614-2241.1.3.86

Maloney, E. A., Waechter, S., Risko, E. F., & Fugelsang, J. A. (2012). Reducing the sex difference in math anxiety: The role of spatial processing ability. *Learning and Individual Differences, 22*(3), 380-384. doi:10.1016/j.lindif.2012.01.001

Mandler, G., & Kremen, I. (1958). Autonomic feedback: A correlational study. *Journal of Personality, 26*(3), 388-399. doi:10.1111/j.1467-6494.1958.tb01594.x

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *The Journal of Abnormal and Social Psychology, 47*(2), 166-173.

Marques, A. H., Silverman, M. N., & Sternberg, E. M. (2010). Evaluation of stress systems by applying noninvasive methodologies: Measurements of neuroimmune biomarkers in the sweat, heart rate variability and salivary cortisol. *Neuroimmunomodulation, 17*(3), 205-208. doi:10.1159/000258725

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.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. D. Phye (Ed.), *Educational psychology series. Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131-198). San Diego, CA: Academic Press.

Martin, B. (1961). The assessment of anxiety by physiological behavioral measures. *Psychological Bulletin, 58*(3), 234. doi:10.1037/h0045492

Matthews, G., Hillyard, E. J., & Campbell, S. E. (1999). Metacognition and maladaptive coping as components of test anxiety. *Clinical Psychology & Psychotherapy, 6*(2), 111-125. doi:10.1002/(SICI)1099-0879(199905)6:2<111::AID-CPP192>3.0.CO;2-4

Matthews, K. A., Salomon, K., Brady, S. S., & Allen, M. T. (2003). Cardiovascular reactivity to stress predicts future blood pressure in adolescence. *Psychosomatic Medicine, 65*(3), 410-415. doi:10.1097/01.PSY.0000057612.94797.5F

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion, 5*(2), 175. doi:10.1037/1528-3542.5.2.175

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209-237. doi:10.1080/02699930802204677

McCraty, R., Atkinson, M., Tiller, W. A., Rein, G., & Watkins, A. D. (1995). The effects of emotions on short-term power spectrum analysis of heart rate variability. *The American Journal of Cardiology, 76*(14), 1089-1093. doi:10.1016/S0002-9149(99)80309-9

McCraty, R., & Zayas, M. A. (2014). Cardiac coherence, self-regulation, autonomic stability, and psychosocial well-being. *Frontiers in Psychology, 5*, 1090. doi:10.3389/fpsyg.2014.01090

McDonald, A. S. (2001). The Prevalence and Effects of Test Anxiety in School Children. *Educational Psychology, 21*(1), 89-101. doi:10.1080/01443410020019867

McGlynn, F. D., Bichajian, C., Giesen, J. M., Rullan, C. M., & Pulver, L. (1981). Factorial study of component procedures in desensitization treatment of test anxiety among college students. *Psychological Reports, 49*(2), 351-362. doi:10.2466/pr0.1981.49.2.351

McIlroy, D. (2000). An evaluation of the factor structure and predictive utility of a test anxiety scale with reference to students' past performance and personality indices. *British Journal of Educational Psychology, 70*, 17-32. doi:10.1348/000709900157949

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2*, 201-218. doi:10.1207/s15366359mea0204_1

Molenaar, P. C., & 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

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A Meta-Analytic Path Analysis of the Internal/External Frame of Reference Model of Academic Achievement and Academic Self-Concept. *Review of Educational Research, 79*(3), 1129-1167. doi:10.3102/0034654309337522

Montgomery, G. K. (1977). Effects of performance evaluation and anxiety on cardiac response in anticipation of difficult problem solving. *Psychophysiology, 14*(3), 251-257. doi:10.1111/j.1469-8986.1977.tb01170.x

Moore, L. J., Vine, S. J., Wilson, M. R., & Freeman, P. (2012). The effect of challenge and threat states on performance: An examination of potential mechanisms. *Psychophysiology, 49*(10), 1417-1425. doi:10.1111/j.1469-8986.2012.01449.x

Moore, L. J., Vine, S. J., Wilson, M. R., & Freeman, P. (2014). Examining the antecedents of challenge and threat states: The influence of perceived required effort and support availability. *International Journal of Psychophysiology, 93*(2), 267-273. doi:10.1016/j.ijpsycho.2014.05.009

Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology, 73*(4), 541-555. doi:10.1037/0022-0663.73.4.541

Morris, L. W., & Liebert, R. M. (1970). Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance. *Journal of Consulting and Clinical Psychology, 35*(3), 332-337. doi:10.1037/h0030132

Morse, D. R., Schacterle, G. R., Furst, M. L., Esposito, J., & Zaydenburg, M. (1982). Stress, relaxation and saliva: Relationship to dental caries and its prevention, with a literature review. *Annals of Dentistry, 42*(2), 47-54.

Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. *British Journal of Educational Psychology Monograph Series II, 12*, 71-87.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (8 ed.). Los Angeles, CA: Muthén & Muthén.

Myrtek, M., Dieterle, W., & Brügner, G. (1990). Psychophysiological response patterns to variations of the experimental load of a reaction time task. *Journal of Psychophysiology, 4(3)*, 209-220.

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, 104*, 1033-1053. doi:10.1037/a0027697

Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology, 75*(1), 77-86. doi:10.1348/096317902167658

Naveh-Benjamin, M. (1991). A comparison of training programs intended for different types of test-anxious students: Further support for an information-processing model. *Journal of Educational Psychology, 83*, 134-139. doi:10.1037/0022-0663.83.1.134

Nett, U. E., Bieg, M., & Keller, M. M. (2017). How much trait variance is captured by measures of academic state emotions? *European Journal of Psychological Assessment*.

Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with boredom in school: An experience sampling perspective. *Contemporary Educational Psychology, 36*, 49-59. doi:10.1016/j.cedpsych.2010.10.003

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 111*(2), 646-651. doi:10.1073/pnas.1321664111

Nummenmaa, L., Hari, R., Hietanen, J. K., & Glerean, E. (2018). Maps of subjective feelings. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 115*(37), 9198-9203. doi:10.1073/pnas.1807390115

Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The single-item math anxiety scale: an alternative way of measuring mathematical anxiety. *Journal of Psychoeducational Assessment, 32*, 306-317. doi:10.1177/0734282913508528

Obrist, P. A. (1976). The cardiovascular-behavioral interaction—As it appears today. *Psychophysiology, 13*(2), 95-107. doi:10.1111/j.1469-8986.1976.tb00081.x

OECD. (2004). *Messages from PISA 2000.The Final Summary Report from the PISA 2000 Survey*. Paris, France: OECD.

OECD. (2016). *PISA 2015 assessment and analytical framework: Science, reading, mathematic and financial literacy*. Paris: OECD.

Ohle, A., & McElvany, N. (2015). Teachers' diagnostic competences and their practical relevance. Special Issue Editorial. *Journal for Educational Research Online, 7*(2), 5-10.

Onwuegbuzie, A. J., Johnson, R. B., & Collins, K. M. T. (2010). A framework for assessing legitimation in mixed research: Implications for the field of stress and coping. In K. M. T. Collins, A. J. Onwuegbuzie, & Q. G. Jiao (Eds.), *Toward a broader understanding of stress and coping: Mixed methods approaches.* (pp. 3-29). Charlotte, NC: IAP Information Age Publishing.

Parfitt, G., Hardy, L., & Pates, J. (1995). Somatic anxiety and physiological arousal: Their effects upon a high anaerobic, low memory demand task. *International Journal of Sport Psychology, 26*(2), 196-213.

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

Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotions. In P. A. Alexander, R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 561-579). Abingdon, England: Routledge.

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*, 115-135. doi:10.1037/a0013383

Pekrun, R., Goetz, T., & Frenzel, A. C. (2006). *Academic Emotions Questionnaire - Mathematics (AEQ-M) - User's manual*. Munich, Germany: University of Munich, Department of Psychology.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36-48. 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, 17*(3), 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. doi:10.1207/S15326985EP3702_4

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 11-20). New York, NY: Routledge.

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

Perry, R. P., & Hamm, J. (2017). An attribution perspective on competence and motivation: Theory and treatment interventions. In C. S. D. A. J. Elliot, & D. S. Yeager (Ed.), *Handbook of competence and motivation: Theory and application* (Vol. 2006, pp. 61–84). New York, NY: The Guilford Press.

Perry, R. P., Hladkyj, S., Pekrun, R. H., & Pelletier, S. T. (2001). Academic control and action control in the achievement of college students: A longitudinal field study. *Journal of Educational Psychology, 93*(4), 776.

Podsakoff, N. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.

Pollard, T. M. (1995). Use of cortisol as a stress marker: Practical and theoretical problems. *American Journal of Human Biology, 7*(2), 265-274. doi:10.1002/ajhb.1310070217

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185-227. doi:10.1080/00273170701341316

Putwain, D. (2007). Researching academic stress and anxiety in students: Some methodological considerations. *British Educational Research Journal, 33*(2), 207-219. doi:10.1080/01411920701208258

Putwain, D. W., Connors, L., & Symes, W. (2010). Do cognitive distortions mediate the test anxiety–examination performance relationship? *Educational Psychology, 30*(1), 11-26. doi:10.1080/01443410903328866

Qu, Y., & Telzer, E. H. (2017). Cultural differences and similarities in beliefs, practices, and neural mechanisms of emotion regulation. *Cultural Diversity and Ethnic Minority Psychology, 23*(1), 36-44. doi:10.1037/cdp0000112

Raphelson, A. C. (1957). The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *The Journal of Abnormal and Social Psychology, 54*(1), 13. doi:10.1037/h0041374

Raudenbush, S., Bryk, A., Cheong, Y., Congdon, R., & du Toit, M. (2011). HLM (Version 7.01)[Computer Software]. Lincolnwood, IL: Scientific Software International: Inc.

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

Reeve, C. L., Bonaccio, S., & Charles, J. E. (2008). A policy-capturing study of the contextual antecedents of test anxiety. *Personality and Individual Differences, 45*(3), 243-248. doi:10.1016/j.paid.2008.04.006

Reeve, J. (2014). *Understanding motivation and emotion* (5 ed.). Hoboken, NJ: Wiley.

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

Robinson, M. D., & Sedikides, C. (2009). Traits and the self: Toward an integration.

Rosen, L. D., & Maguire, P. (1990). Myths and realities of computerphobia: A meta-analysis. *Anxiety Research, 3*(3), 175-191.

Rubin, D. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592.

Ruthig, J. C., Perry, R. P., Hall, N. C., & Hladkyj, S. (2004). Optimism and attributional retraining: Longitudinal effects on academic achievement, test anxiety, and voluntary course withdrawal in college students. *Journal of Applied Social Psychology, 34*(4), 709-730. doi:10.1111/j.1559-1816.2004.tb02566.x

Saito, M., & Nakamura, Y. (1995). Cardiac autonomic control and muscle sympathetic nerve activity during dynamic exercise. *The Japanese Journal of Physiology, 45*(6), 961-977. doi:10.2170/jjphysiol.45.961

Salend, S. J. (2012). Teaching students not to sweat the test. *Phi Delta Kappan, 93*(6), 20-25.

Sarason, I. G. (1972). Experimental approaches to test anxiety: Attention and the uses of information. *Anxiety: Current Trends in Theory and Research, 2*, 383-403.

Sarason, I. G. (1978). The Test Anxiety Scale: Concept and research. In C. D. Spielberger & I. G. Sarason (Eds.), *Stress and Anxiety* (pp. 193-216). Washington, DC: Hemisphere.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46*(4), 929-938. doi:10.1037/0022-3514.46.4.929

Sarason, S. B., Davidson, K., Lighthall, F., & Waite, R. (1958). A test anxiety scale for children. *Child Development*, 105-113. doi:10.2307/1126274

Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. *Review of Personality & Social Psychology, 5*, 37-63.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 317). Hilsdale, NJ: Erlbaum.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York, NY: Oxford University Press.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*, 695-729. doi:10.1177/0539018405058216

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23*(7), 1307-1351. doi:10.1080/02699930902928969

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research*. New York, NY: Oxford University Press.

Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies, 4*, 79-106. doi:10.1023/a:1023661322862

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review, 61*(2), 81 - 88. doi:10.1037/h0054570

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3 ed.). Los Angeles, CA: Sage.

Schutz, P. A., & Davis, H. A. (2000). Emotions and self-regulation during test taking. *Educational Psychologist, 35*, 243-256. doi:10.1207/S15326985EP3504_03

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Methods, 21*(2), 206-221. doi:10.1037/met0000062

Schwartz, E. B., Granger, D. A., Susman, E. J., Gunnar, M. R., & Laird, B. (1998). Assessing Salivary Cortisol in Studies of Child Development. *Child Development, 69*(6), 1503-1513. doi:10.1111/j.1467-8624.1998.tb06173.x

Schwarzer, R., & Jerusalem, M. (1992). Advances in anxiety theory: A cognitive process approach. In K. A. Hagtvet & T. B. Johnsen (Eds.), *Advances in test anxiety research* (pp. 2-31). Lisse, Netherlands: Swets & Zeitlinger Publishers.

Scollon, C. N., Prieto, C.-K., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strength and weaknesses *Assessing well-being* (pp. 157-180). Dordrecht: Springer.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research, 4*, 27-41. doi:10.1080/08917779108248762

Selye, H. (1976). The stress concept. *Canadian Medical Association Journal, 115*(8), 718.

Selye, H. (1976). Stress without distress *Psychopathology of human adaptation* (pp. 137-146): Springer.

Setz, C., Arnrich, B., Schumm, J., Marca, R. L., Tr, G., #246, . . . Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable EDA device %J Trans. Info. Tech. Biomed. *14*(2), 410-417. doi:10.1109/titb.2009.2036164

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology, 5*, 1040. doi:10.3389/fpsyg.2014.01040

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology, 96*(5), 1055. doi:10.1037/a0023322

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol., 4*, 1-32. doi:10.1146/annurev.clinpsy.3.022806.091415

Shuman, V., Clark-Polner, E., Meuleman, B., Sander, D., & Scherer, K. R. (2017). Emotion perception from a componential perspective. *Cognition and Emotion, 31*(1), 47-56. doi:10.1080/02699931.2015.1075964

Shuman, V., & Scherer, K. R. (2014). Concepts and structures of emotions. In R. Pekrun & E. A. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 13-35). New York, NY: Taylor & Francis.

Singh, J. P., Larson, M. G., Tsuji, H., Evans, J. C., O'Donnell, C. J., & Levy, D. (1998). Reduced heart rate variability and new-onset hypertension. *Hypertension, 32*(2), 293-297. doi:10.1161/01.HYP.32.2.293

Sloan, R., Shapiro, P., Bagiella, E., Boni, S., Paik, M., Bigger, J., . . . Gorman, J. (1994). Effect of mental stress throughout the day on cardiac autonomic control. *Biological Psychology, 37*(2), 89-99. doi:10.1016/0301-0511(94)90024-8

Spangler, G., Pekrun, R., Kramer, K., & Hofmann, H. (2002). Students' emotions, physiological reactions, and coping in academic exams. *Anxiety, Stress & Coping, 15*(4), 413-432. doi:10.1080/1061580021000056555

Spielberger, C. D., & Gonzalez, H. P. (1980). *Preliminary professional manual for the test anxiety inventory: TAI*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Anton, W. D., Algaze, B., & Ross, G. K. (1980). *Test anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.

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

Spielberger, C. D., & Vagg, P. R. (1995). *Test anxiety: A transactional process model*. Philadelphia, PA: Taylor & Francis.

Spodick, D. H. (1993). Survey of selected cardiologists for an operational definition of normal sinus heart rate. *The American Journal of Cardiology, 72*(5), 487-488.

Stern, R. M., & Higgins, J. D. (1969). Perceived somatic reactions to stress: Sex, age and familial occurrence. *Journal of Psychosomatic Research, 13*(1), 77-82. doi:10.1016/0022-3999(69)90022-1

Szafranski, D. D., Barrera, T. L., & Norton, P. J. (2012). Test anxiety inventory: 30 years later. *Anxiety, Stress, & Coping, 25*(6), 667-677. doi:10.1080/10615806.2012.663490

Taub, M. (2018). *Using multi-channel data to assess, understand, and support metacognition with advanced learning technologies.* (79), ProQuest Information & Learning. Retrieved from http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3dpsyh%26AN%3d2018-21183-175%26site%3dehost-live

Taylor, J., & Deane, F. P. (2002). Development of a short form of the Test Anxiety Inventory (TAI). *The Journal of General Psychology, 129*(2), 127-136. doi:10.1080/00221300209603133

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*, 21-27. doi:10.1037/0033-2909.116.1.21

Tong, E. M., Bishop, G. D., Enkelmann, H. C., Why, Y. P., Diong, S. M., Khader, M., & Ang, J. (2007). Emotion and appraisal: A study using ecological momentary assessment. *Cognition and Emotion, 21*, 1361-1381. doi:10.1080/02699930701202012

Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science, 23*(6), 466-470. doi:10.1177/0963721414550706

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. *Psychological Assessment, 21*(4), 457-462. doi:10.1037/a0017653

Tryon, G. S. (1980). The measurement and treatment of test anxiety. *Review of Educational Research, 50*(2), 343-372. doi:10.2307/1170150

Turner, J. R., Carroll, D., Hanson, J., & Sims, J. (1988). A comparison of additional heart rates during active psychological challenge calculated from upper body and lower body dynamic exercise. *Psychophysiology, 25*(2), 209-216. doi:10.1111/j.1469-8986.1988.tb00990.x

Uchino, B. N., Berg, C. A., Smith, T. W., Pearce, G., & Skinner, M. (2006). Age-related differences in ambulatory blood pressure during daily stress: Evidence for greater blood pressure reactivity with age. *Psychology and Aging, 21*(2), 231. doi:10.1037/0882-7974.21.2.231

Uchino, B. N., Holt-Lunstad, J., Bloor, L. E., & Campo, R. A. (2005). Aging and cardiovascular reactivity to stress: longitudinal evidence for changes in stress reactivity. *Psychology and Aging, 20*(1), 134 - 143. doi:10.1037/0882-7974.20.1.134

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The Relation Between Self-Beliefs and Academic Achievement: A Meta-Analytic Review. *Educational Psychologist, 39*(2), 111-133. doi:10.1207/s15326985ep3902_3

Van Yperen, N. W. (2007). Performing well in an evaluative situation: The roles of perceived competence and task-irrelevant interfering thoughts. *Anxiety, Stress, and Coping, 20*(4), 409-419. doi:10.1080/10615800701628876

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

Wanous, J. P., & Hudy, M. J. (2001). Single-item reliability: A replication and extension. *Organizational Research Methods, 4*(4), 361-375. doi:10.1177/109442810144003

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. M.-D. Lisi & R. D. Lisi (Eds.), *Biology, society, and behavior: The development of sex differences in cognition* (Vol. 21, pp. 93-124). Westport, CT: Ablex Publishing.

Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology, 84*(3), 552-569. doi:10.1016/j.biopsycho.2010.01.017

Wilhelm, F. H., & Roth, W. T. (2001). The somatic symptom paradox in DSM-IV anxiety disorders: Suggestions for a clinical focus in psychophysiology. *Biological Psychology, 57*(1-3), 105-140. doi:10.1016/S0301-0511(01)00091-6

Wilhelm, P., Perrez, M., & Pawlik, K. (2012). Conducting research in daily life: A historical review. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 62-86). New York, NY: The Guilford Press.

Wilson, T. D., & Gilbert, D. T. (2005). Affective Forecasting: Knowing What to Want. *Current Directions in Psychological Science, 14*(3), 131-134. doi:10.1111/j.0963-7214.2005.00355.x

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2004). 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

Zanstra, Y. J., & Johnston, D. W. (2011). Cardiovascular reactivity in real life settings: Measurement, mechanisms and meaning. *Biological Psychology, 86*(2), 98-105. doi:10.1016/j.biopsycho.2010.05.002

Zeidner, M. (1990). Does test anxiety bias scholastic aptitude test performance by gender and sociocultural group? *Journal of Personality Assessment, 55*, 145-160. doi:10.1207/s15327752jpa5501&2_14

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

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 165-184). San Diego, CA: Elsevier Academic Press.

Zeidner, M. (2014). Anxiety in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 265-288). New York, NY: Routledge.

Zeidner, M., & Matthews, G. (2005). Evaluation anxiety: Current theory and research. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 141-163). New York, NY: Guilford Press.

Zeidner, M., & Matthews, G. (2011). *Anxiety 101*. New York, NY: Springer.

Zuckerman, M. (1960). The development of an affect adjective check list for the measurement of anxiety. *Journal of Consulting Psychology, 24*(5), 457.

Zuckerman, M., & Spielberger, C. D. (2015). *Emotions and Anxiety: New Concepts, Methods, and Applications*. New York, NY: Psychology Press.

Zung, W. W. (1976). SAS, self-rating anxiety scale *ECDEU Assessment Manual for Psychopharmacology, Revised Edition* (pp. 337-340). Rockville, Maryland.

6 Index of Tables

7 Index of Figures