

---

---

What Can They Tell Us?

**Understanding the Meaning and Application of Physiological  
Measures in Achievement Situations**

---

---



**Inaugural-Dissertation**

zur Erlangung des Doktorgrades der Philosophie  
an der Ludwig-Maximilians-Universität München

vorgelegt von

**Miriam Wunsch**

aus Ingolstadt

2026

Referentin: Prof. Dr. Anne C. Frenzel, Ludwig-Maximilians-Universität München

Korreferent: Prof. Dr. Lars-Erik Malmberg, University of Oxford

Tag der mündlichen Prüfung: 11.12.2025

---

## Acknowledgements

Thank you to my supervisors and co-authors for challenging and supporting me, to the lovely team who are so much more than just colleagues, and to all the smart, inspiring, and simply wonderful friends who have been with me on this journey. Thank you, Jake, my best friend.

And thank you to my family, I would not be here without you.

It has been a wonderful time. This is only the beginning.



## **Zusammenfassung**

### **(German Summary)**

Die vorliegende Arbeit beschäftigt sich mit der Bedeutung und Anwendung physiologischer Maße des autonomen Nervensystems in Leistungssituationen. Die Verwendung solcher physiologischen Maße gewinnt im Bereich der pädagogischen Psychologie zunehmend an Bedeutung, da diese Maße gegenüber dem häufig verwendeten Selbstbericht einige Vorteile bringen (z.B. Pekrun & Marsh, 2022). Zum einen können Selbstberichtsmaße dadurch beeinträchtigt werden, dass eine Person nicht ausreichend dazu in der Lage ist, ihren eigenen Zustand zu erkennen. Außerdem können auch Verzerrungen im Berichten dieses Zustands auftreten (Paulhus & Vazire, 2007). Zum anderen sind Selbstberichtsmaße häufig nicht kontinuierlich anwendbar, sondern an den Einsatz zu bestimmten Zeitpunkten gebunden. Dies kann zu Verzerrungen dadurch führen, dass Zustände retrospektiv berichtet werden. Außerdem kann der Selbstbericht die Situation selbst unterbrechen, was zu Beeinträchtigungen von Lernen und Leistung führen, oder die erfassten Konstrukte selbst beeinflussen kann (Chen et al., 2024; Johar & Sackett, 2018; Quinio & Lam, 2021). Physiologische Maße können dagegen kontinuierlich während einer Leistungssituation angewendet werden, ohne die Situation zu unterbrechen und ohne die Aufmerksamkeit wiederholt auf die zu untersuchenden Konstrukte zu lenken (Pekrun, 2023b). Besonders diese Möglichkeit zur kontinuierlichen Anwendung könnte die Verwendung physiologischer Maße im Sinne von Mikro-Adaptivität von Lern- und Testumgebungen ermöglichen. Das heißt, dass im Verlauf einer Lern- oder Leistungssituation ungünstige Veränderungen erkannt werden können und darauf reagiert werden kann (Plass & Pawar, 2020).

Besonders Maße des autonomen Nervensystems bieten die genannten Vorteile, da sie im Vergleich zu anderen physiologischen Maßen effizient und einfach anzuwenden sind (Ishikawa, 2023). Das autonome Nervensystem besteht aus zwei zentralen Teilen: dem sympathischen („*fight or flight*“) und dem parasympathischen Nervensystem („*rest and digest*“; Waxenbaum et al., 2023), deren Aktivität durch verschiedene Maße erfasst werden kann. Die Messung der Hautleitfähigkeit gibt einen Einblick in die Aktivität des sympathischen Nervensystems, Herzrate in sowohl die Aktivität des sympathischen als auch des parasympathischen Nervensystems und Herzratenvariabilität indiziert parasympathische Aktivität (Wehrwein et al., 2016). Obwohl die Verwendung dieser Maße einen methodischen Fortschritt in der pädagogischen Psychologie darstellen kann und bereits auf verschiedene Weise eingesetzt wurde, ist wenig über ihre Bedeutung und ihre Anwendung in Leistungssituationen bekannt.

Wenn Zusammenhänge zwischen physiologischen Zuständen und psychologischen Prozessen etabliert werden sollen, muss klar zwischen Zusammenhängen auf der inter- und der intraindividuellen Ebene unterschieden werden. Interindividuelle Zusammenhänge zwischen Variablen beschreiben, wie Unterschiede zwischen Personen auf einer Variable mit Unterschieden zwischen Personen auf einer anderen Variable zusammenhängen. Intraindividuelle Zusammenhänge beschreiben hingegen, wie Veränderungen auf einer Variable innerhalb einer Person mit Veränderungen auf einer anderen Variable zusammenhängen. Da sich inter- und intraindividuelle Assoziationen von Variablen voneinander unterscheiden können, ist es in der Analyse und Interpretation solcher Assoziationen essentiell, die beiden Analyseebenen klar voneinander zu trennen (Hoffman, 2015; Murayama et al., 2017). In bestehenden theoretischen Arbeiten wird darauf hingewiesen, dass Zusammenhänge zwischen psychologischen und physiologischen Zuständen in erster Linie auf der intraindividuellen Ebene zu finden sind (Mauss et al., 2024; Pekrun, 2023b). Dabei bleibt jedoch unklar, ob davon ausgegangen wird, dass Assoziationen auf der interindividuellen Ebene nicht existent, oder lediglich schwer zu erkennen sind. Das erste Leitthema der vorliegenden Arbeit ist es daher, die Unterschiede zwischen inter- und intraindividuellen Zusammenhängen von physiologischen und psychologischen Zuständen in Leistungssituationen zu beleuchten.

In früheren Studien wurden bereits Zusammenhänge zwischen den physiologischen Zuständen des autonomen Nervensystems und psychologischen Konstrukten wie Kognition, Motivation und Emotion gefunden. Die bestehende Evidenz ist eher begrenzt und zum Teil inkonsistent; dennoch deuten die Befunde darauf hin, dass tatsächlich ein Zusammenhang zwischen physiologischen Maßen und psychologischen Prozessen besteht. Hierbei stellt es allerdings für die sinnvolle Anwendung dieser Maße ein Problem dar, wenn physiologische Veränderungen nicht klar auf einen bestimmten psychologischen Prozess zurückgeführt werden können (Richter & Slade, 2017). Ein präziseres Verständnis der Bedeutung physiologischer Zustände in Leistungssituationen wäre daher für Theorie, Forschung und Praxis von großem Erkenntnisgewinn. In der vorliegenden Arbeit wird daher ein Modell vorgeschlagen, welches bisherige Theorien und Befunde aus den Bereichen der Emotion, Motivation und Kognition integriert, um die Bedeutung physiologischer Prozesse in Leistungssituationen präziser zu definieren. In diesem Modell werden physiologische Veränderungen im autonomen Nervensystem an der Schnittstelle von Emotion und Motivation verortet, wo die Konstrukte die Funktion teilen, adaptives Verhalten durch die Mobilisierung von Ressourcen für kognitive Prozesse zu fördern. Das zweite Leitthema der vorliegenden Arbeit ist eine Evaluation dieses

Modells, mit dem Ziel, ein genaueres Verständnis der Bedeutung physiologischer Zustände in Leistungssituationen zu etablieren.

Schließlich geht aus den bisherigen Befunden hervor, dass Studien verschiedene Indikatoren des autonomen Nervensystems verwendet, und sich teilweise Unterschiede zwischen den Indikatoren gezeigt haben. Obwohl dies möglicherweise darauf hindeutet, dass die verschiedenen Maße unterschiedliche Prozesse indizieren könnten, konnte hierzu bisher kein klares Muster identifiziert werden. Das dritte Leitthema der vorliegenden Arbeit ist daher eine explorative Betrachtung möglicher Unterschiede in der Bedeutung verschiedener Indikatoren des autonomen Nervensystems.

Die vorliegende Arbeit widmet sich den beschriebenen Themen, indem in zwei Studien quantitative Daten in einer Leistungssituation erhoben wurden. Die Teilnehmenden (Studie 1:  $N = 90$ ; Studie 2:  $N = 73$ ) waren erwachsen und überwiegend weibliche Studierende. Beide Studien fanden im Labor statt, wo die Teilnehmenden am Computer mehrere Blöcke mit jeweils vier Items einer numerischen Denkaufgabe bearbeiteten (sechs Blöcke in Studie 1, fünf Blöcke in Studie 2). In Studie 1 passte sich bei der Hälfte der Blöcke die Aufgabenschwierigkeit adaptiv an die Leistung der Teilnehmenden an, während in Studie 2 alle Blöcke nicht adaptiv, d.h. für alle Personen gleich, waren. Während der Aufgabenbearbeitung wurden Hautleitfähigkeit und kardiovaskuläre Aktivität aufgezeichnet, woraus jeweils zwei Parameter für jede Modalität abgeleitet wurden (Hautleitfähigkeitsreaktion und Hautleitfähigkeitslevel, Herzrate und Herzratenvariabilität). Nach jedem Aufgabenblock füllten die Teilnehmenden einen Selbstberichtsfragebogen aus, in welchem sie unter anderem ihr subjektives Erleben verschiedener Leistungsemotionen sowie ihre Wahrnehmung von Schwierigkeit und Wert der Aufgabe berichteten. Die aus den wiederholten Messungen resultierenden Daten ermöglichten die Analyse verschiedener inter- und intraindividuelle Beziehungen. Diese Analysen wurden im Rahmen von drei empirischen Arbeiten durchgeführt, welche im Folgenden beschrieben werden.

Die erste empirische Arbeit untersucht Zusammenhänge zwischen den vier physiologischen Variablen (Hautleitfähigkeitsreaktion, Hautleitfähigkeitslevel, Herzrate und Herzratenvariabilität) und der selbstberichteten Intensität von sechs Leistungsemotionen (Freude, Stolz, Ärger, Frustration, Angst und Langeweile). Bivariate Korrelationen auf der inter- und intraindividuellen Ebene zeigten, dass Zusammenhänge zwischen physiologischen Zuständen und selbstberichteten Leistungsemotionen in erster Linie auf der intraindividuellen Ebene bestehen. Dabei wurden für Freude, Stolz und Angst aktivierende Muster festgestellt, d.h. ein verstärktes Erleben dieser Emotionen ging mit einer erhöhten physiologischen

Aktivierung einher. Für Ärger, Frustration und Langeweile zeigte sich dagegen ein gegensätzliches, deaktivierendes Muster. Auffällig waren auch die intraindividuellen Befunde zum Hautleitfähigkeitslevel, welches ein gegensätzliches Muster zur Hautleitfähigkeitsreaktion zeigte, obwohl beide Maße sympathische Aktivierung indizieren sollten.

In der zweiten empirischen Arbeit wurden physiologische Zustände nicht nur mit Leistungsemotionen, sondern auch mit den motivationalen Konstrukten von Leistungsmotivation und Anstrengung in Verbindung gesetzt. Das Ziel hierbei war es, herauszufinden, welches der Konstrukte am engsten mit den physiologischen Maßen in Zusammenhang steht. Dazu wurde aus den zugrunde liegenden Theorien abgeleitet, dass alle drei Konstrukte – Leistungsemotionen, Leistungsmotivation und Anstrengung – von einer Interaktion aus der wahrgenommenen Schwierigkeit und dem wahrgenommenen Wert einer Aufgabe vorhergesagt werden sollten. Darauf aufbauend wurden Annahmen aufgestellt, wie diese Interaktion physiologische Prozesse (Hautleitfähigkeitsreaktion, Herzrate und Herzratenvariabilität) vorhersagen sollte, wenn diese Prozesse mit dem jeweiligen psychologischen Konstrukt in Verbindung stehen. Diese Annahmen wurden dann mit den tatsächlich beobachteten Mustern verglichen. In Mehrebenen-Regressionsmodellen mit einem variablen Intercept zeigten sich quadratische Beziehungen zwischen der wahrgenommenen Schwierigkeit der Aufgabe und allen physiologischen Maßen, teilweise moderiert durch den wahrgenommenen Wert. Hautleitfähigkeitsreaktion und Herzrate waren bei sehr hoher und sehr niedriger wahrgenommener Schwierigkeit eher niedrig, während die Herzratenvariabilität bei mittlerer wahrgenommener Schwierigkeit ihr Minimum erreichte. In Übereinstimmung mit früheren Arbeiten zur *Motivational Intensity Theory* (Brehm & Self, 1989; Gendolla, 2025) wiesen diese beobachteten Muster der physiologischen Aktivierung die stärkste Ähnlichkeit zu den Annahmen auf, die für das Konstrukt der Anstrengung abgeleitet wurden. Für die Hautleitfähigkeitsreaktion blieb über den quadratischen Effekt hinaus auch ein linearer Effekt der Schwierigkeit bestehen, was möglicherweise auf einen feinen Unterschied zwischen den verschiedenen Maßen hindeutet.

Die dritte empirische Arbeit untersucht, wie sich die Verwendung eines adaptiven Leistungstests auf den physiologischen und emotionalen Zustand der Teilnehmenden auswirkt. Die Hautleitfähigkeitsreaktion und die selbst berichteten Leistungsemotionen wurden jeweils über die drei adaptiven und die drei nicht-adaptiven Testblöcke aggregiert. Lineare Mehrebenen-Regressionsmodelle mit einem variablen Intercept zeigten, dass auf der intraindividuellen Ebene die Hautleitfähigkeitsreaktion während des adaptiven Tests höher war als während des nicht-adaptiven Tests. Hinsichtlich der selbstberichteten Leistungsemotionen

zeigte sich kein signifikanter Haupteffekt der Testart. Stattdessen wurde eine Interaktion zwischen Testart und relativer wahrgenommener Schwierigkeit festgestellt. Diese zeigt, dass in dem Test, der als leichter bewertet wurde, mehr Freude und Stolz und weniger Angst, Ärger, Frustration und Langeweile auftraten.

Die Ergebnisse aus allen drei empirischen Analysen liefern gemeinsam zentrale Erkenntnisse zu den Leitthemen der vorliegenden Arbeit – (1) die Unterschiede zwischen Analyse-Ebenen, (2) die Bedeutung physiologischer Zustände in Leistungssituationen und (3) mögliche Unterschiede zwischen den Indikatoren. Hinsichtlich des ersten Leitthemas der Analyse-Ebene zeigen die Ergebnisse beider Studien eine erhebliche interindividuelle Varianz in den physiologischen Maßen. Eine unzureichende Korrektur für solche großen interindividuellen Unterschiede könnte einen Grund dafür darstellen, warum sich auf der interindividuellen Ebene keine substantiellen Zusammenhänge zwischen physiologischen Zuständen und psychologischen Prozessen zeigten. Außerdem unterstreichen die Befunde der ersten empirischen Arbeit die Notwendigkeit, klar zwischen Zusammenhängen zwischen Variablen auf der inter- und der intraindividuellen Ebene zu trennen, da diese sich substantiell voneinander unterscheiden können.

Hinsichtlich der Bedeutung physiologischer Zustände in Leistungssituationen liefern die vorliegenden Ergebnisse ein konsistentes Bild. In Unterstützung des vorgeschlagenen konzeptionellen Modells legen sie nahe, dass Aktivität im autonomen Nervensystem mit der Mobilisierung von Ressourcen einhergeht: Zum einen zeigen die Befunde aus der zweiten empirischen Arbeit eine Übereinstimmung der Muster für Anstrengung und physiologische Aktivierung. Des Weiteren deuten auch die intraindividuellen Aktivierungsmuster für Freude, Stolz und Angst aus der ersten empirischen Arbeit darauf hin, dass physiologische Aktivierung dann besonders hoch ist, wenn eine Aufgabe als wertvoll und einigermaßen kontrollierbar wahrgenommen wird. In diesem Fall stellt die Investition von Ressourcen eine adaptive Reaktion dar. Die als deaktivierend identifizierten Emotionen Wut, Frustration und Langeweile treten hingegen auf, wenn eine Aufgabe entweder als nicht wertvoll wahrgenommen wird und/oder nur in sehr hohem oder sehr geringem Maße kontrollierbar ist, d. h. dann, wenn die Investition von Ressourcen unnötig oder ineffektiv ist. Schließlich stützt auch der Haupteffekt des adaptiven Testens auf die sympathische Aktivierung die These, dass physiologische Aktivierung die Mobilisierung von Ressourcen widerspiegelt: Teilnehmende waren dann mehr physiologisch aktiviert, wenn sich die Aufgabenschwierigkeit eines Tests an ihr Leistungsniveau anpasste, d.h. dann, wenn sie kontinuierlich gefordert, aber nicht unter- oder überfordert und deshalb weniger engagiert waren.

Ein solches engeres Verständnis physiologischer Veränderungen als die Mobilisierung von Ressource in einer Leistungssituation impliziert auch, dass physiologische Maße alleine nicht genug sind, um einen umfassenden Einblick in die emotionalen und motivationalen Zustände von Personen zu gewinnen. Wie sich in den Ergebnissen zeigte, ist physiologische Aktivierung oder Deaktivierung nicht spezifisch für eine bestimmte emotionale Valenz oder Aufgabenwahrnehmung. Daher müssen physiologische Maße mit anderen Informationsquellen in Verbindung gesetzt werden, um einen umfassenden Einblick in psychologische Prozesse während einer Leistungssituation zu gewinnen.

Schließlich zeigt eine explorative Betrachtung der vorliegenden Ergebnisse, dass sich verschiedene Indikatoren tatsächlich darin unterscheiden, wie sie mit psychologischen Variablen zusammenhängen. Auf Basis der beobachteten Muster ist es möglich, dass sympathische und parasympathische Aktivierung mit unterschiedlichen Aspekten der Ressourcenmobilisierung zusammenhängen. Sympathische Aktivierung könnte eher die Bereitschaft für das Engagement mit einer Aufgabe widerspiegeln, während parasympathische Aktivierung eher mit den Anforderungen der Aufgabe zusammenhängen könnte. Obwohl es sich hierbei nur um eine mögliche Interpretation handelt, zeigen die Befunde, dass die beiden Bestandteile des autonomen Nervensystems als eigenständige Informationsquellen betrachtet werden sollten. Darüber hinaus unterstreichen die unerwarteten Befunde für das Hautleitfähigkeitslevel die Notwendigkeit eines tieferen Verständnisses der einzelnen Maße und betonen, dass verschiedene Indikatoren nicht gegeneinander austauschbar sind.

Aus den vorliegenden Ergebnissen lassen sich mehrere theoretische und praktische Schlussfolgerungen ableiten: Die Befunde legen nahe, den Fokus auf intraindividuelle Zusammenhänge zu legen, physiologische Maße als Indikatoren für Ressourcenmobilisierung zu verstehen, sowie dass unterschiedliche Indikatoren möglicherweise verschiedene Nuancen der Ressourcenmobilisierung widerspiegeln. Die Ergebnisse und die resultierende Diskussion werfen darüber hinaus wichtige Fragen für Theorie und Empirie auf. Zum einen ist eine klarere theoretische Konzeptualisierung physiologischer Zustände und ihrer Beziehung zu anderen Konstrukten nötig. Außerdem sollte ein genaueres Verständnis der einzelnen Indikatoren angestrebt werden, sowie die Identifikation von Möglichkeiten, verschiedene Indikatoren oder Modalitäten miteinander zu integrieren, um einen umfassenderen Einblick in psychologische Prozesse in Leistungssituationen zu erlangen. Insgesamt leistet die vorliegende Arbeit einen wertvollen Beitrag zu einem vertieften Verständnis der Bedeutung und Anwendung physiologischer Maße in Leistungssituationen und stellt damit einen wichtigen Schritt auf dem Weg zu ihrem breiteren Einsatz in Forschung und Praxis dar.

### Extended Summary

The central aim of the present work is to advance our understanding of the meaning and application of physiological measures of the autonomic nervous system in achievement situations. In the field of educational psychology, there is an increasing demand for the use of physiological measures, as these measures may enable a more comprehensive understanding of individuals' states by overcoming limitations associated with the more commonly used self-report measures (e.g., Pekrun & Marsh, 2022). First, a central issue with self-report measures is their reliance on the individual's ability to accurately recognize and report their own states, which may be impaired by the inaccessibility of certain internal processes and by biases that occur in the reporting process (Paulhus & Vazire, 2007). Second, self-report measures in achievement situations are typically tied to specific time points, which may introduce retrospective reporting biases and interrupt task execution. Such interruptions divert attention from the achievement activity to the construct of interest, potentially impairing learning and performance and altering the assessed construct (Chen et al., 2024; Johar & Sackett, 2018; Quinio & Lam, 2021). Physiological measures can overcome these limitations of self-report, as they are more objective and can be applied continuously throughout a situation without interrupting it or repeatedly directing the individual's attention to the constructs of interest (Pekrun, 2023b). Such a continuous application may also enable micro-adaptivity of learning or testing environments by detecting unfavorable real-time changes in an individual's state and allowing a teacher or algorithm to intervene (Plass & Pawar, 2020). Consequently, physiological measures constitute a methodological advancement in educational psychology research and practice.

Measures of the autonomic nervous system are particularly well-suited for these applications in research and practice. The autonomic nervous system consists of two main branches, being the sympathetic (*fight or flight*) and the parasympathetic branch (*rest and digest*; Waxenbaum et al., 2023). Electrodermal activity indicates sympathetic activation, whereas cardiovascular measures reflect activation in both branches; specifically, heart rate is influenced by both branches, while heart rate variability reflects parasympathetic activation (Wehrwein et al., 2016). Although these measures bring the aforementioned benefits over self-report measures and have already been employed in various ways in educational psychology, little is known about their exact meaning and application in achievement situations.

When aiming to establish associations between physiological states and psychological processes, it is crucial to distinguish between associations at the between- and within-person levels. Associations at the between-person level describe how differences between individuals

on one variable relate to differences between them on another variable, whereas within-person associations describe the relation of changes on two variables within the same person. As associations at the two levels can substantially and truly differ from one another, it is a central distinction to be made in both analysis and interpretation (Hoffman, 2015; Murayama et al., 2017). Previous theoretical work has suggested that associations between psychological and physiological states may primarily occur at the within-person level (Mauss et al., 2024; Pekrun, 2023b), but it remains unspecified whether such associations at the between-person level are absent or merely challenging to detect. This motivates the first guiding goal of the present work, which is to examine how associations between physiological states and psychological processes differ at the between- and within-person levels.

Previous work has linked physiological states in the autonomic nervous system to cognitive, motivational, and emotional variables. Although the evidence is somewhat sparse and ambiguous, it generally appears that relationships exist with all three psychological constructs. However, the existence of relations to different constructs may hinder the meaningful application of physiological measures in research and practice, as changes in physiological states are not clearly attributable to a specific psychological process (Richter & Slade, 2017). Therefore, theory, research, and practice would benefit from a narrower understanding of the meaning of physiological states in achievement situations. Based on theoretical and empirical prior work, I have developed a conceptual model that places physiological activation at the intersection of emotion and motivation, where the two constructs overlap in their function to mobilize resources for the cognitive processes required for a given achievement task. The second goal topic of the present work is to evaluate this model and establish a more specific understanding of the meaning of physiological states in achievement situations.

Finally, prior studies have used various indicators of autonomic activation, revealing differences in their relations to psychological constructs; however, no clear patterns could be identified. Therefore, the present work follows the third, more exploratory goal, of inspecting differences between different indicators of autonomic activation in the patterns they exhibit in their relation to psychological variables and possibly identifying nuanced differences in their meanings within achievement situations.

The present work addresses the outlined goals by synthesizing across the findings from three empirical works, which are based on the data from two laboratory studies. Both studies included adult participants (Study 1:  $N = 90$ ; Study 2:  $N = 73$ ), who were largely female university students. In the laboratory, they completed several blocks of a computerized

numerical reasoning tasks (six blocks in Study 1, five blocks in Study 2), containing four items each. In Study 1, half of the task blocks were adaptive to participants' performance, whereas in Study 2, all blocks were non-adaptive. While working on the task, participants' electrodermal and cardiovascular activity were recorded. From the data, two indices were derived from each modality, namely skin conductance response and skin conductance level for electrodermal activity, and heart rate and heart rate variability for cardiovascular activity. After each task block, participants filled in a pen-and-paper self-report questionnaire, assessing various variables, including their subjective experience of different achievement emotions, as well as their perception of task difficulty and value. This repeated-measures data enables the investigation of between- and within-person relationships of physiological states in the autonomic nervous system with emotional experiences, task perceptions, and task characteristics, which is done in three empirical works.

The first empirical work investigates associations of four physiological variables (skin conductance response, skin conductance level, heart rate, and heart rate variability), with the self-reported intensity of six achievement emotions, (enjoyment, pride, anger, frustration, anxiety, boredom). These associations are explored by calculating bivariate correlations at both the between-person and within-person levels, separately for each study. The findings show that associations between physiological states and self-reported achievement emotions were primarily present at the within- rather than the between-person level. Further, at the within-person level, activating patterns, that is, an increased emotional experience being accompanied by increased physiological activation, emerged for enjoyment, pride, and anxiety. Instead, opposing deactivating patterns emerged for anger, frustration, and boredom. Further, unexpected findings for skin conductance level showed patterns opposite to those of skin conductance response, despite both being indicators of sympathetic activation.

The second empirical work used data from Study 2 to explore how physiological states relate not only to achievement emotions, but also to the motivational constructs of achievement motivation and effort. To examine how physiological states link to these constructs, we identified that according to their underlying theories, each construct is predicted by an interaction between perceived task difficulty and value. Hence, we modelled how an interaction between self-reported task difficulty and value predicted skin conductance response, heart rate, and heart rate variability on a within-person level. We then compared the observed patterns to those expected for each construct, assuming that physiological activation is linked to the respective construct. At the within-person level, multilevel regression models with a random intercept revealed negative quadratic relationships between perceived task difficulty and skin

conductance response and heart rate, as well as a positive quadratic relationship between perceived difficulty and heart rate variability. Further, an additional negative linear effect of difficulty on skin conductance response remained. Either the linear or quadratic effects of perceived difficulty were moderated by perceived value, suggesting that higher levels of value amplified the effects of task difficulty on physiological states. In line with prior work around the Motivational Intensity Theory (Gendolla, 2025), the observed patterns for physiological activation resembled most closely the assumptions derived for the construct of effort, suggesting a close link between the two.

The third empirical work used data from Study 1 to investigate how the use of an adaptive, compared to a non-adaptive achievement test affects test-takers' physiological states and their emotional experiences. Skin conductance response and self-reported achievement emotions were aggregated across the three adaptive, and the three non-adaptive task blocks, respectively. Multilevel linear regression models with a random intercept showed that on the within-person level, skin conductance response was higher during the adaptive compared to the non-adaptive test. No main effect of test type emerged for self-reported achievement emotions. Instead, test type interacted with relative perceived difficulty in predicting self-reported achievement emotions, with more intense enjoyment and pride in the test that was perceived as easier, and more intense anxiety, anger, frustration, and boredom, in the test that was perceived as more difficult. These findings show that the adaptivity of the test elicited higher sympathetic activation, whereas in terms of valence the adaptive test could have been accompanied by either more positive or more negative achievement emotions depending on relative perceived difficulty.

Findings of the three individual empirical investigations are synthesized to derive central insights into the three main themes of the present work, being (1) the level of analysis, (2) the meaning of physiological states in achievement situations, and (3) possible differences between indicators. Regarding the level of analysis, both samples exhibited considerable between-person variability in physiological parameters, underscoring the need for an adequate baseline correction in between-person analyses (Insight 1). Inadequate baseline correction may hinder the detection of between-person associations, which could explain the inconsistent links observed between physiological measures and self-reported achievement emotions at the between-person level. Secondly, the findings show that associations of physiological states with psychological processes vary across the between- and within-person level (Insight 2). Not only did the within-person level show a more substantial pattern of associations between physiological measures and self-reported achievement emotions, but also the associations that

reached statistical significance on the between-person level were opposing in direction to those on the within-person level. These differences clearly underline the importance of distinguishing between the two levels in analysis and interpretation.

Regarding the meaning of physiological states, the present findings yield a consistent picture in support of the proposed conceptual model, placing physiological states at the intersection of emotion and motivation in the function to mobilize resources. All findings align with an understanding of autonomic arousal as the mobilization of resources (Insight 3), which is linked to both emotional and motivational constructs. Specifically, in line with work around Motivational Intensity Theory (Gendolla, 2025), the similarity of patterns for effort and physiological activation suggests that autonomic activation corresponds to the mobilization of resources, increasing with task difficulty until exceeding a point where the investment of more resources is not beneficial. The findings on achievement emotions further complement this picture: The achievement emotions enjoyment, pride, and anxiety occur when a task is perceived as valuable and somewhat controllable, making the investment of resources an adaptive response, which aligns with the finding of activating within-person patterns for these emotions. Instead, anger, frustration, and boredom occur when a task is either not valued, or very high or low in controllability, that is, when investing resources is unnecessary or ineffective, matching the deactivating patterns found for these emotions. Finally, the main effect of adaptive testing on sympathetic activation supports the view that physiological activation reflects resource mobilization: Activation was higher when a task matched the individual's ability level – demanding resources without surpassing the point where investment stops – compared to non-adaptive tests, where most items were either too easy or too difficult, reducing invested resources. This narrower understanding of physiological states in an achievement situation implies that individual physiological measures need to be complemented by other measures to gain a comprehensive insight into individuals' emotional and motivational states (Insight 4). As seen from the present set of findings, patterns of physiological activation or deactivation alone are unable to indicate emotional valence or specific task perceptions, for example, activation is observed both in enjoyment and anxiety, and deactivation both when a task is perceived as easy and difficult. Hence, in their application, physiological measures require triangulation with other sources of information that can provide an adequate context.

Finally, an exploratory view of the present findings indeed shows that indicators of autonomic activation differ in their patterns of associations with psychological variables and, therefore, possibly in their meaning (Insight 5). Based on the observed patterns, I suggest that sympathetic and parasympathetic activation correspond to different aspects of resource

mobilization, with sympathetic activation reflecting the readiness to engage with a task, and parasympathetic activation relating more to the resources mobilized depending on task demands. Although tentative, the observed differences underline the importance of distinguishing between the two independent autonomic branches. Further, unexpected findings emerged regarding the parameter of skin conductance level, showing correlational patterns opposite to those of skin conductance response. While the cause remains unclear, this discrepancy underscores the need for a deeper understanding of each measure and emphasizes that different indicators cannot be used interchangeably.

The present findings yield central insights advancing our understanding of the meaning and application of physiological measures in achievement situations. Several theoretical and practical implications can be drawn, including an emphasis on within-person associations, the suggestion to adopt a narrower understanding of physiological states in achievement contexts by integrating the constructs of emotion, motivation, and cognition, as well as the acknowledgement that different indicators may reflect distinct processes. Additionally, the findings and discussion raise important questions for future research, including how to precisely conceptualize physiological states and their relation to other constructs, as well as the goal of better understanding individual autonomic indicators and combining them with each other or with other modalities to achieve a more comprehensive understanding of physiological states in achievement contexts. Overall, the present work enhances our understanding of the meaning and application of physiological measures in achievement settings, constituting an important step toward their meaningful use in research and practice.

**Table of Contents**

<b>Acknowledgements</b> .....	<b>iii</b>
<b>Zusammenfassung</b> .....	<b>v</b>
<b>Extended Summary</b> .....	<b>xi</b>
<b>Table of Contents</b> .....	<b>xvii</b>
<b>INTRODUCTION</b> .....	<b>21</b>
<b>USING MEASURES OF THE AUTONOMIC NERVOUS SYSTEM IN EDUCATIONAL PSYCHOLOGY</b> .....	<b>25</b>
2.1 Benefits of Physiological over Self-Report Measures .....	27
2.1.1 <i>Unbiased Insights</i> .....	28
2.1.2 <i>Unobtrusive Real-Time Measurement</i> .....	29
2.2 Physiological Measures of the Autonomic Nervous System .....	30
2.2.1 <i>Electrodermal Activity</i> .....	33
2.2.2 <i>Cardiovascular Activity</i> .....	35
<b>PSYCHOLOGICAL CORRELATES OF AUTONOMIC MEASURES IN THE EDUCATIONAL CONTEXT</b> .....	<b>37</b>
3.1 Level of Analysis.....	39
3.2 Findings on Associations between Autonomic States and Psychological Constructs in the Achievement Context.....	43
3.2.1 <i>Cognition</i> .....	43
3.2.2 <i>Motivation</i> .....	45
3.2.3 <i>Emotion</i> .....	49
3.2.4 <i>Synthesis and Conceptual Model</i> .....	54
3.3 Aims of the Present Work.....	56
<b>DATA USED IN THE PRESENT WORK</b> .....	<b>59</b>
4.1 Laboratory Setup .....	61
4.2 Software iMotions .....	61
4.3 Hardware .....	62
4.4 Processing of Physiological Signals.....	62
4.4.1 <i>Electrodermal Activity</i> .....	62
4.4.2 <i>Cardiovascular Activity</i> .....	63
4.5 Study Design .....	63
4.5.1 <i>Baseline and Relaxation Phases</i> .....	64
4.5.2 <i>Validation Phase</i> .....	64
4.5.3 <i>Numerical Reasoning Task</i> .....	65
4.5.4 <i>Self-Report Questionnaire</i> .....	66
<b>OVERVIEW OF EMPIRICAL WORK</b> .....	<b>69</b>
5.1 Empirical Work 1: Physiological States and Self-reported Achievement Emotions.....	72
5.2 Empirical Work 2: Difficulty x Value Perceptions .....	74
5.3 Empirical Work 3: Effects of Adaptive Testing.....	76
<b>EMPIRICAL WORK 1</b> .....	<b>79</b>
6.1 Abstract.....	81
6.2 Educational Impact and Implications Statement.....	81
6.3 Introduction .....	82

## Table of Contents

---

6.3.1 Achievement Emotions .....	83
6.3.2 Measures of Autonomic Nervous System Activity.....	84
6.3.3 Between-Person versus Within-Person Approaches to Investigating Links between Autonomic Processes and the Subjective Experience of Achievement Emotions .....	86
6.3.4 The Present Research.....	89
6.4 Study 1.....	89
6.4.1 Method.....	89
6.4.2 Results .....	94
6.4.3 Discussion .....	98
6.5 Study 2.....	99
6.5.1 Method.....	99
6.5.2 Results .....	101
6.5.3 Discussion .....	103
6.6 General Discussion.....	106
6.6.1 Limitations and Implications for Future Research.....	108
6.6.2 Theoretical and Practical Implications.....	109
6.7 Conclusion.....	110
6.8 References .....	111
6.9 Supplemental Materials.....	115
<b>EMPIRICAL WORK 2 .....</b>	<b>121</b>
7.1 Introduction .....	121
7.1.1 Theoretical Perspectives .....	123
7.1.2 Psychophysiological States in Achievement Situations.....	125
7.1.3 Perceived Difficulty and Value of Success Predicting Physiological States .....	127
7.1.4 The Present Research.....	129
7.2 Method .....	131
7.2.1 Participants .....	131
7.2.2 Procedure .....	131
7.2.3 Numerical Reasoning Task.....	132
7.2.4 Measures .....	133
7.2.5 Analyses.....	134
7.2.6 Openness and Transparency .....	135
7.3 Results .....	135
7.3.1 Descriptives.....	135
7.3.2 Hierarchical Multilevel Linear Modelling .....	137
7.4 Discussion .....	142
7.4.1 Limitations and Future Research .....	145
7.4.2 Implications.....	145
7.5 References .....	146
7.6 Appendix .....	149
<b>EMPIRICAL WORK 3 .....</b>	<b>151</b>
8.1 Abstract.....	153
8.2 Introduction .....	153
8.2.1 Principles of Adaptive Testing.....	154

## Table of Contents

---

8.2.2 Achievement Emotions and Their Link with Test Difficulty .....	154
8.2.3 Psychophysiological Arousal .....	155
8.2.4 Subjective Experiences of Achievement Emotions .....	157
8.2.5 The Present Study.....	159
8.3 Methods.....	161
8.3.1 Ethics Statement and Data Transparency .....	161
8.3.2 Participants.....	161
8.3.3 Procedure .....	162
8.3.4 Measures .....	162
8.3.5 Analyses.....	165
8.4 Results .....	165
8.4.1 Descriptive Statistics and Manipulation Check.....	165
8.4.2 Effects of Test Type on Psychophysiological Arousal (H1).....	168
8.4.3 Effects of Test Type and Ability on Self-Reported Achievement Emotions (H2a) .....	168
8.4.4 Effects of Test Type and Relative Perceived Difficulty on Self-Reported Achievement Emotions (H2b) .....	169
8.5 Discussion.....	170
8.5.1 Effects of Adaptive versus Fixed-Item Testing on Psychophysiological Arousal .....	172
8.5.2 Effects of Adaptive Versus Fixed-Item Testing on Subjective Emotional Experience.....	172
8.5.3 Limitations and Directions for Future Research.....	175
8.6 Conclusion.....	177
8.7 References .....	177
<b>GENERAL DISCUSSION .....</b>	<b>181</b>
9.1 Insights Gained on Physiological Measures in an Achievement Situation .....	183
9.1.1 Level of Analysis.....	183
9.1.2 Meaning of Autonomic Changes in an Achievement Situation .....	189
9.1.3 Differences between Autonomic Measures.....	197
9.2 Strengths and Limitations of the Present Work.....	201
9.3 Considerations for Theory and Practice .....	205
9.3.1 Considerations around the Level of Analysis.....	205
9.3.2 Considerations around the Meaning of Physiological States .....	207
9.3.3 Considerations around the Differences between Indicators .....	208
9.4 Outlook.....	209
<b>CONCLUSION .....</b>	<b>211</b>
<b>REFERENCES .....</b>	<b>215</b>



---

# 1

---

## INTRODUCTION

---

Miriam Wunsch



In a learning or testing situation, there is a lot more happening within an individual than the mere acquisition and reproduction of knowledge. Just like any other situation in life, educational contexts are the setting for a multitude of psychological processes beyond cognition, including a broad range of emotional and motivational experiences: We can become curious during an interesting lecture, or anxious before a presentation. We can put all our time into studying for an exam, or we can sit in class and not care about algebra. Such emotional and motivational processes not only occur alongside of, but in a reciprocal relation with learning and performance (Camacho-Morles et al., 2021; Liu & Hou, 2018; Pekrun et al., 2017; Putwain et al., 2025; Vu et al., 2024). For example, understanding the concepts taught in a lesson might motivate us to study more for an upcoming test, leading to better performance. Alternatively, receiving a bad grade may increase our anxiety during the next test, resulting in another bad grade.

Given this central role of emotion and motivation, we aim to design learning and testing environments in an emotionally and motivationally supportive way, thereby maximizing learning outcomes and enabling individuals to show their full potential in assessments. Designing such supportive environments requires a thorough understanding of emotion and motivation in achievement situations, which depends on the availability of reliable and valid measures to capture these processes. These measures are needed to both deepen the theoretical understanding and to evaluate the effectiveness of design choices or intervention once developed for practical application. Typically, emotion and motivation research has employed self-report measures, which has led to many meaningful findings in the field of educational psychology. However, the use of self-report poses methodological disadvantages and can yield inaccurate measurements, such as when an individual is not sufficiently able to recognize their current state, they consciously or unconsciously fail to report it accurately, or when the self-report measure interrupts or influences the situation (Paulhus & Vazire, 2007; Pekrun, 2023b; Quinio & Lam, 2021). Drawing on the notion that mind and body are connected, research in educational psychology has begun to address these limitations of self-report by exploring the use of physiological measures: Supplementing or even replacing self-report with physiological measures may allow us to gain a more objective, unobtrusive real-time insight into an individual's emotion and motivation (Martin et al., 2023; Pekrun, 2023b). In practice, this may enable emotional and motivational micro-level adaptivity of learning and testing environments, that is, adapting learning or performance tasks to changes in the individual's emotional or motivational state (Azevedo et al., 2017; Plass & Pawar, 2020). For example, if a physiological measure could detect unfavourable changes in real-time, such as increases in frustration or a

decline in motivation, an adaptive test system could respond by presenting a simpler task to sustain a beneficial emotional and motivational state. While the use of physiological measures promises such unbiased and unobtrusive insight into an individual's state, possibly enabling emotional and motivational micro-level adaptivity of learning and testing environments in the future, this is still an emerging field in educational psychology. To date, little established knowledge exists on how to meaningfully apply physiological measures. Specifically, unresolved issues include methodological considerations around the appropriate level of analysis, as well as major gaps in understanding which physiological measures indicate which psychological processes in an achievement situation. Addressing these open questions is an essential next step towards the possibility of meaningfully employing physiological measures in educational research and practice.

In the present work, I aim to advance our understanding of the meaning and application of physiological measures, particularly those of the autonomic nervous system, in achievement situations. Following this introduction, Chapter 2 provides the necessary context and background by outlining the relevance of physiological measures in educational psychology and explaining the measures of electrodermal and cardiovascular activity used in this work. Chapter 3 narrows the focus to the central aim of this work. The chapter begins by highlighting the crucial distinction between associations on the between- and within-person levels. Next, the chapter provides an overview of existing knowledge on cognitive, emotional, and motivational correlates of physiological activation in an achievement context. These considerations around the level of analysis and psychological correlates of physiological states set the stage for the three guiding themes of the present work, which are (1) the level of analysis, (2) the meaning of physiological states, and (3) possible differences between indicators. Chapter 4 details how the quantitative data for the present work have been collected, followed by Chapter 5, which provides an overview of the three empirical works and contextualizes them within the previously derived research questions. The subsequent Chapters 6, 7, and 8 present these three empirical works. Finally, Chapter 9 provides a general discussion that synthesizes the findings from the three empirical works into five central insights relevant to the previously derived open questions about the use of physiological measures in achievement situations. The chapter critically discusses the limitations of the present work before concluding with implications for research and practice informed by the insights gained. Altogether, the present work aims to make a meaningful contribution to our understanding of the use of physiological measures in educational psychology research and practice.

---

# 2

---

## **USING MEASURES OF THE AUTONOMIC NERVOUS SYSTEM IN EDUCATIONAL PSYCHOLOGY**

---

Miriam Wunsch



Educational psychology, which is concerned with the broad areas of learning and teaching, follows two main goals: To “enhance theoretical knowledge” and to “improve educational practice” (O’Donnell et al., 2025, p. 5). These goals are not independent of one another, but instead, improvements in educational practice rely on a thorough understanding of the processes involved in learning and teaching, such as emotion and motivation. Achieving such an understanding and evaluating attempts to improve educational practice requires the availability of valid and reliable research methods to capture psychological processes in learners and teachers. The first part of this chapter outlines how physiological measures might enhance the assessment of these processes by overcoming the limitations of self-report. Building on the potential of physiological measures, the second part of the chapter focuses on measures of the autonomic nervous system, specifically electrodermal and cardiovascular activity, explaining their biological foundations and measurement methods. Altogether, this chapter aims to provide context for the present work by highlighting the relevance of exploring the use of physiological measures in educational psychology and providing foundational knowledge on the specific measures used.

### **2.1 Benefits of Physiological over Self-Report Measures**

In educational psychology, research to date has commonly relied on self-report measures to obtain insights into an individual’s mind and body (Pekrun, 2023b). In self-report measures, participants are prompted to either directly or indirectly report on a construct of interest about themselves (Paulhus & Vazire, 2007). There are various verbal or written forms of self-report, such as questionnaires, interviews, or think-aloud techniques (Pekrun, 2020; Quinio & Lam, 2021). Self-report is popular among researchers as an efficient and scalable way of collecting data that can provide nuanced insights into all constructs and processes that are consciously accessible to the individual, such as their cognition, behaviour, affect, bodily sensations, etc. (Pekrun, 2020). However, self-report is subject to well-known limitations, such as its reliance on an individual’s accurate perception and reporting of their own states, and the necessity of directing the individual’s attention away from the situation and onto the constructs of interest (Johar & Sackett, 2018; Paulhus & Vazire, 2007; Quinio & Lam, 2021). Physiological measures have the potential to overcome these limitations, as they provide an unbiased measurement, and many can be applied unobtrusively in real-time during learning or testing situations (Martin et al., 2023; Martin & Burns, 2023; Pekrun, 2023b).

### 2.1.1 Unbiased Insights

The first central advantage of physiological measures over self-report is their unbiased assessment. In contrast, self-report measures, such as the Achievement Emotions Questionnaire, rely on an individual to be able to consciously access and correctly report on a construct of interest (“I get bored.”), or indicators thereof (“While working, it seems like this task will never be over.”; Pekrun et al., 2011, 2023). Firstly, the validity of a self-report measure can diminish if an individual struggles to access information about their internal processes in a sufficiently nuanced manner (Paulhus & Vazire, 2007; Pekrun, 2023b). For example, individuals have been found to differ in their emotional differentiation, meaning the extent to which they can distinguish between similar emotional experiences (Kashdan et al., 2015): When asking someone with low emotional differentiation about the extent to which they feel angry and frustrated, they might struggle to provide accurate information on these two questions, as they might not be able to distinguish between these different emotions. Secondly, even if the processes of interest were recognized sufficiently by the individual, several distortions, i.e., biases, can occur in the reporting process. Biases can occur as response styles, meaning an individual’s general tendency to answer in a certain way regardless of the content, like the tendency to agree with statements (*acquiescence response style*) or the tendency to select only extreme or only moderate values on a scale (*extreme/moderacy response style*). Biases can also occur as response sets, in which case they are only specific to the particular situation or content, but cause a distorted assessment nonetheless (Paulhus & Vazire, 2007; Ravazi, 2001). For example, when participants lack the motivation to adequately engage in a given assessment, they may exhibit *satisficing* behaviour, meaning they would provide plausible, yet inaccurate responses due to a lack of effort necessary to provide accurate information (Barge & Gehlbach, 2012; Quinio & Lam, 2021). One example of these mental “shortcuts to conserve effort” (Barge & Gehlbach, 2012, p. 182) that respondents with a lack of motivation might take is *anchoring heuristics*. In that case, the answer given to an initial item serves as an anchor for responding in a similar way to either similar or any subsequent items, even if the accurate response would deviate (Gehlbach & Barge, 2012). Another well-known bias distorting the accurate reporting of internal states that can either occur as a response style or as a response set is *socially desirable responding*, where individuals either unconsciously (self-deception) or consciously (impression management) respond to self-report in a way that seems desirable than giving the accurate response (Paulhus & Vazire, 2007; Quinio & Lam, 2021). Altogether, this exemplary outline of biases in self-report measures highlights possible issues originating from their reliance on an individual being aware of and accurately reporting their internal processes.

Physiological measures can address these biases, as they do not require the individual to report any information themselves. Hence, their measurement is independent of the individuals' ability to consciously access certain information, independent of verbal cues and response scales, and independent of a person's commitment to the assessment. Further, while physiological states can be influenced intentionally, such as through breathing (Laborde et al., 2022; Van Diest et al., 2014), it appears rather unlikely that individuals would consciously apply such techniques to manipulate an assessment: Firstly, with physiological responses, it does not seem obvious what (socially) desirable states would be, and secondly, it would require knowledge of and experience with these techniques to be able to apply them in parallel to an achievement task. Hence, compared to self-report measures, physiological measures can offer a more objective, unbiased insight into an individual's state during an achievement situation (Martin et al., 2023; Pekrun, 2023b).

### ***2.1.2 Unobtrusive Real-Time Measurement***

The second central advantage of many physiological measures over self-report measures is the possibility of applying them continuously and unobtrusively. Often, self-report measures can only be employed at specific moments in time, such as questionnaires or interviews, in which individuals are asked to report on concurrent or past experiences. Reporting retrospectively about past experiences bears the risk of *memory* and *hindsight bias*, meaning that the accuracy of responses is distorted either by over- or understating an earlier state or experience (“memory-experience gap”, e.g., Anvari et al., 2024), or by the knowledge that has been gained between the experience and the assessment (Quinio & Lam, 2021). For instance, an individual may retrospectively – in hindsight – report more intense enjoyment and less intense anxiety than they actually experienced during the test, once they have learned that the test outcome was successful.

Some self-report measurement tools are designed in a way that they can be administered in parallel to a situation, such as think-aloud techniques (e.g., Solomon, 1995), or the use of a joystick to provide continuous real-time reports (e.g., Krieglmeier & Deutsch, 2010; Lizdek et al., 2012). While such a concurrent assessment minimizes the risk of memory or hindsight biases, these self-report measures, even when attempting to be minimally interruptive, still require the individual to allocate some attention to the recognition and reporting of internal processes. Directing attention to certain experiences has been found to impact these experiences, such as faster declines in negative emotions when they are reported repeatedly (Johar & Sackett, 2018). This “disturbance in a system due to measurement” (p. 293) is termed the *observer effect* (Johar & Sackett, 2018) and limits the generalizability of findings obtained

using self-report measures to situations without self-report assessments. Furthermore, if applied in a practical learning or assessment context, repeated or continuous self-reports might draw attention away from the task, with these interruptions possibly impairing learning or performance (Chen et al., 2024; Eatchel et al., 2012; Foroughi et al., 2015).

Many physiological measures can mitigate these issues, as they can continuously record data in real-time, without diverting the individual's attention from the situation or task toward the construct of interest. Hence, they can avoid biases associated with retrospective reporting and interruptions of the situation. Arguably, observer effects might still be possible, as, due to ethical and practical considerations, the individual will be aware that certain physiological measurements are taking place, possibly eliciting anxiety, curiosity, or annoyance and drawing some attention away from the task. However, compared to self-report measures, the application of physiological measures still appears less likely to impact the psychological processes of interest, as the individual will not be exposed to repeated verbal cues that draw attention to these processes. Nevertheless, it seems advisable to employ subtle measurement tools where possible (e.g., screen-based eye tracking instead of glasses) and to incorporate habituation phases to minimize possible remaining observer effects.

Altogether, in comparison to self-report assessments, physiological measures can be applied continuously in real-time while the situation or task of interest is taking place, with minimal to no interruption of the situation and distortion of psychological processes. Hence, their use promises to enhance the validity, reliability, and generalizability of research findings, and to facilitate practical applications that capture emotional and motivational states in an achievement situation in real-time without compromising learning or performance outcomes.

### **2.2 Physiological Measures of the Autonomic Nervous System**

Having established the potential benefits of using physiological measures in educational psychology, the next question is which measures to use. Several methods have been successfully employed in educational psychology research, including the measurement of hormone levels, brain activation, and activation in the autonomic nervous system (Pekrun, 2023b). Additionally, the role of physical activity (e.g., exercise) has been explored; however, it appears to be more often examined as a predictor of educational and psychological outcomes, rather than as an indicator of underlying psychological processes (e.g., Heemskerk et al., 2023).

Hormonal changes have been used to measure stress levels in test-takers and teachers (e.g., Graham et al., 2023; Jögi et al., 2023; Ringeisen et al., 2019), as certain biomarkers, particularly cortisol, are indicative of a physiological stress response (Špiljak et al., 2022). While assessing such biomarkers of stress can provide the desired unbiased insights into an

individual's state, these biomarkers are commonly measured in individuals' saliva, requiring repeated sampling and analysis, which does not allow for an unobtrusive real-time measurement. Brain activation, instead, can be measured continuously throughout a situation without requiring the individual's attention and provides an unbiased insight into their physiological state. In the domain of educational neuroscience, numerous patterns relating brain activation to learning-relevant processes have been explored, using measures such as electroencephalography, magnetic resonance imaging, or functional near-infrared spectroscopy (Feiler & Stabio, 2018; Luk & Christodoulou, 2024). For example, Wei and Jin (2024) have used EEG to explore the effects of test anxiety on attentional control and Takeuchi et al. (2014) employed MRI to investigate the neurological basis of achievement motivation. While such neuroimaging techniques can provide meaningful, continuous insights into physiological processes in the brain, their application is often expensive, requires a rather intricate and potentially obtrusive setup (e.g., an EEG cap with cables), and can be bound to a specific location (e.g., an MRI scanner). Measures of the autonomic nervous system are free from the disadvantages of hormonal markers and brain imaging techniques, as they can be applied continuously, are comparatively affordable, and easy to use in various setups and locations (Ishikawa, 2023).

The autonomic nervous system (ANS) is part of the peripheral nervous system and “regulates involuntary physiological processes including heart rate, blood pressure, respiration, digestion, and sexual arousal” (Waxenbaum et al., 2023, Introduction section). This implies that, unlike functions in the somatomotor nervous system, we cannot directly control processes driven by the ANS (Wehrwein et al., 2016). The ANS has important functions, which Levenson (2014) summarized as activation, coordination, and communication: In its activation function, the ANS sets free physiological resources to promote our ability to successfully deal with occurring events. In its coordinating function, the ANS is a way of a bi-directional exchange of information between the brain and the body, with both afferent (i.e., body to brain) and efferent (i.e., brain to body) pathways (Harrison et al., 2013; Waxenbaum et al., 2023). Finally, the ANS serves a communication function, as certain changes in the ANS are visible to the outside, conveying information about an individual's state to others (Levenson, 2014).

Following the distinction introduced by Langley in 1900, the ANS is commonly divided into three distinct branches, based on their anatomical and functional differences (Harrison et al., 2013; Wehrwein et al., 2016). The first, probably least known, subsystem is the enteric nervous system, which is responsible for digestive processes (Waxenbaum et al., 2023). Most literature on the autonomic nervous system (e.g., Hamill et al., 2012) focuses on the other two,

more well-known ANS branches, being the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). Activation in the SNS has been associated with a state of activity (*fight or flight*) in the body, whereas activation in the PNS is associated with a state of relaxation (*rest and digest*; Waxenbaum et al., 2023). Despite their seemingly contrasting functions, activation in the SNS and PNS is independent of one another, leading to different activation patterns being possible in the autonomic space. In coactivation, both the SNS and PNS exhibit increased activation, whereas in coinhibition, activation in both systems decreases. Furthermore, there can be uncoupled (para)sympathetic activation/withdrawal, meaning that activation in one system increases or decreases while the other remains constant, or reciprocal (para)sympathetic activation/withdrawal, meaning that activation increases in one system and decreases in the other (Berntson et al., 1994). As activation in the SNS and PNS can show these different patterns, it is beneficial to measure activation in both branches when aiming to gain a comprehensive picture of an individual's ANS activation.

All organs are innervated by either the SNS, the PNS, or both (Kreibig & Gendolla, 2014). The SNS is the larger of the two systems, as it “innervates nearly every living tissue in the body” (Introduction section), whereas the PNS only innervates certain body regions (Waxenbaum et al., 2023). In those organs that are innervated by both the SNS and the PNS, their influences can either be antagonistic, complementary, synergistic, or independent of one another. When working antagonistically, as in the heart, activation in the SNS and PNS have opposite effects; for example, SNS activation accelerates the heart rate, while PNS activation decelerates it. When the two systems work complementarily, activation in each system contributes to the overall function, such as in salivary secretion, where SNS and PNS activation both contribute to the production of different types of saliva. In synergistic functioning, SNS and PNS complement each other in a coordinated manner, for example, in pupil dilation, where the SNS initiates pupil dilation, and the PNS subsequently modulates pupil contraction (Wehrwein et al., 2016). Given their distinct patterns of innervation in different body regions and unique effects on these organs, different physiological functions can be measured to capture different facets of ANS activation.

According to Ishikawa (2023), the most commonly employed measures of the ANS in psychological studies are those of electrodermal activity, cardiovascular activity, pupil dilation, and blink rate. In their overview of studies, it appears that pupil dilation and blink rate have mainly been used in the context of cognitive variables, such as attention, cognitive load, and information processing. Electrodermal and cardiovascular activity have additionally been

investigated in several studies related to emotions and stress (Ishikawa, 2023), which is why these measures are the focus of the present work.

### ***2.2.1 Electrodermal Activity***

Interest in electrodermal activity (EDA) as a psychophysiological measure emerged in the 1880s, and it has since become one of the most widely employed measures in psychophysiological research (Boucsein, 2012). Other terms have been used to describe EDA, such as galvanic skin response or psychogalvanic reflex; however, the use of the term EDA has been recommended, as “activity” is deemed more fitting to describe the phenomenon than “response” or “reflex” (Boucsein, 2012; Dawson et al., 2007). EDA describes “autonomic changes in the electrical properties of the skin” (Braithwaite et al., 2015, p. 3), which are mediated by the activity of eccrine sweat glands (Boucsein, 2012; Christopoulos et al., 2019). The more these glands fill with sweat, the less resistant the skin is in the conductance of electrical impulses (Dawson et al., 2007). As eccrine sweat glands are only innervated by the SNS, EDA is commonly used as an indicator of sympathetic activation without parasympathetic influences (Ishikawa, 2023; Wehrwein et al., 2016). However, it is worth noting that the influence of the SNS on eccrine sweat glands is mediated by cholinergic transmitters, in contrast to other functions that the SNS influences via adrenergic transmitters. Hence, patterns found for EDA might differ from those for other indicators of SNS activity (Kreibig & Gendolla, 2014).

EDA is typically measured on the palms, as they have a high density of eccrine sweat glands and smooth, hairless skin (Boucsein, 2012; Boucsein et al., 2012; Dawson et al., 2007). It can be measured either endosomatically or exosomatically: The endosomatic measurement does not involve an external electric current, but instead a single electrode to measure changes in the skin’s electric potentials. The exosomatic method involves two electrodes, through which either a direct or an alternating current is applied (Boucsein et al., 2012). The resulting signal indicates how well the current is conducted through the skin, typically measured in microSiemens (Dawson et al., 2007). This overall signal can be further divided into its tonic and phasic components (Braithwaite et al., 2015). The tonic component, also referred to as skin conductance level (SCL), represents the more slowly changing part of the signal, which is thought to reflect “general changes in autonomic arousal” (Braithwaite et al., 2015, p. 4) and usually measured in microSiemens (Boucsein et al., 2012). The phasic component, referred to as the skin conductance response (SCR), instead describes short-term changes that occur on top of the tonic signal (Braithwaite et al., 2015). These phasic changes can occur without an obvious reason (nonspecific; NS-SCR) or can be elicited by certain stimuli (event-related; ER-SCR)

with a latency of one to three (Braithwaite et al., 2015; Dawson et al., 2007) or five seconds (Christopoulos et al., 2019). Since SCR and SCL are driven by different neural mechanisms (Nagai et al., 2004), they appear to be rather independent of one another (Dawson et al., 2007). Even within the phasic signal, where features such as frequency, amplitude, rise time, or latency, can be extracted (Boucsein, 2012), these different features are only moderately correlated, suggesting that they may reflect different functions (Dawson et al., 2007), while little is known about the nature of these different functions.

There are some inconsistencies in the literature regarding the exact definition and measurement of SCL: In Dawson et al. (2007) for example, it seems that SCL was defined as the overall EDA signal, whereas Braithwaite et al. (2015) and Boucsein (2012) defined SCL as the remaining signal when removing the phasic components. Furthermore, Boucsein et al. (2012) provided a brief introduction into the debate of whether NS-SCR are a measure of tonic activity: As they are not elicited by certain stimuli, they might more of a tonic indicator of general autonomic arousal. Braithwaite et al. (2015) instead clearly assigned them to the phasic signal. Possibly, the assignment of NS-SCR to the phasic or the tonic signal depends on the study context: When investigating responses to certain stimuli, ER-SCR are the main focus and NS-SCR are part of the tonic “background music”. Instead, when interested in autonomic arousal over a longer-lasting situation, such as an achievement task, without clearly defined, response-eliciting stimuli, it appears impossible to distinguish between ER-SCR and NS-SCR. In these cases, it seems reasonable to consider all SCR as the phasic outcome of interest (as done for example in Kiuru, Malmberg, et al., 2022; Roos et al., 2023).

Altogether, differences in the meaning between SCR and SCL, as well as between the various indicators of SCR are largely underexplored, limiting the ability to make an evidence-based decision on what measure to use. In the present work, we therefore employ both measures and inspect their differences, operationalizing SCR as the frequency of peaks in the phasic signal, in line with similar work (Kiuru, Malmberg, et al., 2022; Roos et al., 2023), and SCL as the tonic EDA signal in the absence of phasic changes (Boucsein, 2012; Braithwaite et al., 2015).

Interestingly, it has been found that certain individuals are hyporesponsive in their EDA, meaning that they show hardly any or no changes in their EDA related to external occurrences. This has been found to be particularly prevalent for persons with certain mental health problems, such as schizophrenia or depression, but also in healthy individuals (Dawson et al., 2007; Ikezawa et al., 2012; Venables & Mitchell, 1996). To deal with these cases, some studies using EDA have implemented procedures to identify and exclude non-responders, such as if a

person would not show any SCR to a certain number of stimuli (e.g., Zimmer & Richter, 2023). For the present research, we employed a validation protocol, which is described in Section 4.5.2, to identify and exclude non-responders.

### ***2.2.2 Cardiovascular Activity***

Cardiovascular activity (CVA) refers to processes related to our heart, which on average beats around 100,000 times every day to keep the blood circulating in our body. Each of these cardiac cycles consists of a systole, that is, the ventricular contraction, and a diastole, that is, the ventricular relaxation (Shaffer et al., 2014). The heart is innervated by both the SNS and the PNS, with the SNS accelerating and the PNS decelerating its activity (Wehrwein et al., 2016). The influence of the PNS is almost immediate, whereas SNS-driven changes can have a latency of up to five seconds (Shaffer et al., 2014).

A number of different features around CVA can and have been extracted in achievement contexts, such as heart rate (HR), heart rate variability (HRV), blood pressure, or pre-ejection period (e.g., Fairclough & Ewing, 2017; Kreibig et al., 2012; Roos et al., 2021). The present work focusses on HR and HRV. HR refers to the number of heart beats per minute, and changes thereof can be detected in a few seconds (Ishikawa, 2023). As HR is subject to antagonistic influences of the SNS and the PNS, changes in HR could be driven by either an increase in activation in one or a decrease in activation in the other system or both (Kreibig & Gendolla, 2014). For the same reason, an absence of changes in HR does not indicate an absence of changes in ANS activation, but instead, states of coactivation or coinhibition of the two ANS branches would also result in an overall unchanged HR (Weissman & Mendes, 2021). Therefore, HR does not allow for inferences on SNS or PNS activity specifically, but instead shows the “net effect” (p. 2) between SNS and PNS influences on the heart (Shaffer et al., 2014)

Another commonly measured feature of CVA is HRV, which describes variations in the time intervals between heartbeats (Ishikawa, 2023; Laborde et al., 2017; Lu et al., 2009; Quintana & Heathers, 2014) and is majorly driven by parasympathetic influences (Chapleau & Sabharwal, 2011). Parameters of HRV can be either frequency domain or time domain metrics: Frequency domain metrics decompose the overall signal into different frequency bands, with certain frequencies being associated with certain physiological processes (Ishikawa, 2023). For example, high frequency bands between 0.14 and 0.40 Hz are linked to parasympathetic influences, and low frequency bands between 0.04 and 0.15 Hz are linked with the baroreflex, which regulates blood pressure (Shaffer et al., 2014). Time domain metrics instead are calculated based on the variance in the intervals between heartbeats (Ishikawa, 2023). Common time domain metrics of HR are the SDNN, quantifying the standard deviation of the interbeat

interval, the pNN50, identifying the percentage of successive intervals that differed from each other by more than 50ms, or the RMSSD, referring to the root mean square of successive differences between heart beats (Shaffer et al., 2014). Although these measures are largely correlated, RMSSD is often preferred (Kleiger et al., 2005; Shaffer et al., 2014) and has therefore been selected for the present work. It can reliably be calculated from recordings as short as one minute (Laborde et al., 2017), with higher values on the RMSSD, indicating higher HRV, indicating higher levels of activation in the PNS (Kleiger et al., 2005).

Besides electrocardiography, in which several electrodes are attached to the individual's torso, photoplethysmography (PPG) poses a good way to measure CVA (Shaffer et al., 2014). The use of PPG is popular due to its ease of application and cost-effectiveness (Allen et al., 2021), requiring only a single sensor that can be attached to the individual's finger or earlobe (Lu et al., 2009). This sensor works by emitting near-infrared light and capturing its reflection from the capillaries. That way, it detects changes in blood volume driven by heartbeats (Allen et al., 2021; Betz et al., 2021). While electrocardiography had commonly been considered the better measurement tool for HRV in particular, this notion has changed with advances in PPG technology (Shaffer et al., 2014). In fact, Lu et al. (2009) found HRV measures derived from PPG to be almost identical to those obtained via electrocardiography. Notably, those results were obtained using a PPG sensor attached to the earlobe, whereas measurement at the fingertip might be more susceptible to artifacts in the signal caused by hand movements (Lu et al., 2009). Hence, in the present work, using a PPG sensor, we ensured that we minimized the risk of movement artifacts. Additionally, we utilized specialized HRV analysis software to obtain valid and reliable HR and HRV values from the PPG signal.

---

# 3

---

## **PSYCHOLOGICAL CORRELATES OF AUTONOMIC MEASURES IN THE EDUCATIONAL CONTEXT**

---

Miriam Wunsch



Having provided an overview of the relevance of the present work, as well as different measures of the ANS, with a focus on EDA and CVA, this next chapter specifically outlines the theoretical background regarding the central aim of the present work: To advance our understanding of the meaning and application of physiological states in achievement situations. The first part of this chapter introduces an important methodological distinction to be considered when exploring associations between physiological and psychological variables, namely the distinction between associations at the between- and the within-person level. Distinguishing between the two levels of analysis is crucial in drawing the right conclusions from findings and in gaining a comprehensive understanding of relationships between variables (Murayama et al., 2017). The second part of this chapter dives deeper into previous findings on associations between ANS measures and psychological variables. In their summary of psychological correlates of ANS measures, Mendes (2016) categorized the psychological variables into four categories: cognitive, emotional, motivational, and stress-related constructs. The following section adopts this distinction, but omits stress-related variables, given that stress often seems to already denote a certain bodily response to a threat to homeostasis (Del Giudice et al., 2018; Goldstein & Kopin, 2007) rather than a possible psychological correlate of physiological states.

#### **3.1 Level of Analysis**

All measures we take from a person, whether self-reported, physiological, or using any other modality, are potentially influenced by two sources of variance: between-person and within-person variation (Schuurman, 2023). Between-person variation refers to time-invariant differences between individuals, such as some individuals generally having a lower HR than others, or some individuals generally being more anxious than others. In addition, for many variables, there is also within-person variation, meaning that values of the same person vary over time (Hoffman, 2015; Murayama et al., 2017). This variation can either be driven by a systematic change over time, e.g., anxiety increasing over the course of a school year, or by fluctuations around a constant mean, e.g., HR being higher or lower in some situations than in others (Hoffman, 2015). When assessing individuals, both cross-sectionally and longitudinally, each unique measurement is a blend of these two different sources of variance (Schuurman, 2023).

Just like the variance of variables, covariances between variables also occur at either or both the between- and the within-person level. At the between-person level, covariances indicate how differences between individuals on one variable relate to differences between individuals on another variable (Murayama et al., 2017); thus, they refer to the question of

“Who?” (Frenzel et al., 2020). For example, when investigating the association of HR and anxiety, the between-person level covariance would indicate whether general differences in HR between individuals relate to differences in anxiety between these individuals (*Who* has a higher HR? – Those *who* are more anxious?). At the within-person level, covariances indicate how differences between different timepoints within the same individual on one variable relate to differences between these timepoints on another variable (Murayama et al., 2017); thus, they refer to the question of “When?” (Frenzel et al., 2020). For example, at the within-person level, an association of HR with anxiety would indicate whether a person showing higher or lower levels of HR in a given moment compared to another timepoint, also experiences higher or lower levels of anxiety in that moment compared to another timepoint (*When* do they have a higher HR? – *When* they are more anxious?). Similar to the measurement itself, cross-sectional covariances are a blend of covariances at both the between- and within-person levels (Schuurman, 2023).

Covariances between variables at different levels are independent of and can may differ from one another, as they can be driven by distinct phenomena (Hoffman, 2015). Murayama et al. (2017) illustrate this with the example of reaction time and accuracy on a task. At the between-person level, there may be a negative relationship between the variables, as individuals with higher task-relevant abilities might be both faster and more accurate in their task execution. In contrast, their within-person relationship is likely to be positive: if reaction time increases, i.e., an individual works more slowly than they usually do, they may achieve higher accuracy than they would otherwise. Instead, if they work faster than usual, their accuracy might suffer. This example illustrates that between-person and within-person relationships can differ substantially from one another.

Only when very strict conditions are met will the results be ergodic, that is, equivalent at the between- and within-person levels. The first condition, homogeneity, means that the same model applies to every person, for example, in terms of mean values and variance of within-person fluctuation. The second condition is stationarity, meaning that parameters, such as means or variances, remain constant over time (Molenaar & Campbell, 2009; Schuurman, 2023). These conditions for ergodicity are rarely met in psychological research, as common phenomena such as autoregressions, group differences, and mean or cyclic trends in variables violate the strict assumptions, leading to non-equivalence of between- and within-person associations (Voelkle et al., 2014). Hence, although often done by researchers, it is not possible to draw inferences on within-person associations from between-person findings and vice versa. Instead, interpretations can and should be confined to the level that a study’s design analysis

actually address (Murayama et al., 2017; Pekrun, 2023b; Schuurman, 2023; Voelkle et al., 2014).

No level of analysis is inherently superior to the other. If a theory specifies the level it pertains to, this level can be the focus of research and practice grounded in that theory. Alternatively, instead of focusing on one analysis level, inspecting both between- and within-person associations can provide complementary insights and contribute to a more comprehensive understanding of relations between variables (Zawadzki et al., 2017). Psychological theories commonly make assumptions on the within-person level (Murayama et al., 2017), specifically those on emotion and motivation (Pekrun & Marsh, 2022), and those describing the function of physiological processes in psychological constructs (Pekrun, 2023b). For example, it has been argued that theories on the associations between physiological states and emotional experience primarily refer to within-person level associations (Mauss et al., 2005, 2024). However, while these theoretical considerations emphasize a focus on the within-person level, they do not clearly state whether associations at the between-person level are considered implausible, irrelevant, or simply more difficult to detect.

When investigating associations between physiological and psychological variables at the between-person level, a central question is how to account for the large physiological differences between individuals that are due to overarching biological factors unrelated to the achievement context. A pure between-person approach would simply relate differences between individuals in their physiological parameters to differences in their emotional experiences in the achievement situation (e.g., “Does someone with a higher HR experience more anxiety during the test than someone with a lower HR?”). However, this relation would severely lack construct symmetry (Pekrun, 2023b): While a physiological variable like HR is highly person-specific across contexts, due to a multitude of general biological factors, anxiety during a test is specific to that particular situation. The large between-person differences in HR unrelated to the achievement situation might mask or confound its associations with the much more situation-specific differences in anxiety during a test (Mauss et al., 2024). Therefore, substantial associations between physiological and situation-specific psychological variables seem unlikely to be detected using such a purely between-person approach. Instead, in the present theoretical and research context, between-person perspectives seem to refer to the question of whether differences in the intensity of the physiological response to a task relate to differences in psychological variables between individuals. For example, instead of comparing individuals’ observed HR during the test, one might ask whether the person whose HR increased more during the test, compared to their normal HR, experienced more anxiety than the person whose

HR barely differed from their normal HR while working on the test. To quantify the intensity of a physiological response to a task by controlling for pre-existing differences in physiological parameters, studies typically include a baseline phase or use a first task or practice block as the baseline (Horvers et al., 2021). While this approach is still between-person in the sense that it relates differences between individuals on one variable (i.e., intensity of response to a task) to differences between individuals on another variable (e.g., emotional experience), the physiological variable is operationalized as a within-person response. However, there is no clear consensus in the literature on how to perform such a correction for baseline differences (Horvers et al., 2021). Meanwhile, for within-person approaches, general differences between individuals in physiological parameters do not pose an issue, as within-person analyses focus solely on the relation between values of the same individual.

Repeated-measures data allow for the investigation of both between- and within-person level associations, and are commonly analysed using either structural equation modelling, particularly when involving latent variables, or multilevel linear modelling, when investigating observed variables (Hoffman, 2015). For the present work, the focus is on observed variables, hence, on multilevel regression models. Multilevel regression models, also known as hierarchical models or mixed-effects models (Hoffman & Walters, 2022), separate between- and within-person variance (Schuurman, 2023) by accounting for different sources of dependencies that are present in repeated-measures data (Hoffman, 2015). Firstly, one source of dependency can be differences in mean levels between individuals, making the data points of one individual more similar to each other than to those of other individuals. For example, one person could have a generally higher test anxiety than another person, making all their test anxiety measurements higher than the other person's. Multilevel regression models can account for this source of dependency through random intercepts, allowing for variance in the intercepts of individuals' regression lines (Hoffman, 2015). Secondly, associations between variables may also differ between individuals. For example, test anxiety might strongly affect performance for some individuals, but not for others. Multilevel regression models can account for these differences through random slopes, that is, by allowing for variance in within-person associations or effects (Hoffman, 2015). In the present work, within-person associations between variables are investigated using random intercepts and fixed slopes. This approach is adequate given the number of participants and measurement time points, while providing important insights into the presence and direction of associations between variables at the within-person level.

### **3.2 Findings on Associations between Autonomic States and Psychological Constructs in the Achievement Context**

Following this central methodological consideration, the second part of this chapter provides an overview of previous findings on associations between physiological measures of activation in the ANS and the psychological constructs of cognition, motivation, and emotion in an achievement context. While cognition, motivation, and emotion are theoretically distinguishable constructs, it is not possible to fully separate them due to their “holistic, integrated nature” (Pekrun, 2023a, p. 229). Hence, the separate considerations of each construct are followed by a synthesis that yields a conceptual model of the relation of the three constructs and the role of physiological activation in it.

#### ***3.2.1 Cognition***

Although the empirical investigations in the present work focus primarily on emotional and motivational states, a brief look into the cognitive correlates of physiological measures is indispensable for a deeper understanding of their meaning in an achievement context. Cognitive processes encompass a range of functions that can be broadly summarized as attention, memory, perception, language, and thinking (Malim, 1994). Research in the field of cognition has investigated associations between these functions and physiological states, typically focusing on cardiovascular measures (Ishikawa, 2023) and using smaller samples (e.g.,  $N = 11$  in Tatsumi et al., 2025;  $N = 10$  in Portnova et al., 2023;  $N = 26$  in Mackersie & Calderon-Moultrie, 2016). Studies can be broadly categorized into two main groups: Some studies, also originating from the achievement context, relate individuals’ task performance to their physiological states. Others, typically from the field of cognition, use variations in task type and/or task demands to examine how they affect physiological states. The following outlines exemplary findings from both approaches to provide insight into the current state of knowledge on the association between cognitive processes and physiological states.

#### ***Physiological States and Performance***

Studies relating task performance to physiological states in achievement situations have typically approached this association on the between-person level. In Horvers et al.’s (2021) review of the correlates of EDA in achievement settings, most included studies found a positive association between measures of EDA and different performance outcomes, such as exam grades or learning gains. Similarly, Harley et al. (2019) found a positive correlation between diagnostic performance and SCR (but not SCL). Other findings contrast such a positive association between sympathetic activation and performance, like results by Huber and Bannert (2023) showing that SCL negatively predicted learning performance, whereas HR was not

associated with learning performance. Furthermore, most studies used in the review by Forte et al. (2019) found that higher HRV was associated with better performance in various cognitive domains, such as memory, attention, language, or processing speed. Hence, previous findings seem somewhat contradictory, with some suggesting better cognitive performance in more activated (higher SNS activation), and others in more relaxed states (higher PNS activation).

These inconsistencies might at least partly be due to between-person results being “smushed effects” (Hoffman & Walters, 2022, p. 667) of both between- and within-person mechanisms, which can differ from one another. This possible difference is demonstrated in a study by Duschek et al. (2009): They found that individuals with higher respiratory sinus arrhythmia, which is linked to HRV and indicates higher PNS activation, made fewer mistakes on an attention task. Conversely, when examining the change from baseline to task, those individuals who showed a larger decrease in PNS activation between baseline and task made fewer mistakes. Duschek et al. (2009) interpreted this pattern based on previous findings showing that higher resting HRV is associated with a higher potential for cardiovascular reactivity. Thus, Duschek et al.'s (2009) findings are a prime example that associations can differ on the between- and within-person level: Those individuals with a generally high HRV might be able to show a better performance on cognitive tasks (positive association on the between-person level), because a larger decrease in HRV facilitates better performance (negative association on the within-person level). Altogether, it appears that higher sympathetic and lower parasympathetic activation might be associated with better performance on achievement tasks. However, findings do not convey a clear picture, possibly due to a lack of sufficient separation of between- and within-person associations.

#### ***Physiological States and Task Demands***

Studies relating physiological states to the demands of cognitive tasks further support the notion that higher sympathetic and lower parasympathetic are associated with enhanced cognitive processing, in particular on the within-person level. For example, in a study by Chang and Huang (2012), participants worked on three tasks with different attentional demands in counterbalanced order. While individuals' SCL did not differ significantly between the tasks, the cardiovascular measures showed that the more attention-demanding the task was, the higher the individual's sympathetic activation and the lower their parasympathetic activation (Chang & Huang, 2012). Similarly, in Luque-Casado et al. (2016), participants worked on three tasks with different levels of cognitive demands, namely a psychomotor vigilance (lowest demands), a duration discrimination (medium demands), and an n-back task (highest demands). Again, they found lower levels of HRV, i.e., decreased parasympathetic activation, the more

cognitively demanding the task was. The same pattern of increased SNS and decreased PNS activation with increasing task difficulty was observed for a speech repetition (Mackersie & Calderon-Moultrie, 2016), as well as a working memory task (Mandrick et al., 2016). Altogether, findings consistently show higher levels of sympathetic and lower levels of parasympathetic activation within a person the more cognitively demanding a task was.

Results by Portnova et al. (2023) extend on these findings by demonstrating that physiological activation is only positively related to cognitive task demands up to a certain point of task difficulty. In their study, participants were working on six blocks of a change detection task, with increasing levels of difficulty from the first to the sixth block. Over the first three blocks, increases in task difficulty were accompanied by increased sympathetic and decreased parasympathetic activation, in line with the previously reported findings. Then, however, this pattern largely reversed from the third to the sixth block, with sympathetic activation decreasing and parasympathetic activation increasing as task difficulty further increased. This pattern suggests that changes in physiological arousal are not directly driven by changes in the cognitive demands of a task, but another influential factor must play a role in this relationship. This factor might be the motivational construct of (mental) effort, which determines how many of one's limited cognitive resources are allocated to a given task (Silvestrini & Gendolla, 2019) and has been considered as a mediator between task demands and the involvement of cognitive processes (Shenhav et al., 2017). Although influencing cognition, (mental) effort and its possible association with physiological states will be discussed in the following section as a motivational variable.

#### ***Summary***

Previous findings on the association between physiological states and cognitive processes suggest that within an individual, a higher involvement of cognitive processes, facilitating better task performance, is accompanied by higher sympathetic and lower parasympathetic activation. While the cognitive demands of a task are one aspect that determines the involvement of cognitive processes and the accompanying physiological arousal, motivational variables may play a central role in this association.

#### ***3.2.2 Motivation***

Although definitions of motivation vary between research fields (Braver et al., 2014), their core can be described as “mental processes that shape the goal direction, intensity, and persistence of behavior” (Pekrun, 2023a, p. 227). Accordingly, motivation has a directional and an activation function (Salamone & Correa, 2024), meaning that it both guides the individual what goal to strive for (choice) and how intensely to do so (effort; Silvestrini & Gendolla, 2019).

Both functions have received attention in educational psychology, with different theories focusing on one aspect or the other.

#### ***Directional Function***

The Situated Expectancy x Value Theory (EVT; Eccles & Wigfield, 2020; Wigfield & Eccles, 2000) of achievement motivation provides a theoretical framework to understand an individual's decision to direct their behaviour towards the engagement with an achievement task. According to EVT, achievement motivation is driven by perceptions of expectancy and value: Expectancy refers to the individual's perception of being able to successfully master the task, whereas value refers to the perceived benefit or importance of doing well on the task. Perceived value can have different sources that can add to each other, such as intrinsic value, i.e., enjoying the task, attainment value, i.e., personal importance of the task, utility value, i.e., usefulness of successful task completion for a higher goal, and relative costs, i.e., losses through engaging with the task that diminish overall value (Eccles & Wigfield, 2020). Expectancies and value interact in predicting the decision to engage with a task or subject, as well as intentions for future choices relating to it. Specifically, both expectancy and value are assumed to have a positive effect on achievement motivation, given that the other variable is non-zero. The effect of each variable is amplified by higher values of the other variable; that means that the higher the perceived value, the more strongly expectancies drive achievement motivation, and vice versa (Nagengast et al., 2011).

In line with the view that one function of autonomic activation is to facilitate responses to situational demands (Levenson, 2014), the decision to engage with an achievement task might be accompanied by increased sympathetic and decreased parasympathetic activation. This should be the case regardless of whether this motivation is driven by the perception of the task as a challenge, aiming to achieve success, or as a threat, aiming to avoid failure (Chalabaev et al., 2009). In a between-person study by Seery et al. (2009), participants were assigned to one of three conditions – they could either gain money through successful task performance (challenge), lose money through unsuccessful performance (threat), or received no incentives (control). Based on the EVT, in both the challenge and threat conditions, participants' perceived value, and thereby their achievement motivation, should be higher than in the control condition without any consequences (Eccles & Wigfield, 2020). Seery et al. (2009) indeed found higher levels of sympathetic arousal, indicated by HR and pre-ejection period in both the threat and challenge conditions compared to the control condition, suggesting that higher levels of achievement motivation were accompanied by higher physiological activation. Similarly, in a study with high-school students, Kiuru, Malmberg, et al. (2022) reported an interaction at the

between-person level between perceived mastery and value in predicting physiological states: When perceived value was high, mastery positively predicted SCR and negatively predicted HRV, indicating that appraisal combinations eliciting high achievement motivation were indeed accompanied by higher sympathetic and lower parasympathetic activation. However, contrary to the assumptions of the EVT, when perceived value was low, these relationships were inverted, with higher expectancies being accompanied by lower sympathetic and higher parasympathetic activation. No such interaction patterns were found on the within-person level (Kiuru, Malmberg, et al., 2022).

Altogether, to my knowledge, research linking physiological states to the directional function of motivation in achievement contexts remains limited. Some findings suggest that appraisal combinations associated with higher achievement motivation, i.e., the tendency to engage with a task, are accompanied by higher sympathetic and lower parasympathetic arousal at the between-person level. However, more empirical investigations are needed to assess the robustness of this finding, in particular at the within-person level.

#### ***Activation Function***

The activation function of motivation is commonly referred to as effort. Effort describes the intensity of the individual's behaviour in aiming to achieve a goal (Capa & Audiffren, 2009; Inzlicht et al., 2018; Silvestrini & Gendolla, 2019). It serves the important function of managing and conserving our limited cognitive resources (Shenhav et al., 2017): When faced with a task, we do not automatically expend all the cognitive resources we have, as this would quickly deplete them. Instead, effort determines the number of resources mobilized (Gendolla et al., 2019; Silvestrini & Gendolla, 2019). The necessity of effort to achieve success on a task depends on the demands of the task and the extent to which the required actions are automated processes for the individual (Shenhav et al., 2021). If a task is easy and/or highly automated, such as counting from one to ten, little effort is required to complete it successfully. The more a behaviour is difficult and/or not automated, such as solving algebraic equations, the more effort is required to perform the necessary cognitive processes (Shenhav et al., 2017). This also implies that the amount of effort for a successful performance on a given task can differ between individuals, depending on how automatized an action is for them. For example, a mathematics student would likely require less effort to solve a complex equation than someone who does not regularly apply mathematics, as the processes are more automated for them. Some authors have considered effort to be the mediator between potential and actual performance (Shenhav et al., 2017), whereas others do not adopt this view of a mediating role, but instead consider effort as the input into and performance as the output of a task (Silvestrini & Gendolla, 2019). However,

despite these differences in the exact model, the shared idea appears to be that, especially on less automated/more difficult tasks, effort largely influences performance outcomes through the allocation of cognitive resources to the task.

Motivational Intensity Theory (MIT; Brehm & Self, 1989) provides a theoretical framework for the factors influencing the exertion of effort. The MIT builds on the principle of resource conservation, meaning that individuals strive not to mobilize more energy than needed for a given task. Followingly, the amount of mobilized energy, also known as effort, should rise with rising task demands. Additionally, it would be wasteful to invest energy in a task that is impossible to solve, or to expend more energy on a task than its benefits would justify. Hence, when reaching a certain level of task difficulty where either success is impossible, or where the necessary effort is not justified by the perceived value of the task, effort drops to and remains at zero for all higher levels of task difficulty (Brehm & Self, 1989; Gendolla, 2025; Gendolla et al., 2019; Kreibig & Gendolla, 2014; Richter et al., 2016).

In the context of MIT, the application of cardiovascular measures has a long tradition, building on the premise that sympathetic activation indicates effort (Richter et al., 2016). For example, in a study by Richter et al. (2008), individuals were assigned to one of four conditions of a memory task, being either low, moderate, or high difficulty, or a level of difficulty where the task was impossible to solve. In this between-person design, they found support for the assumptions of the MIT, as individuals in the more difficult condition showed higher sympathetic activation than those the easy condition, while in the condition where success was impossible, sympathetic activation was lower. A number of similar studies have been conducted since, strengthening the assumption of a close relationship between effort and physiological activation, in particular, the cardiovascular measure of the sympathetically driven pre-ejection period (Albinet et al., 2024; Gendolla, 2025). Also, the previously mentioned within-person findings by Portnova et al. (2023) showing a positive relation between the cognitive demands of a task and sympathetic arousal only up to a certain point, match the assumptions of the MIT for effort. Other studies explicitly investigated the association between physiological activation and effort by relating self-reported mental effort to physiological states. For example, in a between-person study, Maier et al. (2003) found positive correlations between various measures of cardiovascular activity, including HR, and self-reported task engagement. Similarly, in a within-person approach, Radtke et al. (2025) found a positive association between participants' self-reports of how much effort they invested due to being bored and their EDA (notably, only in the hard, not the easy task). Altogether, both findings from the between- and the within-person level support the assumption that higher levels of effort are accompanied by increased

sympathetic and decreased parasympathetic activation. This “action-ready bodily state” (Radulescu et al., 2015, p. 238) aligns with the purpose of effort being the mobilization of resources to achieve a goal (Gendolla et al., 2019; Silvestrini & Gendolla, 2019).

#### ***Summary***

Altogether, both the directional and the activation function of motivation seem to be associated with changes in physiological states. Specifically, higher physiological activation has been linked to appraisal patterns indicating a greater likelihood of engaging with a task, as well as to the number of resources mobilized to support that engagement. Especially in the context of the MIT, autonomic measures, in particular pre-ejection period, are considered to reflect effort, that is, the resources mobilized for a task (Albinet et al., 2024; Gendolla, 2025).

#### ***3.2.3 Emotion***

Emotions are episodes that are short-lived and relate to an event or situation that is relevant to the individual (Pekrun, 2023a; Shuman & Scherer, 2014). There appears to be a certain consensus among theorists that emotions comprise multiple components (Lange et al., 2020). For example, Mauss et al. (2024) pose that the three systems involved in an emotion are experience, behaviour, and physiology; Lange et al. (2020) list subjective feelings, cognition, action tendencies, expression, and physiology; and Shuman and Scherer (2014) speak of affective, motor, and physiological components, as well as action tendencies, and appraisals. Despite some differences, these sets of presumed components largely overlap: They all involve an affective component, meaning a certain subjective experience or “feeling”, which is considered the defining feature of an emotion (Pekrun, 2023a). Further, they all mention behaviour, either in terms of expression or in the sense of a motivational tendency. And finally, all three of the exemplary definitions include physiology as a component of emotions.

Depending on the context and the subject eliciting the emotion, they can be classified into different categories. In educational settings, typically occurring groups of emotions include achievement, epistemic, and social emotions (Pekrun, 2024a). The present work focuses on achievement emotions, for which the Control Value Theory of achievement emotions (CVT; Pekrun, 2006, 2024a) provides a theoretical framework. Achievement emotions are defined as emotions that relate to “activities or outcomes where performance is judged according to competence-based standards of quality” (Pekrun, 2024a, p. 83). The most typical examples from an educational context are exams or other forms of written or oral assessments, but also doing homework, revising before a test, or being asked a question in class can evoke achievement emotions. According to the CVT, appraisals of control and value determine which emotion is experienced to what intensity (Pekrun, 2006): Control describes the perception of

being able to influence the achievement activity and/or outcomes. Value has both a qualitative side, specifying whether it is pertains to the value of achieving success vs. the value of avoiding failure, and a quantitative side, reflecting the amount of perceived value. Depending on a combination of these appraisals, different emotions arise, such as enjoyment in high control and positive value, anxiety in medium control and negative value, and frustration in low control and either positive or negative value (Pekrun, 2006). Furthermore, the quantitative aspect of value generally intensifies achievement emotions, as well as the effect of control on these emotions (Shao et al., 2020). An exception is boredom, which is particularly high when perceived value is low (Pekrun, 2006). Like emotions generally, achievement emotions are also considered to comprise affective, cognitive, motivational, behavioural, and physiological components (Pekrun, 2006; Pekrun et al., 2023).

#### ***Emotion Coherence***

The consideration of different emotion components leads to a far-reaching, long-discussed issue in the field of psychology: The question of emotion coherence, which Mauss et al. (2005) have defined as “the coordination, or association, of a person’s experiential, behavioral, and physiological responses as the emotion unfolds over time” (p. 175). For example, when assuming coherence in the emotional response components, we might expect that narrowed eyes, clenched fists, and a fast-beating heart always accompany the feeling of anger. However, already in the 1960s and since, emotion researchers have found the links between the different emotion components to be less pronounced than expected, which led to the now common consideration of the different emotion components being “loosely coupled” (Lang, 1988, p. 177, as cited in Mauss et al., 2005). According to the more recent integrated view of emotion coherence by Mauss et al. (2024), which combines previous theories and findings, coherence is not an all-or-nothing characteristic but can instead vary, for example, between different individuals (e.g., Van Doren et al., 2021), or between different combinations of emotion components (e.g., Constantinou et al., 2023; Evers et al., 2014). Similarly, for achievement emotions, their components are also assumed to be “loosely coupled” (Pekrun et al., 2023, p. 146). Achievement emotions containing all components may be viewed as prototypes, while actually the boundaries between the different emotions are rather gradual, and components other than the affective core may also be absent (Pekrun, 2023a; Pekrun et al., 2023).

A particular focus of the present work is on the potential coherence between affective experiences and physiological states. Different emotion theories pose different assumptions about the underlying mechanisms of such links: According to the afferent view, bodily changes

act as an informant of the emotional experience; in the efferent view, the emotional experience elicits physiological processes; and appraisal theories assume both afferent and efferent pathways between the autonomic nervous system and the emotional experience (Kreibig & Gendolla, 2014). Independent of the underlying mechanism, coherence between the subjective experience of emotions and changes in the body might serve an evolutionary function by eliciting a physiological state that facilitates necessary behaviours in the situation (Evers et al., 2014; Levenson, 2014). For example, when facing the stereotypical bear, a racing heart accompanying fear is rather adaptive, allowing us to run faster. Indeed, Kreibig (2010) concluded in their extensive review that previous findings on associations between measures of the autonomic nervous system and the emotional experience are in line with the idea that physiological activation prepares for an adequate behavioural response. However, findings for the same emotion are not always homogenous. For example, approach-oriented anger was found to be activating, whereas withdrawal-oriented anger was observed to be deactivating (Kreibig, 2010; Stemmler et al., 2007), possibly suggesting that physiological states are more closely related to the adaptive response than the emotion itself. It is also worth noting that affective experiences and physiological states are influenced by a multitude of other variables, making a perfect correspondence between them unlikely (Kreibig & Gendolla, 2014). The next section focuses on a potential coherence between physiological states and the affective experience of achievement emotions, which are the specific subtype of emotions central to the present work.

#### ***Physiology and Achievement Emotions***

Achievement emotions can be classified along three dimensions: valence, arousal, and object focus. Valence denotes whether an emotion is positive/pleasant or negative/unpleasant; arousal describes whether the emotion is physiologically activating or deactivating; and object focus specifies whether the emotion refers to an activity or outcome, occurring either in the past (retrospective), present (concurrent), or future (anticipatory). For example, enjoyment would be classified as a positive, activating emotion referring to a concurrent activity; anxiety as a negative, activating emotion referring to a prospective outcome; and relief as a positive, deactivating emotion referring to a retrospective outcome (Pekrun, 2024a). All three dimensions contribute to an individual's functioning through their motivational consequences: Valence and object focus steer the individual in the direction of approach or avoidance behaviour towards a certain target, while activation or deactivation facilitate effort or withdrawal in these behaviours through a conducive physiological state (Pekrun et al., 2023). Thus, also in the achievement

context, the physiological emotion component is assumed to serve the purpose of facilitating adaptive behaviours.

Various studies have explored associations between the affective experience of achievement emotions and physiological states. Thereby, different studies have investigated different sets of achievement emotions, resulting in rather sparse evidence for each emotion (e.g., pride, hopelessness, anxiety, and shame in Harley et al., 2019; boredom, excitement, calmness, and anxiety in Ketonen et al., 2023; enjoyment, hope, fear, hopelessness, and anxiety in Kiuru, Malmberg, et al., 2022).

Anxiety poses an exception, receiving the most interest: For example, Martin et al. (2021) found a positive correlation at the between-person level for trait science anxiety and mean EDA of high school students during a science test. Similarly, Eidlin Levy and Rubinsten (2021) found SCR amplitudes to correlate positively with both math trait and state anxiety on the between-person level in university students during an arithmetic task, and Harley et al. (2019) found a positive between-person correlation between anxiety and the SCL of medical students during a medical simulation task. Also in their meta-analysis, which primarily included between-person studies, Roos et al. (2021) found moderate correlations between test anxiety with SCR and HR. In contrast, Strohmaier et al. (2020) reported no significant between-person correlation between SCR and maths state anxiety, and also Kiuru, Malmberg, et al. (2022) and Kiuru, Trög, et al. (2022) found no associations between self-reported anxiety and neither SCR, HR, nor HRV on the between-person level. On the within-person level, instead, Kiuru, Malmberg, et al. (2022) found a positive correlation between anxiety and HR, but no association with SCR or HRV. However, in their second publication using the same sample but a different achievement task (Kiuru, Trög, et al., 2022), also the correlation between anxiety and HR was non-significant. Finally, a study by Roos et al. (2023) involved high school students working on six blocks of a math task and repeatedly reporting on different anxiety components, while their SCR was being recorded. The resulting correlations show significant negative associations between SCR and the cognitive and affective component of anxiety, no association with the self-reported physiological component, and a positive association with the motivational component. The latter supports the idea that physiological changes in anxiety support motivational tendencies, whereas the negative associations with cognitive and affective anxiety are surprising when assuming coherence between the different response systems. However, it also needs to be noted that Roos et al. (2023) only reported overall bivariate correlations, without accounting for the nested structure of the data, meaning that the results likely are

“smushed effects” (p. 667) that do not allow for clear insights into either between- or within-person associations (Hoffman & Walters, 2022).

Other achievement emotions besides anxiety have received less systematic attention. On the between-person level, Kiuru et al. have found significant negative associations between enjoyment and SCR and HR (Kiuru, Malmberg, et al., 2022), a negative association between HRV and hope (Kiuru, Malmberg, et al., 2022; Kiuru, Trög, et al., 2022), and a negative association between SCR and hopelessness (Kiuru, Trög, et al., 2022). All remaining between-person correlations between SCR, HR, and HRV with enjoyment, hope, anger, anxiety, fear, hopelessness (Kiuru, Malmberg, et al., 2022), and surprise (Kiuru, Trög, et al., 2022) were non-significant. Further, Harley et al. (2019) found a positive between-person correlation of SCL with shame, but not with hopelessness or pride. Finally, Huber and Bannert (2023) found higher levels of HR in those participants who were not bored during a task. On the within-person level, Kiuru, Malmberg, et al. (2022) found more significant correlations than on the between-person level, with HR correlating negatively with self-reported enjoyment and pride, and positively with anger, fear, hopelessness, and – as previously mentioned – anxiety. In addition, enjoyment also correlated positively with HRV. This seems to suggest a pattern of lower levels of HR accompanying a more intense experience of positive emotions, and higher HR in more intense negative emotions. However, again, these correlations were all non-significant in the analysis for another task with the same sample (Kiuru, Trög, et al., 2022). This discrepancy might suggest a high volatility of findings, as correlations do not seem to replicate reliably even within the same sample. Another possible reason for this discrepancy may have been the low number of blocks in Kiuru, Trög, et al. (2022), as two task blocks appear insufficient to obtain reliable within-person correlations. Finally, Ketonen et al. (2023) found boredom to be associated with lower HR, and excitement to be associated with higher HR and lower HRV, whereas calmness and anxiety were unrelated to both cardiovascular measures. Notably, they used an experience sampling approach where data was collected in both school-related and non-school-related situations. Taken together, on the within-person level, there are considerable differences in the findings between studies, such as Kiuru, Malmberg, et al. (2022) finding that the positive, activating achievement emotions enjoyment and pride negatively correlated with HR, whereas in Ketonen et al. (2023), the also positive, activating emotion of excitement correlated positively with the same physiological variable.

**Summary.** Altogether, findings show some significant associations of physiological measures with the self-reported affective experience of achievement emotions at both the between- and the within-person level. However, as Horvers et al. (2021) concluded in their

systematic review, “there was no clear agreement in their results” (p. 19). For anxiety, the evidence base is somewhat bigger, and, besides some controversial findings, seems to generally support the notion of anxiety as an activating emotion. This aligns with the idea that physiological activation in emotions generally, and achievement emotions in particular, serves the function of facilitating adaptive behaviour, such as investing effort into a task when experiencing anxiety (Pekrun, 2024a; Pekrun et al., 2023). For other achievement emotions, the use of different sets of self-report and physiological variables in different studies does not allow for definite conclusions. Results differ between and even within the same samples, underlining the need for a bigger evidence base and the replication of findings. Further, most studies have not separated between- and within-person covariances between variables, although, as demonstrated in Kiuru, Malmberg, et al. (2022) associations at the two levels may differ.

#### ***3.2.4 Synthesis and Conceptual Model***

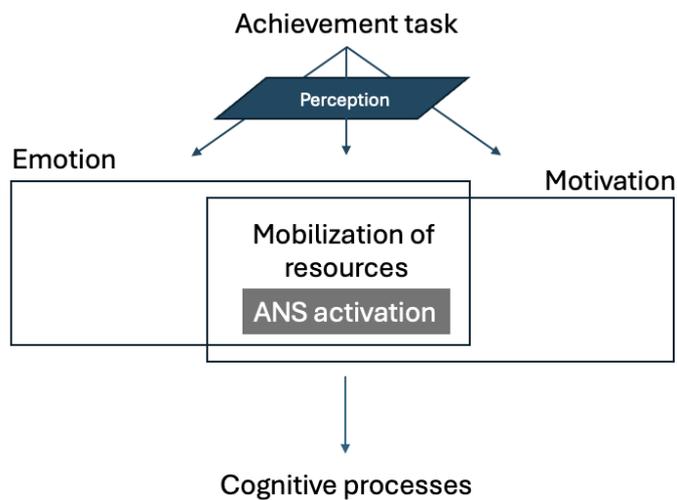
Already in the separate overviews of the three constructs, it has become apparent that it is impossible to fully separate cognition, motivation, and emotion. While there is no part-whole relationship between the constructs, they overlap and are deeply entangled with one another: In the prototypical case, emotions include both a motivational and a cognitive component, just like motivation involves a cognitive representation of a goal state and an affective desire (Pekrun, 2023a). As outlined, associations between all three constructs – cognition, motivation, and emotion – and physiological states have been found. However, such a one-to-many mapping may hinder the application of physiological measures in research and practice, leaving it uncertain whether an observed physiological state is driven by the intended construct or by a different process (Richter & Slade, 2017). Hence, to meaningfully employ physiological measures in educational psychology, it is essential to narrow down their meaning to a more specific function or process. This more precise understanding will, in turn, clarify why and how physiological changes in an achievement situation relate to broader constructs such as cognition, motivation, and emotion.

The present work aims to identify a narrower meaning of physiological states in an achievement situation using an integrated approach that acknowledges the overlap between cognition, motivation, and emotion. Based on the previously outlined theories and findings, I have developed the conceptual model depicted in Figure 1. In line with considerations by Salamone and Correa (2024) stating that emotion and motivation both include arousal elements, it places physiological activation at the intersection of motivation and emotion, where these constructs overlap in their function to mobilize resources for cognitive processing in an achievement situation. Firstly, it should be noted that this model and the following

considerations regarding the meaning of physiological states pertain only to processes occurring during an achievement situation, like a learning or assessment task. Secondly, it should be noted that the presented model does not claim to be a comprehensive summary of all constructs and processes involved. For example, for the sake of clarity, the model only refers to cognitive processes relevant to task completion, excluding other aspects of cognition, such as prototypical thoughts or mental goal representations, that overlap with the emotion and motivation constructs (Pekrun, 2023a). Hence, rather than being a comprehensive theoretical model, this conceptual framework integrates various constructs to enhance our understanding of the specific role of physiological states in achievement situations.

**Figure 1**

*Proposed Conceptual Model*



*Note.* ANS = Autonomic nervous system.

The starting point of the model is the achievement task itself, which impacts the individual’s emotion and motivation through their perceptions of control/expectancy and perceived value (Eccles & Wigfield, 2020; Pekrun, 2024a). Emotion and motivation are considered distinct constructs; however, they share the function of promoting adaptive behaviour by mobilizing energy for an upcoming task (Miron & Brehm, 2012; Pekrun et al., 2023; Salamone & Correa, 2024). Furthermore, the perception of task difficulty also directly influences the mobilization of resources, as it provides information on the number of resources required for successful task completion (Brehm & Self, 1989). The mobilization of resources, which has also been termed effort (Gendolla et al., 2019), then directly influences the cognitive processes performed with the goal of successful task completion, and thereby performance on

the task (Shenhav et al., 2017). This mobilization of resources happens through physiological changes in the autonomic nervous system, which can promote an “action-ready bodily state” (Radulescu et al., 2015, p. 238) through sympathetic activation and parasympathetic deactivation.

The proposed model integrates theoretical and empirical work on cognition, motivation, and emotion. In contrast to previous findings, it does not consider physiological states as an outcome or a part of either construct, but acknowledges their entanglement, aligning with the call for greater theoretical integration (Pekrun, 2024b). The empirical works reported in the following chapters each cover a part of this proposed model, with the aim of gaining a better understanding of the meaning of physiological states in achievement situations.

### **3.3 Aims of the Present Work**

Despite the growing use of physiological measures in educational psychology research and their potential advantages over self-report measures, several open questions regarding their meaning and application remain. The present work aims to advance our understanding of physiological states in achievement situations by exploring three main themes.

The first theme is the distinction between the between- and within-person level of analysis. The beginning of this chapter (Section 3.1) has introduced the distinction between the two levels of analysis, as well as its central role in the analysis and interpretation of findings. It has been argued that associations between psychological and physiological processes might primarily emerge at the within-person level (Mauss et al., 2024; Pekrun, 2023b), and indeed different patterns of associations were observed between the levels when they were separated (Kiuru, Malmberg, et al., 2022). Hence, in the present context, acknowledging this distinction may be crucial for identifying patterns. The present work therefore explores the often-neglected distinction between associations at the between- and within-person levels for psychological, particularly emotional, and physiological variables in achievement situations.

The second guiding theme of the present work is to gain a more specific understanding of the meaning of psychological states in achievement situations. Based on the entanglement of the three constructs of cognition, emotion, and motivation, which have all been linked to physiological states in previous work, I have proposed a conceptual model where physiological states are located at the intersection of emotion and motivation, serving the function of mobilizing resources for a task at hand. This proposition is examined from different angles by relating physiological responses to affective states, as well as emotionally and motivationally relevant task appraisals and task characteristics.

The third theme of the present work is the exploration of differences between autonomic indicators in their relations with psychological processes. From the overview of existing literature, it has become apparent that different studies have employed different physiological measures of the autonomic nervous system, with findings varying across indicators. While there is a general understanding of what physiological processes the different measures indicate (e.g., EDA indicates sympathetic activation, HRV indicates parasympathetic activation), we lack a detailed understanding of what – possibly different – psychological processes they relate to. Based on the use of two indicators each of EDA and CVA, the present work provides preliminary insights into possible differences in the specific meanings of these different indicators in the achievement context. The outlined key issues are addressed by combining the findings from three empirical works, using the data from two data collections.



---

# 4

---

## **DATA USED IN THE PRESENT WORK**

---

Miriam Wunsch



For the present work, data were collected in two laboratory-based studies. The laboratory setting ensured control over potentially confounding influences and environmental variables that could compromise data quality, such as distracting surroundings or very high or low room temperatures. Having conducted two data collections brings two main benefits: Firstly, we were able to build on our experiences from the first study to make improvements to the design or setup for the second study. Secondly, conducting two studies with similar designs in the same environment allows for a comparison of findings. The following sections detail the lab setup, the software and hardware used, the data processing procedures, and the study design. While these aspects were largely consistent across both studies, differences are highlighted where applicable.

### **4.1 Laboratory Setup**

The lab is located in a university building at a central location in Munich, making it easily accessible for both student and non-student participants. The room is around 23 m<sup>2</sup> in size and contains several desks with computers, as well as a large table and a wardrobe for storage. The participant's desk, which was located facing a wall to minimize distractions, had a 24-inch screen with a resolution of 1920 x 1200. A webcam, as well as a keyboard and mouse for the participant were attached to the screen. A foam block was placed on either the left or right side of the keyboard for participants to rest their non-dominant hand with the sensors on. The experimenter was sitting at another desk with a screen, separated from the participant by a white, tall wooden partition wall in Study 1 and a grey, shoulder-high movable partition wall in Study 2. Thus, while the participant was aware of the experimenter's presence in the room, they were separated by a visual barrier. In Study 1, the participant was sitting on a wooden chair that was not height-adjustable. In Study 2, this was replaced by a height-adjustable office chair. The participant's screen was connected via an HDMI cable to the experimenter's computer.

### **4.2 Software iMotions**

The experimenter computer was running the software iMotions (2022, Version 9), which allows the collection, synchronization, and – to a limited extent – analysis of various data streams. The producer of this software is the Denmark-based company iMotions A/S, which is an internationally active subsidiary company of Smart Eye AB. According to their website, iMotions has “an estimated 50% market share in the early-stage software market for multimodal behavioral research” (iMotions, 2025), serving both academic and corporate clients. In comparison to other technical solutions, such as recording EDA data directly in Biopac Systems Inc.'s AcqKnowledge software, the use of the iMotions software brings the main advantages of

being able to present stimuli directly in the software, and to simultaneously collect multimodal data in one tool with an automatic synchronization of these data streams. The data streams collected in the present studies were electrodermal activity, cardiovascular activity, and facial expression. Since the latter was not used in the present work, only the recording and analysis of electrodermal and cardiovascular activity, as well as the self-report measures, are described in the following.

### **4.3 Hardware**

Both electrodermal and cardiovascular activity were recorded using a BioNomadix Wireless PPG and EDA Transmitter (Biopac Systems, Inc., 2025a), which was attached to the participant's non-dominant wrist. One cable connected the transmitter wristband to a photoplethysmography (PPG) sensor on the tip of the participant's index or ring finger. Another cable connected the transmitter to two adhesive electrodes with liquid gel and an Ag/AgCl contact, which were placed on the participant's palm (Shimmer, 2025). Following recommendations by Boucsein et al. (2012), the electrodes were attached a few minutes before the start of the recording, allowing the gel to warm and soften. Both the PPG and EDA signals were transmitted to a Biopac MP160 receiver (Biopac Systems, Inc., 2025b) with a Biopac Bionomadix PPG and EDA Amplifier (Biopac Systems, Inc., 2025c) and recorded in the iMotions software. In Study 1, we collected the physiological data at a frequency of 500 Hz, which was reduced to 200 Hz in Study 2 to simplify data storage and accelerate data processing. Considering that, for the present work, the physiological data were aggregated to match the resolution of the self-report measures, I believe that this reduction in recording frequency did not impair data quality while allowing for more efficient data processing.

### **4.4 Processing of Physiological Signals**

#### ***4.4.1 Electrodermal Activity***

The EDA signal, which, in its raw form, indicates the conductance of an electric current between the two electrodes in microSiemens, was processed directly in the iMotions software with the built-in EDA Peak Detection Algorithm. The algorithm separates the EDA signal into its phasic and tonic components and provides aggregated peak metrics for specific stimuli or time frames. We carefully compared the software's default settings to recommendations in the literature and maintained most default setting, except for the Peak Amplitude Threshold which we adjusted from the default of .005 mS to the more commonly recommended .01 mS (Boucsein, 2012).

Following the application of the peak detection algorithm, iMotions provided a dataset that contained a value in microSiemens for both the phasic and the tonic signal, as well as a binary variable indicating whether a peak was detected in the phasic signal. Further, in a different export option, the number of peaks per minute was already aggregated for certain stimuli or time periods of interest. I validated this aggregated measure by manually performing the presumed underlying calculation for several exemplary data snippets – dividing the number of peaks by the duration of the respective stimulus – and found that this manual calculation matched the aggregated values. Hence, for SCR, I used the already aggregated values for peaks per minute provided by iMotions. For SCL, there was no built-in aggregation of the tonic EDA data available in iMotions. Therefore, I performed this aggregation in R, using a for-loop that imported the data set for each participant separately, calculated the mean tonic EDA level for each task block, and saved the respective values in a new data frame.

### ***4.4.2 Cardiovascular Activity***

For the cardiovascular activity data collected using a PPG sensor, iMotions (in its current version) only provides limited analysis options: While it provides a HR value for each measurement timepoint, there currently is no algorithm to aggregate across these timepoints, and no HRV measures can be calculated from PPG data. Further, in Study 1 the PPG signal was falsely labelled as electrocardiography data, meaning that while the raw signal was correctly recorded, no export of cardiovascular parameters was possible for Study 1. Hence, we chose to use the software Kubios HRV Scientific (Tarvainen et al., 2014) for the analysis of the PPG data. The Kubios software utilizes a pulse wave detector for analyzing the raw signal, which employs a matched filtering approach: In the first step, the algorithm detects the initial pulses and derives a template for the typical pulse wave from them. It then compares each part of the recorded signal to this template, classifying a pattern as a pulse wave when it sufficiently resembles the template. Further, the software detects artifacts in the form of ectopic, i.e., irregular, beats and corrects them using interpolation. It also applies a smoothness priors method to remove slow, temporal trends, and excludes overly noisy parts of the signal from the analysis (Kubios Oy, 2024). From this processed and corrected signal, values of HR and HRV were exported for each data segment of interest.

## **4.5 Study Design**

The two studies were similar in their design, as both contained a baseline and validation phase, several blocks of a numerical reasoning task, and self-report questionnaires after each

task block. In the following, I outline each of these components and explain the similarities and differences between the two studies.

### ***4.5.1 Baseline and Relaxation Phases***

The baseline measurement was preceded by the instruction to relax and breathe normally, while watching a blue bar fill with colour that indicated the remaining baseline duration. In Study 1, the baseline measurement was one minute long, whereas it was three minutes in Study 2. The duration was increased, as similar studies reported baselines of several minutes (Horvers et al., 2021), and extending the baseline might have improved the quality of this measurement. In Study 1, the baseline was presented at the start of the procedure, before the validation and the testing phases. Instead, in Study 2, it was placed right before the start of the testing phase. The decision to move the baseline was based on the goal of capturing participants' physiological activation right before the phase of interest.

Additionally, based on the findings from Study 1, it appeared that physiological arousal levels in participants were particularly high at the beginning of the study, possibly due to the novelty of the laboratory situation. To mitigate temporal trends and obtain a baseline measurement free from the high initial arousal, we introduced a relaxation phase at the beginning of Study 2. This relaxation phase consisted of a breathing exercise with slow breaths and long exhales, intended to increase activation in the PNS and decrease activation in the SNS (Gerritsen & Band, 2018). During the relaxation phase, participants were instructed to inhale deeply for five seconds and exhale for eight seconds, guided by a circle on the screen that expanded and contracted in a synchronized breathing rhythm. They repeated this cycle five times before being guided into the validation phase.

### ***4.5.2 Validation Phase***

The validation phase, which was placed between baseline and task in Study 1, and between relaxation phase and baseline in Study 2, aimed to identify EDA non-responders. As these individuals do not show meaningful changes in their EDA in response to stimuli, including their EDA data to explore associations of EDA with psychological processes would have confounded the validity of our findings. Hence, their EDA data were omitted.

The inclusion of a separate validation phase before the achievement task enabled the identification of non-responders without relying on their on-task data. In Study 1, the validation phase consisted of two physical tasks, namely holding one's breath for ten seconds and biting one's tongue for ten seconds. Following Study 1, the information gained from this validation paradigm was combined with that from an earlier study, in which the validation paradigm also

included emotional stimuli. From the combined data, it became apparent that the physical stimulus of holding one's breath and an anxiety-inducing emotional stimulus were the most predictive of whether an individual would be a non-responder in the subsequent test phases. Hence, for Study 2, the validation paradigm was adapted to contain an anxiety-inducing video of a person climbing a tall building, and the physiological act of holding one's breath.

For both studies, individuals who did not show any SCR during the two validation stimuli were classified as non-responders and excluded from the EDA analyses. In Study 1, 19 out of 89 (21%) individuals were affected, whereas in Study 2, only two out of 70 (3%)<sup>1</sup> individuals were affected. These different ratios of non-responders suggest that we might have been overly exclusive in Study 1, using the first version of the validation protocol. However, I believe it was good research practice to adhere to the previously defined exclusion criteria for Study 1, but to adjust the validation paradigm for Study 2 based on the new insights gained.

### ***4.5.3 Numerical Reasoning Task***

The task in both studies consisted of numerical reasoning items, in which participants were required to identify patterns in number sequences and fill in the missing one or two numbers accordingly. The different items required different cognitive operations, such as to identify parallel sequences or to detect complex progressive coefficient patterns (for an overview of all cognitive operations, see Loe et al., 2018). The items were developed using an automated item generator and evaluated in an adult sample of 570 participants, resulting in a Rasch-scaled difficulty estimate for each of the final 49 items (Loe et al., 2018). We used these Rasch-scaled difficulty estimates to construct our tests.

In Study 1, there were six blocks of four items each, with three blocks being adaptive to the participant's performance and three being the same for all participants. In the non-adaptive blocks, item difficulty increased from the first through the twelfth item. The remaining 37 items constituted the item bank for the adaptive test, in which item difficulties varied depending on the participant's performance. In Study 2, no adaptive test blocks were included. The number of blocks was reduced to five in order to maintain an acceptable study duration despite the addition of the relaxation phase. The first three blocks of Study 2 were equivalent to the non-adaptive blocks from Study 1, meaning that item difficulty increased from the first to the twelfth item. For the remaining two blocks, we selected items in a way that the fourth

---

<sup>1</sup> Numbers for non-responder count for Study 1 differ across subsequent reports because the order of exclusion steps differed (Empirical Work 2 removed data unusable for other reasons first, which included two non-responders → 17 non-responders; Empirical Work 3 removed non-responders first → 19 non-responders). Also overall  $N$  might vary across reports for both studies, depending on whether individuals with unusable data were counted.

block started with a rather easy item, before item difficulty again increased continuously until the final twentieth item.

In both studies, there was a break after the third item block. Participants were instructed to relax before starting the second test (Study 1) or before continuing the test (Study 2). They could end the break in a self-directed way whenever they were ready by clicking the respective button on the screen.

### ***4.5.4 Self-Report Questionnaire***

After each item block, a self-report questionnaire was administered in pen-and-paper format. The first two questions assessed the arousal and valence dimension of the current emotional experience using the self-assessment manikin (Bradley & Lang, 1994), complemented by verbal cues for each response option. Further, we assessed participants' agreement to statements indicating the experience of distinct states, namely enjoyment, pride, anxiety, anger, stress, frustration, and boredom, and their perception of task difficulty and challenge. In Study 2, we additionally included an item assessing their mastery perception, i.e., how well they believe they are currently doing, as well as their value perception, asking how important success on the task is to them. Of those items, the self-reported achievement emotions are used in Empirical Works 1 and 3, except for stress, which is not typically considered an achievement emotion. The items assessing perceived difficulty and perceived value are used in the analyses for Empirical Work 2, as they align best with the constructs of interest. For an overview of the self-report items used, see Table 1. Data from the other self-report items, from an online pre-questionnaire before the laboratory sessions, as well as from a baseline self-report questionnaire and a cooldown phase after the procedure in Study 2, are not relevant to the present work, but have already been or can be used for other analyses.

**Table 1***Self-Report Items Used in Present Work*

<b>Variable</b>	<b>Item (English/German)</b>	<b>Scale</b>
Enjoyment	I am enjoying this. / Es macht mir Spaß.	Please choose the option that describes best how you are currently feeling. / Bitte gib an, wie sehr Folgendes auf dich zutrifft.
Pride	I feel proud. / Ich fühle mich stolz.	1 = not true at all / stimmt gar nicht
Anger	I feel angry. / Ich fühle mich ärgerlich.	2 = rather not true / stimmt kaum
Boredom	I feel bored. / Ich fühle mich gelangweilt.	3 = somewhat true / stimmt etwas
Frustration	I feel frustrated. / Ich fühle mich frustriert.	4 = mostly true / stimmt weitgehend
Anxiety	I feel tense and nervous. / Ich fühle mich angespannt und nervös.	5 = completely true / stimmt genau
Perceived Difficulty	To me the tasks are... / Die Aufgaben sind für mich...	1 = very easy / sehr leicht
		2 = rather easy / eher leicht
		3 = neither easy nor difficult / weder leicht noch schwer
		4 = rather difficult / eher schwer
		5 = very difficult / sehr schwer
Perceived Value	How important is it to you to do well on the test? / Wie wichtig ist es dir, im Test gut zu sein?	1 = not important at all / überhaupt nicht wichtig
		5 = very important / sehr wichtig



---

# 5

---

## OVERVIEW OF EMPIRICAL WORK

---

Miriam Wunsch

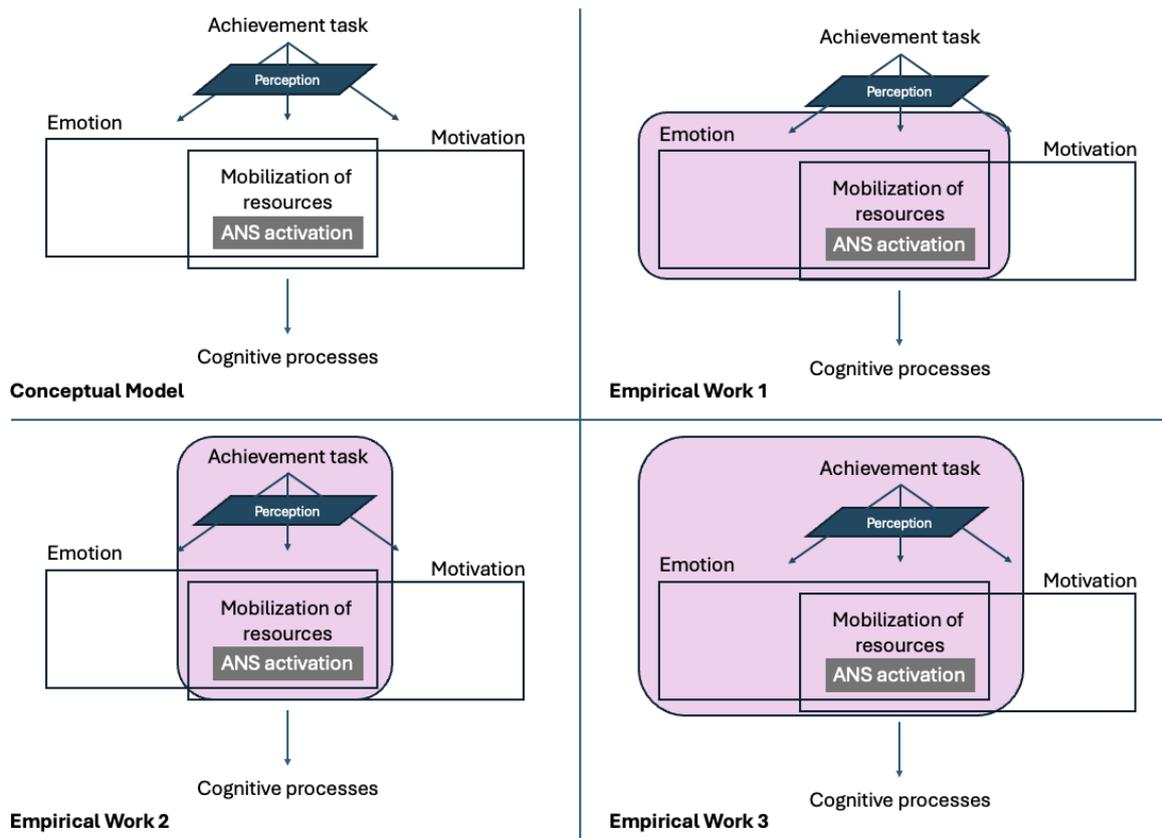


Each empirical work (EW) presented in the following three chapters is a self-contained investigation of a particular research question, using data from one or both described studies. Beyond their individual research questions, it is the combination of their findings that yields valuable insights into the meaning and application of physiological measures in achievement situations. Specifically, the present work explores three central themes: the level of analysis, the meaning of physiological changes, and potential differences between indicators.

Insights on the first theme, the level of analysis, are primarily provided by EW1, reporting intraclass correlations for all four physiological measures – before and after baseline correction – along with associations between these measures and achievement emotions at both the between- and within-person levels. The intraclass correlations provide insights into the proportion of between- and within-person variance in the physiological measures into whether the applied baseline correction has reduced the potentially confounding large individual differences unrelated to the achievement situation. Additionally, the correlational findings of EW1 show potential differences in the associations of physiological measures with self-reported achievement emotions at the between- and within-person levels.

The second theme, the meaning of physiological changes in an achievement context, is addressed in all three empirical works, each covering a part of the proposed conceptual model (see Figure 2): EW1 explored how the affective experience of achievement emotions relates to physiological changes; EW2 investigated the association between emotionally and motivationally relevant task perceptions and physiological states; and findings from EW3 can be used to compare the effect of task characteristics on physiological and self-reported emotional states. Synthesizing across the three sets of findings enables an evaluation of the proposed model that locates physiological states at the intersection of emotion and motivation in the function of mobilizing resources for a task.

The third research theme of the present work, potential differences between indicators of autonomic activation, is addressed in the findings of EW1 and EW2. Both employed various indicators of autonomic activation, namely SCR, HR, and HRV in EW2, and additionally SCL in EW1. This allows for the inspection of differences between indicators, such as between indicators of sympathetic and parasympathetic activation, and between indicators of the same autonomic branch, like SCR and SCL. The exploratory inspection of differences between indicators offers preliminary insights into possible differences in their meaning within an achievement situation. The following sections provide a brief overview of each of the three empirical works, outlining their theoretical approach, research questions, and methodological aspects.

**Figure 2***Overview of Empirical Works Covering Parts of the Proposed Conceptual Model*

*Note.* ANS = Autonomic nervous system.

### 5.1 Empirical Work 1: Physiological States and Self-reported Achievement Emotions

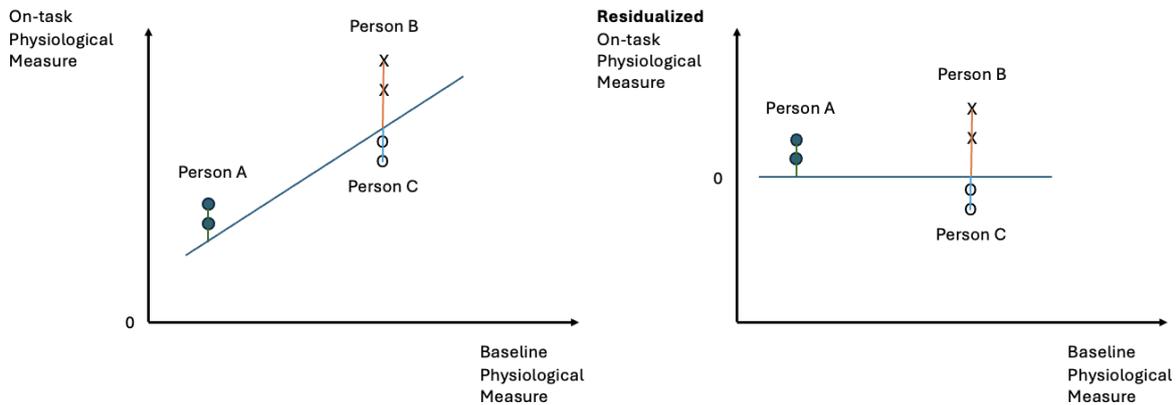
In the first empirical work, we investigated the question of whether and how physiological states (SCR, SCL, HR, HRV) relate to the subjective, affective experience of the achievement emotions (enjoyment, pride, anger, anxiety, frustration, boredom). While grounded in the broader discussion of emotion coherence (e.g., Mauss et al., 2024), this work focused on the specific category of achievement emotions, using CVT (Pekrun, 2006, 2024a) as the guiding theoretical framework. We examined bivariate correlations between physiological states and self-report measures of achievement emotions at the between- and within-person levels. Doing so, we aimed to gain insights into differences between the two levels of analysis, and to evaluate the consideration of affect and physiology as “loosely coupled” (Pekrun et al., 2023, p. 146) components of achievement emotions, as well as their theoretical classification into activating and deactivating emotions (Pekrun, 2024a). As this work employed a rather exploratory approach, no specific hypotheses were formulated.

EW1 contains two sets of results, using data from Study 1 and Study 2, respectively. This two-study approach allowed us to replicate our findings in a second sample, thereby examining their robustness and ensuring that the use of adaptive testing in Study 1 did not confound findings on the associations between physiological and affective states.

In the analyses, we accounted for potentially large baseline differences between individuals in their physiological parameters, by baseline-correcting each physiological data point. Specifically, we employed a residualizing approach, predicting the physiological values across all individuals and all blocks from the baseline assessment, and extracting the residuals for each data point from that regression line. Each resulting value indicated how much the data point deviated from what would have been expected based on the individual's baseline value (see Figure 3 for an illustration). This approach accounts for baseline differences in the physiological parameters between individuals, and additionally – in contrast to approaches simply subtracting the baseline from on-task values – considers the possibility that changes between baseline and task might be more (or less) pronounced depending on the baseline level.

Using the residualized values, the primary statistical analysis in this work involved calculating multilevel bivariate correlations for each pair of physiological measures and self-reported achievement emotions. At the between-person level, these correlations indicate whether general differences between individuals on a physiological parameter, considering their baseline level, related to differences between them in the intensity of their affective experience of an achievement emotion. Instead, the correlations at the within-person level indicate whether an individual who exhibited a higher or lower value in a physiological parameter during one block, compared to another block, also reported a more or less intense experience of the respective achievement emotion in that block, compared to the other block.

Based on the CRediT system, my contributions to this empirical work were in the areas of conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, visualization, and writing.

**Figure 3***Illustration of Residualizing Approach to Baseline Correction*

*Note.* The example includes two assessment timepoints for three individuals, different individuals indicated by different shapes. In the first step (left) a regression line is calculated predicting on-task physiological values from the baseline. In the second step (right) residuals to the regression are extracted for each measurement timepoint for each person, making on-task values of individuals of different baseline levels more comparable. For example, when only considering raw values, Person A would have had lower on-task physiological values than Person C. However, after correcting for the higher baseline level of Person C, they have lower baseline-corrected on-task values than Person A. When two individuals have the same baseline, like Person B and Person C, the order remains unchanged – Person B still has higher baseline-corrected values than Person C.

## 5.2 Empirical Work 2: Difficulty x Value Perceptions

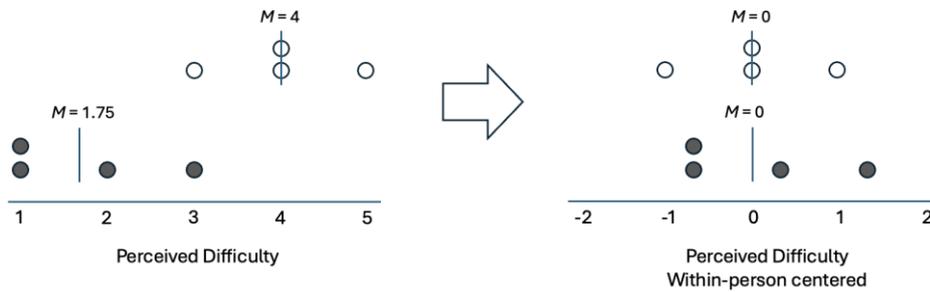
In the second empirical work, we investigated how the emotionally and motivationally relevant perceptions of task difficulty and value relate to physiological states on a within-person level. While we did not include a direct measure of emotional and motivational states, we aimed to explore which of the constructs of achievement emotions, achievement motivation, and effort is most closely related to physiological changes in an achievement situation. Our approach to doing so was based on the principle of predictive validity: In the narrower sense, predictive validity means that if a measure indicates the intended construct, it should predict outcome variables in the same way that the underlying construct would (Richter & Slade, 2017). In EW2, we follow a somewhat inverted reasoning of this principle, posing that if physiological changes are indicative of achievement emotions, achievement motivation, or effort, they should be predicted by task appraisals in the same way that these constructs are. Specifically, according

to CVT (Pekrun, 2024a), EVT (Eccles & Wigfield, 2020), and MIT (Brehm & Self, 1989), all three constructs should be predicted by appraisals of task difficulty and value. Drawing on the three theoretical frameworks, we derived predictions on how difficulty and value appraisals would relate to physiological states if those states reflected the corresponding psychological constructs and then compared the derived predictions with the patterns observed in the data.

The self-report questionnaire in Study 1 did not contain a measure of perceived value; this item was added in Study 2 specifically to enable this investigation. Hence, only data from Study 2 were used in EW2. In EW2, we utilized the physiological measures of SCR, HR, and HRV, as these were found to exhibit more meaningful and consistent patterns than SCL in EW1.

Based on the findings of EW1, the second empirical work focused only on associations at the within-person level. We employed a multilevel regression model with random intercepts. As random intercepts only account for interindividual differences in outcome, but not in predictor variables, we followed the recommendation of person-mean centring the predictor variables (Hoffman & Walters, 2022). In this procedure, a person's mean value on the predictor variables, that is, perceived difficulty and value, is subtracted from each of their individual values, thereby centring the individual values around their mean (see Figure 4 for illustration; Algina & Swaminathan, 2011). Consequently, the resulting centred values no longer reflect absolute values on the respective scale of the predictor variable (e.g., a perceived difficulty of "4"), but instead indicate the individual's deviation from their own mean on the predictor variable at that time point (e.g., "1.5 units higher than their average perceived difficulty"). This procedure removes the between-person variance contained in each measurement time point of the predictor variables, while also improving the interpretability of the intercept and interaction terms by assigning a meaning – the person's average – to a value of zero on the predictor variable (Hoffman & Walters, 2022). Doing so enabled us to investigate the question of how difficulty and value perceptions predicted physiological states from a purely within-person perspective.

My contributions to this empirical work included conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, visualization, and writing.

**Figure 4***Illustration of Person-Mean Centring*

*Note.* The example includes four assessment timepoints of perceived difficulty for two individuals; different individuals are indicated by colours (white, grey). Mean values are calculated for each person (left) and then subtracted from each assessment time point (right).

### 5.3 Empirical Work 3: Effects of Adaptive Testing

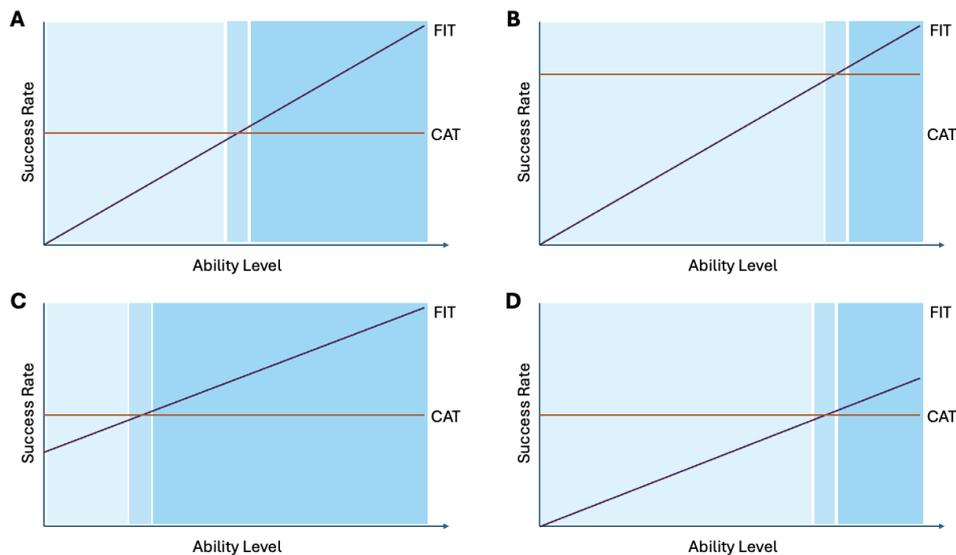
In the third empirical work, using the data from Study 1, we investigated whether and how the adaptivity of a test affects physiological states, as well as the affective experience of achievement emotions. Considering the increasing use of adaptive testing systems, this exploration is crucial to ensure that the increased efficiency promised by adaptive tests does not come at the expense of test-takers' experiences and, thereby, possibly their performance. Reviewing previous work on this question, we identified that studies often appeared to focus on features of the adaptive test and how they influence the emotional experience compared to a non-adaptive test, while neglecting the central role of features of the non-adaptive test in this comparison. Specifically, based on CVT (Pekrun, 2024a), we identified that the experience of success, indicating control, should be a central predictor of how individuals emotionally respond to a test, given a constant perceived value. Hence, differences in the emotional response to two tests should depend on the differences in their success rates, which is a parameter that results from a combination of the test-taker's ability and the overall difficulty of the test. Figure 5 illustrates how different combinations of test difficulty between an adaptive and a non-adaptive (fixed-item) test determine at what ability level an individual should perceive either one or the other test as more difficult. Based on this reasoning, ability level and perceived difficulty were explored as possible moderators of the effect of test type on the subjective experience of joy, pride, anger, anxiety, frustration, and boredom. Meanwhile, we hypothesized a main effect of adaptive testing on sympathetic activation, with higher SCR in the adaptive test, due to the better fit between individuals' ability level and test difficulty, maintaining higher engagement in all individuals than the non-adaptive test.

To investigate both the main effect of test type on SCR and the interactions between test type and ability/perceived difficulty, multilevel linear regression models were used. Again, a random intercept was included to account for differences in SCR, as well as in the subjective experience of different achievement emotions. While the main focus of this individual empirical work was the effect of adaptive testing, the results also provide insights into the meaning of physiological changes in an achievement situation, allowing for a comparison of how task characteristics differentially affect physiological and subjective emotional states.

According to the CRediT statement, my contributions to this empirical work were in the areas of conceptualization, data curation, formal analysis, investigation, methodology, project administration, visualization, and writing.

### Figure 5

#### *Visualization of Relative Success Rates on Adaptive and Fixed-Item Test*



*Note.* FIT = Fixed-item test, CAT = Computerized adaptive test. The different panels illustrate how combinations of test difficulty determine at what ability levels which test has a higher or lower success rate (light blue shade: FIT easier/higher success rate than CAT, middle area: similar difficulty/success rate on both tests, darker blue shade; FIT more difficult/lower success rate than CAT). When success rate on CAT is set higher (A  $\rightarrow$  B), individuals of a broader ability range will have a higher success rate on the CAT than the FIT. When success rate on the FIT is set higher (A  $\rightarrow$  C), individuals of a broader ability range will have a higher success rate on the FIT than on the CAT. When success rate on FIT is set lower (A  $\rightarrow$  D), individuals of a broader ability range will have a higher success rate on the CAT than the FIT.



---

# 6

---

## EMPIRICAL WORK 1

### PHYSIOLOGICAL STATES AND SUBJECTIVE EMOTIONAL EXPERIENCES IN AN ACHIEVEMENT SITUATION: BETWEEN- AND WITHIN-PERSON LINKS

---

Miriam Wunsch  
Anne C. Frenzel  
Franziska Eckerskorn  
Reinhard Pekrun  
Ji-Young Min  
Nadyanna M. Majeed

**Reference:**

Wuensch, M., Frenzel, A. C., Eckerskorn, F., Pekrun, R., Min, J.-Y., & Majeed, N. M. (under review). Physiological states and subjective emotional experiences in an achievement situation: Between- and within-person links.

The chapter printed here is the manuscript version as of October 15<sup>th</sup>, 2025; this manuscript version has not been peer reviewed.



## 6.1 Abstract

Achievement emotions are defined as multi-faceted constructs including subjective and physiological components. In two studies with adult participants (Study 1:  $N=87$ ; Study 2:  $N=68$ ), we investigated whether the subjective experience of six achievement emotions (enjoyment, pride, anger, anxiety, frustration, and boredom) was associated with indicators of physiological arousal (skin conductance response, skin conductance level, heart rate, heart rate variability) on a between- and within-person level. In both studies, participants completed several blocks of a numerical reasoning test while their physiological states were measured. After each block, they reported their emotional experience. Multilevel correlation analysis revealed no substantial correlational pattern between self-reported emotions and indicators of physiological arousal on the between-person level in either of the studies. Instead, on the within-person level, increases in the subjective experience of joy and pride (Studies 1 and 2) as well as anxiety (Study 2) were accompanied by increases in sympathetic and decreases in parasympathetic activation. Despite presumably being activating emotions, anger and frustration were associated with decreases in sympathetic activation (Studies 1 and 2). Boredom was accompanied by decreases in sympathetic (Studies 1 and 2) and increases in parasympathetic activation (Study 1). These findings show that the subjective experience of achievement emotions and physiological states are coupled within persons across moments, but individuals' physiological states do not allow inferences about their subjective emotional experiences relative to others. The results further suggest that physiological measures can supplement, yet not replace, self-report measures of achievement emotions in educational psychology research and practice.

*Keywords:* achievement emotion, psychophysiology, electrodermal activity, cardiovascular activity, multilevel analysis

## 6.2 Educational Impact and Implications Statement

This study finds that changes in a person's electrodermal and cardiovascular activity are linked to changes in their subjective experience of emotions in a testing situation. However, it is not possible to compare the electrodermal or cardiovascular activity of different individuals to draw inferences on differences in their subjective emotional experience or vice versa. Hence, educational interventions aiming to adapt learning environments to learners' emotional experiences using physiological measures should focus on changes within individuals rather than differences between them.

### 6.3 Introduction

Think back to that last important presentation, submission, or exam. What may come to your mind might be feelings of excitement or anxiety, possibly accompanied by a racing heart and sweaty hands. While we have all experienced such changes in both the emotional experience and physiological processes in achievement situations, we are yet to understand whether and how these changes are linked to each other. In the broader literature on emotions, most theories acknowledge that physiological changes constitute one facet of emotions, despite disagreeing on their exact role in the process (Lange et al., 2020; Mauss et al., 2024; Scherer, 2009). Similarly, achievement emotions are defined to involve a physiological component that is associated with the subjective emotional experience (Pekrun, 2006; Pekrun et al., 2023). This conception manifests in the use of items such as “Worry about not completing the material makes me sweat” for measuring learning-related anxiety, or “My heart beats faster with joy” for test-related joy in the seminal measurement tool for achievement emotions, the Achievement Emotions Questionnaire (Pekrun et al., 2011). Building on the multi-component view of emotions, studies increasingly supplement classical self-report measures of achievement emotions with psychophysiological assessments to gain a more comprehensive and less biased insight into emotional processes in learning and achievement situations (Pekrun, 2023). However, the empirical evidence on links between self-reported achievement emotions and psychophysiological variables is mixed. These inconsistencies are likely due to the use of different physiological measures, consideration of different selected emotions, and a lack of distinction between relations on the between- and within-person level.

The present work, therefore, seeks to systematically explore whether and how the subjective experience of achievement emotions, as indicated in self-report measures, relates to physiological variables. In two studies, participants were working on a demanding mathematics-related achievement task while we continuously recorded their physiological states via several common psychophysiological variables (skin conductance response, skin conductance level, heart rate, and heart rate variability). Additionally, participants repeatedly reported their subjective experience of discrete achievement emotions (enjoyment, pride, anger, anxiety, frustration, and boredom). We investigated associations between physiological variables and self-reported achievement emotions both on the between-person level (i.e., whether differences between individuals on one variable relate to differences on another variable) and on the within-person level (i.e., whether changes within the same person on one variable relate to changes on another variable).

### **6.3.1 Achievement Emotions**

Achievement emotions are defined as emotions that relate to “activities and outcomes that are judged according to competence-based standards of quality” (Pekrun, 2024, p. 83). In line with broader emotion theories, achievement emotions are considered to involve various subsystems, namely affective, cognitive, motivational, expressive, and physiological processes (Pekrun, 2006). These components are assumed to be “loosely coupled” (Pekrun et al., 2023, p. 146). Pekrun (2024) further suggested that different achievement emotions can be classified according to a three-dimensional framework: The first dimension within this framework is valence, describing whether the emotion is pleasant/positive or unpleasant/negative. The second dimension, arousal, describes whether the emotion is physiologically activating or deactivating. The third dimension, object focus, indicates whether the emotion refers to an achievement activity or an achievement outcome (i.e., success or failure). Accordingly, enjoyment would be classified as a positive-activating activity emotion, pride as a positive-activating outcome emotion, anger and frustration as negative-activating activity emotions, anxiety as a negative-activating outcome emotion, and boredom as a negative-deactivating activity emotion (Pekrun, 2024). While sometimes considered an activating emotion, we believe that frustration may also be deactivating, especially in low-stakes achievement situations that do not involve severe consequences of success or failure.

The arousal dimension of achievement emotions is defined in terms of physiological processes: Activating emotions involve an increase, deactivating emotions a decrease in physiological arousal (Harley et al., 2019; Pekrun et al., 2023) – irrespective of their valence (i.e., whether they are pleasant or unpleasant). Hence, subjectively more intense experiences of activating emotions, such as enjoyment, pride, anger, anxiety, or frustration, are expected to be associated with higher physiological arousal, while subjectively more intense experiences of the deactivating emotion of boredom should be linked to reduced physiological arousal.

Based on these deliberations, measurement of physiological processes is gaining popularity in research on achievement emotions. In comparison to self-report, these measures are considered capable of capturing unconscious processes, unbiased by memory or social desirability (Houtveen & De Geus, 2009; Pekrun, 2023; Wilhelm & Grossman, 2010). Further, they might constitute a less obtrusive way to assess individuals’ varying states during achievement situations (Pekrun, 2023). Several studies have employed physiological measures in addition to self-report measures of achievement emotions, yielding mixed results on how different physiological variables relate to subjective, self-reported intensities of these emotions. Two common sets of physiological variables used in this context are measures of electrodermal

and cardiovascular activity (Roos et al., 2021), which are both controlled by the autonomic nervous system. In the following, we provide a brief overview of the autonomic nervous system and describe each of these measures before outlining previous findings on their relationship with self-reported achievement emotions.

### ***6.3.2 Measures of Autonomic Nervous System Activity***

The autonomic nervous system consists of three subsystems: the sympathetic, the parasympathetic, and the enteric nervous system. While the enteric nervous system is located in the gastrointestinal tract, the sympathetic and parasympathetic nervous systems control functions throughout the entire body (for a comprehensive overview, see, e.g., Wehrwein et al., 2016). The sympathetic nervous system is considered the “fight or flight” system and is mainly associated with states of activity and arousal. The parasympathetic nervous system is considered the “rest and digest” system, associated with states of relaxation and recovery (Weissman & Mendes, 2021).

Some organs are only innervated by either the sympathetic or the parasympathetic nervous system, whereas others are influenced by both subsystems. In the latter case, these influences can either be antagonistic (with sympathetic and parasympathetic activity having opposite effects) or synergistic (with sympathetic and parasympathetic activity having complementary effects; Wehrwein et al., 2016). Despite seemingly being associated with opposite processes – the sympathetic nervous system with activation and the parasympathetic nervous system with relaxation – they are not two ends of one dimension but rather two distinct dimensions. Changes in these dimensions can occur independently of each other (Berntson et al., 1994). In the present study, we measured activity in both the sympathetic and the parasympathetic nervous system by employing measures of electrodermal and cardiovascular activity to best capture the complex interplay between the two systems in relation to the subjective experience of achievement emotions.

#### ***Electrodermal Activity***

Electrodermal activity is defined as “autonomic changes in the electrical properties of the skin” (Braithwaite et al., 2015, p. 3). These are driven by changes in the production of sweat by eccrine sweat glands that are located in the subdermal layer: The more sweat is being produced by these glands, the better electric signals can travel across the skin (Christopoulos et al., 2019; Dawson et al., 2007). These eccrine sweat glands are innervated by the sympathetic nervous system only, making measures of electrodermal activity a valuable indicator of sympathetic activity without any parasympathetic influences (Kreibig & Gendolla, 2014).

Electrodermal activity is commonly measured by attaching two electrodes to the skin, usually to the palm of the non-dominant hand. A weak direct current is applied, enabling the measurement of how the current is conducted through the skin (for details on calculations, see Boucsein et al., 2012). The resulting signal can be divided into the tonic and the phasic components. The tonic component, also referred to as skin conductance level (SCL), is considered “a constantly moving baseline” (Braithwaite et al., 2015, p. 5) in which changes occur relatively slowly (Christopoulos et al., 2019; Dawson et al., 2007). The phasic component, in contrast, shows fast changes in response to stimuli – hence it is termed the skin conductance response (SCR). These responses form peaks in the electrodermal activity signal on top of the tonic component. Various features of these peaks can be extracted, such as frequency, amplitude, or latency after a stimulus (Boucsein et al., 2012; Christopoulos et al., 2019; Dawson et al., 2007). Both SCR and SCL are assumed to be positively influenced by sympathetic arousal, yet, surprisingly little is known about the different meanings of the two measures, beyond the notion that the skin conductance level reflects a general state of arousal, whereas the skin conductance response reacts more rapidly to certain events in a situation (Braithwaite et al., 2015; Nagai et al., 2004). SCR and SCL have been used, for example, in research on collaborative work (Ahonen et al., 2018), appraisals of feedback (Kreibig et al., 2012), and other learning-related contexts (see review by Horvers et al., 2021). Based on the use of both SCR and SCL in educational psychology research, we investigated both measures regarding their relationships with the subjective experience of achievement emotions.

### ***Cardiovascular Activity***

Cardiovascular activity is subject to the influences of both the sympathetic and the parasympathetic nervous system. Their effects on the cardiovascular system are antagonistic in that “sympathetic nerve activity exerts an excitatory effect and parasympathetic nerve activity exerts an inhibitory effect on cardiac function” (Wehrwein et al., 2016, p. 1268). For heart rate (HR), usually defined as the number of heart beats per minute, this means that sympathetic and parasympathetic influences are inseparable: An increase in HR can be driven by either increased sympathetic activity, decreased parasympathetic activity, or both. In instances of coactivation or coinhibition where sympathetic and parasympathetic activity increase or decrease simultaneously, these contrasting influences would result in an absence of observable changes in HR (Berntson et al., 1994; Weissman & Mendes, 2021).

Another measure of cardiovascular activity is heart rate variability (HRV), which describes changes in the duration of intervals between successive heartbeats (Laborde et al., 2017; Quintana & Heathers, 2014). As HRV is mainly subject to parasympathetic influences, it

is often used as an indicator of parasympathetic activity (Chapleau & Sabharwal, 2011; Kleiger et al., 2005). Common measures of HRV quantify variability in the intervals between heartbeats, with higher variability indicating higher levels of parasympathetic activity (Kleiger et al., 2005; Laborde et al., 2017). Altogether, the use of both electrodermal and cardiovascular parameters allows for a comprehensive assessment of both the sympathetic and the parasympathetic functions of the autonomic nervous system, thus providing fertile ground for investigating associations with the subjective experience of achievement emotions.

### ***6.3.3 Between-Person versus Within-Person Approaches to Investigating Links between Autonomic Processes and the Subjective Experience of Achievement Emotions***

The question of how autonomic processes relate to the subjective experience of achievement emotions can be approached in different ways. One can use a between-person approach by comparing how differences between individuals in one variable, such as a physiological measure, relate to differences between individuals on another variable, such as a self-reported achievement emotion. For example, a between-person study may ask if individuals who report being more anxious about a test than others are also characterized by higher heart rates and skin conductance responses during test situations. Because psychophysiological baseline values differ between individuals, physiological measures are usually controlled for baseline physiological activity (e.g., Harley et al., 2019; Kiuru, Malmberg, et al., 2022), although studies differ in their approach of doing this (Horvers et al., 2021).

In contrast, studies employing a within-person approach investigate how changes in one variable within an individual relate to changes in another variable within the same individual. For example, such studies may explore whether differences in anxiety at different timepoints during a test, such as more anxiety in the beginning than in the end, are linked with corresponding differences in HR, such as higher HR in the beginning compared to the end. As such, by selecting between- and within-person approaches, researchers target different research questions that can each be highly relevant in their own right – basically, between-person studies ask “who,” while within-person studies ask “when” (see also Frenzel et al., 2020). As patterns on the between- and within-person level can differ, it is crucial to interpret the results according to the respective approach, instead of interpreting findings from a between-person study in a within-person way and vice versa (Molenaar & Campbell, 2009; Murayama et al., 2017; Voelkle et al., 2014). A lack of a clear distinction between associations on the two levels might explain the mixed findings on associations between physiological and self-report measures of emotions (Pekrun, 2023; Zawadzki et al., 2017). The present work acknowledges the

importance of this distinction by exploring associations both on the between- and the within-person level.

#### ***Associations on the Between-Person Level***

A handful of studies inspected the between-person links between psychophysiological measures and the subjective experience of achievement emotions, with mixed results. In a study by Harley et al. (2019), 37 medical students completed a diagnostic reasoning simulation task. Participants' normalized SCL was positively related to their self-reported shame and anxiety, while not showing any significant associations with hopelessness or pride. Another study involved a sample of 190 high-school students completing two different sets of tasks: non-verbal reasoning (Kiuru, Trög, et al., 2022), and math and reading tasks (Kiuru, Malmberg, et al., 2022). In the Kiuru, Trög, et al. (2022) study, the authors reported significant negative between-person correlations of SCR with hopelessness and of HRV with hope. In the Kiuru, Malmberg, et al. (2022) study, there was also a significant negative between-person correlation between HRV and hope. In contrast to the non-verbal reasoning task, however, SCR was uncorrelated with hopelessness, while HR and SCR were negatively correlated with enjoyment (Kiuru, Malmberg, et al., 2022). In neither of the tasks, any significant between-person correlations were found between SCR, HR, or HRV and self-reported anger, anxiety, or fear (Kiuru, Malmberg, et al., 2022; Kiuru, Trög, et al., 2022).

Overall, on the between-person level, some positive and negative associations between physiological processes and self-reported achievement emotions have been found, whereas most investigated relationships appeared to be non-significant. Even within the same sample on two different tasks, there were discrepancies in the between-person correlations, suggesting that associations between physiological measures and self-reported achievement emotions on the between-person level are weak and inconsistent at best. This is consistent with the broader emotion literature, arguing that links between the subjective emotional experience and physiological states happen at the within-person level, which does not necessarily mean that the same holds true for the between-person level (Mauss et al., 2024; Zawadzki et al., 2017).

#### ***Associations on the Within-Person Level***

Pekrun (2023) argued that “physiological processes occur within persons, and theories on their functions for motivation, emotions, and engagement typically refer to within-person mechanisms” (p. 7). This calls for a within-person approach to explore associations between the subjective momentary experience of achievement emotions and physiological states. Based on the arousal dimension from the three-dimensional taxonomy of achievement emotions, we would expect that, within the same individual, a more intense experience of activating emotions

would be accompanied by higher physiological arousal, whereas a more intense experience of deactivating emotions would go along with lower physiological arousal (Pekrun et al., 2023). For the deactivating emotion of boredom, an experience-sampling study by Ketonen et al. (2023) and a laboratory study by Huber and Bannert (2023) supported this idea by showing a negative association between HR and self-reported boredom on the within-person level.

Furthermore, the papers by Kiuru and colleagues also provided a broad range of insights into associations between physiological and self-report measures on the within-person level. They employed a non-verbal reasoning task consisting of two blocks (Kiuru, Trög, et al., 2022), as well as a math and reading task consisting of four blocks (Kiuru, Malmberg, et al., 2022), allowing them to explore the within-person covariation between changes in physiology and self-reported emotions. Despite the inclusion of a broad range of emotions (enjoyment, hope, anger, anxiety, fear, hopelessness, and surprise) and various physiological measures (SCR, HR, HRV; Kiuru, Malmberg, et al., 2022; Kiuru, Trög, et al., 2022), most within-person correlations were not statistically significant. Of the correlations that reached statistical significance, some were in line with the pattern we would expect based on the arousal dimension of achievement emotions. Those included positive correlations between HR and anger, anxiety, and fear (Kiuru, Malmberg, et al., 2022), as well as a negative correlation between HRV and surprise (Kiuru, Trög, et al., 2022). However, the authors also reported negative within-person correlations of HR with enjoyment and hope, and positive correlations between HR and hopelessness, and between HRV and enjoyment (Kiuru, Malmberg, et al., 2022), which are not in line with expectations.

Altogether, studies on the within-person level paint a mixed picture of whether and how physiological measures relate to the subjective experience of achievement emotions. While some results corroborate the notion that the arousal dimension of achievement emotions, as conceptualized in taxonomies of these emotions (Pekrun et al., 2023), might determine the direction of associations of physiological states and the affective experience, some findings do not align with this view. In the two sets of results by Kiuru et al., it further stands out that most within-person associations between physiological and self-report measures were non-significant. One reason for this likely was the small number of within-person observation (four in Kiuru, Malmberg, et al., 2022; two in Kiuru, Trög, et al., 2022). Clearly, we are still lacking a comprehensive understanding of whether and how psychophysiological measures relate to the subjective experience of achievement emotions, in particular on the within-person level.

### **6.3.4 The Present Research**

The present research addresses the lack of understanding of the relations between psychophysiological processes and the subjective experience of achievement emotions. To investigate these relations the present two studies are characterized by four core features: (1) The use of a repeated-measures design with multiple assessments across the course of an achievement situation to model associations on the between-person and within-person level; (2) inclusion of multiple physiological variables (SCR, SCL, HR, HRV); (3) inclusion of multiple discrete achievement emotions (enjoyment, pride, anger, anxiety, frustration, and boredom); and (4) replication across two independent samples.

The present studies thus provide insights into whether and to what extent physiology and subjective experience can be considered as two “loosely coupled” (Pekrun et al., 2023, p. 146) components of achievement emotions. Further, the present findings contribute to a better understanding of whether and how physiological measures can be used to supplement or substitute self-report measures of achievement emotions in educational research and practice.

## **6.4 Study 1**

In Study 1, we employed a repeated-measures design with six blocks of a numerical reasoning task, repeatedly assessing both physiological states and the subjective experience of achievement emotions. This approach enabled the investigation of relations between physiological and self-report measures on both the between- and within-person levels.

### **6.4.1 Method**

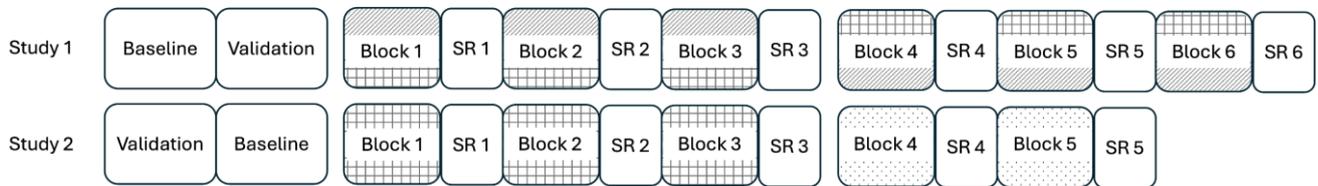
#### ***Participants***

We recruited participants via university mailing lists and social media postings. There were no restrictions on participation except the requirement of being at least eighteen years old. After removing three participants with fully missing data, the final sample comprised  $N = 87$  participants for whom self-report and at least one of the physiological measures were available and valid (68% female, 30%, 2% diverse). Their mean age was 26.79 years ( $SD = 8.78$  years,  $Min = 18$ ,  $Max = 77$ ), and 84% of them were university students. Based on pre-defined quality criteria for the physiological measures, certain measures and measurement time points were excluded for some participants. A detailed overview of the reasons for exclusion and the exclusion process can be found in Supplemental Materials Section 6.9.2. The final dataset included 417 blocks from 70 individuals for the analyses of electrodermal activity and 487 blocks from 85 individuals for the analyses of cardiovascular activity.

### *Procedure*

Participants were welcomed to the lab and signed a consent form. Subsequently, the sensors of electrodermal and cardiovascular activity were attached to their non-dominant hand. Participants were instructed to sit comfortably in front of a computer screen and move as little as possible throughout the experiment to avoid movement artifacts. The experimenter visually inspected the quality of the physiological signals. Upon ensuring their functioning, the experimenter asked the participants to follow the instructions provided on the screen.

The procedure (see Figure 1, upper panel) started with a baseline measurement of one minute, where participants were asked to relax and breathe normally. Next, they were guided through two validation tasks that ensured the validity of the electrodermal activity measure (see Supplemental Materials Section 6.9.1). Upon completion, participants received the instructions for the numerical reasoning task, where they were informed that they would perform two tests of numerical reasoning ability with the same item type. They were further informed that these items measured a component of intelligence and that they would receive feedback on their total test score at the end, to boost the importance of doing well in the task. After an explanation of the task and the presentation of an example, participants completed six task blocks, presented as three blocks in the first and three blocks in the second test, with four items each. Each item block was followed by the instruction to fill in a self-report questionnaire in a printed booklet on the table in front of them. Upon completion of the self-report questionnaire, they proceeded with the next item block on the computer screen. After the third block and self-report questionnaire, participants could take a break for as long as they wanted, to then continue with the test. Finally, after finishing the last block and questionnaire, participants were told how many of the items they had solved correctly, debriefed, and awarded either monetary compensation or participant credits for their participation.

**Figure 1***Overview of Procedure for the Separate Studies*

*Note.* SR = Self-Report Questionnaire. In Study 1, either blocks 1 to 3 were adaptive to performance (dashed shading) and blocks 4 to 6 were non-adaptive (squared shading), or vice versa. For Study 2, all blocks were non-adaptive. In this study, items used in blocks 1 to 3 were identical to those used in the non-adaptive blocks in Study 1 (squared shading); for blocks 4 and 5, new items were added (dotted shading).

### ***Numerical Reasoning Task***

Participants completed blocks of four numerical reasoning items. The items were developed by Loe et al. (2018) with known item difficulties on a Rasch scale. They were presented on a computer screen via the Concerto platform (The Psychometrics Centre, n.D.). The items consisted of number sequences, in which the participants had to identify the underlying mathematical pattern and complete the sequence with one or two numbers (see Figure 2 for a sample item). Participants were instructed to solve the tasks without taking notes and to type in their answers before pressing a button to proceed to the next item. Further, they were told that if they did not know the correct answer, they could insert “X” and proceed to the next item. There was no time limit for the completion of the tasks.

The average difficulty of items varied between participants, as this study involved a within-subject experimental variation. In three of the six blocks, the item difficulties were adaptive to the participants' performance. In the other three blocks, items followed a fixed order with increasing item difficulty (see Figure 1). Whether the first or last three blocks were adaptive to performance was counter-balanced across participants (43 participants started with the adaptive, 44 with the fixed difficulty order blocks). For the present study, the experimental variation regarding adaptive and non-adaptive item presentation is not of interest, as there is no reason to assume that the adaptivity of the task would impact underlying associations between the physiological and affective components of achievement emotions.

**Figure 2***Sample Item of Numerical Reasoning Task*

What number(s) follow in the sequence?							
11	28	56	73	146	163	<input type="text"/>	<input type="text"/>

*Note.* This item is an example and has not been used in the study. The correct answers are 326 and 343 (relation between subsequent numbers: +17, \*2, +17, \*2, etc.).

**Measures**

**Electrodermal and Cardiovascular Activity.** Electrodermal and cardiovascular activity were measured using the BioNomadix wristband connected to a Biopac MP160 receiver and Biopac Bionomadix 2CH GSR/EDA Amplifier. The wristband and sensors were attached to the participant's non-dominant hand to minimize movement artifacts. Two electrodes were attached to the palm, measuring electrodermal activity. A photoplethysmography sensor was attached to the participant's fingertip, measuring cardiovascular activity via changes in peripheral blood-volume pressure. The data was collected at a frequency of 500 Hz, and the data streams were transmitted and synchronized in the software iMotions (2022). In case of unexpected events that might cause artifacts in the data (e.g. interruptions, loud noises, talking, sneezing), the raw data was inspected, and in case of visible artifacts, the respective time window was removed from the data. Next, the raw data were processed to yield the parameters of interest for the present research question.

For electrodermal activity, we applied a peak detection algorithm as provided by the iMotions software (see Supplemental Materials Section 6.9.3 for parameters). This algorithm separated the phasic and tonic signals and provided us with an aggregated measure of peaks per minute in the SCR for each item block and the baseline. For the tonic signal, aggregation was done manually by averaging the tonic signal across all measurements in the period of interest, resulting in one value of SCL for the baseline period and each block.

For cardiovascular activity, the raw data were exported into the software Kubios HRV Scientific, where it was corrected for artifacts and episodes of low data quality (Kubios Oy, 2024; Tarvainen et al., 2014). We then extracted HR, measured as beats per minute, for the baseline and each block. Further, we extracted the root mean square of successive differences as a measure of HRV, which is a commonly used time-domain based indicator of parasympathetic activity in short-term settings (Laborde et al., 2017). Settings used for these analyses can be found in Supplemental Materials Section 6.9.3.

**Subjective Experience of Achievement Emotions.** After each block, participants were asked to report their subjective emotional experience (“Please indicate how much the following applies to you right now”). They indicated their agreement on a five-point Likert scale (1 = not true at all, 2 = hardly true, 3 = somewhat true, 4 = mostly true, 5 = completely true) to the following statements: “I am enjoying this” (enjoyment), “I feel proud” (pride), “I feel angry” (anger), “I feel bored” (boredom), “I feel frustrated” (frustration), “I feel tense and nervous” (anxiety).

### ***Data Analysis***

The four physiological measures were corrected for baseline differences between individuals. We did this by running a simple linear regression model, with all blocks of SCR, SCL, HR, and HRV being predicted by their respective baseline values. Next, we extracted the residuals for the predicted variables, which indicate the deviation from the expected value based on the baseline assessment. These baseline-corrected values for SCR, SCL, HR, and HRV were used to calculate between- and within-person correlations with the self-reported achievement emotions of enjoyment, pride, anger, anxiety, frustration, and boredom. To do so, each pair of variables was entered separately into the `multilevel.cor` function of the `misty` package (Version 0.6.8; Yanagida, 2020) in R (Version 4.4.0; R Core Team, 2024). We used the maximum likelihood estimator with Huber-White robust standard errors to account for non-normal distributions. For all results, we applied a significance level of  $\alpha = .05$ .

### ***Ethics***

This research was conducted in line with the APA ethical standards. Participation in the study was voluntary, and all participants provided informed consent.

### ***Transparency and Openness***

All missing data and exclusions of participants from the analyses are described in detail in Supplemental Materials Section 2. The data and analysis code for both studies are available on OSF ([https://osf.io/hkq9c/?view\\_only=26794c6e282e4426b99fd613298fab30](https://osf.io/hkq9c/?view_only=26794c6e282e4426b99fd613298fab30)). The data from Study 1 have previously been used to investigate the effects of adaptive versus fixed-item testing on emotional and physiological states. The data from Study 2 have previously been used to investigate trait self-concept as a predictor of individuals’ emotional and physiological states on a between-person level. The studies were not pre-registered. Artificial intelligence (ChatGPT-4; OpenAI, 2025) has been utilized to develop and refine code for the data analysis in R and to improve language and style in the present manuscript.

### **6.4.2 Results**

#### ***Descriptives***

The numerical reasoning items used in the task were solvable, yet challenging for participants, as indicated by an average solution rate of 2.07 out of four items per block. Participants reported experiencing a range of emotions during the test, with average values across all individuals and blocks ranging between two and three on the five-point scale for enjoyment, pride, anger, anxiety, and frustration. Meanwhile, the average endorsement of the boredom item was more towards the lower end of the scale. All emotions showed considerable between- and within-person variance. For enjoyment and anxiety, 70% and 75% of the variance was attributable to differences on the between-person level, respectively. For all other emotions, this proportion ranged between 53% and 58% (see Table 1).

To ensure that participants' emotional experience was related to their achievement, we inspected correlations between task performance, that is, the number of correctly solved items, and self-reported achievement emotions on both the between- and within-person level (Table 2). Anxiety was not correlated with task performance on either level, and boredom showed a small yet significant negative correlation with task performance only on the within-person level. For all other emotions, substantial links between performance and subjective emotional experience were found: Participants who performed better overall also reported higher levels of the positive emotions of enjoyment and pride, and lower levels of the negative emotions of anger and frustration. The same pattern also appeared on the within-person level, with individuals reporting significantly higher levels of enjoyment and pride, and lower levels of the negative emotions of anger and frustration on blocks where they solved more items correctly compared to other blocks. These correlations highlight that the reported emotions are related to achievement during the task, supporting their classifications as achievement emotions. Correlations between the self-reported emotions are reported in Supplemental Materials 6.9.4.

For the physiological measures SCR, SCL, HR, and HRV, descriptive information is displayed in Table 1. Their intercorrelations are reported in Supplemental Materials 6.9.4. Considerable proportions of the variance in the baseline-corrected physiological measures could be accounted for by between-person differences, as indicated by high intraclass correlations representing the proportion of between-person relative to the total variance. The intraclass correlations showed that SCL was the psychophysiological measure with the least within-person variance (see Table 1). However, there also clearly was substantial within-person variability across blocks for all psychophysiological measures and self-reported emotions, hence inspection into their covariance on the within-person level was reasonable.

**Table 1***Descriptive Information for Variables in Study 1*

	<i>n</i> <sub>Blocks</sub>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>ICC1</i>
Score	519	2.07	1.25	0.00	4.00	.19
<b>SCR</b>						
Raw	417	5.60	3.20	0.00	15.15	.73
Residualized	417	0.00	2.59	-5.79	8.73	.59
<b>SCL</b>						
Raw	417	2.26	1.57	0.09	7.30	.96
Residualized	417	0.00	0.66	-1.53	2.99	.76
<b>HR</b>						
Raw	492	79.85	11.43	53.41	124.30	.89
Residualized	487	0.00	5.73	-16.04	37.93	.58
<b>HRV</b>						
Raw	492	39.91	22.77	7.19	187.91	.83
Residualized	487	0.00	14.04	-45.83	150.58	.55
Enjoyment	519	2.85	1.31	1.00	5.00	.70
Pride	518	2.21	1.14	1.00	5.00	.55
Anger	518	2.17	1.12	1.00	5.00	.53
Anxiety	519	2.66	1.16	1.00	5.00	.75
Frustration	519	2.75	1.21	1.00	5.00	.55
Boredom	519	1.68	0.95	1.00	5.00	.58

*Note.* ICC1 was obtained using the statsBy function of the psych package in R (Version 2.4.6.26; Revelle, 2024). For the physiological variables, descriptives are reported for the raw values as well as the baseline-corrected (residualized) values, which are used in the analyses. SCR = skin conductance response in peaks per minute, SCL = skin conductance level in microSiemens, HR = heart rate in beats per minute, HRV = heart rate variability as the root mean square of successive differences.

**Table 2***Correlations between Task Performance and Emotions in Study 1*

	Between-Person			Within-Person		
	<i>r</i>	<i>SE</i>	<i>p</i>	<i>r</i>	<i>SE</i>	<i>p</i>
Enjoyment	<b>.56</b>	0.10	< .001	<b>.23</b>	0.05	< .001
Pride	<b>.47</b>	0.15	.001	<b>.12</b>	0.05	.015
Anger	<b>-.48</b>	0.12	< .001	<b>-.11</b>	0.05	.036
Anxiety	-.05	0.14	.736	-.06	0.05	.258
Frustration	<b>-.46</b>	0.14	< .001	<b>-.14</b>	0.05	.003
Boredom	-.16	0.14	.251	<b>-.12</b>	0.05	.024

*Note.* **Bold** font highlights statistically significant ( $p < .05$ ) correlations. Task performance is the number of correctly solved items out of four per block.

### ***Bivariate Correlations between Physiology and Self-Report***

Table 3 displays the bivariate between-person and within-person correlations of the physiological measures and the self-report variables. On the between-person level, none of the bivariate correlations between the baseline-corrected physiological measures and the emotions reached statistical significance ( $p > .05$ ). Thus, there appears to be no substantive pattern of between-person relationships between individuals' baseline-corrected physiological states and their subjective experience of achievement emotions indicated via self-report.

On the within-person level (Table 3, Figure 3), there were several significant correlations between physiological measures and self-reported achievement emotions. The positive emotions of enjoyment and pride correlated positively with SCR and HR, and negatively with HRV. Further, enjoyment and pride correlated negatively with SCL, although this parameter was statistically significant only for enjoyment. Self-reported anger correlated negatively with SCR; the associations with SCL, HR, and HRV were non-significant. Frustration correlated negatively with SCR and positively with SCL, but did not show significant associations with HR and HRV. Last, boredom correlated negatively with SCR and HR, and positively with HRV. All correlations were small to moderate in size.

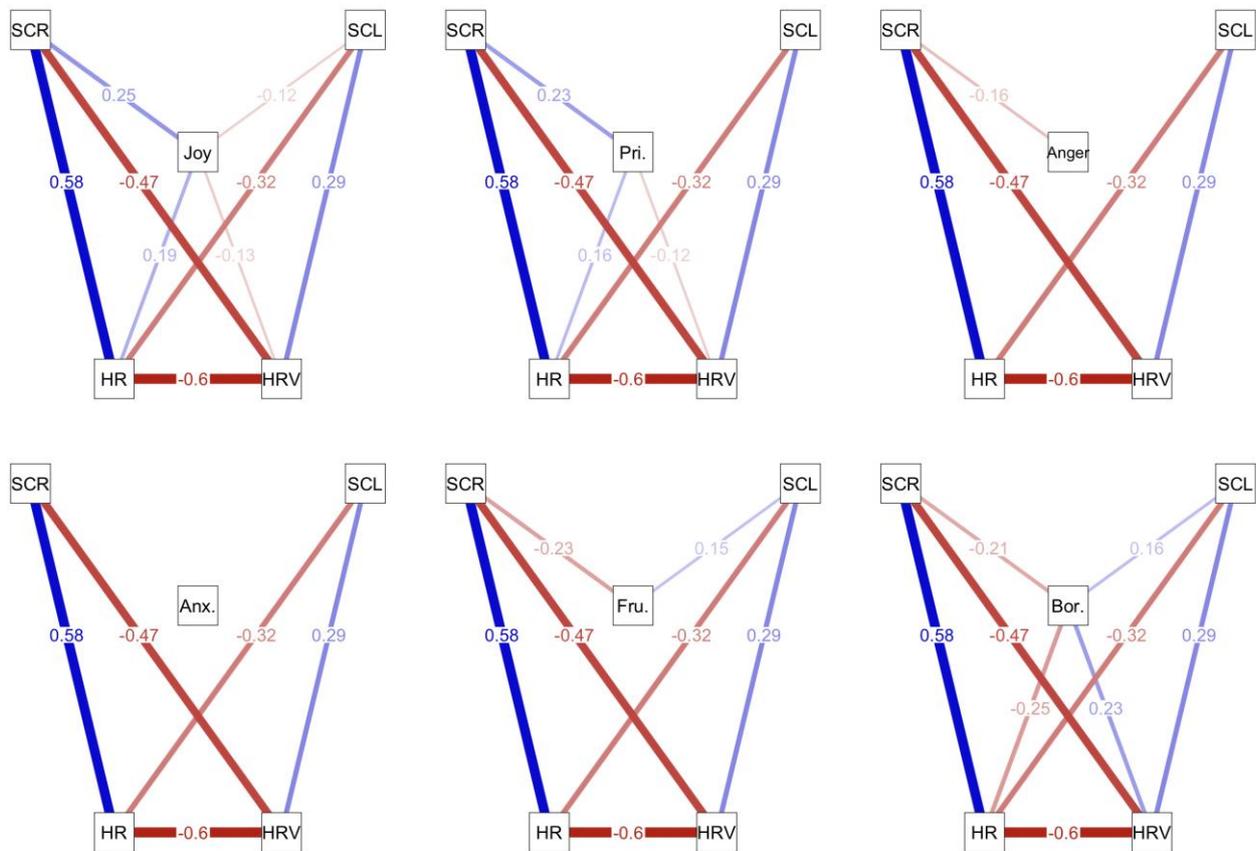
**Table 3**  
*Correlations between Physiological Measures and Self-Reported Achievement Emotions in Study 1*

Emotion	SCR			SCL			HR			HRV		
	<i>r</i>	<i>SE</i>	<i>p</i>	<i>r</i>	<i>SE</i>	<i>p</i>	<i>r</i>	<i>SE</i>	<i>p</i>	<i>r</i>	<i>SE</i>	<i>p</i>
	Between-Person											
Enjoyment	-.01	0.15	.922	.01	0.12	.966	-.01	0.13	.932	-.19	0.12	.115
Pride	-.19	0.13	.143	-.04	0.12	.750	.01	0.12	.944	-.07	0.09	.453
Anger	.06	0.13	.629	.23	0.13	.072	.16	0.11	.131	-.02	0.11	.838
Anxiety	.11	0.13	.390	.17	0.11	.126	.12	0.12	.317	-.02	0.11	.833
Frustration	.08	0.15	.587	.16	0.11	.121	.09	0.12	.424	.04	0.10	.697
Boredom	-.11	0.14	.465	-.11	0.13	.430	-.11	0.10	.279	.17	0.16	.311
	Within-Person											
Enjoyment	<b>.25</b>	0.06	<.001	<b>-.12</b>	0.05	.025	<b>.19</b>	0.05	<.001	<b>-.13</b>	0.04	.002
Pride	<b>.23</b>	0.07	<.001	-.11	0.07	.085	<b>.16</b>	0.05	.002	<b>-.12</b>	0.06	.040
Anger	<b>-.16</b>	0.06	.011	.08	0.07	.243	-.03	0.06	.659	.02	0.06	.662
Anxiety	-.01	0.08	.848	-.06	0.06	.349	.07	0.07	.336	-.11	0.07	.137
Frustration	<b>-.23</b>	0.05	<.001	<b>.15</b>	0.06	.012	-.08	0.05	.128	.08	0.06	.176
Boredom	<b>-.21</b>	0.05	<.001	<b>.16</b>	0.06	.006	<b>-.25</b>	0.07	<.001	<b>.23</b>	0.08	.002

*Note.* **Bold font:**  $p < .05$ . SCR = skin conductance response; SCL = skin conductance level; HR = heart rate; HRV = heart rate variability

**Figure 3**

*Within-Person Bivariate Correlations between each Emotion and the Physiological Measures in Study 1*



*Note.* Only statistically significant ( $p < .05$ ) correlations are displayed. SCR = skin conductance response; SCL = skin conductance level; HR = heart rate; HRV = heart rate variability.

### 6.4.3 Discussion

Using repeated physiological and self-report measurements during an achievement situation, the findings of this study provide an insight into the association between physiological parameters and the subjective experience of achievement emotions on both the between-person and the within-person level. The results underline the notion that these associations are mainly evident on the within-person rather than the between-person level (Mauss et al., 2024; Pekrun, 2023). Of the physiological variables, SCR showed the strongest within-person associations with self-reported emotions. Generally, SCR, HR, and HRV exhibited a consistent correlational pattern for enjoyment and pride: In blocks in which participants reported more intense positive emotions (enjoyment and pride), they showed higher sympathetic activation in terms of more frequent SCR, and relatively faster HR, alongside parasympathetic deactivation, indicated by lowered HRV. Conversely, more intense boredom was related to lower levels of SCR and HR

and higher levels of HRV, indicating sympathetic deactivation and parasympathetic activation, which aligns with the conceptualization of boredom as a deactivating emotion. Further, the evidence for elevated physiological arousal for the supposedly negative-activating emotions of anxiety, frustration, and anger on the within-person level was weak. Anxiety did not exhibit any significant correlations with the sympathetic or parasympathetic activation parameters. Frustration and anger, if anything, seemed to be linked with sympathetic *deactivation*, as indicated by lower SCR. There were no systematic within-person links between these emotions and HR or HRV. Notably, the results for SCL were not consistent with those for the other physiological measures: SCL demonstrated small yet significant negative associations with enjoyment, alongside positive associations with frustration and boredom. This pattern contrasted with that of SCR for the same emotions, which is surprising, given that both measures are assumed to increase with rising sympathetic activity. A potential limitation of this study was its experimental design, involving half of the test blocks being adaptive to participants' performance. Furthermore, the sample in this first study was rather heterogeneous in terms of age and occupation.

### 6.5 Study 2

In Study 2, we aimed to replicate the findings of Study 1 by again employing a repeated-measures design and inspecting between- and within-person associations of physiological states and the subjective experience of achievement emotions. Study 2 addressed the limitations of Study 1 by using a more homogenous sample and consistently using the same set of items with a fixed order for all participants.

#### 6.5.1 Method

##### *Participants*

In comparison to the broader sample in Study 1, the sample of Study 2 was restricted to female-only second-year university students enrolled in a primary school teaching course. The final sample for whom at least one physiological measure and self-report data were available consisted of  $N = 68$  individuals. Their mean age was 20.34 years ( $SD = 3.65$  years,  $Min = 18$  years,  $Max = 44$  years). Following the exclusion process detailed in Supplemental Materials Section 6.9.2, the final data included 328 blocks from 66 individuals for electrodermal activity and 329 blocks from 67 individuals for cardiovascular activity.

##### *Procedure*

The procedure in Study 2 was similar to Study 1, with slight changes to the order of presentation (see Figure 1, lower panel). In Study 2, the procedure started with a modified

version of the validation protocol (see Supplemental Materials Section 6.9.1), followed by an additional guided breathing task to reduce participants' initial arousal due to the novel situation in the laboratory. After this, there was a baseline measurement of three minutes, again with the instruction to relax and breathe normally. The instructions for the numerical reasoning task were the same as in Study 1, except that the task was not presented as two separate tests, but rather as one test. The procedure of the numerical reasoning test was the same as in Study 1, that is, a self-report questionnaire after each block and a break after the third block. Further, in Study 2, there were five instead of six item blocks. As in Study 1, after the last block, participants were informed about how many of the items they had solved correctly by the experimenter. Subsequently, they were debriefed and awarded either monetary compensation or participant credits for their participation.

### ***Numerical Reasoning Task***

The numerical reasoning items were again based on Loe et al. (2018). They were presented directly in the data collection software iMotions (2022). Participants completed five blocks of four items each, with the item difficulties being the same for all participants (instead of adaptive to their performance as in Study 1). The first three blocks were equivalent to the non-adaptive blocks of Study 1, meaning that item difficulty increased from the first to the twelfth item. The fourth block (after the break) started with an easier item, and then item difficulty again increased throughout the two final blocks. All instructions were the same as in Study 1.

### ***Measures***

All physiological and self-report measures were collected and processed in the same way as in Study 1. The only difference was that in Study 2, the physiological data were collected at a frequency of 200 Hz, which reduced data storage requirements and speeded up processing.

### ***Data Analysis***

Following the same procedure as in Study 1, we first corrected the physiological measures for their baseline level, using the residuals of a simple linear regression model. Next, we calculated bivariate between- and within-person correlations with the `multilevel.cor` function of the `misty` package (Version 0.6.8; Yanagida, 2020) in R (Version 4.4.0; R Core Team, 2024), using the maximum likelihood estimator with Huber-White robust standard errors. For all results, we applied a significance level of  $\text{Alpha} = .05$ .

### **6.5.2 Results**

#### ***Descriptives***

Descriptive information is provided in Table 4. Similar to Study 1, participants solved on average two out of four items per block, indicating that the task was challenging yet solvable for them. Participants experienced a range of emotions with mean values for enjoyment, pride, anger, anxiety, and frustration between 2.13 and 3.13 on the five-point scale. Again, boredom was experienced to a lesser extent, with a mean value towards the lower end of the scale. The ICCs for the self-reported emotions were comparable to those in Study 1, ranging between 50% and 67% of variance being attributable to between-person differences.

All self-reported achievement emotions, except for anxiety, showed significant correlations with the number of correctly solved items (Table 5). The positive emotion enjoyment was positively associated with the test scores on the between- and within-person level; pride and test scores were significantly positively associated only on the within-person level. The negative emotions of anger, frustration, and boredom were negatively associated with the test scores on both the between- and within-person levels. This again underlines the assumption that the reported emotions are related to individuals' achievement on the task, supporting their classification as achievement emotions.

Table 4 displays descriptive information for the physiological measures. Again, most of the variance in SCL, HR, and HRV could be accounted for by between-person differences, with SCL again being the physiological measure with the smallest within-person variance. In contrast, for SCR, only 42% of the total variance was attributed to between-person differences, leaving a considerable amount of within-person variance (see Table 4). Intercorrelations between the self-reported achievement emotions and between the physiological measures are reported in Supplemental Materials Section 6.9.4.

#### ***Bivariate Correlations between Physiology and Self-Report***

On the between-person level, the bivariate correlations again showed hardly any significant associations between physiological and self-report measures ( $p > .05$ ). Exceptions were a significant negative correlation of anxiety with SCR, and the positive correlations of anxiety and frustration with HRV (see Table 6). Thus, similar to Study 1, no substantial pattern of correlations on the between-person level emerged.

On the within-person level, all four physiological measures were significantly correlated with three or more of the self-reported achievement emotions (see Table 6, Figure 4). The pattern of results largely replicated the pattern found in Study 1. Enjoyment and pride correlated positively with SCR and HR, and negatively with SCL and HRV. Anger again correlated

negatively with SCR and HR. In contrast to Study 1, anxiety significantly correlated positively with SCR and HR, and negatively with HRV. Frustration and boredom correlated negatively with SCR and HR, and frustration additionally correlated positively with SCL.

**Table 4***Descriptive Information for Variables in Study 2*

	<i>n</i> <sub>Blocks</sub>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>ICC1</i>
Score	337	1.95	1.32	0.00	4.00	.11
SCR						
Raw	328	6.22	3.05	0.00	14.52	.67
Residualized	328	0.00	2.31	-5.35	7.77	.42
SCL						
Raw	328	2.87	1.61	0.39	8.66	.95
Residualized	328	0.00	0.69	-1.86	2.47	.72
HR						
Raw	333	84.95	11.81	57.41	124.08	.91
Residualized	329	0.00	5.68	-14.02	18.93	.61
HRV						
Raw	333	37.70	19.85	8.99	120.65	.91
Residualized	329	0.00	9.87	-27.19	29.61	.62
Enjoyment	337	3.13	1.17	1.00	5.00	.67
Pride	337	2.26	1.14	1.00	5.00	.62
Anger	337	2.13	1.22	1.00	5.00	.58
Anxiety	337	2.55	1.09	1.00	5.00	.63
Frustration	337	2.58	1.28	1.00	5.00	.50
Boredom	335	1.53	0.88	1.00	5.00	.67

*Note.* ICC1 was obtained using the statsBy function of the psych package in R (Version 2.4.6.26; Revelle, 2024). For the physiological variables, descriptives are reported for the raw values as well as the baseline-corrected (residualized) values, which are used in the analyses. SCR = skin conductance response in peaks per minute, SCL = skin conductance level in microSiemens, HR = heart rate in beats per minute, HRV = heart rate variability as the root mean square of successive differences.

**Table 5***Correlations between Task Performance and Emotions in Study 2*

	Between-Person			Within-Person		
	<i>r</i>	<i>SE</i>	<i>p</i>	<i>r</i>	<i>SE</i>	<i>p</i>
Enjoyment	<b>.61</b>	0.15	< .001	<b>.52</b>	0.04	< .001
Pride	.33	0.18	.073	<b>.58</b>	0.04	< .001
Anger	<b>-.37</b>	0.19	.044	<b>-.38</b>	0.05	< .001
Anxiety	-.23	0.20	.248	-.05	0.07	.455
Frustration	<b>-.51</b>	0.17	.002	<b>-.57</b>	0.04	< .001
Boredom	<b>-.50</b>	0.14	< .001	<b>-.13</b>	0.05	.001

*Note.* **Bold** font highlights statistically significant ( $p < .05$ ) correlations. Task performance is the number of correctly solved items.

### 6.5.3 Discussion

The results of this study largely replicated the Study 1 finding that associations between physiological and self-report indicators are mainly present on the within- rather than the between-person level, with two notable exceptions. First, we observed a counter-expected negative correlation on the between-person level between self-reported anxiety and SCR. Second, in this study, there was a rather consistent pattern for HRV on the between-person level, showing medium-sized positive links between HRV and the negative emotions of anxiety, frustration, and anger, and negative links with the positive emotions of enjoyment and pride. Even if those correlations only reached statistical significance for anxiety and frustration, there appears to be a pattern implying that those individuals who experienced the test situation as more unpleasant reacted with a stronger parasympathetic activation than those individuals who experienced the situation as more pleasant. Importantly, however, there was no corresponding pattern on the other three physiological measures. Since we only find this pattern in Study 2, and just for one of the physiological indicators, we are hesitant to interpret this substantially without further replication. Nevertheless, the findings provide a hint towards HRV possibly being a noteworthy psychophysiological marker also on the between-person level in an achievement context.

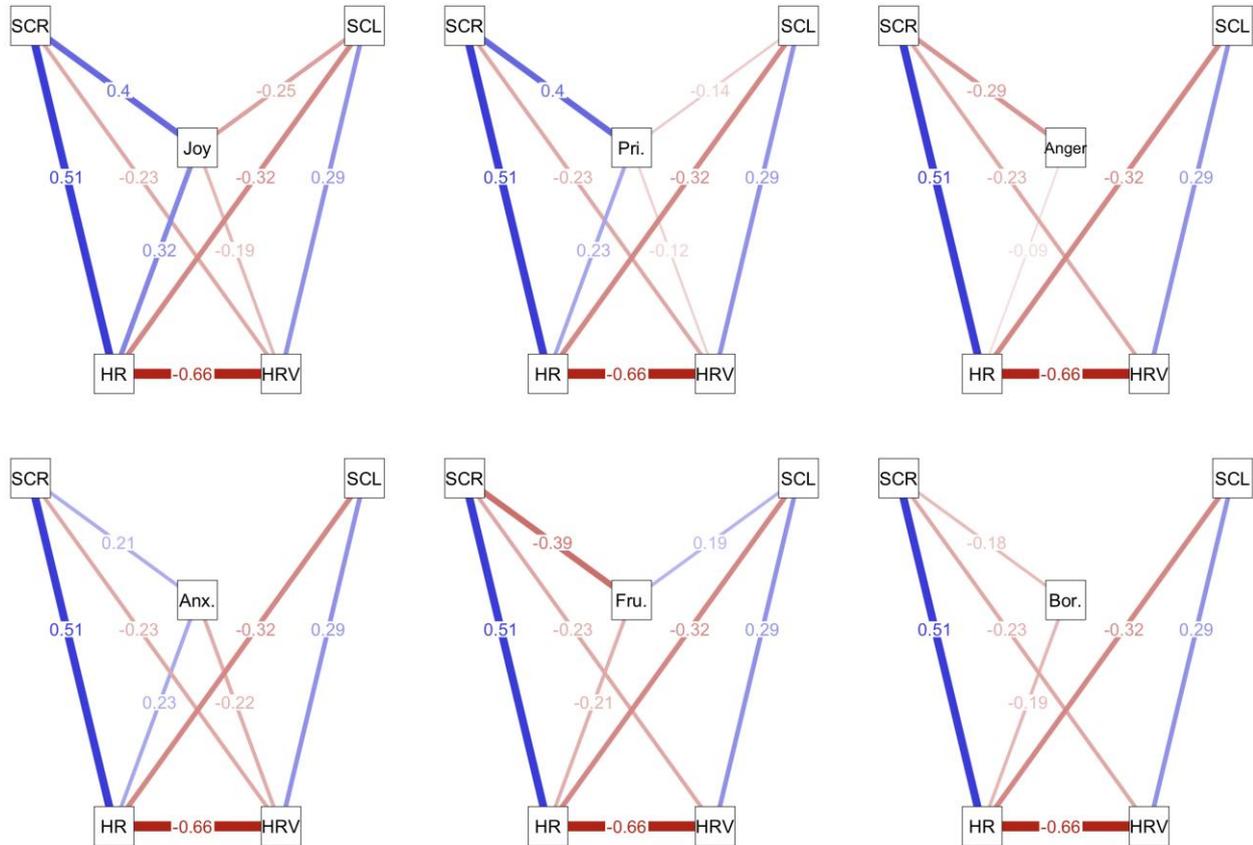
**Table 6**  
*Correlations between Physiological Measures and Self-Reported Achievement Emotions in Study 2*

Emotion	SCR			SCL			HR			HRV		
	<i>r</i>	<i>SE</i>	<i>p</i>	<i>r</i>	<i>SE</i>	<i>p</i>	<i>r</i>	<i>SE</i>	<i>p</i>	<i>r</i>	<i>SE</i>	<i>p</i>
	Between-Person											
Enjoyment	-.04	0.15	.790	.02	0.10	.841	-.01	0.14	.943	-.21	0.12	.079
Pride	-.15	0.11	.203	-.08	0.11	.473	.04	0.16	.816	-.24	0.12	.052
Anger	-.10	0.13	.435	-.15	0.12	.225	-.17	0.13	.193	.24	0.13	.071
Anxiety	<b>-.26</b>	0.12	.031	-.02	0.13	.857	-.03	0.13	.830	<b>.37</b>	0.13	.004
Frustration	-.07	0.15	.663	-.08	0.11	.443	-.05	0.15	.737	<b>.38</b>	0.15	.009
Boredom	.02	0.14	.871	-.05	0.11	.651	-.12	0.10	.239	.00	0.10	.974
	Within-Person											
Enjoyment	<b>.40</b>	0.05	< .001	<b>-.25</b>	0.07	< .001	<b>.32</b>	0.05	< .001	<b>-.19</b>	0.05	< .001
Pride	<b>.40</b>	0.05	< .001	<b>-.14</b>	0.06	.023	<b>.23</b>	0.06	< .001	<b>-.12</b>	0.05	.014
Anger	<b>-.29</b>	0.04	< .001	.10	0.06	.082	<b>-.09</b>	0.04	.018	-.02	0.05	.736
Anxiety	<b>.21</b>	0.08	.005	-.06	0.08	.416	<b>.23</b>	0.08	.002	<b>-.22</b>	0.09	.013
Frustration	<b>-.39</b>	0.05	< .001	<b>.19</b>	0.07	.003	<b>-.21</b>	0.05	< .001	.11	0.07	.115
Boredom	<b>-.18</b>	0.07	.007	-.02	0.06	.691	<b>-.19</b>	0.08	.022	.09	0.05	.086

*Note.* **Bold font:**  $p < .05$ . SCR = skin conductance response; SCL = skin conductance level; HR = heart rate; HRV = heart rate variability.

**Figure 4**

*Within-Person Bivariate Correlations between each Emotion and the Physiological Measures in Study 2*



*Note.* Only statistically significant ( $p < .05$ ) correlations are displayed. SCR = skin conductance response; SCL = skin conductance level; HR = heart rate; HRV = heart rate variability.

On the within-person level, SCR again was the physiological measure with the most consistent and largest correlations with the subjective experience of achievement emotions. Stronger enjoyment and pride were again associated with higher sympathetic activation and parasympathetic deactivation. Stronger experiences of anger, frustration, and boredom seemed to be accompanied by sympathetic deactivation in terms of reduced SCR and HR, whereas no correlations were found for these emotions with HRV as a marker of parasympathetic activity. Notably, in this study, we additionally found the expected pattern for anxiety. More intense experience of anxiety was accompanied by sympathetic activation and parasympathetic deactivation. The results for SCL again were unexpected. The significant correlations of SCL with enjoyment, pride, and frustration were in the opposite direction from those of SCR.

## 6.6 General Discussion

The present research systematically investigated whether and how different physiological measures relate to the subjective experience of various achievement emotions. Possible relations between variables were inspected both on a between-person and on a within-person level, as both analyses provide different insights (Frenzel et al., 2020; Hamaker, 2023; Murayama et al., 2017). To disentangle the two levels, we used repeated-measures data from two studies, in which participants were working on several blocks of a numerical reasoning task. The results from both studies provide insights into whether and how physiological measures relate to the subjective, self-reported experience of achievement emotions.

First, the results clearly show differences between associations on the between- compared to the within-person level in both studies. On the between-person level, out of 24 observed correlations (6 emotions x 4 physiological measures), none (Study 1) or only three correlations (Study 2) reached statistical significance ( $p < .05$ ). As such, our studies show that baseline-corrected physiological measures and self-reported achievement emotions are not systematically correlated on the between-person level. This finding implies that differences between individuals' physiological arousal levels across a given limited timeframe (such as working on an achievement task for 30-40 minutes) do not allow for insights into differences in their subjective emotional experience, and vice versa, even after controlling for baseline differences in physiological measures. In other words, if an individual shows higher physiological arousal across a given timeframe (such as working on an achievement task for 30-40 minutes) than another individual, this does, for example, not allow for conclusions on whether they are experiencing higher levels of anxiety than the other person.

In contrast, on the within-person level, most of the correlations between physiological and self-report variables were statistically significant, albeit mostly small in size. Hence, within individuals, changes in their subjective emotional experiences during an achievement situation are systematically linked to changes in their psychophysiological responses. This difference in findings depending on the level of analysis aligns with previous results by Kiuru, Malmberg, et al. (2022), who also only found significant correlations between heart rate and self-reported emotions on the within-person but not the between-person level. The findings support the general notion of emotional episodes or states as “interrelated, synchronized changes” in different subsystems (Scherer, 2005, p. 697). They also confirm that such patterns are not necessarily reflected in corresponding associations among emotion components on the between-person level (Mauss et al., 2005; Pekrun, 2023).

More specifically, our findings of statistically significant, weak to moderate relations between physiological and self-report measures on the within-person level align with the multi-component view on achievement emotional states. Indeed, we found that in achievement situations, within-person changes in the subjective emotional experience are “loosely coupled” (Pekrun et al., 2023, p. 146) with physiological changes in the ANS. The small to moderate size of these associations also aligns with previous considerations, stating that strong or even perfect correspondence between the processes is unlikely, as both processes are influenced by a multitude of variables (Kreibig & Gendolla, 2014; Pekrun, 2023).

Second, the direction of the associations on the within-person level provides important insights into the relation between physiological processes and the subjective experience of achievement emotions. Pekrun (2024) proposed that achievement emotions can be classified according to three dimensions: valence, arousal, and object focus. According to the arousal dimension, an increased experience of activating emotions would be accompanied by an increase in physiological arousal, whereas deactivating emotions would be associated with reduced physiological arousal (Pekrun, 2024). According to the taxonomy, five of the six discrete emotions assessed in the present work would have been classified as activating (joy, pride, anger, anxiety, frustration), whereas only boredom would be classified as a deactivating emotion (Pekrun, 2006, 2024; Pekrun et al., 2023). In line with these theoretical assumptions, the activating emotions of enjoyment, pride, and, in Study 2, anxiety showed a pattern of sympathetic activation and parasympathetic deactivation accompanying a more intense subjective emotional experience. Further, again in line with theory and some previous findings (Huber & Bannert, 2023), our within-person bivariate correlations showed sympathetic deactivation, and in Study 1 also parasympathetic activation, being associated with higher levels of boredom.

Counter to expectations, a pattern of sympathetic deactivation appeared in both studies for the presumably activating emotions of anger and frustration. This pattern might be specific to the achievement context, as in broader emotion research, anger is commonly found to correlate positively with SCR, SCL, HR, and negatively with HRV (Kreibig, 2010). Furthermore, our finding of a sympathetically deactivating pattern for anger challenges the assumptions of the arousal dimension in the three-dimensional taxonomy of achievement emotions by Pekrun et al. (2023), which classifies anger and frustration as activating emotions. Possibly, the arousal dimension as presented in the literature so far rather represents the subjective experience of activation, which might not always correspond to the physiological

one. This view would be consistent with the non-significant correlation between SCR and the self-reported physiological component of test anxiety as reported by Roos et al. (2023).

Of the four physiological variables that we investigated, SCR stood out as the one showing the most and the largest associations with the subjective experience of achievement emotions, being significantly associated with all self-report measures in Study 2 and all but one in Study 1. In contrast, our results for SCL as the second measure of electrodermal activity were far less consistent and clear. For the few subjective emotional experiences where correlations were found, the pattern was, in fact, opposing to that of SCR, showing negative correlations with emotions when SCR indicated sympathetic activation, and vice versa – even though both measures are generally understood as indicators of sympathetic activity (Boucsein et al., 2012; Nagai et al., 2004). Our unexpected findings regarding SCL underscore the need for further investigation and clarification of the meaning of the measure of SCL as a state indicator of sympathetic activity. Clearly, these findings also suggest that measures of SCR and SCL cannot be used interchangeably in educational psychology research and practice.

Altogether, based on the three measures of SCR, HR, and HRV, our results suggest that an increase in an individual's experience of the positive achievement emotions of enjoyment and pride is accompanied by an increase in sympathetic and a decrease in parasympathetic arousal, possibly as a function of alertness and engagement with the task driven by these emotions. The same pattern of sympathetic activation and parasympathetic deactivation was found for the negative emotion of anxiety in one of the two studies. In contrast, an increased experience of the negative emotions of anger, frustration, and boredom goes along with a decrease in sympathetic arousal, possibly as a function of disengagement from the task (see Corcoran et al., 2024 for findings on lower heart rate accompanying disengagement).

### ***6.6.1 Limitations and Implications for Future Research***

The present work has some limitations, which need to be considered in interpreting the findings and can be used to derive directions for future research. First, the data used for the present studies was collected in a highly controlled laboratory environment, which ensured high data quality. However, the laboratory-based achievement situation was also rather artificial. While the subjective value for participants was increased through an ego-threat induction by informing them that the task measures a component of intelligence and that they would receive feedback on their performance, the laboratory task may not have fully resembled a real-life, high-stakes achievement situation. Therefore, the present findings should be replicated in real-life, high-stakes testing situations.

Second, although we could replicate findings across two separate studies, the generalizability of the results needs to be investigated. Specifically, both samples were recruited from the same geographic area, and participants had a relatively high level of education. This is common in studies employing physiological measures, but suggests that the findings should be replicated with samples showing different demographic characteristics.

Third, the present study focused on a set of six achievement emotions, including enjoyment, pride, anger, anxiety, frustration, and boredom. While we believe that these emotions cover a considerable part of the typical array of emotional experiences in achievement situations, future research could benefit from including a larger variety of achievement emotions. Specifically, future work could include a broader variety of positive and negative emotions classified as deactivating (Pekrun et al., 2023), beyond the single emotion of boredom.

Finally, it needs to be noted that the within-person associations investigated in the present study represented aggregates of within-person associations across participants. As such, our results reflect associations between physiological and self-report measures for an average, albeit hypothetical, person (Epskamp et al., 2022). However, this pattern might not necessarily apply to every individual. For example, on the aggregated level, there appears to be a small- to medium-sized positive within-person correlation between SCR and the subjective experience of enjoyment. However, there might be individuals for whom this correlation is quite large, whereas for others, it might be very small. Future research can delve into between-person differences in these within-person associations by allowing them to vary across individuals, and then explore personal or situational variables as predictors of the strength of these associations.

### ***6.6.2 Theoretical and Practical Implications***

For the practical application of physiological measures in the context of achievement emotions, the present results clearly show the importance of the level of analysis: On the between-person level, physiological variables and subjective emotional experience seem to be mostly unrelated, suggesting that studies would benefit from employing a within-person perspective when aiming to investigate relations between physiological states and subjective emotional experience. Similarly, practical applications that aim to adapt learning environments to learners' affective states (Plass & Pawar, 2020) using physiological measures would benefit from employing a responsive, individualized approach, where learning environments respond dynamically to within-person changes (Harley et al., 2017).

Due to the rather small size of the within-person correlations observed in the present, as well as the non-specificity of physiological patterns in mapping on certain discrete emotions, we conclude that physiological measures are a good way to supplement, but not substitute, self-

report measures of the subjective emotional experience. Based on the present findings, physiological measures might be a way to gain an insight into how the achievement situation affects the individual more broadly, that is, whether they become activated or tune out, while self-reports of the subjective experience can specify this experience more closely using different emotion categories.

Comparing the different physiological measures used in the present study (SCR, SCL, HR, and HRV) suggests that while all of them showed bivariate correlations with self-reported achievement emotions, SCR is the most closely related to the subjective experience. Possibly, this is due to SCR being the measure with the highest level of within-person variability, whereas the other measures showed considerably fewer within-person changes. Nevertheless, we follow recommendations using several physiological measures, as only a combination of various measures allows for a comprehensive insight into physiological states of both the sympathetic and the parasympathetic nervous system (Kreibig & Gendolla, 2014; Weissman & Mendes, 2021).

### **6.7 Conclusion**

The present research contributes to our understanding of the associations between physiological states and the subjective experience of achievement emotions. The results underline the importance of investigating these associations in a situated manner. Specifically, our findings across two studies consistently suggest that there are significant correlations between self-reported achievement emotions and physiological arousal, primarily on the within-person level in achievement settings. The associations were most pronounced for SCR, compared to the measures of SCL, HR, and HRV. Further, we found that certain self-reported emotions, like anger, which are considered activating in the general emotion literature, showed negative within-person correlations with sympathetic arousal in an achievement situation. Our findings also provide a foundation for future research by raising new questions. Answering these questions will further enhance our comprehension of the meaning of physiological states in achievement emotions and the application of their measurement in research and practice.

## 6.8 References

- Ahonen, L., Cowley, B. U., Hellas, A., & Puolamäki, K. (2018). Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific Reports*, *8*(1), 3138. <https://doi.org/10.1038/s41598-018-21518-3>
- Berntson, G. G., Cacioppo, J. T., Quigley, K. S., & Fabro, V. T. (1994). Autonomic space and psychophysiological response. *Psychophysiology*, *31*(1), 44–61. <https://doi.org/10.1111/j.1469-8986.1994.tb01024.x>
- Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Fillion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for psychological experiments*. <https://www.birmingham.ac.uk/documents/colleges-les/psych/saal/guide-electrodermal-activity.pdf>
- Chapleau, M. W., & Sabharwal, R. (2011). Methods of assessing vagus nerve activity and reflexes. *Heart Failure Reviews*, *16*(2), 109–127. <https://doi.org/10.1007/s10741-010-9174-6>
- Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2019). The body and the brain: Measuring skin conductance responses to understand the emotional experience. *Organizational Research Methods*, *22*(1), 394–420. <https://doi.org/10.1177/1094428116681073>
- Corcoran, A. W., Le Coz, A., Hohwy, J., & Andriillon, T. (2024). *When your heart isn't in it anymore: Cardiac correlates of task disengagement*. bioRxiv. <https://doi.org/10.1101/2024.06.21.599851>
- Dawson, M. E., Schell, A. M., Fillion, D. L., & Berntson, G. G. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology* (3rd ed., pp. 157–181). Cambridge University Press. <https://doi.org/10.1017/CBO9780511546396.007>
- Epskamp, S., Hoekstra, R. H. A., Burger, J., & Waldorp, L. J. (2022). Longitudinal design choices: Relating data to analysis. In A.-M. Isvoranu, S. Epskamp, L. J. Waldorp, & D. Borsboom (Eds.), *Network Psychometrics with R* (1st ed., pp. 157–168). Routledge. <https://doi.org/10.4324/9781003111238-12>
- Frenzel, A. C., Fiedler, D., Marx, A. K. G., Reck, C., & Pekrun, R. (2020). Who enjoys teaching, and when? Between- and within-person evidence on teachers' appraisal-emotion links. *Frontiers in Psychology*, *11*, 1092. <https://doi.org/10.3389/fpsyg.2020.01092>
- Hamaker, E. L. (2023). The within-between dispute in cross-lagged panel research and how to move forward. *Psychological Methods*. <https://doi.org/10.1037/met0000600>
- Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science*, *47*(2), 151–180. <https://doi.org/10.1007/s11251-018-09480-z>
- Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International*

- Journal of Artificial Intelligence in Education*, 27(2), 268–297.  
<https://doi.org/10.1007/s40593-016-0126-8>
- Horvers, A., Tombeng, N., Bosse, T., Lazonder, A. W., & Molenaar, I. (2021). Detecting emotions through electrodermal activity in learning contexts: A systematic review. *Sensors*, 21(23), 7869. <https://doi.org/10.3390/s21237869>
- Houtveen, J. H., & De Geus, E. J. C. (2009). Noninvasive psychophysiological ambulatory recordings: Study design and data analysis strategies. *European Psychologist*, 14(2), 132–141. <https://doi.org/10.1027/1016-9040.14.2.132>
- Huber, K., & Bannert, M. (2023). What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements. *Journal of Computers in Education*, 10(1), 189–215. <https://doi.org/10.1007/s40692-022-00228-w>
- Ikezawa, S., Corbera, S., Liu, J., & Wexler, B. E. (2012). Empathy in electrodermal responsive and nonresponsive patients with schizophrenia. *Schizophrenia Research*, 142(1–3), 71–76. <https://doi.org/10.1016/j.schres.2012.09.011>
- iMotions* (Version 9.3). (2022). [Software].
- Ketonen, E. E., Salonen, V., Lonka, K., & Salmela-Aro, K. (2023). Can you feel the excitement? Physiological correlates of students' self-reported emotions. *British Journal of Educational Psychology*, 93(S1), 113–129. <https://doi.org/10.1111/bjep.12534>
- Kiuru, N., Malmberg, L.-E., Eklund, K., Penttonen, M., Ahonen, T., & Hirvonen, R. (2022). How are learning experiences and task properties associated with adolescents' emotions and psychophysiological states? *Contemporary Educational Psychology*, 71, 1–16. <https://doi.org/10.1016/j.cedpsych.2022.102095>
- Kiuru, N., Trög, A.-S., Pasanen, M., Tourunen, A., Mikkonen, J., Ahonen, T., & Penttonen, M. (2022). Associations between adolescents' subjectively experienced emotions and psychophysiological reactions in achievement situations. *Merrill-Palmer Quarterly*, 68(1), 39–71. <https://doi.org/10.1353/mpq.2022.0003>
- Kleiger, R. E., Stein, P. K., & Bigger, J. T. (2005). Heart rate variability: Measurement and clinical utility. *Annals of Noninvasive Electrocardiology*, 10(1), 88–101. <https://doi.org/10.1111/j.1542-474X.2005.10101.x>
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Kreibig, S. D., & Gendolla, G. H. E. (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). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780203148211.ch31>
- Kreibig, S. D., Gendolla, G. H. E., & Scherer, K. R. (2012). Goal relevance and goal conduciveness appraisals lead to differential autonomic reactivity in emotional responding to performance feedback. *Biological Psychology*, 91(3), 365–375. <https://doi.org/10.1016/j.biopsycho.2012.08.007>
- Kubios Oy. (2024). *Kubios HRV Scientific User's Guide 2024*. <https://www.kubios.com/downloads/HRV-Scientific-Users-Guide.pdf>
- Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research – Recommendations for experiment planning, data

- analysis, and data reporting. *Frontiers in Psychology*, 8, 213.  
<https://doi.org/10.3389/fpsyg.2017.00213>
- Lange, J., Dalege, J., Borsboom, D., van Kleef, G. A., & Fischer, A. H. (2020). Toward an integrative psychometric model of emotions. *Perspectives on Psychological Science*, 15(2), 444–468. <https://doi.org/10.1177/17456916198950>
- Loe, B., Sun, L., Simonfy, F., & Doebler, P. (2018). Evaluating an automated number series item generator using linear logistic test models. *Journal of Intelligence*, 6(2), 20. <https://doi.org/10.3390/jintelligence6020020>
- 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), 175–190. <https://doi.org/10.1037/1528-3542.5.2.175>
- Mauss, I. B., Zerwas, F. K., Wilhelm, F. H., & John, O. P. (2024). Coherence of emotional response systems: Theory, measurement, and benefits. In *Advances in Experimental Social Psychology* (Vol. 69, pp. 59–149). Elsevier. <https://doi.org/10.1016/bs.aesp.2023.11.002>
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(2), 112–117. <https://doi.org/10.1111/j.1467-8721.2009.01619.x>
- Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. In D. W. Putwain & K. Smart (Eds.), *BJEP Monograph Series II: Part 12 The Role of Competence and Beliefs in Teaching and Learning*. British Psychological Society. <https://doi.org/10.53841/bpsmono.2017.cat2023.6>
- Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. R., & Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: A physiological account of a “default mode” of brain function. *NeuroImage*, 22(1), 243–251. <https://doi.org/10.1016/j.neuroimage.2004.01.019>
- OpenAI. (2025). *ChatGPT* (June 16 version) [Large language model]. <https://chat.openai.com/chat>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R. (2023). Mind and body in students’ and teachers’ engagement: New evidence, challenges, and guidelines for future research. *British Journal of Educational Psychology*, 93(S1), 227–238. <https://doi.org/10.1111/bjep.12575>
- Pekrun, R. (2024). Control-Value Theory: From achievement emotion to a general theory of human emotions. *Educational Psychology Review*, 36(3), 1–36. <https://doi.org/10.1007/s10648-024-09909-7>
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students’ learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1), 36–48. <https://doi.org/10.1016/j.cedpsych.2010.10.002>
- Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., Van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology*, 124(1), 145–

178. <https://doi.org/10.1037/pspp0000448>
- Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, 52(3), 275–300.  
<https://doi.org/10.1080/15391523.2020.1719943>
- Quintana, D. S., & Heathers, J. A. J. (2014). Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in Psychology*, 5, 805.  
<https://doi.org/10.3389/fpsyg.2014.00805>
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing* [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W. (2024). *psych: Procedures for Psychological, Psychometric, and Personality Research* (Version R package version 2.4.6) [Software]. Northwestern University.  
<https://CRAN.R-project.org/package=psych>
- Roos, A.-L., Goetz, T., Krannich, M., Donker, M., Bieleke, M., Caltabiano, A., & Mainhard, T. (2023). Control, anxiety and test performance: Self-reported and physiological indicators of anxiety as mediators. *British Journal of Educational Psychology*, 93(S1), 72–89. <https://doi.org/10.1111/bjep.12536>
- Roos, A.-L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review*, 33(2), 579–618. <https://doi.org/10.1007/s10648-020-09543-z>
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729. <https://doi.org/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.  
<https://doi.org/10.1080/02699930902928969>
- Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, 113(1), 210–220.  
<https://doi.org/10.1016/j.cmpb.2013.07.024>
- The Psychometrics Centre. (o. J.). *Concerto Adaptive Testing Platform*. University of Cambridge. Retrieved April 2, 2024, from <https://www.psychometrics.cam.ac.uk/newconcerto>
- Venables, P. H., & Mitchell, D. A. (1996). The effects of age, sex and time of testing on skin conductance activity. *Biological Psychology*, 43(2), 87–101.  
[https://doi.org/10.1016/0301-0511\(96\)05183-6](https://doi.org/10.1016/0301-0511(96)05183-6)
- Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, 49(3), 193–213. <https://doi.org/10.1080/00273171.2014.889593>
- Wehrwein, E. A., Orer, H. S., & Barman, S. M. (2016). Overview of the anatomy, physiology, and pharmacology of the autonomic nervous system. *Comprehensive Physiology*, 1239–1278. <https://doi.org/10.1002/cphy.c150037>
- Weissman, D. G., & Mendes, W. B. (2021). Correlation of sympathetic and parasympathetic nervous system activity during rest and acute stress tasks. *International Journal of Psychophysiology*, 162, 60–68. <https://doi.org/10.1016/j.ijpsycho.2021.01.015>

- Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, *84*(3), 552–569.  
<https://doi.org/10.1016/j.biopsycho.2010.01.017>
- Yanagida, T. (2020). *misty: Miscellaneous Functions „T. Yanagida“* (0.7.1) [Software].  
<https://doi.org/10.32614/CRAN.package.misty>
- Zawadzki, M. J., Smyth, J. M., Sliwinski, M. J., Ruiz, J. M., & Gerin, W. (2017). Revisiting the lack of association between affect and physiology: Contrasting between-person and within-person analyses. *Health Psychology*, *36*(8), 811–818.  
<https://doi.org/10.1037/hea0000466>

## 6.9 Supplemental Materials

### 6.9.1 Electrodermal Activity Validation Protocol

In both studies, participants completed a validation protocol which was used to identify and exclude the electrodermal activity signal of non-responders, i.e. individuals who show no reactions in their electrodermal activity in response to stimuli (for information on non-responders, see e.g., Ikezawa et al., 2012; Venables & Mitchell, 1996). In Study 1, these tasks were to hold one’s breath for ten seconds and to bite one’s tongue for ten seconds. In Study 2, the validation stimulus of biting one’s tongue was replaced by the presentation of an anxiety-inducing short video of a person climbing a tall building. This was presented first, followed by the instruction to hold one’s breath for ten seconds.

Both the physical acts and the anxiety induction would be expected to elicit a skin conductance response in individuals (Bari et al., 2018; Kreibig, 2010). Accordingly, we classified individuals as non-responders, if they did not show any skin conductance response in response to either of the two stimuli and omitted their two electrodermal activity measures (skin conductance response and skin conductance level) from the analyses.

### 6.9.2 Exclusions and Missing Data

#### *Study 1*

Of the 90 participants tested in Study 1, for one individual, no physiological data was recorded, and for two persons, their study number was the same, meaning it was impossible to match their physiological to their self-report data. These three persons were excluded from all analyses. Furthermore, one person only completed four blocks, and for one person, physiological data could not be used for the last block. Further, 17 non-responders were identified based on the electrodermal activity validation protocol and their skin conductance response and skin conductance level not used for any analyses. Last, for 27 blocks of twelve individuals, the cardiovascular activity data (heart rate, heart rate variability) was excluded due to low data quality (defined as more than 5% of corrected beats, as recommended in Kubios

Oy, 2024), as well as the baseline measure for one individual. In the self-report data, there was missing information for one person on the last block for pride and anger.

### ***Study 2***

Of the 73 individuals participating in Study 2, no physiological data was available for three individuals. The study numbers of two persons were the same, making it impossible to match physiological and self-report data. Followingly, five individuals were excluded from all analyses. Three persons did not complete the last block and questionnaire. Based on the electrodermal activity validation protocol, skin conductance response and skin conductance level measures were omitted from two individuals. Last, four blocks of four individuals were excluded from the cardiovascular activity data (heart rate, heart rate variability) due to low data quality (more than 5% of corrected beats), as well as the baseline measure for one other individual, which led to the exclusion of the cardiovascular data from this person as no baseline-correction could be performed. In the self-report data, boredom was missing from two individuals at the third time point.

#### ***6.9.3 Analysis Settings for Physiological Measures***

Skin conductance response and skin conductance level were calculated with the Peak Detection Algorithm in iMotions (2022) with the following parameters:

- Phasic filter length: 8000 milliseconds
- Lowpass filter cutoff frequency: 5 Hz
- Peak Onset Threshold: 0.01 microSiemens
- Peak Offset Threshold: 0 microSiemens
- Peak Amplitude Threshold: 0.01 microSiemens
- Minimum Peak Duration: 500 ms

Heart rate and heart rate variability were extracted using the following settings in the software Kubios Scientific HRV (Kubios Oy, 2024; Tarvainen et al., 2014):

- Quality and beat correction: medium
- RR time series interpolation rate: 4 Hz
- RR interval detrending: smoothness priors
- Smoothing parameter: 500
- Cutoff frequency: 0.035 Hz

#### 6.9.4 Correlations between Study Variables

Multilevel bivariate correlations between the self-reported achievement emotions and physiological measures, respectively, were calculated by entering each combination separately into the `multilevel.cor` function of the `misty` package (Version 0.6.8; Yanagida, 2020) in R (Version 4.4.0; R Core Team, 2024), using the maximum likelihood estimator with Huber-White robust standard errors. For the physiological measures skin conductance response (SCR), skin conductance level (SCL), heart rate (HR), and heart rate variability (HRV), baseline-corrected (residualized) values were used. Within-person correlations are reported below, between-person correlations above the diagonal. Bold font indicates  $p < .05$ , values in brackets denote standard errors.

##### *Study 1: Correlations between Baseline-Corrected Physiological Measures*

	SCR	SCL	HR	HRV
SCR	–	<b>.49</b> (0.11)	<b>.26</b> (0.13)	.20 (0.13)
SCL	–.07 (0.09)	–	.12 (0.1)	.11 (0.14)
HR	<b>.58</b> (0.04)	<b>–.32</b> (0.1)	–	–.26 (0.14)
HRV	<b>–.47</b> (0.04)	<b>.29</b> (0.06)	<b>–.60</b> (0.04)	–

*Note.* **Bold** font highlights statistically significant ( $p < .05$ ) correlations.

##### *Study 2: Correlations between Baseline-Corrected Physiological Measures*

	SCR	SCL	HR	HRV
SCR	–	<b>.63</b> (0.08)	.21 (0.15)	<b>–.29</b> (0.12)
SCL	–.02 (0.1)	–	.10 (0.13)	.01 (0.12)
HR	<b>.51</b> (0.05)	<b>–.32</b> (0.07)	–	<b>–.57</b> (0.09)
HRV	<b>–.23</b> (0.08)	<b>.29</b> (0.07)	<b>–.66</b> (0.04)	–

*Note.* **Bold** font highlights statistically significant ( $p < .05$ ) correlations.

*Study 1: Correlations between Self-Reported Emotions*

	Joy	Pride	Anger	Anxiety	Frustration	Boredom
Joy	–	<b>.59</b> (0.09)	<b>-.45</b> (0.11)	<b>-.24</b> (0.11)	<b>-.63</b> (0.08)	<b>-.51</b> (0.08)
Pride	<b>.59</b> (0.06)	–	-.15 (0.12)	<b>-.33</b> (0.11)	<b>-.47</b> (0.11)	-.12 (0.11)
Anger	<b>-.42</b> (0.06)	<b>-.38</b> (0.07)	–	<b>.39</b> (0.11)	<b>.71</b> (0.06)	.09 (0.14)
Anxiety	<b>-.19</b> (0.07)	<b>-.21</b> (0.08)	<b>.30</b> (0.07)	–	<b>.59</b> (0.10)	-.20 (0.11)
Frustration	<b>-.50</b> (0.05)	<b>-.53</b> (0.05)	<b>.59</b> (0.04)	<b>.35</b> (0.09)	–	.09 (0.11)
Boredom	<b>-.37</b> (0.05)	<b>-.25</b> (0.06)	<b>.15</b> (0.07)	.08 (0.09)	<b>.22</b> (0.07)	–

*Note.* **Bold** font highlights statistically significant ( $p < .05$ ) correlations.

*Study 2: Correlations between Self-Reported Emotions*

	Joy	Pride	Anger	Anxiety	Frustration	Boredom
Joy	–	<b>.57</b> (0.08)	<b>-.42</b> (0.11)	<b>-.39</b> (0.11)	<b>-.68</b> (0.08)	<b>-.55</b> (0.07)
Pride	<b>.60</b> (0.04)	–	<b>-.31</b> (0.11)	<b>-.48</b> (0.10)	<b>-.55</b> (0.09)	<b>-.18</b> (0.09)
Anger	–	<b>-.38</b> (0.06)	–	<b>.47</b> (0.12)	<b>.76</b> (0.06)	<b>.42</b> (0.13)
Anxiety	<b>.45</b> (0.07)	–	<b>-.05</b> (0.08)	.09 (0.10)	<b>.76</b> (0.07)	.06 (0.10)
Frustration	–	<b>-.55</b> (0.06)	<b>.66</b> (0.05)	<b>.21</b> (0.07)	–	<b>.40</b> (0.11)
Boredom	–	<b>-.10</b> (0.09)	<b>.17</b> (0.09)	<b>-.30</b> (0.12)	<b>.10</b> (0.08)	–

*Note.* **Bold** font highlights statistically significant ( $p < .05$ ) correlations.

**6.9.5 References Supplemental Materials**

- Bari, D. S., Aldosky, H. Y. Y., Tronstad, C., Kalvøy, H., & Martinsen, Ø. G. (2018). Electrodermal responses to discrete stimuli measured by skin conductance, skin potential, and skin susceptance. *Skin Research and Technology*, *24*(1), 108–116. <https://doi.org/10.1111/srt.12397>
- Ikezawa, S., Corbera, S., Liu, J., & Wexler, B. E. (2012). Empathy in electrodermal responsive and nonresponsive patients with schizophrenia. *Schizophrenia Research*, *142*(1–3), 71–76. <https://doi.org/10.1016/j.schres.2012.09.011>
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Kubios Oy. (2024). *Kubios HRV scientific user's guide 2024*. <https://www.kubios.com/downloads/HRV-Scientific-Users-Guide.pdf>
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing* [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. <https://doi.org/10.1016/j.cmpb.2013.07.024>
- Venables, P. H., & Mitchell, D. A. (1996). The effects of age, sex and time of testing on skin conductance activity. *Biological Psychology*, *43*(2), 87–101. [https://doi.org/10.1016/0301-0511\(96\)05183-6](https://doi.org/10.1016/0301-0511(96)05183-6)
- Yanagida, T. (2020). *misty: Miscellaneous Functions „T. Yanagida“* (S. 0.7.1) [Software]. <https://doi.org/10.32614/CRAN.package.misty>



---

# 7

---

## EMPIRICAL WORK 2

### DIFFICULTY AND VALUE PERCEPTIONS PREDICT PHYSIOLOGICAL STATES IN AN ACHIEVEMENT SITUATION: A WITHIN-PERSON STUDY

---

Miriam Wunsch  
Lars-Erik Malmberg  
Anne C. Frenzel  
Franziska Eckerskorn  
Reinhard Pekrun

#### **Reference:**

Wuensch, M., Malmberg, L.-E., Frenzel, A.C., Eckerskorn, F., & Pekrun, R. (in prep.).  
Difficulty and value perceptions predict physiological states in an achievement situation: A  
within-person study.

The chapter printed here is the manuscript version as of October 15<sup>th</sup>, 2025; this manuscript  
version has not been peer reviewed.



## 7.1 Introduction

In educational psychology research, the use of physiological measures is gaining popularity. Compared to self-report measures, which are subject to several biases, physiological measures promise an unobtrusive, more objective insight into an individual's state during an achievement situation (Pekrun, 2023b). While, in theory, the use of physiological measures has the potential to advance our assessment of individuals' states in achievement situations in both research and practice (Martin et al., 2022), little is known about the specific meaning of psychophysiological changes in achievement situations.

Grounded in control-value theory (CVT; Pekrun, 2024), situated expectancy-value theory (SEVT; Eccles & Wigfield, 2020), and motivational intensity theory (MIT; Brehm & Self, 1989), we seek to contribute to a deeper understanding of the meaning of physiological states in achievement situations by investigating whether and how emotionally and motivationally relevant task appraisals are associated with physiological arousal. Specifically, we investigate within-person effects of perceived difficulty, perceived value of success, and their interaction on electrodermal and cardiovascular activity during a numerical reasoning task.

### 7.1.1 Theoretical Perspectives

The present work draws on three established theories that have identified perceived task difficulty and perceived value of success as important predictors of an individual's state in an achievement situation, namely the CVT of achievement emotions (Pekrun, 2024), the SEVT of achievement motivation (Eccles & Wigfield, 2020), and the MIT on the exertion of effort (Brehm & Self, 1989).

#### *Control-Value Theory (CVT)*

The CVT (Pekrun, 2006, 2024) describes how combinations of control and value appraisals predict the occurrence and intensity of different achievement emotions. Perceived control refers to the subjective judgment of whether one can exert an influence over a given achievement activity or outcome. Perceived value denotes the extent to which the activity is perceived as desirable (positive activity value) or undesirable (negative activity value), and to what extent it is important to achieve success (positive value of success) or to avoid failure (negative value of failure). In addition, CVT distinguishes between the value of approaching or avoiding activities and outcomes for their own sake (intrinsic value) or because they generate consequences (extrinsic value). Generally, higher levels of control are accompanied by more intense positive and less intense negative emotions, whereas low perceived control elicits more intense negative and less intense positive emotions (Pekrun et al., 2023). This association is

moderated by perceived value, in that high levels of perceived value amplify the effect of perceived control (Shao et al., 2020).

### ***Situated Expectancy-Value Theory (SEVT)***

The SEVT (Eccles & Wigfield, 2020) describes how expectancy and perceived value interact to determine motivation in achievement situations. Specifically, the theory states that expectancy and value influence choices between different achievement activities (Wigfield & Eccles, 2000), but also “the persistence with which [students] pursue those activities, the intensity of their engagement in them, and their performance on them” (Wigfield et al., 2015, p. 658). The expectancy construct is similar to control in the CVT, with the main difference being that perceptions of control can also refer to past outcomes, whereas expectancies refer to future outcomes only (Pekrun, 2024). The SEVT differentiates between four facets of value, including intrinsic, attainment, and utility value as well as perceived costs (Eccles & Wigfield, 2020). Together, these qualitatively different types of value make up an individual’s overall perception of the importance of an achievement task, which moderates the effect of expectancies on achievement motivation: When perceived value is high, expectancies positively predict achievement motivation, whereas at low levels of value, this relationship is nonexistent (Nagengast et al., 2011, 2013).

### ***Motivational Intensity Theory (MIT)***

According to the MIT (Brehm & Self, 1989), the amount of effort an individual invests in a task depends on how difficult the individual perceives the task to be and how valuable success is to them. Task difficulty determines the amount of effort required to succeed in the task, and the individual aims to mobilize only the necessary number of resources to achieve success, but not more. Hence, task difficulty is positively related to effort. This positive relationship only persists up to a certain point of perceived difficulty, when either the demands of the task exceed the perceived value of success or when the task becomes impossible to solve. At this point of perceived difficulty, effort drops to zero, as any further investment of effort would waste valuable resources (Richter et al., 2016). By implication, if this threshold is reached very quickly due to low perceived value of success, there would be hardly any relation between perceived difficulty and effort, whereas if this point is reached only at very high levels of difficulty due to perceived value being high, the association between task difficulty and effort would be positive for most of the range of perceived difficulty.

### ***Synthesis of Theoretical Perspectives***

Although they address different outcomes, the three theories overlap in the variables they deem influential and their assumptions regarding these influences. First, both control in

CVT and expectancy in SEVT refer to the individual's perception of whether they can master a task at hand. An individual's perception of task difficulty, a central construct in the MIT, plays a crucial role in this appraisal: When a task is perceived as easy, the individual likely experiences high levels of control and has high success expectancies. In contrast, when a task is perceived as more difficult, perceptions of control and expectancy diminish. Thus, control and expectancy should be inversely related to the perception of task difficulty, connecting the three theories.

Second, all three theories attribute a moderating role to the perception of value: If perceived value is high, perceptions of control, expectancy, and difficulty should be strongly related to achievement emotions, achievement motivation, and effort, respectively. In contrast, when value is low, the individual will not experience intense achievement emotions (except for boredom), strong motivation, or invest much effort, regardless of perceived control, expectancy, or difficulty. Building on these three theories that identified the interaction between perceived task difficulty and perceived value of success as key determinants of an individual's state in achievement situations, we investigate whether this interaction also predicts different measures of physiological arousal to enhance our understanding of the meaning of physiological changes in achievement situations.

### ***7.1.2 Psychophysiological States in Achievement Situations***

In contrast to most self-report measures, psychophysiological measures can be applied continuously without interrupting an achievement situation, are less susceptible to biases, and may provide more objective insights by not requiring individuals to be aware of and report their current state (Pekrun, 2023b). While some studies have used other physiological measures, like cortisol level (e.g., Graham et al., 2023), educational research, for the most part, has focused on measures of activity in the autonomic nervous system (e.g., Harley et al., 2019; Kiuru et al., 2022; Roos et al., 2023; Strohmaier et al., 2020). The autonomic nervous system comprises two branches with distinct functions: Activation in the sympathetic nervous system is commonly associated with states of *fight or flight*, whereas activation in the parasympathetic nervous system is associated with states of *rest and digest* (Ishikawa, 2023; Wehrwein et al., 2016; Weissman & Mendes, 2021). Despite their apparently opposing functions, activation in the two branches can vary independently of one another (Berntson et al., 1994).

Electrodermal activity is a commonly used indicator of sympathetic activation. It is driven by the sweat production in eccrine sweat glands, which increases with sympathetic activation, thereby making the skin more conducive for electric currents (Boucsein, 2012). Two

indicators of electrodermal activity commonly used in educational research are skin conductance level (SCL) and skin conductance response (SCR). SCL is a tonic, rather slowly moving signal, whereas SCR describes phasic, rapid changes that occur as peaks on top of the SCL (Boucsein et al., 2012; Braithwaite et al., 2015).

Cardiovascular activity also indicates activation in the autonomic nervous system. Heart rate (HR), that is, the frequency of heart beats, is influenced by both sympathetic activation, which accelerates HR, and parasympathetic activation, which decelerates HR. As sympathetic and parasympathetic activation can occur independently of one another, HR alone does not allow for inferences on activation in the individual branches. For example, a decrease in HR could be driven by either sympathetic deactivation, parasympathetic activation, or both (Wehrwein et al., 2016; Weissman & Mendes, 2021). Heart rate variability (HRV) describes the variation in the intervals between heartbeats and is uniquely positively influenced by parasympathetic activation (Chapleau & Sabharwal, 2011; Kleiger et al., 2005; Laborde et al., 2017).

So far, studies have found significant associations between physiological arousal and various cognitive (e.g., attentional demands; Chang & Huang, 2012), motivational (e.g. effort; Albinet et al.; 2024), and emotional variables (e.g., achievement emotions; Kiuru et al., 2022). On the one hand, these findings demonstrate that psychological states are indeed related to physiological states in an achievement situation. On the other hand, relating physiological activation to multiple psychological constructs, which are also entangled with one another (Pekrun, 2023a), hinders their use as indicators of psychological states, as observed changes are not clearly attributable to a specific process (Richter & Slade, 2017).

The present research, therefore, aims to gain a more specific understanding of the meaning of psychological states in an achievement situation by exploring their relation to the constructs of achievement emotions, achievement motivation, and effort. To do so, we build on the common predictors of all three constructs, being perceived difficulty and perceived value of success. Specifically, in the following, we derive assumptions from CVT, SEVT, and MIT, on how the interaction between perceived difficulty and perceived value of success predicts physiological arousal, assuming that physiological states reflect achievement emotions, achievement motivation, or effort, respectively. Comparing these assumptions to the patterns observed in the data will contribute to a better understanding of what psychological processes are most closely related to changes in physiological states in achievement situations.

### ***7.1.3 Perceived Difficulty and Value of Success Predicting Physiological States***

Depending on the theoretical framework, different assumptions can be derived regarding how perceived difficulty and perceived value of success might predict physiological arousal.

#### ***Control-Value Theory of Achievement Emotions***

The CVT proposes that, given high perceived value of success, perceived control facilitates positive emotions, whereas a perceived lack of control prompts negative emotions. Assuming that difficulty negatively influences control, it follows that when perceived value is high, lower difficulty should promote positive emotions, such as enjoyment, and higher difficulty should promote negative emotions, such as anxiety. Considering the classification of these emotions as activating (Pekrun, 2024), this leads to the assumption of a U-shaped link between perceived difficulty and physiological arousal, driven by intense activating positive emotions at low difficulty, and intense activating negative emotions at high difficulty. At moderate difficulty, we expect arousal levels to be lower, with both positive and negative emotional responses being less intense (see Figure 1). These assumptions align with exploratory within-person findings by Roos et al. (2023), who showed that in instances where anxiety was the dominant emotion, that is, when control is generally rather low, lower perceived control was accompanied by higher levels of SCR – possibly, *even higher* perceived difficulty was accompanied by *even higher* levels of activating anxiety. In contrast, when joy was the dominant emotion, there was a non-significant, yet positive association between perceived control and SCR, suggesting that in rather controllable conditions, *even lower* difficulty might elicit *even higher* levels of activating enjoyment. These findings support the assumption of a U-shaped relation between perceived difficulty and sympathetic activation given high perceived value of success. Consequently, for parasympathetic activation, we would expect the opposite pattern, an inverted U-shaped relation.

Alternatively, when perceived value is low, deactivating boredom is thought to be prompted (Pekrun, 2006, 2024). Boredom can be generated by both underchallenge (low difficulty) and overchallenge (high difficulty; Pekrun & Goetz, 2024). As such, when perceived value of success is low, physiological arousal should generally be lower due to the occurrence of deactivating boredom, and particularly low at the extreme ends of perceived difficulty. This results in the assumption of a rather flat, inverted U-shaped curve for sympathetic, and U-shaped curve for parasympathetic arousal linking perceived difficulty and physiological arousal at low perceived value (Figure 1).

It should be noted that these derived assumptions only pertain to particularly high and low levels of perceived value of success, with boredom presumably only occurring when perceived value is low. However, when considering value as a continuous rather than a dichotomous variable, it seems possible that boredom due to under- or overchallenge may also occur to some degree even when the perceived value of success is relatively high but not extremely so. For example, while boredom may not arise in a very easy final exam, given its extremely high value, boredom may arise during a very easy practice exam, even though perceived value is still relatively, but not extremely, high. Hence, although the previously derived assumptions align with propositions of the CVT for particularly high and low perceived value, they do not account for the possibility of certain deactivating responses at the extreme ends of perceived difficulty even under relatively high perceived value.

### ***Situated Expectancy-Value Theory of Achievement Motivation***

Considering that one function of autonomic activation is to facilitate adaptive responses to situational demands (Levenson, 2014), physiological arousal may be directly linked to achievement motivation, reflecting the choice of and persistence in engaging with a task (Wigfield et al., 2015). Assuming that physiological arousal reflects achievement motivation, and difficulty is negatively linked with expectancy, the following assumptions can be derived from SEVT: Given high perceived value of success, difficulty should negatively predict sympathetic and positively predict parasympathetic arousal, whereas at low perceived value of success, no relation should be observed between perceived difficulty and physiological activation. In other words, we expect particularly high levels of sympathetic and particularly low levels of parasympathetic activation at appraisal combinations of low perceived difficulty and high perceived value of success, as a way of facilitating engagement with the task (see Figure 1).

These assumptions partially align with the between-person findings by Kiuru et al. (2022), who observed that at high perceived value, mastery (i.e., expectancy) was positively related to SCR and negatively related to HRV. However, unexpectedly, low perceived value not only weakened but even inverted this relation, which does not align with expectations derived from SEVT. Further, these associations were only found at the between- but not at the within-person level. The present work, which focuses on the within-person level, will contribute to a better understanding of whether such within-person associations are truly absent.

### ***Motivational Intensity Theory***

According to MIT, the effort invested in a task increases alongside perceived task difficulty up to a certain point where either perceived value is not high enough to justify the

investment of more resources, or the task is deemed impossible to solve. In research on MIT, the use of physiological – in particular cardiovascular – measures as indicators of effort is well-established (Albinet et al., 2024; Gendolla, 2025). For example, Richter et al. (2008) found that higher task difficulty was accompanied by higher sympathetic activation, as long as the task was solvable, whereas activation was low in a condition where the task was impossible to solve. Transferring this to a continuous assessment of perceived difficulty and perceived value of success, the function resembles a triangular shape, with a linear increase in effort alongside an increase in perceived difficulty, followed by a vertical drop of effort to zero at a certain level of task difficulty. The point of perceived difficulty at which this drop occurs is defined by perceived value (as long as the task is solvable; for illustration, see Gendolla, 2025).

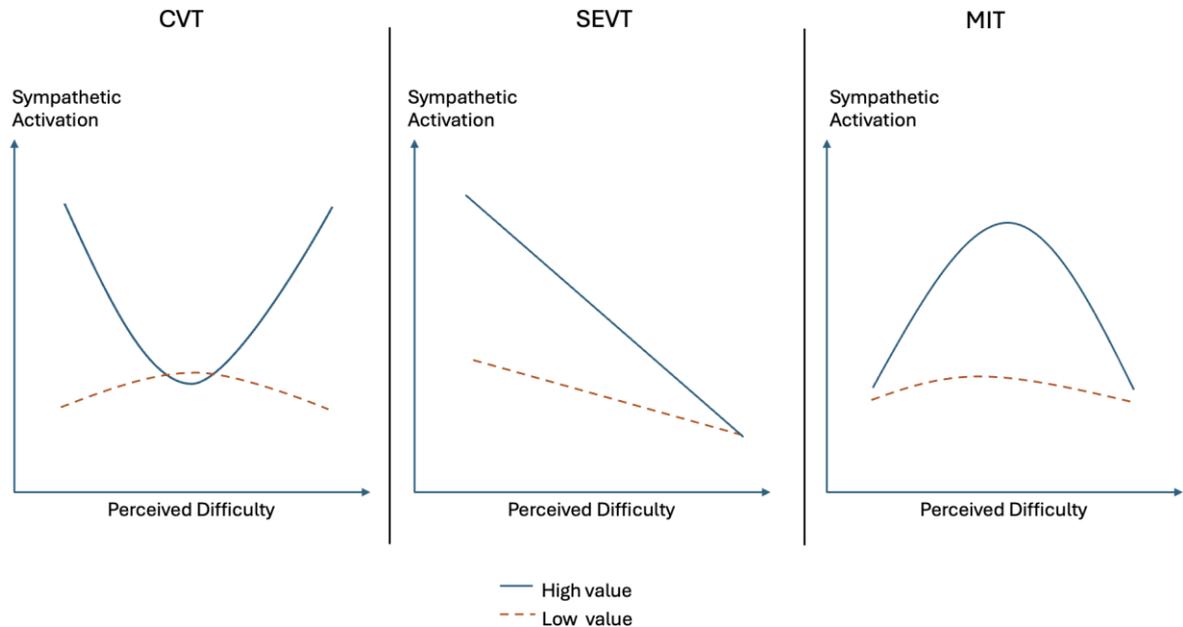
For the present purpose, we suggest approximating the relation between perceived difficulty and physiological arousal derived from the MIT with an inverted U-shaped function for sympathetic arousal, and a U-shaped function for parasympathetic arousal: We expect an increase in sympathetic and a decrease in parasympathetic arousal with increasing perceived difficulty, until a certain point is reached after which sympathetic activation decreases, and parasympathetic activation increases. The shape of this curve is influenced by perceived value. If perceived value is high, the point of inflection of the curve is reached at higher perceived difficulty, and thus a higher level of arousal. In contrast, at low perceived value, individuals will stop investing effort in the task, already at a low perceived difficulty, resulting in a much flatter inverted U-shaped curve for sympathetic and U-shaped curve for parasympathetic activation (see Figure 1).

### ***7.1.4 The Present Research***

The present research aims to contribute to a better understanding of the meaning of physiological states in an achievement situation by investigating how appraisals of difficulty and value of success interact in predicting physiological arousal. From three well-established theories in the achievement context, we have derived assumptions of possible predictive patterns, depending on whether physiological arousal most closely reflects emotional arousal, achievement motivation, or effort. By comparing the observed patterns with the theoretically derived assumptions, the present study contributes to a clearer understanding of what psychological processes the physiological measures are indicative of in achievement situations, which is crucial for their meaningful application in educational research and practice.

**Figure 1**

*Visualization of Assumptions for Sympathetic Activation Derived from Theories*



*Note.* Only patterns for sympathetic activation are depicted for clarity, patterns for parasympathetic activation are assumed to be inverted. CVT = Control-Value Theory for Achievement Emotions, SEVT = Situating Expectancy x Value Theory of Achievement Motivation, MIT = Motivational Intensity Theory for Effort.

Most previous studies have employed between-person approaches, meaning that they investigated whether differences between individuals in one variable (e.g., perceived difficulty) related to differences between individuals in another variable (e.g., heart rate). However, since theories on links between physiological and psychological variables commonly refer to intraindividual mechanisms (Pekrun, 2023b), the present study focuses on the within-person level. This means that we investigate whether and how changes in a variable between two or more timepoints relate to changes on another variable at these timepoints within the same person. In line with previous research in the field, we employ measures of the autonomic nervous system. Specifically, we use SCR as a measure of electrodermal activity, as well as HR and HRV as measures of cardiovascular activity. Employing different physiological measures that capture activity in both the sympathetic and parasympathetic branches allows for a more comprehensive picture of an individual's autonomic state.

Based on the different assumptions of the theories regarding the association between perceived difficulty and physiological arousal, we employ a two-step approach. In the first step,

we investigate the assumptions derived from SEVT, specifically a negative linear relationship between perceived difficulty and sympathetic activation, and a positive linear relationship between perceived difficulty and parasympathetic activation. In the second step, we add a quadratic term to the model, accounting for the assumptions derived from CVT and MIT. In line with the assumptions of all three theories, we additionally investigate whether both the linear and the quadratic term are moderated by value, with the curve being flatter or even inverted at lower perceived value of success. Following this approach, we aim to contribute to a clearer picture of the meaning of changes in physiological arousal over the course of an achievement situation.

### **7.2 Method**

#### ***7.2.1 Participants***

Of the initial sample of 73 participants, physiological data was not available for three persons, and two datasets were not separable due to identical participant numbers. After excluding these five individuals, the final sample consisted of  $N = 68$  individuals. Their mean age was 20.34 years ( $SD = 3.65$  years,  $Min = 18$  years,  $Max = 44$  years). As the data was collected in a study on pre-service teachers (see Eckerskorn et al., under revision), all participants were female students enrolled in a primary education teaching course. To ensure the validity of the physiological data, we followed the pre-defined exclusion process detailed in the Appendix, resulting in a final dataset with 327 blocks from 66 individuals for electrodermal and 332 blocks from 68 individuals for cardiovascular activity.

#### ***7.2.2 Procedure***

Upon their arrival at the lab, participants signed a consent form, and the experimenter attached the electrodes and sensors for the physiological measurement to their non-dominant hand. All following instructions were presented on a computer screen. First, participants were guided through a validation paradigm, which was used to ensure the quality of their electrodermal activity data (see Appendix). Next, there was a short breathing exercise with the purpose of calming down initial arousal caused by the novel laboratory situation, followed by a three-minute baseline measurement in which participants were instructed to sit still and breathe normally. After the baseline assessment, participants received the instructions for the numerical reasoning task, accompanied by two examples. They were informed that this test measured numerical reasoning ability, which constituted one facet of intelligence, and that they would receive feedback on their performance after completing the task. Participants then worked on the task, divided into five blocks with four items each. After each block, participants

were instructed to fill in a pen-and-paper self-report questionnaire that was placed on the table in front of them. Between the third and fourth block, there was a short break in which participants were instructed to take a breath and relax before proceeding to the second part of the test. There was no time limit on the break, and participants could end the break and return to the test by clicking the respective button. Upon completion of the fifth block and questionnaire, the experimenter informed participants how many of the twenty items they had solved correctly and debriefed them regarding the purpose of the study. Participants received 15 Euros as compensation for their study participation.

### ***7.2.3 Numerical Reasoning Task***

The numerical reasoning task comprised five blocks with four items each. The items were developed by Loe et al. (2018) and presented in the iMotions software (iMotions, 2022). The items were selected in a way that item difficulty increased across the first three blocks, then dropped at the beginning of the fourth block, before increasing again until the end of block five. This variation in item difficulty across the five blocks ensured sufficient variation in perceived difficulty within each participant.

Each item consisted of a number sequence that followed a certain pattern (Figure 2). Participants had to identify the pattern and complete the sequence by filling in the one or two missing numbers without taking notes. They typed in their answers using the keyboard and could proceed to the next item by clicking the respective button, without the possibility of returning to previous items. In case they wanted to proceed to the next item without answering the current one, they could insert an “X” instead of a solution. There was no time limit to the task, and feedback on participants’ performance was only provided after the completion of the whole procedure.

### **Figure 2**

#### *Example Item Numerical Reasoning Task*

27      12      36      21      63           

*Note.* This item is for exemplary purposes only and has not been used in the study. Correct answers are 48 and 144 (Pattern:  $-15, *3, -15, *3, \dots$ ).

### 7.2.4 Measures

#### *Electrodermal and Cardiovascular Activity*

Electrodermal and cardiovascular activity were captured using a BioNomadix wristband connected to a Biopac MP160 receiver and Biopac Bionomadix 2CH GSR/EDA Amplifier. The data was collected at a frequency of 200Hz and recorded in the software iMotions (iMotions, 2022). For the measurement of electrodermal activity, two electrodes were attached to the palm of the non-dominant hand, while cardiovascular activity was captured using a photoplethysmography sensor attached to one fingertip. Following the recommendations by Boucsein et al. (2012), experimenters took note of interruptions that might influence the physiological data (e.g., loud noises, talking, or sneezing), and visually inspected the raw signal afterwards to remove the respective time window in case of visible artifacts.

To extract features of the electrodermal activity signal, we used the peak detection algorithm implemented in the iMotions software (see Table 1 for parameters), which separates the signal into its tonic and phasic components. SCR for the baseline and each block was captured as the peaks per minute in the phasic signal.

For cardiovascular activity, the signal was further processed in the software Kubios HRV Scientific, which corrects the raw data for artifacts and instances of low data quality (Kubios Oy, 2024; Tarvainen et al., 2014, see Table 2 for parameters). The software then provided a measure of HR in beats per minute for the baseline and each block. Similarly, it provided the root of the mean square of successive differences as a measure of HRV, which is a commonly used indicator of parasympathetic arousal (Laborde et al., 2017).

**Table 1**

*Parameters used in Peak Detection Algorithm of the iMotions Software*

---

Phasic Filter Length	8000 ms
Lowpass Filter Cutoff Frequency	5 Hz
Peak Onset Threshold	0.01 microSiemens
Peak Offset Threshold	0 microSiemens
Peak Amplitude Threshold	0.01 microSiemens
Minimum Peak Duration	500 ms

---

**Table 2***Parameters used for Extraction of Heart Rate (Variability) in Kubios Scientific HRV Software*

Quality and beat correction	Medium
RR time series interpolation rate	4 Hz
RR interval detrending	Smoothn priors
Smoothing parameter	500
Peak Amplitude Threshold	0.035 Hz

***Perceived Difficulty and Value***

The self-report questionnaires presented after each block consisted of single-item indicators to minimize interruptions to the task. Perceived task difficulty was assessed on a five-point scale (“To me, the tasks are ...”; 1 = very easy, 2 = rather easy, 3 = neither easy nor difficult, 4 = rather difficult, 5 = very difficult). Further, participants indicated their subjective value of success on a five-point scale (“How important is it for you to do well on the test?”; 1 = not important at all to 5 = very important). In accordance with existing recommendations (e.g., Hoffman & Walters, 2022), both difficulty and value were centered within each individual by subtracting the person’s average of all five blocks from each of their ratings. Thus, the person-centered values for each block denote a person’s deviance in this block from their personal mean. For example, a person-centered difficulty rating of one on a block would indicate that the person perceived this block as one unit more difficult than this person’s average difficulty perception across all blocks.

***7.2.5 Analyses***

To investigate how difficulty and value perceptions relate to physiological states on a within-person level, we employed a hierarchical multilevel regression model approach with a random intercept for each of the three outcome variables (SCR, HR, HRV). In a first step, the respective outcome was predicted by fixed effects of person-centered difficulty, person-centered value, and an interaction between them. In a second step, a quadratic term of person-centered difficulty and its interaction with person-centered value were added as fixed effects to the set of predictors. For each of the three physiological outcome variables, the following equation denotes the final model on the within-person level (PWC = Person-wise centred):

**Step 1:**  $Y_{ij} = b_0 + b_1\text{Difficulty}_{\text{PWCij}} + b_2\text{Value}_{\text{PWCij}} + b_3\text{Difficulty}_{\text{PWCij}} \times \text{Value}_{\text{PWCij}} + u_{0j} + e_{ij}$

**Step 2:**  $Y_{ij} = b_0 + b_1\text{Difficulty}_{\text{PWCij}} + b_2\text{Value}_{\text{PWCij}} + b_3\text{Difficulty}_{\text{PWCij}} \times \text{Value}_{\text{PWCij}} + b_4(\text{Difficulty}_{\text{PWCij}})^2 + b_4(\text{Difficulty}_{\text{PWCij}})^2 \times \text{Value}_{\text{PWCij}} + u_{0j} + e_{ij}$

The analyses were conducted in R, using the lmer function of the lme4 package (Bates et al., 2015). A significance level of  $\alpha = .05$  was used.

### **7.2.6 Openness and Transparency**

The study was not preregistered. Data from this study have previously been used to investigate between- and within-person correlations of physiological measures and self-report measures of achievement emotions (Study 2 in Wuensch et al., under revision), and the between-person effect of pre-service teachers' mathematics trait self-concept on their emotional and physiological state (Eckerskorn et al., under revision).

## **7.3 Results**

### **7.3.1 Descriptives**

Table 3 displays descriptive statistics for the variables across all blocks, as well as for each of the five task blocks. The descriptive values for the three physiological measures show decreases in sympathetic activation and increases in parasympathetic activation across the five blocks, with average values of SCR and HR decreasing and HRV increasing over time.

The task seemed to be adequately difficult for our participants, as on average, around half of the items per block were solved correctly. Further, the average overall rating of perceived difficulty was above the mid-point of the scale, and differences between the blocks in average perceived difficulty aligned with the average task performance on the respective block. The average perceived value was also above the mid-point of the scale, underlining that participants perceived a successful task performance as rather important, possibly due to the ego-threat induced by informing them that the test was measuring intelligence and announcing feedback. Average perceived value decreased over time.

Table 4 displays the within-person correlations between the variables. The three physiological variables correlated as expected based on their physiological meaning: SCR and HR, which are both positively influenced by sympathetic activation, correlated positively with each other, and negatively with HRV as an indicator of parasympathetic activation. Further, perceived difficulty and value correlated negatively with each other, meaning that in blocks where a person perceived the difficulty as higher compared to the other blocks, they judged it less important to do well on the task compared to the other blocks. Perceived difficulty correlated negatively with SCR and HR. Value correlated positively with SCR and HR, and negatively with HRV. Task performance positively correlated with SCR, HR, and value, and negatively with perceived difficulty. The within-person correlations also show significant

correlations with time for all variables, with SCR, HR, value, and task performance decreasing over time, while perceived difficulty and HRV increased over time.

We also calculated the intraclass correlation for each variable (Table 4), which indicates how much of the variance is due to between-person differences and within-person variation. These intraclass correlations demonstrate that for all physiological measures, a substantial proportion of the overall variance was attributed to between-person differences (67–91%) compared to within-person differences between blocks. This means that participants largely differed from one another in their overall levels of the physiological measures, but each individual also demonstrated some variance in those measures across the five task blocks. Subjective difficulty and task performance, in contrast, showed substantial within-person variance. It is worth noting that for value, again, most variance was attributable to between-person differences. Half of the sample (33 participants) did not show any within-person variance in their value ratings. Hence, value ratings only showed limited change over the course of the specific achievement situation our participants were confronted with; this limits the possibility of detecting within-person effects of value and interactions with perceived difficulty in the present sample.

**Table 3**

Means (SD) for all variables across all blocks and separately for each block.

	SCR	HR	HRV	Perceived Difficulty	Perceived Value	Task performance
Overall	6.22 (3.05)	84.95 (11.81)	37.70 (19.85)	3.62 (1.04)	3.25 (1.08)	1.95 (1.32)
Block						
1	8.29 (2.99)	87.9 (13.3)	34.6 (18.4)	2.87 (0.93)	3.60 (1.01)	2.90 (1.04)
2	6.66 (2.92)	86.2 (11.9)	35.6 (19.5)	3.66 (0.80)	3.50 (1.00)	2.03 (1.05)
3	5.30 (2.93)	84.7 (11.5)	37.1 (20.6)	4.25 (0.70)	3.18 (1.09)	0.85 (0.89)
4	5.76 (2.57)	83.2 (11.2)	40.8 (20.8)	3.00 (1.08)	3.15 (1.11)	2.88 (1.04)
5	5.05 (2.73)	82.6 (10.4)	40.5 (19.8)	4.28 (0.75)	3.00 (1.11)	1.08 (0.97)
ICC1	.67	.91	.91	.22	.77	.11

*Note.* ICCs reflect the percentage of overall variance that is attributable to between-person differences. ICCs were calculated using the statsBy function of the psych package (Version 2.4.6.26; Revelle, 2024) in R. SCR = Skin conductance response, HR = Heart rate, HRV = Heart rate variability.

**Table 4***Within-person Correlations between Variables*

	SCR	HR	HRV	Perceived Difficulty	Perceived Value	Task Perf.
HR	<b>.51</b>					
HRV	<b>-.23</b>	<b>-.66</b>				
Perc. Difficulty	<b>-.47</b>	<b>-.19</b>	.03			
Perc. Value	<b>.30</b>	<b>.35</b>	<b>-.28</b>	<b>-.27</b>		
Task Perf.	<b>.44</b>	<b>.21</b>	<b>-.02</b>	<b>-.74</b>	<b>.21</b>	
Time	<b>-.66</b>	<b>-.66</b>	<b>.49</b>	<b>.36</b>	<b>-.46</b>	<b>-.35</b>

*Note.* Within-person correlations were calculated using the statsBy function of the psych package (Version 2.4.6.26; Revelle, 2024) in R. SCR = Skin conductance response, HR = Heart rate, HRV = Heart rate variability, Time = Number of the block from 0 to 4. For **bold** values  $p \leq .001$ .

### 7.3.2 Hierarchical Multilevel Linear Modelling

The results of the hierarchical multilevel linear models with random intercepts for each physiological outcome variable are displayed in Table 5. In a first step, person-centered difficulty and value, as well as their interaction, were entered as predictors. In a second step, a quadratic term of person-centered difficulty and its interaction with person-centered value were added as predictors.

For SCR, the first model showed significant within-person effects of difficulty and value, as well as a significant linear interaction between them (Table 5), which is visualized in the first panel of Figure 3. The interaction effect shows that the negative association between perceived difficulty and SCR is amplified by higher levels of perceived value and dampened at lower levels of perceived value. Adding the quadratic term of difficulty in interaction with perceived value of success significantly improved the model fit, indicated by an ANOVA. In the second model, again, the main effects of perceived difficulty and value, as well as their linear interaction, were statistically significant. Additionally, the quadratic term of perceived difficulty was significant, illustrated in the second panel of Figure 3: At the lower end of the

difficulty scale, SCR shows a slight increase with increasing difficulty, before dropping more steeply with higher levels of person-centered perceived difficulty.

For HR, the first model only showed a significant positive effect of value, but neither a main effect of perceived difficulty nor an interaction between perceived difficulty and value. Again, adding the quadratic term of difficulty and its interaction with value significantly improved the model fit. In the second model, there are significant effects of perceived difficulty and value, and most notably a significant interaction between the quadratic term for difficulty and the perception of value (Table 5). Figure 4 illustrates the two models for HR, showing how the addition of the quadratic term enabled us to detect associations between perceived difficulty and HR that were not visible in the first model, including only linear terms. The second model shows an inverted U-shaped relation between perceived difficulty and HR at high person-centered values, whereas this parabola becomes flatter with lower levels of perceived value and seems to even slightly invert its shape at low levels of perceived value compared to the individual's average.

Similarly, for HRV, the first model only showed a negative main effect of perceived value (Table 5, first panel of Figure 5). The model fit improved significantly in the second step, with the model still showing a significant negative main effect of perceived value, as well as a significant effect of the quadratic term of perceived difficulty. The second panel of Figure 5 illustrates this quadratic term, showing that HRV is particularly high at both low and high perceived difficulty. Although the interaction with perceived value did not reach statistical significance, the figure shows a similar trend to the model for HR, with the parabola being flatter and even slightly inverted at lower levels of person-centered perceived value.

Regarding the assumptions derived from the three different theories, the quadratic terms of perceived difficulty resemble most closely the assumptions derived for effort; the additional linear term for SCR additionally the ones for achievement motivation. Finally, as stated in all three theories, higher perceived value of success generally amplified effects of perceived difficulty on physiological arousal.

**Table 5**

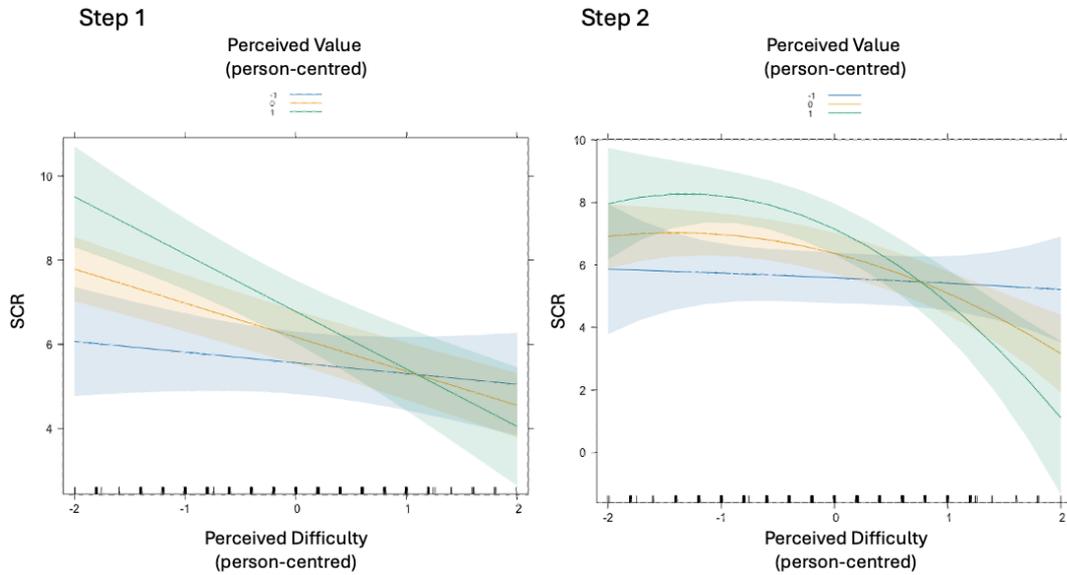
*Hierarchical Multilevel Linear Regression Analysis with a Random Intercept (unstandardized coefficients)*

SCR						
	Step 1			Step 2		
	Est.	SE	<i>p</i>	Est.	SE	<i>p</i>
Intercept	<b>6.17</b>	0.33	< .001	<b>6.37</b>	0.33	< .001
Perceived Difficulty	<b>-0.81</b>	0.10	< .001	<b>-0.94</b>	0.11	< .001
Perceived Value	<b>0.61</b>	0.19	.001	<b>0.78</b>	0.24	< .001
Difficulty * Value	<b>-0.56</b>	0.24	.024	<b>-0.78</b>	0.26	.003
Quadr. Perceived Difficulty				<b>-0.33</b>	0.12	.008
Quadr. Difficulty * Value				-0.32	0.20	.111
AIC		1385.3			1380.5	
BIC		1408.1			1410.8	
LogLik		-686.7			-682.23	
Model Comparison			$\chi^2 (2) = 8.87, p = .012$			
HR						
	Step 1			Step 2		
	Est.	SE	<i>p</i>	Est.	SE	<i>p</i>
Intercept	<b>84.95</b>	1.39	< .001	<b>85.21</b>	1.40	< .001
Perceived Difficulty	-0.42	0.23	.073	<b>-0.68</b>	0.25	.007
Perceived Value	<b>2.20</b>	0.41	< .001	<b>2.98</b>	0.54	< .001
Difficulty * Value	-0.65	0.54	.229	-0.91	0.57	.113
Quadr. Perceived Difficulty				-0.44	0.28	.108
Quadr. Difficulty * Value				<b>-1.13</b>	0.44	.012
AIC		2021.7			2017.5	
BIC		2044.6			2047.9	
LogLik		-1004.9			-1000.8	
Model Comparison			$\chi^2 (2) = 8.24, p = .016$			

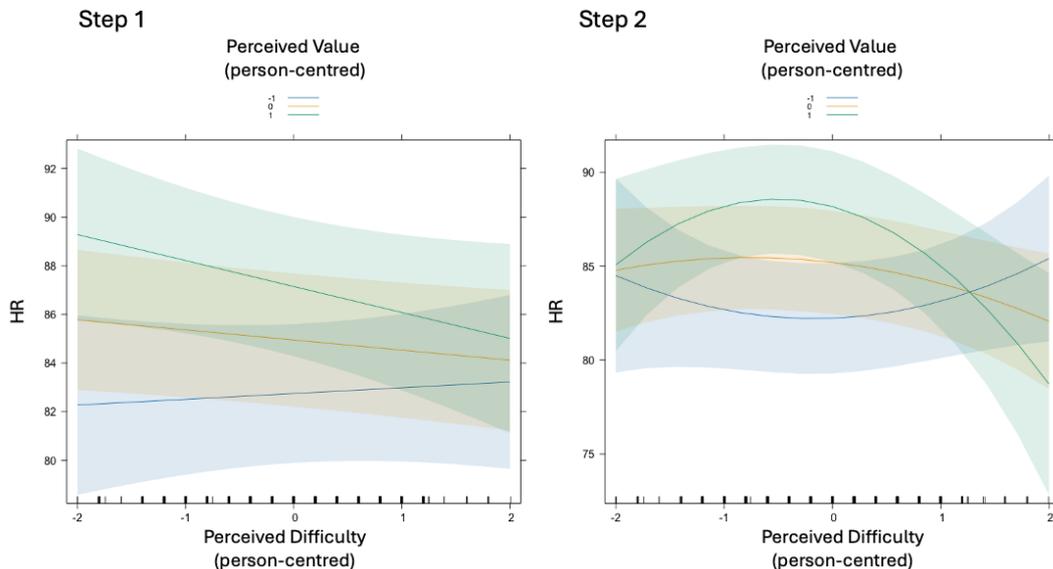
**Table 5 (continued)**

	HRV					
	Step 1			Step 2		
	Est.	<i>SE</i>	<i>p</i>	Est.	<i>SE</i>	<i>p</i>
Intercept	<b>37.82</b>	2.32	< .001	<b>37.22</b>	2.34	< .001
Perceived Difficulty	-0.30	0.41	.458	0.16	0.44	.720
Perceived Value	<b>-3.48</b>	0.73	< .001	<b>-4.41</b>	0.95	< .001
Difficulty*Value	0.43	0.96	.653	1.08	1.01	.288
Quadr. Perceived Difficulty				<b>1.02</b>	0.49	.038
Quadr. Difficulty * Value				1.50	0.79	.056
AIC		2393.2			2390.0	
BIC		2416.1			2420.5	
LogLik		-1190.6			-1187.0	
Model Comparison				$\chi^2 (2) = 7.18, p = .028$		

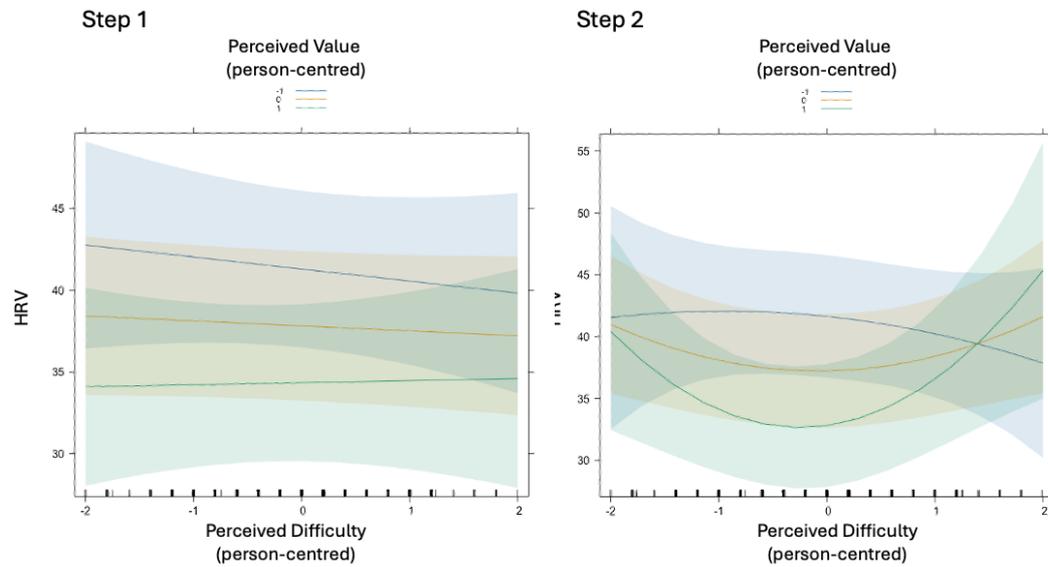
*Note.* SCR = Skin conductance response; HR = Heart rate; HRV = Heart rate variability.

**Figure 3***Visualization of Models for SCR*

*Note.* Step 1 included linear interaction between perceived difficulty and value of success, Step 2 additionally quadratic term for perceived difficulty in interaction with perceived value of success. SCR = Skin conductance response in peaks per minute.

**Figure 4***Visualization of Models for HR*

*Note.* Step 1 included linear interaction between perceived difficulty and value of success, Step 2 additionally quadratic term for perceived difficulty in interaction with perceived value of success. HR = Heart rate in beats per minute.

**Figure 5***Visualization of Models for HRV*

*Note.* Step 1 included linear interaction between perceived difficulty and value of success, Step 2 additionally quadratic term for perceived difficulty in interaction with perceived value of success. HRV = Heart rate variability in root mean square of successive differences.

**7.4 Discussion**

With the present study, we aimed to contribute to a deeper understanding of the meaning of changes in physiological arousal throughout an achievement situation. We drew on the CVT of achievement emotions (Pekrun, 2024), the SEVT of achievement motivation (Eccles & Wigfield, 2020), and the MIT on effort (Brehm & Self, 1989) to derive hypotheses on how perceived difficulty and perceived value of success would predict physiological states in an achievement situation, given that physiological arousal indicates achievement emotions, achievement motivation, and effort, respectively. We tested these hypotheses at the within-person level for different indicators of physiological arousal, including SCR, HR, and HRV, using a two-step hierarchical multilevel linear modeling approach. First, we included only a linear term, and second, a quadratic term for perceived difficulty, both in interaction with perceived value of success.

For all three outcome variables, including a quadratic term for perceived difficulty, significantly improved the model fit. Specifically, for HR and HRV, models that included the quadratic term revealed significant relationships between these physiological variables and perceived difficulty, whereas models that only included the linear term suggested no such

association. Thus, the exclusive focus on linear terms might have been a reason why previous studies have not found within-person effects of difficulty-related variables on physiological indicators (e.g., linear models for mastery in Kiuru et al., 2022).

For SCR and HRV, the final models yielded significant quadratic main effects of perceived difficulty, and for HR, this quadratic effect was significant in interaction with perceived value. Specifically, the quadratic effects on the within-person level showed that sympathetic activation was lower and parasympathetic activation was higher at both very low and very high levels of perceived task difficulty compared to moderate levels of difficulty. This quadratic pattern aligns most closely with assumptions derived for the construct of effort. According to MIT, effort increases with increasing task demands to mobilize as much energy as needed for a successful task completion, until a certain point is reached and effort drops (Brehm & Self, 1989). In the context of MIT, a close alignment between autonomic activation – in particular, sympathetic cardiovascular activity – has already been established (Gendolla, 2025; Richter et al., 2016). The observed quadratic relationships between perceived difficulty and all three physiological indicators given high perceived value of success further support this link between effort and autonomic activation.

In addition to the quadratic term, suggesting a link with effort, the final model for SCR additionally showed a pronounced linear relationship of perceived difficulty in interaction with perceived value of success. Thus, the results for SCR show an additional alignment with the assumptions derived from the SEVT on achievement motivation by exhibiting a linear negative relation between perceived difficulty and physiological arousal, given a high perceived value of success. This combination of findings for the different indicators may suggest that sympathetic activation (SCR) is more related to the motivation to engage with the task, whereas parasympathetic activation (affecting HR and HRV) might be more closely related to the mobilization of effort depending on task demands.

The effects of perceived difficulty on all three indicators oppose the hypotheses we derived from CVT. Building on findings by Roos et al. (2023), we assumed that individuals would be particularly physiologically aroused at both low and high perceived difficulty, given high perceived value of success, due to an intense experience of positively activating, or negatively activating emotions, respectively. As noted earlier, these assumptions may have overlooked the occurrence of deactivating emotions even at higher levels of perceived value when difficulty is either very low or very high. As the average ratings of perceived value of success demonstrate, participants did care about doing well on the task, but the perceived value was not extraordinarily high. Hence, boredom due to over- or underchallenge may have

occurred at the extreme ends of the difficulty scale, even when person-centered value was relatively high. Further, the derived assumptions may have overlooked the deactivating emotion of hopelessness, which may have replaced anxiety when difficulty was particularly high (Pekrun, 2006). Overall, the contrasting findings compared to our assumptions derived from CVT could stem from potential theoretical misspecifications, neglecting the occurrence of deactivating emotions at both very high and very low perceived difficulty, even at rather high perceived value.

Regarding perceived value of success, all three theories (CVT, SEVT, MIT) ascribed a moderating role, in the sense that effects of perceived difficulty on the respective outcome would be weak to non-existent if perceived value was low. In the present study, only about half of the participants showed within-person changes in perceived value, which limited our ability to detect within-person effects. Nevertheless, perceived value was a significant moderator for the linear effect of perceived difficulty on SCR and for the quadratic term of perceived difficulty on HR. Despite only being close to statistical significance, the illustration of the final model for HRV also suggests that the quadratic effect of perceived difficulty on HRV changes depending on the level of perceived value. The observed moderating effect of value was generally in line with the expectations of all three theories: The higher perceived value compared to the person's average perception of value, the more pronounced the effect of perceived difficulty on the respective outcome. Conversely, at lower levels of perceived value of success, perceived difficulty had a much smaller influence on the respective outcome. At particularly low levels of value, the effects of difficulty on HR and HRV even appear to be slightly inverted, suggesting that especially parasympathetic activation was rather lower at the extreme ends of the difficulty scale. This is not in line with assumptions derived from CVT, which assumes particularly intense boredom and, therefore, parasympathetic activation at low values and particularly high and low difficulty levels. However, as these inversions of effects are not very pronounced and evidence at the extreme ends of both the difficulty and value scale is rather sparse, future research is needed to replicate and more closely investigate such potentially inverted effects on the within-person level.

Overall, the present results provide valuable insights into the meaning of physiological arousal in an achievement situation. They demonstrate that changes in physiological arousal are indeed related to individuals' task appraisals, both in terms of perceived difficulty and perceived value. Perceived difficulty showed a quadratic relationship with all three indicators of autonomic activation, demonstrating higher levels of physiological arousal at moderate levels of perceived task difficulty, rather than at high or low levels. While these patterns suggest a

close alignment between physiological activation and effort, as previously demonstrated in the realm of MIT (Gendolla, 2025), sympathetic activation might additionally show some correspondence to achievement motivation in the sense of readiness to engage with a task. Finally, in line with expectations, higher perceived value amplified the effect of perceived difficulty on the physiological outcome variables.

#### ***7.4.1 Limitations and Future Research***

First, some features of the present study might limit the generalizability of our results. While the use of a homogenous sample can be considered an advantage as it reduces demographic factors confounding the results, especially in the physiological data, the present findings need to be replicated in other samples, specifically including male participants. Similarly, the laboratory setting of the study allowed us to control environmental factors and thereby reduce unwanted noise in the data; however, it might limit the generalizability to real-life settings, which would need to be explored in future research.

Second, it needs to be noted that it is not possible to infer causality from our results, due to the correlational nature of the present study. Thus, while, the present findings can show that certain appraisal combinations are accompanied by certain physiological arousal patterns, future studies may target the potential causal linkages across appraisals and physiological arousal. To do so, multi-timepoint repeated measures designs spanning larger timeframes, with larger naturally occurring within-person variance in value of success and task difficulty, or within-person experimental designs actively manipulating individuals' appraisals in randomized order.

Last, it needs to be considered that the present results are aggregated within-person effects, meaning that while the patterns found here seem to be present in the average individual, they might not be the same for each person. Instead, for some persons, certain effects or the interaction between appraisals may be more or less pronounced. Future research is needed to first identify whether within-person effects of appraisals on physiological states differ between individuals, and if so, what demographic or trait variables predict those differences.

#### ***7.4.2 Implications***

The present results support the possibility of using physiological measures to gain an unobtrusive, unbiased insight into an individual's state during an achievement situation. For example, indicators of physiological arousal could be used to evaluate how engaging different learning or testing materials are. Further, a real-time measurement of physiological arousal may also be applied to detect undesirable changes in learners' states and adapt a learning task

accordingly. For example, a drop in sympathetic activation and an increase in parasympathetic activation may suggest that the individual is beginning to disengage from the task, which might be counteracted by presenting an easier item, a motivation intervention, or allowing for a break. However, the development of such adaptive systems requires more investigation, for example, to be able to distinguish between an increase in parasympathetic arousal due to disengagement from one due to a task being perceived as easy. Altogether, the present study advances our understanding of the meaning of physiological changes in an achievement situation, and points out directions for future research that will bring us closer to the meaningful application of physiological measures in educational research and practice.

## 7.5 References

- Albinet, C. T., De Faria, C., & Causse, M. (2024). Cardiac Pre-Ejection Period to Index Motivation and Effort Mobilization in Cognitive Studies: A Critical Narrative Review. *Journal of Psychophysiology*, *38*(2), 81–101. <https://doi.org/10.1027/0269-8803/a000332>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). <https://doi.org/10.18637/jss.v067.i01>
- Berntson, G. G., Cacioppo, J. T., Quigley, K. S., & Fabro, V. T. (1994). Autonomic space and psychophysiological response. *Psychophysiology*, *31*(1), 44–61. <https://doi.org/10.1111/j.1469-8986.1994.tb01024.x>
- Boucsein, W. (2012). *Electrodermal Activity*. Springer US. <https://doi.org/10.1007/978-1-4614-1126-0>
- Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for psychological experiments*. <https://www.birmingham.ac.uk/documents/colleges/psych/saal/guide-electrodermal-activity.pdf>
- Brehm, J. W., & Self, E. A. (1989). The Intensity of Motivation. *Annual Review of Psychology*, *40*(1), 109–131. <https://doi.org/10.1146/annurev.ps.40.020189.000545>
- Chang, Y.-C., & Huang, S.-L. (2012). The influence of attention levels on psychophysiological responses. *International Journal of Psychophysiology*, *86*(1), 39–47. <https://doi.org/10.1016/j.ijpsycho.2012.09.001>
- Chapleau, M. W., & Sabharwal, R. (2011). Methods of assessing vagus nerve activity and reflexes. *Heart Failure Reviews*, *16*(2), 109–127. <https://doi.org/10.1007/s10741-010-9174-6>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 1–13.

- <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Eckerskorn, F., Frenzel, A. C., Wünsch, M., Pekrun, R., Brings, L., & Kleine, M. (under revision). *Confronted with the “necessary evil”: The impact of pre-service elementary school teachers’ mathematics self-concept on emotional and physiological responses to mathematics exposure.*
- Gendolla, G. H. E. (2025). The intensity of mental effort: “The heart does not lie.” *Synthese*, 205(6), 231. <https://doi.org/10.1007/s11229-025-05043-1>
- Graham, M. C., Husman, J., Pekrun, R., Villanueva, I., & Christensen, D. (2023). The dynamic experience of taking an examination: Ever changing cortisol and expectancy for success. *British Journal of Educational Psychology*, 93(S1), 195–210. <https://doi.org/10.1111/bjep.12521>
- Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science*, 47(2), 151–180. <https://doi.org/10.1007/s11251-018-09480-z>
- Hoffman, L., & Walters, R. W. (2022). Catching Up on Multilevel Modeling. *Annual Review of Psychology*, 73(1), 659–689. <https://doi.org/10.1146/annurev-psych-020821-103525>
- iMotions* (Version 9.3). (2022). [Computer software].
- Ishikawa, M. (2023). Measuring the autonomic nervous system as a window into the mind and brain: A selective review. *European Psychologist*, 28(2), 67–82. <https://doi.org/10.1027/1016-9040/a000500>
- Kiuru, N., Malmberg, L.-E., Eklund, K., Penttonen, M., Ahonen, T., & Hirvonen, R. (2022). How are learning experiences and task properties associated with adolescents’ emotions and psychophysiological states? *Contemporary Educational Psychology*, 71, 1–16. <https://doi.org/10.1016/j.cedpsych.2022.102095>
- Kleiger, R. E., Stein, P. K., & Bigger, J. T. (2005). Heart rate variability: measurement and clinical utility. *Annals of Noninvasive Electrocardiology*, 10(1), 88–101. <https://doi.org/10.1111/j.1542-474X.2005.10101.x>
- Kubios Oy. (2024). *Kubios HRV Scientific User’s Guide 2024*. <https://www.kubios.com/downloads/HRV-Scientific-Users-Guide.pdf>
- Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, 08. <https://doi.org/10.3389/fpsyg.2017.00213>
- Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emotion Review*, 6(2), 100–112. <https://doi.org/10.1177/1754073913512003>
- Loe, B., Sun, L., Simonfy, F., & Doeblner, P. (2018). Evaluating an automated number series item generator using linear logistic test models. *Journal of Intelligence*, 6(2), 20. <https://doi.org/10.3390/jintelligence6020020>
- Martin, A. J., Malmberg, L., Pakarinen, E., Mason, L., & Mainhard, T. (2022). The potential of biophysiology for understanding motivation, engagement and learning experiences. *British Journal of Educational Psychology*, 93(S1), 1–9. <https://doi.org/10.1111/bjep.12584>
- Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the “×” out of expectancy-value theory?: A psychological mystery, a substantive-methodological synergy, and a cross-national generalization.

- Psychological Science*, 22(8), 1058–1066. <https://doi.org/10.1177/0956797611415540>
- Nagengast, B., Trautwein, U., Kelava, A., & Lüdtke, O. (2013). Synergistic effects of expectancy and value on homework engagement: The case for a within-person perspective. *Multivariate Behavioral Research*, 48(3), 428–460. <https://doi.org/10.1080/00273171.2013.775060>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R. (2023a). Dissecting the elephant: Cognition, emotion, and motivation as distinct but intertwined entities. In M. Bong, J. Reeve, & S. Kim (Eds.), *Motivation Science* (1st ed., pp. 227–231). Oxford University Press New York. <https://doi.org/10.1093/oso/9780197662359.003.0038>
- Pekrun, R. (2023b). Mind and body in students' and teachers' engagement: New evidence, challenges, and guidelines for future research. *British Journal of Educational Psychology*, 93(S1), 227–238. <https://doi.org/10.1111/bjep.12575>
- Pekrun, R. (2024). Control-Value Theory: From Achievement Emotion to a General Theory of Human Emotions. *Educational Psychology Review*, 36(3), 1–36. <https://doi.org/10.1007/s10648-024-09909-7>
- Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., Van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology*, 124(1), 145–178. <https://doi.org/10.1037/pspp0000448>
- Revelle, W. (2024). *psych: Procedures for Psychological, Psychometric, and Personality Research* (Version R package version 2.4.6) [Computer software]. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Richter, M., Friedrich, A., & Gendolla, G. H. E. (2008). Task difficulty effects on cardiac activity. *Psychophysiology*, 45(5), 869–875. <https://doi.org/10.1111/j.1469-8986.2008.00688.x>
- Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). Three Decades of Research on Motivational Intensity Theory. In *Advances in Motivation Science* (Vol. 3, pp. 149–186). Elsevier. <https://doi.org/10.1016/bs.adms.2016.02.001>
- Richter, M., & Slade, K. (2017). Interpretation of physiological indicators of motivation: Caveats and recommendations. *International Journal of Psychophysiology*, 119, 4–10. <https://doi.org/10.1016/j.ijpsycho.2017.04.007>
- Roos, A.-L., Goetz, T., Krannich, M., Donker, M., Bieleke, M., Caltabiano, A., & Mainhard, T. (2023). Control, anxiety and test performance: Self-reported and physiological indicators of anxiety as mediators. *British Journal of Educational Psychology*, 93(S1), 72–89. <https://doi.org/10.1111/bjep.12536>
- Shao, K., Pekrun, R., Marsh, H. W., & Loderer, K. (2020). Control-value appraisals, achievement emotions, and foreign language performance: A latent interaction analysis. *Learning and Instruction*, 69, 101356. <https://doi.org/10.1016/j.learninstruc.2020.101356>
- Strohmaier, A. R., Schiepe-Tiska, A., & Reiss, K. M. (2020). A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety. *Frontline Learning Research*, 8(1), 16–32. <https://doi.org/10.14786/flr.v8i1.427>

- Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. <https://doi.org/10.1016/j.cmpb.2013.07.024>
- Wehrwein, E. A., Orer, H. S., & Barman, S. M. (2016). Overview of the anatomy, physiology, and pharmacology of the autonomic nervous system. *Comprehensive Physiology*, 1239–1278. <https://doi.org/10.1002/cphy.c150037>
- Weissman, D. G., & Mendes, W. B. (2021). Correlation of sympathetic and parasympathetic nervous system activity during rest and acute stress tasks. *International Journal of Psychophysiology*, *162*, 60–68. <https://doi.org/10.1016/j.ijpsycho.2021.01.015>
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. <https://doi.org/10.1006/ceps.1999.1015>
- Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. In R. M. Lerner (Ed.), *Handbook of Child Psychology and Developmental Science* (1st ed., pp. 1–44). Wiley. <https://doi.org/10.1002/9781118963418.childpsy316>
- Wuensch, M., Frenzel, A. C., Eckerskorn, F., Pekrun, R., Min, J.-Y., & Majeed, N. M. (under revision). *Physiological States and Subjective Emotional Experiences in an Achievement Situation: Between- and Within-Person Links*.

## 7.6 Appendix

Of the 73 individuals participating in Study 2, no physiological data was available for three individuals. The study numbers of two persons were the same, making it impossible to match physiological and self-report data. Followingly, five individuals were excluded from all analyses. Three persons did not complete the last block and questionnaire. A validation protocol was implemented to ensure the validity of the electrodermal activity data. In the validation phase, participants watched an anxiety-inducing short video of a person climbing a tall building, followed by the instruction to hold one’s breath for ten seconds. We classified individuals as non-responders, if they did not show any skin conductance response in response to either of the two stimuli and omitted their two electrodermal activity measures (skin conductance response and skin conductance level) from the analyses. Based this validation protocol, skin conductance response and skin conductance level measures were omitted from two individuals. For cardiovascular activity, four blocks of four individuals were excluded from the cardiovascular activity data (heart rate, heart rate variability) due to low data quality (more than 5% of corrected beats, as recommended in Kubios Oy, 2024).

Kubios Oy. (2024). *Kubios HRV scientific user’s guide 2024*. <https://www.kubios.com/downloads/HRV-Scientific-Users-Guide.pdf>



---

# 8

---

## EMPIRICAL WORK 3

### ENJOYABLE FOR SOME, STRESSFUL FOR OTHERS? PHYSIOLOGICAL AND SUBJECTIVE INDICATORS OF ACHIEVEMENT EMOTIONS DURING ADAPTIVE VERSUS FIXED-ITEM TESTING

---

Miriam Wunsch

Anne C. Frenzel

Reinhard Pekrun

Luning Sun

#### Reference:

Wuensch, M., Frenzel, A. C., Pekrun, R., & Sun, L. (2025). Enjoyable for some, stressful for others? Physiological and subjective indicators of achievement emotions during adaptive versus fixed-item testing. *Contemporary Educational Psychology*, 82, 102388.

<https://doi.org/10.1016/j.cedpsych.2025.102388>

Published open access under a CC BY 4.0 licence

(<https://creativecommons.org/licenses/by/4.0/>), no changes were made.



## 8.1 Abstract

In light of the increasing use of computerized adaptive testing, we investigated how adaptive testing impacts test-takers' subjective emotional experiences and their psychophysiological arousal. Applying a within-person design ( $N = 89$ ), we compared participants' affective states while working on an adaptive and a fixed-item test of numerical reasoning ability. During both tests, we continuously recorded participants' skin conductance response. In addition, they filled in a self-report questionnaire after each of the three item blocks per test, assessing discrete achievement emotions (joy, pride, anger, boredom, frustration, and anxiety) and perceived level of task difficulty. As expected, participants showed higher levels of psychophysiological arousal in the adaptive compared to the fixed-item test, indicating that the adaptive test was more stimulating, independent of emotional valence. For subjective achievement emotions, we expected disordinal interaction effects between test type and ability (objective control experience) and between test type and relative perceived difficulty of the two tests (subjective control experience). This was supported for relative perceived difficulty, as participants indeed reported more joy and pride, and less frustration, anxiety, and anger on whichever test they subjectively perceived as easier. Meanwhile, no main effects of test type and no interaction between test type and ability were found. This is in line with the control-value theory and shows that it is not the adaptivity of a test that influences subjective emotional experience, but rather how difficult the adaptive test is perceived by test-takers compared to a fixed-item test. Directions for future research and implications for practice are discussed.

## 8.2 Introduction

The use of computerized adaptive tests is on the rise. Adaptive tests are used more frequently than ever in educational assessments, including high-stakes exams such as the Standardized Aptitude Test in the USA (College Board, 2022) as well as large-scale international assessments like the Programme for International Student Assessment (PISA; OECD, 2022). While there is ample and consistent evidence that the use of adaptive testing yields increased efficiency and accuracy compared to classical fixed-item testing (Frey & Ehmke, 2007), its effects on test-takers' emotional experiences remain underexplored. Given that emotions have been shown to substantially influence test performance (Pekrun et al., 2023) and school-related well-being (Obermeier et al., 2022), it is crucial to ensure that the increased use of adaptive testing does not negatively impact test-takers' emotional situation.

The present study contributes to this goal by employing a within-person, repeated-measures design to explore test-takers' emotional experiences during adaptive testing compared

to classical fixed-item testing, as well as potential interactions of test type with ability and perceived test difficulty. We add to previous research by investigating not only self-reported emotions, but also psychophysiological states as indicated by test-takers' skin conductance response. Doing so allows for a differentiated picture of the effects of adaptive testing on test-takers' affective state.

### ***8.2.1 Principles of Adaptive Testing***

In classical fixed-item testing, each individual is presented with the same set of items of varying difficulty. In adaptive testing, instead, an underlying algorithm continuously estimates participants' ability and selects the next item accordingly (Thompson & Weiss, 2011; Weiss, 2004). Usually, test difficulty is set to 50%, as this maximizes test efficiency and precision (Wise, 2014). This procedure implies that items are selected in a way that the test-taker has a 50% probability of solving them correctly based on the algorithm's current estimate of their ability (Thompson & Weiss, 2011). Following this procedure, adaptive testing is more efficient than fixed-item testing, with tests requiring only 53-57% of items to achieve the same precision of ability estimates as a fixed-item test (Frey & Ehmke, 2007). In practice, adaptivity of a test usually implies that solving an item correctly is followed by the presentation of a more difficult item, while an easier item follows incorrect answers. As a result, individuals with higher ability face relatively difficult items, and individuals with lower ability encounter relatively easy items. In contrast, for fixed-item testing, all participants get the same set of items with identical difficulties. Accordingly, items on a fixed-item test should be easier to solve for individuals with high ability than for those with low ability. As different individuals encounter different sets of items in adaptive testing, item-response-theory is used to obtain a final ability estimate, as opposed to the typical use of sum scores in fixed-item testing (Frey & Ehmke, 2007).

While adaptive testing outperforms fixed-item testing in terms of efficiency, it is crucial to ensure that these advantages do not happen at the expense of test-takers' affective and motivational experiences. The present study compares adaptive and fixed-item testing in terms of test-takers' achievement emotions, specifically psychophysiological arousal and subjective affective experience, driven by differential experiences of test difficulty.

### ***8.2.2 Achievement Emotions and Their Link with Test Difficulty***

The emotions test-takers experience during testing are likely achievement emotions, which are defined as emotions that occur in situations "judged according to competence-based standards of quality" (Pekrun et al., 2023, p. 146). They are "multicomponent processes, with

components loosely coupled” (Pekrun et al., 2023, p. 146), which comprise the subjective affective experience as well as motivational tendencies, expressive behavior, cognitive appraisals, and psychophysiological processes (Pekrun, 2006).

Effects of tests on test-takers’ emotions can be explained using Pekrun’s (2006, 2021, 2024) control-value theory of achievement emotions. This theory posits that subjective control, that is, an individual’s perceived causal influence over actions and outcomes, and subjective value, that is, perceived intrinsic and extrinsic value of the activity or outcome, interact in predicting achievement emotions. According to the theory, positive achievement emotions are prompted by high levels of subjective control and value. Negative achievement emotions are triggered by a lack of control, combined with high value (except for boredom, which should be reduced by high value; Pekrun et al., 2023).

We propose that the use of an adaptive testing format would not have any systematic effect on the perceived value of a given test (e.g., the value of a college entrance exam should be similarly high for an individual, independent of whether it is an adaptive or fixed-item test). Control, in contrast, may differ between an adaptive and a fixed-item test, based on individuals’ perceptions of difficulty while taking the respective test. In fixed-item testing, control perceptions should vary considerably between individuals, depending on their ability: Test-takers with low ability should experience a fixed-item test as relatively hard and, therefore, less controllable, whereas test-takers with high ability should experience the same test as relatively easy, and thus more controllable. In adaptive testing, instead, the difficulty of each item is adapted to the person’s ability, supposedly leading to perceptions of difficulty and corresponding experiences of control being similar for all persons (Betz & Weiss, 1976). We propose that differences in perceived control due to different perceptions of difficulty in adaptive and fixed-item tests drive effects of test type on affective states, with differential effects on the psychophysiological and subjective components of achievement emotions.

### ***8.2.3 Psychophysiological Arousal***

Psychophysiological arousal refers to activation in the autonomic nervous system, which is divided into the sympathetic and parasympathetic branches (Kreibig & Gendolla, 2014). The sympathetic nervous system is activated in situations that stimulate the individual. Activation of this system is related to stress (Weissman & Mendes, 2021), attention (e.g., Zhang et al., 2021), and cognitive load (e.g., Nourbakhsh et al., 2017; Vanneste et al., 2021). Activation of the parasympathetic nervous system is associated with relaxation and recovery (Weissman & Mendes, 2021). Whereas parameters such as heart and respiration rates are influenced by both branches, the eccrine sweat glands are only innervated by the sympathetic nervous system,

making them a good indicator of sympathetic arousal without parasympathetic influences (Braithwaite et al., 2015; Ishikawa, 2023; Kreibig & Gendolla, 2014). Eccrine sweat glands are the root of electrodermal activity (EDA), as changes in sweat production cause fluctuations in the skin's conductance and electric potentials (Christopoulos et al., 2019).

EDA is commonly captured by attaching two electrodes to the skin, usually the palm of the hand, and measuring the level of conductance between them when applying a constant current (Boucsein et al., 2012). The resulting signal can be divided into a tonic and a phasic component: The tonic component, termed skin conductance level (SCL), changes rather slowly over time. The phasic component – the skin conductance response (SCR) – responds to stimuli more quickly, with a latency of around one to four or five milliseconds. Due to SCR showing faster changes in response to stimuli compared to SCL, the present study focuses on SCR. SCR is visible as sudden increases (i.e., peaks) in skin conductance (Boucsein et al., 2012; Christopoulos et al., 2019).

### ***Psychophysiology and Achievement Emotions***

Both SCR and SCL have been shown to positively relate to the intensity of positively and negatively valenced emotions, likely triggered by emotion-inherent action tendencies that prepare the body physiologically for approach or avoidance behaviors (Kreibig, 2010). Within the educational-psychological literature, scattered studies have reported on the relationships between EDA and self-reported emotions. Although we still lack a clear understanding of their association (Horvers et al., 2021), it appears that psychophysiological responses are related, yet not equivalent, to subjective emotional experiences. Hence, the inclusion of SCR as a measure of sympathetic physiological arousal adds a new layer of understanding to the question of how adaptive testing influences test-takers' emotional experiences.

### ***SCR in Adaptive and Fixed-Item Testing***

As outlined earlier, perceptions of difficulty might determine differences in emotions between adaptive and fixed-item testing. Relative to fixed-item testing, taking an adaptive test can either constitute a gain or a loss in terms of the perceived ease of solving problems and related perceptions of control. We propose that in both cases, the adaptive test will be accompanied by a higher SCR than the fixed-item test due to the better fit between test difficulty and the individual's ability level. When test items are too difficult for an individual, they report less effort and more boredom (Asseburg & Frey, 2013). Thus, individuals for whom the fixed-item test is clearly too difficult and some items are unsolvable (i.e., individuals with lower ability) may disengage from the task and show weaker arousal, indicated by a decrease in SCR.

In contrast, on the adaptive test, which better matches their ability level, these individuals might stay engaged with the task, showing higher levels of SCR.

Instead, individuals for whom the fixed-item test is relatively easy (i.e., individuals with higher ability) would encounter more difficult items in the adaptive test. Although harder, the items on the adaptive test are still solvable for these individuals, given that item difficulty adapts to their ability level. Thus, they would not disengage in the adaptive test but instead experience a moderate level of perceived control, which might also increase their arousal compared to the fixed-item test. Taken together, since SCR indicates physiological arousal irrespective of valence, we hypothesize that adaptive testing leads to higher SCR relative to fixed-item testing.

### ***8.2.4 Subjective Experiences of Achievement Emotions***

#### ***Prior Findings***

In addition to psychophysiological arousal, we also investigated the subjective emotional experience in adaptive versus fixed-item testing. In early work in this field, Betz and Weiss (1976) hypothesized an overall positive effect of adaptive testing on test-takers' affective experiences, due to the higher fit between test-takers' ability and test difficulty. They posited that this would lead to high-ability individuals being less bored and low-ability individuals being less stressed and frustrated in adaptive than in fixed-item testing. Intriguingly, Betz and Weiss did not appear to consider the possibility of inverse effects, particularly increased stress and frustration in high-ability individuals. In fact, their initial claim did not hold up in their empirical investigation, as they found a main effect of test type with increased anxiety in adaptive testing for all participants (Betz & Weiss, 1976).

Since these findings were published, considerable advances have been made regarding – now computerized – adaptive testing. One might argue that adaptive testing should have generally beneficial effects on the emotional experience during test taking today. However, more recent research further challenges the notion of a uniformly positive effect of adaptive testing on subjective affective experiences: Two meta-analyses by Akhtar et al. (2022) and Frey et al. (2024) comparing adaptive and fixed-item testing concluded that there were no significant differences in self-reported emotions between the two test types. While Akhtar et al. (2022) focused on the experience of test anxiety, Frey et al. (2024) investigated negative and positive emotions, noting that more studies on distinct emotions are needed for a more in-depth understanding. Both meta-analyses found test difficulty to be a central factor. When the adaptive test was set to an average success rate higher than 50% in individual studies, participants reported less test anxiety (Akhtar et al., 2022) and generally less intense negative emotions (Frey et al., 2024) than in fixed-item testing. However, in these meta-analyses, difficulty was

only assessed with regard to the adaptive test. Not coded was how this feature differed between the adaptive and fixed-item test, that is, whether the adaptive test was more or less difficult than the fixed-item test. With this information missing, it is possible that it is not the adaptivity of an easier adaptive test that drives the more positive emotional experience compared with a fixed-item test, but simply the fact that the adaptive test happened to be easier, and therefore more controllable, than the fixed-item test. This is in line with control-value theory, according to which different perceptions of difficulty of the two tests should determine which of them is accompanied by more intense positive or negative emotional experiences. Therefore, we propose that asking for a main effect of adaptive versus fixed-item testing on subjective emotional experience is too simplified. Instead, the difference should depend on a combination of features of both tests as well as the test-taker.

### ***Considering Ability and Difficulty Perceptions***

To fully grasp the potential effects of adaptive versus fixed-item testing on subjective emotional experiences, we propose that it is essential to consider test-takers' ability and their perceptions of the relative difficulty of the two tests. In fixed-item testing, people with higher ability should perceive items as less difficult than people with lower ability, and hence have a more positive emotional experience (see Goetz et al., 2007, for supporting evidence). In contrast, in adaptive testing, perceptions of difficulty would be independent of test-takers' ability, as item difficulty is adapted to the individual's ability (Betz & Weiss, 1976). These assumptions are supported by Akhtar and Kovacs' (2024) findings showing a significantly positive correlation between individuals' ability and their perception of performing well for a fixed-item test. In contrast, this correlation was non-significant if the test was adaptive. As such, when contrasting adaptive versus fixed-item testing within individuals, test type and personal ability should interact in their effects on test difficulty: For low-ability individuals, the adaptive test would be easier than the fixed-item test; for medium-ability individuals, both tests should be similar in difficulty; and for high-ability individuals, the adaptive test should be harder than the fixed-item test. This central role of relative difficulty, which depends on the interaction of the pre-defined difficulties of the two tests with the individual's ability, might explain the lack of consistency in findings from studies directly comparing adaptive and fixed-item testing: Findings on the effects of test type might take different directions depending on test difficulties and ability levels in the sample. In line with this reasoning, a few studies have considered ability. The results were not consistent: Some studies found that ability was a significant moderator of the relation between test type and value, effort, perceived probability of success, and feelings

of satisfaction (e.g., Betz & Weiss, 1976; Ortner et al., 2013, 2014). Others found no such moderating effect for anxiety as an outcome (e.g., Betz & Weiss, 1976; Ling et al., 2017).

Importantly, these deliberations rest on the assumption that a person's objective ability and resulting differences in test difficulties translate into subjective perceptions of difficulty and control. However, this assumption may not always be correct, especially for adaptive tests. As reported by Ortner et al. (2013), metacognitions may not accurately represent test performance in adaptive testing, as the number of items solved correctly is not an indicator of the final ability estimate. Furthermore, individuals usually do not receive feedback on their performance, which may lead to some holding overly optimistic or pessimistic views of their performance and the controllability of both tests, irrespective of objective difficulty level. As such, individuals might subjectively perceive one test as more difficult than the other, although based on their ability and the resulting objective item difficulties, the opposite would be the case.

Consequently, a possible reason explaining the lack of consistent empirical support for ability as a moderator for the effects of adaptive versus fixed-item testing on affective outcomes might be a discrepancy between objective and subjective difficulty. This is supported by Powell's (1994) finding that not actual, but only perceived performance determined test-takers' preferences for a certain test type. Similarly, perceived performance mediated the relationship between objective performance and metacognitions of difficulty, effort, and satisfaction (Ortner et al., 2013), as well as between objective performance and motivation (notably, however, not between perceived performance and anxiety; Tonidandel et al., 2002). Taken together, it might not be the relative objective difficulty of an adaptive and a fixed-item test that determines emotional experience. Even more important might be the relative perceived difficulty of the two tests. In line with control-value theory, we therefore sought to test the assumption that whichever test the individual perceives as easier, hence more controllable, will elicit a more favorable emotional experience.

### ***8.2.5 The Present Study***

While the superiority of adaptive over fixed-item testing in terms of psychometric efficiency seems undisputed (Frey & Ehmke, 2007), potential effects of adaptive testing on test-takers' affective states are underexplored. The present study aims to contribute to this literature in three ways: First, by complementing the classical mode of inquiry through self-report by a psychophysiological measure; second, by systematically considering the interaction between test type and person characteristics for the subjective emotional experiences; and third, by applying a within-person instead of between-person experimental design.

Self-report is prone to response sets and memory biases, and it covers only the subjective aspect of the emotion process. Nevertheless, to the best of our knowledge, no study on adaptive testing has used psychophysiological measures to date. Based on other studies using skin conductance measures as indicators of achievement emotions, we chose SCR as the psychophysiological outcome of interest, indicating sympathetic arousal (e.g., Kiuru et al., 2022; Roos et al., 2023). We assumed that an adaptive test would be accompanied by higher levels of SCR than a fixed-item test.

Furthermore, previous studies typically considered adaptivity as the only difference between adaptive and fixed-item tests, or at best included either test takers' ability or characteristics of only the adaptive test in their investigations. The existing research neglected the possibility that differences between the tests on features other than adaptivity, such as their difficulties, likely influence emotional responses via control perceptions. Therefore, in the present study, we considered participants' ability and their perceptions of the relative difficulty of the two types of tests as possible moderators of effects of test type on achievement emotions.

Another possible reason for the lack of consistency in empirical findings is the predominant use of between-group experimental designs. These designs are susceptible to a priori-group differences, specifically when samples are small and hence, sampling errors are large (see also Pekrun, 2023). Furthermore, they are not well suited to capture the within-person processes that generate emotions. Therefore, the present study used a within-person experimental design to investigate differences in affective states during a computerized adaptive test (CAT) versus a computerized fixed-item test (FIT). The within-person design ensured that person characteristics were held constant across the two conditions.

Succinctly stated, we tested the following hypotheses:

*Hypothesis 1.* Participants show higher psychophysiological arousal as indicated by SCR in the CAT compared to the FIT, as EDA is independent of emotional valence and the CAT should generally elicit stronger emotional arousal compared to the FIT.

*Hypothesis 2.* For self-reported emotions, we expected no main effects of test type due to the following disordinal interactions with ability and relative perceived difficulty.

*2a.* Ability moderates the effect of test type on self-reported emotions: Higher-ability individuals experience more negative and less positive emotions in the CAT than the FIT (because for them, the CAT should be harder/less controllable). Lower-ability individuals experience more intense positive and less intense negative emotions in the CAT than the FIT (because for them, the FIT should be harder/less controllable).

2b. The effect of test type on self-reported emotions depends on relative perceived difficulty, with more positive and less negative emotions in the test that is perceived as easier by the individual.

At first sight, it may seem counterintuitive to assume a main effect of test type on SCR, but interactions of test type with ability and relative perceived difficulty for self-reported emotions. However, although related, the physiological and subjective components of achievement emotions are not identical. Higher general arousal during the adaptive test does not conflict with specific emotions being experienced at a higher level in the fixed-item test, depending on ability and relative perceived difficulty. The combination of the two hypotheses demonstrates how different features of the emotional experience can be integrated to provide a more comprehensive picture of individuals' emotions during testing.

### **8.3 Methods**

#### ***8.3.1 Ethics Statement and Data Transparency***

The research reported herein was conducted in accordance with the APA ethical standards and has received a formal waiver of ethical approval by the ethics committee of the Faculty of Psychology and Educational Sciences at the University of Munich. Participation in the study was voluntary, and no identifiers that could link individual participants to their results were obtained. All participants provided informed consent. Data and analysis code are available on OSF ([https://osf.io/ys7qx/?view\\_only=e895327efb9a4cab9ce704d1eed2d2ee](https://osf.io/ys7qx/?view_only=e895327efb9a4cab9ce704d1eed2d2ee)).

#### ***8.3.2 Participants***

The study was conducted at a large, research-oriented university in southern Germany. Participants were recruited via university mailing lists and social media postings. Of the  $N = 89$  participants, 60 identified as female, 26 as male, and 2 as diverse. For one participant, demographic information was missing. Age ranged between 18 and 77 years with a mean of 26.57 years ( $SD = 8.72$ )<sup>2</sup>. Of the sample, 84% were students in different undergraduate and graduate degree programs. The remaining 16% were working or retired, with the majority also holding a university degree.

Based on an a priori specified validation protocol, 19 participants were excluded due to being non-responders, that is, not showing reactions to external stimuli in their SCR (for information on non-responders, see e.g., Ikezawa et al., 2012; Venables & Mitchell, 1996).

---

<sup>2</sup> Based on the broad age range, we checked the robustness of our findings in a reduced sample excluding six age-related outliers (defined as values of more than 1.5 interquartile ranges above or below the mean age). All effects related to Hypotheses 1, 2a, and 2b had the same significance and direction as in the full sample. As such, we report results from the full sample.

Furthermore, EDA has not been recorded for two participants due to technical issues, resulting in a sample of 68 persons for the analysis testing H1. For the analyses of self-reported emotions, two participants had to be excluded due to an error in generating participant codes, resulting in a sample of 87 participants for analyses of self-reported emotions testing H2a and H2b.

### **8.3.3 Procedure**

After participants had arrived at the lab and filled in a consent form, the experimenter attached two electrodes and a wristband to their non-dominant hand, which they placed on a foam block on the table. They were asked to follow the instructions on the screen and move as little as possible to avoid movement artifacts in the skin conductance recording. First, they underwent the validation protocol for the EDA recording, in which they were instructed to hold their breath and bite their tongue for ten seconds each. They then received the test instructions with the note to only use the hand without the electrodes. Aiming to increase the perceived value of the test, participants were told that the items measured numerical reasoning ability as one component of intelligence and that they would receive feedback on their performance at the end.

They were informed that there would be two separate tests, but not what the difference entailed. The order of presentation (adaptive first vs. fixed-item first) was counterbalanced across participants. Each test consisted of 12 numerical reasoning items, split into three blocks of four items. Items were not timed. Participants were instructed to enter “X” to proceed if they could not find a correct solution. After the first test, there was a break where participants were instructed to take a breath and relax. They could end the break and continue whenever they wanted by clicking a button on the screen. After the second test, they received information on how many of the total 24 items they had solved correctly. Since there was no time limit on the items and the break, time spent on the two tests, including the self-report ratings and the break, varied between participants. The mean duration was 41.8 minutes ( $SD = 12.9$ ). Upon completion, participants were debriefed on the different test types and the purpose of the study and received either twelve euros or participant credits for their participation.

### **8.3.4 Measures**

#### ***Numerical Reasoning Tests***

Both the CAT and the FIT consisted of twelve items assessing numerical reasoning ability. The items were rows of numbers, in which participants had to identify a pattern and complete the rows with one or two numbers accordingly (see examples in Figure 1). The items were presented on a computer screen. Due to counter-balancing the test order, 45 participants

started with the CAT, and 44 started with the FIT. Both tests were based on the 49 numerical reasoning items generated by Loe et al. (2018). Of those, twelve items of varying difficulty were selected for the FIT in a way that the Rasch-scaled difficulty estimate would be above zero for half of the items and below zero for the other half, with a mean of 0.165. The goal of this selection process was to create an average success rate of 50% in the FIT for our sample, which primarily comprised university students with presumably above-average cognitive abilities. The twelve items of the FIT were presented with increasing difficulty. The remaining 37 items constituted the item pool for the CAT. The first item of the CAT had a Rasch-scaled difficulty of 0.11, and the following eleven items were selected using the maximum Fisher information criterion based on Bayes modal ability estimates with test difficulty set to 50%. Both tests were run on the Concerto Platform (The Psychometrics Centre, n.d.).

### ***Skin Conductance Response***

Participants' skin conductance was measured using two Shimmer EDA electrodes attached to the palm of the non-dominant hand connected to a Biopac BioNomadix wristband. The signal was transmitted to a Biopac MP160 receiver and Biopac Bionomadix 2CH GSR/EDA Amplifier and recorded in the software iMotions (iMotions, 2022) with a frequency of 500 Hz.

### **Figure 1**

#### *Exemplary Numerical Reasoning Items*

Item 1:	8	16	32	64	128	<input type="text"/>	
Item 2:	14	28	23	46	41	<input type="text"/>	<input type="text"/>

*Note.* These items are examples of the type of items, not actual items used in the present study. Solutions are: “256” for Item 1, and “82; 77” for Item 2.

The experimenters continually recorded potential reasons for response artifacts during the experiment, such as participants talking or loud noises. In case of such events and corresponding visible artifacts in the signal, the time periods containing these artifacts were manually removed. Furthermore, the time periods in which participants filled in the self-report questionnaires were removed, so that the skin conductance signal only contained periods when participants were working on the numerical reasoning task. Using the Peak Detection Algorithm

implemented in the iMotions software (see Table 1 for settings), a Peaks Per Minute (PPM) value was calculated for each participant on the CAT and the FIT, respectively.

### ***Self-Reported Achievement Emotions and Relative Perceived Difficulty***

After each item block, participants were instructed to fill in a pen-and-paper questionnaire placed in front of them, assessing their subjective affective state and perception of difficulty. To assess affective state, participants were presented with single-item statements for each emotion: “I am enjoying this/I feel proud/I feel angry/I feel bored/I feel stressed/I feel frustrated/I feel tense and nervous” and asked to indicate their endorsement (“Please choose the option that describes best how you are currently feeling”) on a five-point Likert scale (1 = *completely disagree* to 5 = *completely agree*). A total score for each emotion was obtained by averaging the ratings across all three time points per test. In case of only one missing value per test, the mean score of the remaining two time points was used.

**Table 1**

#### *Settings for Peak Detection Algorithm in iMotions Software*

Phasic Filter Length	8000 ms
Lowpass Filter Cutoff Frequency	5 Hz
Peak Onset Threshold	0.01 microSiemens
Peak Offset Threshold	0 microSiemens
Peak Amplitude Threshold	0.01 microSiemens
Minimum Peak Duration	500 ms

*Note.* We used the default settings implemented in the iMotions software, with the exception of Peak Amplitude Threshold, which we set manually based on recommendations by Boucsein (2012).

Participants’ emotion ratings varied across blocks, likely due to the increasing item difficulty in the FIT. Nevertheless, the ratings were highly consistent in each of the two tests, as indicated by high Cronbach’s Alphas across the three time points per test (CAT: .93/.83/.85/.87/.87/.93; FIT: .91/.85/.77/.88/.82/.92 for joy/pride/anger/boredom/frustration/anxiety, respectively).

At each of the three self-report time points per test, participants were further asked to complete the statement “To me, the tasks are...,” with the five response options: “very easy” (1), “rather easy” (2), “neither easy nor hard” (3), “rather hard” (4), “very hard” (5). From all three time points per test, a mean score of perceived difficulty was calculated for the CAT and FIT, respectively (Cronbach’s Alpha .77 for CAT and .68 for FIT). Based on these mean scores,

a relative perceived difficulty score was calculated for each individual by subtracting the perceived difficulty of the FIT from the perceived difficulty of the CAT. Hence, values of relative perceived difficulty below zero imply that the participant found the CAT to be easier than the FIT, and values above zero that the participant perceived the FIT to be easier than the CAT.

### ***Ability Estimate***

To obtain ability estimates, participants' performance on all 24 items of the two tests was considered. Given that item difficulties were available for all items, a Rasch model could be applied for estimating each participant's ability score, resulting from which items they had solved correctly across both tests. We used the `thetaEst` function of the `catR` package (Magis & Raïche, 2012) to obtain these ability estimates.

### **8.3.5 Analyses**

We accounted for the within-person design by estimating multilevel linear regression models with random intercepts. To test Hypothesis 1 on physiological arousal, we estimated a multilevel linear regression model with PPM as the outcome and test type (CAT vs. FIT) as well as time (first vs. second test) as predictors. To test Hypotheses 2a and 2b, we specified the same multilevel linear regression models, one for each of the discrete emotion scores as outcomes. For Hypothesis 2a, we additionally included ability as well as a term for the cross-level interaction between ability and test type as predictors. For Hypothesis 2b, we added relative perceived difficulty and its interaction with test type as predictors. Our primary focus was on the interaction effects, presumably showing different directions of the effect of test type depending on ability and relative perceived difficulty, respectively. To determine statistical significance, we used  $\alpha = .05$ .

## **8.4 Results**

### ***8.4.1 Descriptive Statistics and Manipulation Check***

Descriptive information for all variables, separately for each test type, can be found in Table 2. On average, participants showed five to six peaks per minute in their SCR while performing the tests. Further, while the items for enjoyment, pride, anger, frustration, and anxiety were endorsed, on average, just below the mid-point of the scale, endorsement was lowest for the boredom items, with a total average towards the lower end of the self-report scale. Participants' Rasch ability estimates based on their performance on all 24 items ranged from -2.08 to 3.24, with a mean of 0.17 ( $SD = 1.09$ ), which is slightly above the population average of 0 based on the item calibration described in Loe et al. (2018). It is worth noting that the

average Rasch-scaled item difficulty was -.12 lower on the CAT than on the FIT. This difference, although small in size, was significant, as indicated by a paired-sample  $t$ -test,  $t(88) = -2.07$ ,  $p = 0.041$ . In line with this finding, test scores (i.e., the number of correctly solved items) were slightly higher on the CAT compared to the FIT,  $t(88) = 2.10$ ,  $p = .039$ . This difference needs to be considered when interpreting the effects of test type on physiological and self-reported outcome variables. Finally, the average item endorsement for the difficulty judgement was just above the mid-point of the scale for both tests. The measure of relative perceived difficulty of the CAT and the FIT varied quite symmetrically around zero, ranging between -1.33 and 1.33 with a mean of 0.05 ( $SD = 0.61$ ).

To gain a better understanding of the relative difficulty perceptions, we split the sample into three subgroups:  $n = 33$  participants who perceived the CAT as easier than the FIT,  $n = 35$  who perceived the FIT as easier than the CAT, and  $n = 19$  for whom the two tests had the same mean perceived difficulty. Descriptively (see Table 3), the average ability level was highest in the group that found the FIT easier and lowest in the group that found the CAT easier. However, the differences in ability scores between the three groups were not statistically significant, as indicated by a one-way ANOVA,  $F(2, 84) = 1.01$ ,  $p = .368$ . This finding supports the notion that a person's objective ability does not necessarily directly translate into their relative subjective experiences of difficulty in the different tests. Regarding objective difficulty, paired-sample  $t$ -tests within each subgroup showed no significant differences between objective difficulty on the CAT and FIT within the group that found the FIT easier,  $t(34) = 1.02$ ,  $p = .314$ , and in the group that perceived the same level of difficulty,  $t(18) = -0.27$ ,  $p = .792$ . In the group that perceived the CAT as easier, however, the objective mean difficulty level was indeed significantly lower on the CAT than on the FIT,  $t(32) = -2.72$ ,  $p = .011$ .

Paired-sample  $t$ -tests comparing scores (i.e., the number of correct responses) on the two tests within each group further revealed that scores were significantly higher on the CAT than on the FIT for the group that perceived the CAT as easier,  $t(32) = 2.43$ ,  $p = .021$ , as well as in the group that perceived the tests as similar in difficulty,  $t(18) = 3.01$ ,  $p = .007$ . In contrast, in the group that perceived the FIT as easier, scores did not differ significantly between the two tests,  $t(34) = -0.61$ ,  $p = .549$ . Altogether, relative perceived difficulty does not seem to consistently follow from either an individual's ability, the actual difficulty of the two tests, or participants' test scores, suggesting that it is a highly specific individual appraisal.

**Table 2***Descriptive Statistics per Test Type*

	CAT		FIT	
	<i>M</i> ( <i>SD</i> )	Min - Max	<i>M</i> ( <i>SD</i> )	Min - Max
PPM ( <i>n</i> = 68)	5.61 (3.15)	0.03 - 12.61	5.23 (2.79)	0.00 - 10.53
Joy ( <i>n</i> = 87)	2.86 (1.22)	1.00 - 5.00	2.84 (1.16)	1.00 - 5.00
Pride ( <i>n</i> = 87)	2.25 (1.01)	1.00 - 5.00	2.20 (0.95)	1.00 - 4.67
Anger ( <i>n</i> = 87)	2.25 (1.00)	1.00 - 5.00	2.11 (0.88)	1.00 - 4.33
Boredom ( <i>n</i> = 87)	1.72 (0.85)	1.00 - 4.67	1.65 (0.88)	1.00 - 5.00
Frustration ( <i>n</i> = 87)	2.77 (1.04)	1.00 - 5.00	2.72 (1.03)	1.00 - 5.00
Anxiety ( <i>n</i> = 87)	2.67 (1.06)	1.00 - 5.00	2.63 (1.10)	1.00 - 5.00
Obj. Difficulty ( <i>n</i> = 89)	0.04 (0.58)	-1.52 - 1.54	0.16 (0.00)	0.16 - 0.16
Test Score ( <i>n</i> = 89)	6.54 (2.19)	2.00 - 12.00	6.12 (2.58)	1.00 - 12.00
Perc. Difficulty ( <i>n</i> = 87)	3.66 (0.72)	1.67 - 5.00	3.61 (0.65)	1.67 - 5.00

*Note.* Descriptive statistics per test type without considering test order. Sample size varies between variables, since data for some variables needed to be excluded for some participants (total sample  $N = 89$ ; 68 included for physiological measure, 87 for self-report measures, full sample for information related to test difficulty and score; see section “Participants”). Objective difficulty was obtained by averaging the difficulty estimates of the 12 items per test. Test score is the number of correctly solved items out of the 12 items per test.

**Table 3***Descriptive Statistics within Subgroups of Relative Perceived Difficulty*

	Average item difficulty on			
	CAT <i>M</i> ( <i>SD</i> )	Ability <i>M</i> ( <i>SD</i> )	Test score <i>M</i> ( <i>SD</i> )	
			CAT	FIT
CAT perceived easier ( <i>n</i> = 33)	-0.11 (0.59)	-0.07 (1.08)	6.24 (2.26)	5.52 (2.54)
Same perceived difficulty ( <i>n</i> = 19)	0.13 (0.64)	0.18 (1.11)	6.79 (1.99)	5.63 (2.52)
FIT perceived easier ( <i>n</i> = 35)	0.08 (0.50)	0.28 (0.98)	6.49 (2.15)	6.69 (2.37)

Next, we explored whether the CAT adaptivity algorithm built into the Concerto Platform (The Psychometrics Centre, n.d.) was indeed adaptive in terms of matching items to the individuals' abilities. To this end, we obtained a correlation between the average objective, Rasch-scaled item difficulty in the CAT, and participants' ability estimate. We expected a strong relationship between the two variables, as the CAT should present more difficult items to examinees with higher levels of ability. Indeed, this correlation was  $r = .86$ , indicating that the CAT adapted the difficulty of the items to the examinee's ability level. Furthermore, as expected, there was a high correlation between ability and test score on the FIT ( $r = .89$ ), showing that participants with higher ability solved more items than those with lower ability. However, counter to our expectation, ability and test score were also strongly related on the CAT ( $r = .93$ ), implying that participants with higher ability still solved more items correctly than participants with lower ability, despite the adaptivity of the test. Thus, in our CAT, item difficulty was indeed adaptive to test-takers' ability level, but still, individuals with higher ability levels solved considerably more items correctly than individuals with lower ability levels.

#### ***8.4.2 Effects of Test Type on Psychophysiological Arousal (H1)***

In Hypothesis 1, we expected higher physiological arousal during the CAT compared to the FIT. We used a multilevel linear regression model with test type and time as predictors and a random intercept per participant. The results support our one-sided hypothesis, with an average of 0.30 PPM less in the FIT compared to the CAT ( $p = .013$ ), controlling for time and allowing for random intercepts. Time itself was also a significant predictor of psychophysiological arousal, with 1.26 PPM less in the second compared to the first test ( $p < .001$ ). The fixed-effects intercept was 6.20 PPM ( $p < .001$ ).

#### ***8.4.3 Effects of Test Type and Ability on Self-Reported Achievement Emotions (H2a)***

Hypothesis 2a posited an interaction between test type and ability in predicting discrete achievement emotions. Table 4 shows the results from the random-intercept multilevel linear regression models with test type, time, ability, and the ability\*test type interaction as predictors of test-takers' emotion scores. As expected, there were no significant main effects of test type on any of the self-reported emotions. Contrary to expectations, the interaction between ability and test type was also not significant for any of the emotions (Table 4). Independently of test type, time (first vs. second test) had a significant effect. Reported joy and anxiety scores were significantly lower, and boredom significantly higher during the second test. Likewise, independent of test type, ability significantly affected the emotions; participants with higher

ability reported significantly more joy and pride, and less frustration and anger. For anxiety and boredom, no relationship with ability could be detected.

#### ***8.4.4 Effects of Test Type and Relative Perceived Difficulty on Self-Reported Achievement Emotions (H2b)***

Hypothesis 2b posited an interaction between test type and relative perceived difficulty in predicting discrete achievement emotions. We used multilevel linear regression with a random intercept for all six self-reported emotion variables. Predictors were test type, relative perceived difficulty, time, and the interaction between test type and relative perceived difficulty. The results are displayed in Table 5. There were no significant main effects of test type or relative perceived difficulty on the emotions (with the exception of frustration, which was higher the more the FIT was perceived easier than the CAT). Time significantly predicted joy, pride, and anxiety (decreasing over time) as well as boredom (increasing over time), whereas frustration and anger remained stable over time. Most importantly, as expected, there were significant interactions between test type and relative perceived difficulty on all emotions except boredom. These interactions are visualized in Figure 2. In line with our hypotheses, the findings indicate that the test perceived as easier was accompanied by more positive and less negative emotions, relative to the other test. These differences were more pronounced with higher differences in perceived difficulty. That is, when perceived difficulty differed only slightly between the two tests, the emotional experience was more similar across the tests than when one test was perceived as much easier or harder than the other.

**Table 4**

*Results of Multilevel Linear Regression Analysis for H2a*

	Joy	Pride	Anxiety	Frustration	Boredom	Anger
Intercept	2.90***	2.29***	2.76***	2.75***	1.57***	2.26***
Test type: FIT	-0.02	-0.06	-0.02	-0.03	-0.06	-0.14
Time: Second test	-0.21**	-0.14	-0.19**	0.10	0.31***	0.05
Ability	0.49***	0.25*	0.01	-0.23*	-0.08	-0.27**
Test type*ability	0.01	0.10	-0.11	-0.13	-0.04	-0.05

*Note.*  $N = 87$ . Coefficients are unstandardized regression coefficients. Effects shown pertain to reference categories “CAT” and “First Test.” \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

**Table 5***Results of Multilevel Linear Regression Analysis for H2b*

	Joy	Pride	Anxiety	Frustration	Boredom	Anger
Intercept	2.98***	2.33***	2.75***	2.70***	1.55***	2.21***
Test type: FIT	-0.04	-0.07	-0.03	-0.02	-0.06	-0.13
Time: Second test	-0.22**	-0.15*	-0.19**	0.11	0.32***	0.06
Relative perceived difficulty	-0.29	-0.12	0.24	0.39*	0.20	0.20
Test Type* Relative perceived difficulty	0.57***	0.54***	-0.24*	-0.66***	-0.21	-0.43***

*Note.*  $N = 87$ . Coefficients are unstandardized regression coefficients. Effects shown pertain to reference categories “CAT” and “First Test.” Relative Perceived Difficulty represents the difference between difficulty on the CAT and the FIT, with values  $< 0$  indicating that the CAT was perceived as easier and  $> 0$  that the FIT was perceived as easier.

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

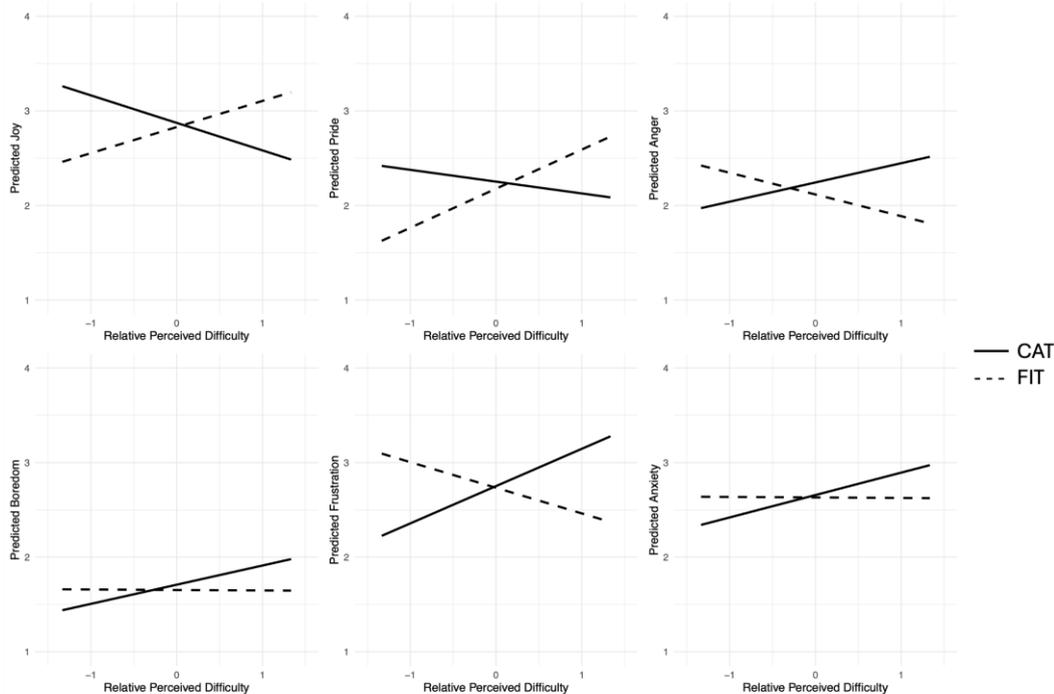
## 8.5 Discussion

Driven by the increasing use of adaptive testing in educational assessment, the present study investigated the question of whether and how adaptive testing influences test-takers’ emotional experiences compared to classical fixed-item testing. As achievement emotions considerably impact performance (Pekrun et al., 2023) and school-related well-being (Obermeier et al., 2022), emotional experiences in adaptive testing need to be investigated to ensure that this more efficient way of testing is not accompanied by undesirable emotional effects on test-takers. To assess test-takers’ affective states, we assessed their psychophysiological arousal (specifically, their SCR), alongside their self-reported discrete achievement emotions (joy, pride, anger, anxiety, frustration, and boredom). Regardless of whether the adaptive or non-adaptive test was administered first, we observed a significant decrease in physiological arousal and self-reported joy, pride, and anxiety, as well as a significant increase in self-reported boredom from the first to the second test. These findings suggest that over the course of a testing situation, physiological arousal and emotional activation decreased, while the experience of boredom as a deactivating emotion increased. To maintain participant engagement throughout a testing situation, it may therefore be advisable to keep tests as short as possible.

Regarding the effects of adaptive compared to non-adaptive testing, in line with expectations, a key finding was that test-takers were more strongly physiologically aroused while working on the adaptive test than during the fixed-item test. However, counter to our expectations, test-takers with low ability did not benefit emotionally from receiving easier tasks in adaptive testing, nor did test-takers with high ability suffer emotionally from adaptive testing presenting them with objectively harder tasks. Yet, we did find support for our hypothesis that the subjective perception of the relative difficulty of the two tests impacted participants' emotional experiences: Whichever test was perceived as easier was accompanied by more joy and pride, and less frustration, anxiety, and anger. Hence, the key message of the present contribution is that while adaptive tests appear to elicit stronger arousal in participants, the subjective emotional experience seems to be driven by subjective perceptions of difficulty independent of the presence or absence of adaptivity. These results alleviate concerns regarding possible adverse effects of adaptive testing on test-takers' affective state.

## Figure 2

*Interactions between Relative Perceived Difficulty and Test Type in Predicting Self-Reported Achievement Emotions*



*Note.* Relative Perceived Difficulty  $< 0$  indicates that the CAT was perceived as easier and  $> 0$  that the FIT was perceived as easier. For boredom, the interaction effect was not significant.

### ***8.5.1 Effects of Adaptive versus Fixed-Item Testing on Psychophysiological Arousal***

To the best of our knowledge, this is the first study to investigate the effect of adaptive testing on test-takers' psychophysiological responses, thereby including an objective indicator of emotional arousal (Pekrun et al., 2023). Specifically, we measured test-takers' SCR as an indicator of sympathetic arousal, which is a process that prepares the individual for action (Rosebrock et al., 2017). In line with our expectations, this valence-independent level of psychophysiological arousal was higher during the adaptive test compared to the fixed-item test. We propose that this main effect of test type on psychophysiological arousal is driven by differential mechanisms depending on the individual's ability level: For individuals with lower ability, the fixed-item test might have exceeded their ability level to an extent that caused them to "switch off" (for similar findings, see Asseburg & Frey, 2013), resulting in generally lower arousal during this test. In contrast, the adaptive test offered them solvable items throughout, keeping them engaged and therefore leading to higher levels of arousal. For individuals with higher ability, the items on both tests were solvable, presumably supporting a certain engagement during both tests. The adaptive test challenged these participants more, thus resulting in higher arousal.

Furthermore, average item difficulty was slightly lower, and test scores were accordingly slightly higher in the CAT than in the FIT. This further supports the notion that the adaptive test being more engaging might result from this test providing both lower- and higher-ability participants with the opportunity to perform well. Since, however, our findings did not show a main effect of test type on the self-reported emotional experience, the higher levels of arousal during the adaptive test might stem from alternative engagement-related processes such as higher levels of attention (Zhang et al., 2021) or cognitive load (Nourbakhsh et al., 2017; Vanneste et al., 2021) rather than differences in emotional experience. This assumption aligns with the finding that more effort may be invested in taking an adaptive compared to a fixed-item test, with effort measured by reaction time (Akhtar & Kovacs, 2024).

### ***8.5.2 Effects of Adaptive Versus Fixed-Item Testing on Subjective Emotional Experience***

Regarding the subjective experience of discrete achievement emotions, we expected interactions between test type and test-takers' characteristics as predictors. Based on control-value theory (Pekrun, 2006), we hypothesized that emotional experiences in adaptive versus fixed-item testing would depend on relative perceived control related to these two tests: Whichever test the individual perceives as more controllable would be accompanied by a more positive and less negative emotional experience. From this perspective, the commonly asked question of whether emotional experiences differ between adaptive versus fixed-item testing

appears too simple, given that the experience would depend on features of the two tests in relation to each other as well as features of the test-taker. We therefore investigated two possible moderators of the effect of test type on self-reported emotions, namely, personal ability and relative perceived difficulty.

Our first interaction hypothesis was not supported, as participants' ability did not significantly interact with test type in predicting any of the discrete emotions. There were only main effects of ability, with higher-ability individuals reporting generally more positive emotions (joy and pride) and less negative emotions (frustration and anger) while working on both tests. Notably, though, we observed no main effects of ability on anxiety or boredom.

For interpreting the lack of an interaction, it needs to be noted that our reasoning that ability would moderate the effect of test type was built on the assumption that the adaptive test would be more difficult for higher-ability individuals. As such, we expected that higher-ability individuals would solve fewer items on the adaptive test than on the fixed-item test. In contrast, for lower-ability individuals, we expected the adaptive test to be easier, with more items solved correctly than on the fixed-item test. This assumption was partly supported, as the CAT indeed provided individuals with higher ability with more difficult items.

However, there was also an unexpected, strong correlation between ability and success rate (i.e., test scores) on the CAT. This means that despite receiving relatively easy items in the CAT, participants with lower ability only solved a few of them correctly, whereas participants with high ability solved many items correctly despite facing relatively difficult items in the CAT. It is unclear whether positive relations between ability and success rates are a common phenomenon in adaptive testing. Some studies reported non-significant correlations (e.g., Ortner et al., 2013; Tonidandel et al., 2002), some did not report success rates (e.g., Ortner et al., 2014), and others found high correlations, similar to the present findings (e.g., Ling et al., 2017). For the present study, it appears that the CAT has been adaptive, but not sufficiently so to achieve a similar level of success for all individuals, regardless of their ability. A possible reason might have been a lack of very easy or very hard items in the item bank. Another possible reason is the relatively low number of items in the test, so that the CAT may have been terminated before settling in at a difficulty level that would correspond to the pre-defined success rate of 50%. Therefore, while ability significantly predicted several emotions as expected, it may not have done so in interaction with test type because the experience of control has not differed as much as expected across the two test types at different levels of ability.

To investigate the second interaction hypothesis, we obtained a measure of relative perceived difficulty, indicating which test was perceived as easier and to what extent. We

observed significant interactions between test type and relative perceived difficulty for all emotions except boredom. The more one test was perceived as easier than the other, the more joy and pride, and the less frustration, anxiety, and anger were experienced on this test compared to the other test. This finding supports the idea that, regardless of adaptivity, an individual's emotional experience is more positive and less negative on whichever test they find easier, likely due to higher levels of perceived control.

Early work on the emotional advantages of adaptive testing (e.g., Betz & Weiss, 1976) has claimed that adaptive testing would be beneficial for the emotional experience due to the reduction of boredom in high-ability individuals and the reduction of anxiety in low-ability individuals. Our findings do not support this assumption. They show that the level of anxiety was related to differences in the perceived difficulty of the two tests rather than their adaptability. For boredom, although the interaction was not statistically significant, the results descriptively show a similar pattern: The more one test was perceived as easier than the other, the less boredom was experienced on this test compared with the other test (for similar findings, see Asseburg & Frey, 2013). However, it is worth noting that boredom was generally very low in the lab setting of the present study (see also Goetz et al., 2023). The low levels of boredom may have been due to the challenging nature of the task and the induction of ego-threat by introducing the test as a measure of intelligence, a highly desirable trait for any individual.

Taken together, our findings indicate that it is not individuals' objectively assessed ability that determines differences in their emotional experience in an adaptive versus a fixed-item test. Rather, their subjective perceptions of difficulty and how they compare between the tests may be driving these differences. This finding corroborates previous studies showing that only perceived, not actual performance is associated with the preference for a certain test (Betz & Weiss, 1976), and that perceived success mediates the effect of actual success on metacognitive experiences like satisfaction (Ortner et al., 2013; Tonidandel et al., 2002).

There was some correspondence between ability and relative perceived difficulty, as participants who found the CAT more difficult than the FIT had higher average ability. However, this difference in mean ability was small and not statistically significant. As such, there were participants with high ability who found the FIT more difficult, as well as some with low ability who found the CAT more difficult.

In sum, our findings on self-reported emotions underline the importance of considering individual perceptions of the test-taker to understand the effects of adaptive testing on emotional experience. Especially the subjective experience of control, indicated by which test is perceived as easier and to what extent, seems to determine how adaptive and fixed-item tests

compare in terms of the emotions they trigger. This finding contradicts the widespread notion that adaptive testing is emotionally beneficial for all test-takers due to a better fit between ability and test difficulty (e.g., Betz & Weiss, 1976).

### ***8.5.3 Limitations and Directions for Future Research***

Several limitations need to be considered in interpreting the findings and can be used to inform directions for future research. Some features of the present study may limit its generalizability to real-life applications. First, the present work was conducted in a lab setting with a sample mainly consisting of university students. While the lab setting allowed us to control variables that impact data quality, it may limit the generalizability of the present findings, in particular to high-stakes testing situations. By informing participants that the task would measure a component of intelligence and that they would receive feedback, we aimed to increase the perceived value of the task and thereby intensify the emotional experience on both tests. Although this instruction may have made the procedure more similar to a real-life testing situation, the setting likely did not fully resemble a high-stakes situation with strongly adverse consequences in the case of failure (e.g., not getting access to a desired study program) or highly desirable consequences in case of success (e.g., getting a desired job offer). In such situations, perceptions of control, and especially a loss or gain of control, might have more profound effects on test-takers' emotional experiences. In particular, when stakes are very high and a test is way too difficult, anxiety might be the dominant emotion, and "switching off," as it was possible in the present context, might be rare.

Second, in the present study, participants were not informed about the adaptivity of the test. In real-life settings, test-takers might be provided with such information, which might lead to perceiving difficult items on an adaptive test not as a loss of control but rather as a sign of having performed well. Third, the study employed numerical reasoning tasks in both tests. Research is needed to replicate the current findings in other domains and establish whether the effects are generalizable across different areas. Fourth, the procedure included both an adaptive and a fixed-item test. This design does not fully mirror typical testing situations, where individuals would rarely encounter both an adaptive and a non-adaptive test. The two tests were administered consecutively, with participants determining the length of the break in between. The break was included to minimize carry-over effects and reinforce the notion in participants that a new test would commence after the break. However, especially participants who experienced negative emotions may have used this break to regulate them, which may have altered their self-report of emotions.

The finding of higher physiological arousal in the adaptive test provides new evidence suggesting that this form of testing is more emotionally arousing for test-takers. The increased arousal was likely triggered by a better balance of task demands and ability that might generate a stimulating level of challenge, increased task engagement, and reduced “switching off.” However, since we were the first to consider psychophysiological reactions to adaptive testing, replication of this effect and further exploration are needed. Specifically, future research could explore how the detected increases in SCR relate to test-takers’ performance and well-being during and after testing.

By considering additional variables, future research could also enhance our understanding of how adaptive testing influences the emotional experience. The present research focused on six common achievement emotions (joy, pride, anger, anxiety, frustration, boredom). While the assessment of these emotions likely covered a significant portion of participants’ emotional experiences, future research could explore how adaptive testing influences other achievement emotions. In addition to assessing a pre-defined set of emotions, this could also be done by including a free-text option to describe the emotional experience. Future research could also include an assessment of perceived control to explore whether control perceptions indeed mediate the effect of test features on physiological and emotional states.

Based on control-value theory, we argued that it would be too simple to assume main effects of adaptive testing on the subjective experience of achievement emotions, as emotional differences between adaptive and fixed-item testing were explained by interactions between individual characteristics and features of the two tests. The findings corroborated this claim by showing that individuals experienced more positive and less negative emotions the more a test appeared easier to them. Surprisingly, neither ability nor objective mean difficulty or test scores could fully explain whether participants would find a certain test easier or not. Research is needed to determine factors that generate differential perceptions of difficulty and could be used as leverage points to positively influence test-takers’ emotional experience. Until these factors are identified, we concur with previous recommendations to increase the success rate in adaptive testing (Asseburg & Frey, 2013). Especially since precision in adaptive testing is still considerably high at success rates of 60% or 70% (Eggen & Verschoor, 2006), it seems sensible to create an adaptive test that allows for relatively low difficulty, thereby providing a more positive emotional experience.

Finally, the present findings represent estimates of within-person effects, which may not necessarily be transferable to a between-person level (Hunter et al., 2024). They also represent

effects with a fixed slope, that is, aggregates of within-person effects. As such, the findings might not hold true for each individual. Future research could use random slopes modeling to investigate generalizability across individuals.

## 8.6 Conclusion

The rising popularity of (computerized) adaptive testing in high-stakes and large-scale assessments raises the question of how adaptive testing affects test-takers' performance and emotional well-being. To the best of our knowledge, the present investigation is the first that used psychophysiological measures to answer this question. The findings show that adaptive testing leads to higher levels of sympathetic arousal compared to fixed-item testing. At the same time, the results imply that neither test format bears a systematic risk of emotionally harming participants. Instead, our results indicate that test-takers' perceptions of difficulty determine how their emotional experience compares between the two types of tests: Whichever test was perceived as easier by the test-taker was accompanied by more positive and less negative emotions. These findings alleviate concerns regarding potential negative effects of adaptivity on test-takers' affective states, thereby encouraging the use of adaptive testing given their psychometric benefits.

## 8.7 References

- Akhtar, H., & Kovacs, K. (2024). Measurement precision and user experience with adaptive versus non-adaptive psychometric tests. *Personality and Individual Differences, 225*, 112675. <https://doi.org/10.1016/j.paid.2024.112675>
- Akhtar, H., Silfiasari, Vekety, B., & Kovacs, K. (2022). The effect of computerized adaptive testing on motivation and anxiety: A systematic review and meta-analysis. *Assessment 30*(5), 1379–1390. <https://doi.org/10.1177/10731911221100995>
- Asseburg, R., & Frey, A. (2013). Too hard, too easy, or just right? The relationship between effort or boredom and ability-difficulty fit. *Psychological Test and Assessment Modeling, 55*(1), 92–104.
- Betz, N. E., & Weiss, D. J. (1976). *Psychological effects of immediate knowledge of results and adaptive ability testing. Research Report 76-4.*
- Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.). Springer US. <https://doi.org/10.1007/978-1-4614-1126-0>
- Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017–1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for psychological experiments.* <https://www.birmingham.ac.uk/documents/colleges-les/psych/saal/guide-electrodermal-activity.pdf>

- Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2019). The body and the brain: Measuring skin conductance responses to understand the emotional experience. *Organizational Research Methods*, 22(1), 394–420. <https://doi.org/10.1177/1094428116681073>
- College, Board. (2022). Transitioning to a digital SAT. <https://professionals.collegeboard.org/pdf/digital-sat-faculty-guide-v3-ada-v0.pdf>
- Eggen, T. J. H. M., & Verschoor, A. J. (2006). Optimal testing with easy or difficult items in computerized adaptive testing. *Applied Psychological Measurement*, 30(5), 379–393. <https://doi.org/10.1177/0146621606288890>
- Frey, A., & Ehmke, T. (2007). Hypothetischer Einsatz adaptiven Testens bei der Überprüfung von Bildungsstandards. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Eds.), *Kompetenzdiagnostik* (pp. 169–184). VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-90865-6\\_10](https://doi.org/10.1007/978-3-531-90865-6_10)
- Frey, A., Liu, T., Fink, A., & König, C. (2024). Meta-analysis of the effects of computerized adaptive testing on the motivation and emotion of examinees. *European Journal of Psychological Assessment*, 40(5), 1–17. <https://doi.org/10.1027/1015-5759/a000821>
- Goetz, T., Bieleke, M., Yanagida, T., Krannich, M., Roos, A.-L., Frenzel, A. C., Lipnevich, A. A., & Pekrun, R. (2023). Test boredom: Exploring a neglected emotion. *Journal of Educational Psychology*, 115(7), 911–931. <https://doi.org/10.1037/edu0000807>
- 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(1), 3–16. <https://doi.org/10.1016/j.lindif.2006.12.002>
- Horvers, A., Tombeng, N., Bosse, T., Lanzonder, A. W., & Molenaar, I. (2021). Detecting emotions through electrodermal activity in learning contexts: A systematic review. *Sensors*, 21(23), 7869. <https://doi.org/10.3390/s21237869>
- Hunter, M. D., Fisher, Z. F., & Geier, C. F. (2024). What ergodicity means for you. *Developmental Cognitive Neuroscience*, 68, 101406. <https://doi.org/10.1016/j.dcn.2024.101406>
- Ikezawa, S., Corbera, S., Liu, J., & Wexler, B. E. (2012). Empathy in electrodermal responsive and nonresponsive patients with schizophrenia. *Schizophrenia Research*, 142(1–3), 71–76. <https://doi.org/10.1016/j.schres.2012.09.011>
- iMotions (Version 9.3). (2022). [Computer software].
- Ishikawa, M. (2023). Measuring the autonomic nervous system as a window into the mind and brain: A selective review. *European Psychologist*, 28(2), 67–82. <https://doi.org/10.1027/1016-9040/a000500>
- Kiuru, N., Malmberg, L.-E., Eklund, K., Penttonen, M., Ahonen, T., & Hirvonen, R. (2022). How are learning experiences and task properties associated with adolescents' emotions and psychophysiological states? *Contemporary Educational Psychology*, 71, 1–16. <https://doi.org/10.1016/j.cedpsych.2022.102095>
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Kreibig, S. D., & Gendolla, G. H. E. (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). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780203148211.ch31>

- Ling, G., Attali, Y., Finn, B., & Stone, E. A. (2017). Is a computerized adaptive test more motivating than a fixed-item test? *Applied Psychological Measurement, 41*(7), 495–511. <https://doi.org/10.1177/0146621617707556>
- Loe, B., Sun, L., Simonfy, F., & Doebler, P. (2018). Evaluating an automated number series item generator using linear logistic test models. *Journal of Intelligence, 6*(2), 20. <https://doi.org/10.3390/jintelligence6020020>
- Magis, D., & Raïche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software, 48*(8), 1–31. <https://doi.org/10.18637/jss.v048.i08>
- Nourbakhsh, N., Chen, F., Wang, Y., & Calvo, R. A. (2017). Detecting users' cognitive load by galvanic skin response with affective interference. *ACM Transactions on Interactive Intelligent Systems, 7*(3), 1–20. <https://doi.org/10.1145/2960413>
- Obermeier, R., Schlesier, J., Meyer, S., & Gläser-Zikuda, M. (2022). Trajectories of scholastic well-being: The effect of achievement emotions and instructional quality in the first year of secondary school (fifth grade). *Social Psychology of Education, 25*(5), 1051–1070. <https://doi.org/10.1007/s11218-022-09726-2>
- OECD. (2022). *PISA 2022 Technical Report*. OECD Publishing. <https://www.oecd.org/pisa/data/pisa2022technicalreport/PISA-2022-Technical-Report-Ch-01-PISA-Overview.pdf>
- Ortner, T. M., Weißkopf, E., & Gerstenberg, F. X. R. (2013). Skilled but unaware of it: CAT undermines a test taker's metacognitive competence. *European Journal of Psychology of Education, 28*(1), 37–51. <https://doi.org/10.1007/s10212-011-0100-7>
- Ortner, T. M., Weißkopf, E., & Koch, T. (2014). I will probably fail: Higher ability students' motivational experiences during adaptive achievement testing. *European Journal of Psychological Assessment, 30*(1), 48–56. <https://doi.org/10.1027/1015-5759/a000168>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R. (2021). Self-appraisals and emotions: A generalized control-value approach. In T. Dicke, F. Guay, H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *Self – a multidisciplinary concept* (pp. 1–30). Information Age Publishing.
- Pekrun, R. (2023). Mind and body in students' and teachers' engagement: New evidence, challenges, and guidelines for future research. *British Journal of Educational Psychology, 93*(S1), 227–238. <https://doi.org/10.1111/bjep.12575>
- Pekrun, R. (2024). Control-value theory: From achievement emotion to a general theory of human emotions. *Educational Psychology Review, 36*(3), 83. <https://doi.org/10.1007/s10648-024-09909-7>
- Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., Van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology, 124*(1), 145–178. <https://doi.org/10.1037/pspp0000448>
- Powell, Z.-H. E. (1994). The psychological impacts of computerized adaptive testing methods. *Educational Technology, 34*(8), 41–47.

- Roos, A.-L., Goetz, T., Krannich, M., Donker, M., Bieleke, M., Caltabiano, A., & Mainhard, T. (2023). Control, anxiety and test performance: Self-reported and physiological indicators of anxiety as mediators. *British Journal of Educational Psychology*, *93*(S1), 72–89. <https://doi.org/10.1111/bjep.12536>
- Rosebrock, L. E., Hoxha, D., Norris, C., Cacioppo, J. T., & Gollan, J. K. (2017). Skin Conductance and Subjective Arousal in Anxiety, Depression, and Comorbidity: Implications for Affective Reactivity. *Journal of Psychophysiology*, *31*(4), 145–157. <https://doi.org/10.1027/0269-8803/a000176>
- The Psychometrics Centre. (n.d.). *Concerto Adaptive Testing Platform*. University of Cambridge. Retrieved 2 April 2024, from <https://www.psychometrics.cam.ac.uk/newconcerto>
- Thompson, N. A., & Weiss, D. A. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation*, *16*(1), 1–9. <https://doi.org/10.7275/WQZT-9427>
- Tonidandel, S., Quiñones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology*, *87*(2), 320–332. <https://doi.org/10.1037/0021-9010.87.2.320>
- Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B. B., Larmuseau, C., Depaepe, F., & Van Den Noortgate, W. (2021). Towards measuring cognitive load through multimodal physiological data. *Cognition, Technology & Work*, *23*(3), 567–585. <https://doi.org/10.1007/s10111-020-00641-0>
- Venables, P. H., & Mitchell, D. A. (1996). The effects of age, sex and time of testing on skin conductance activity. *Biological Psychology*, *43*(2), 87–101. [https://doi.org/10.1016/0301-0511\(96\)05183-6](https://doi.org/10.1016/0301-0511(96)05183-6)
- Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, *37*(2), 70–84. <https://doi.org/10.1080/07481756.2004.11909751>
- Weissman, D. G., & Mendes, W. B. (2021). Correlation of sympathetic and parasympathetic nervous system activity during rest and acute stress tasks. *International Journal of Psychophysiology*, *162*, 60–68. <https://doi.org/10.1016/j.ijpsycho.2021.01.015>
- Wise, S. L. (2014). The utility of adaptive testing in addressing the problem of unmotivated examinees. *Journal of Computerized Adaptive Testing*, *2*(1), 1–17. <https://doi.org/10.7333/1401-0201001>
- Zhang, J., Wang, K., & Zhang, Y. (2021). Physiological characterization of student engagement in the naturalistic classroom: A mixed-methods approach. *Mind, Brain, and Education*, *15*(4), 322–343. <https://doi.org/10.1111/mbe.12300>

---

# 9

---

## GENERAL DISCUSSION

---

Miriam Wunsch



Motivated by the methodological benefits of using physiological measures in educational psychology – namely, their objectivity, unobtrusiveness, and possibility for real-time assessment – the present work aims to advance our understanding of the meaning and application of these measures in achievement situations. Specifically, this work focuses on measures of activation in the autonomic nervous system, which are comparably resource-efficient in their application and have been found to show meaningful associations with psychological processes, including cognition, motivation, and emotion, in an achievement context. Based on an extensive review of existing literature, the present work has identified three central themes related to the use of physiological measures in educational psychology in need for further exploration: 1) The distinction between associations on the between- and within-person level. 2) The meaning of autonomic changes in an achievement situation. 3) Possible differences in the specific meaning of different indicators of autonomic activation. In the following, findings from the presented empirical works are synthesized into five key insights related to these three themes, before outlining the limitations of the present work and deriving implications for future research and educational practice.

### **9.1 Insights Gained on Physiological Measures in an Achievement Situation**

#### **9.1.1 *Level of Analysis***

While between- and within-person approaches can complement one another by providing comprehensive insights into associations between variables (Zawadzki et al., 2017), failing to sufficiently distinguish between the levels in analysis and interpretation can result in inadequate conclusions (Hoffman, 2015; Schuurman, 2023). Previous work on the associations between physiological measures and psychological processes has often overlooked the distinction between the levels of analysis. The present work therefore contributes to a deeper understanding of how the distinction between between- and within-person associations affects the detection and meaning of associations between physiological and psychological states in an achievement context. The two sets of repeated-measures data provide important insights into this issue. Firstly, the intraclass correlations reported in EW1 reveal the presence of substantial between-person variation in the physiological parameters, prompting a discussion on the importance and procedure of adequate baseline correction when performing between-person analyses (Insight 1). Secondly, the findings of EW1 underscore the importance of distinguishing between the two levels in analysis and interpretation, showing that between- and within-person associations between physiological states and self-reported achievement emotions differed considerably, with consistent patterns only at the within-person level (Insight 2).

---

**Insight 1****Considerable between-person variability in physiological parameters underlines the need for adequate baseline correction in between-person approaches.**

---

The intraclass correlations for the two samples, reported in EW1, demonstrate that most of the variance in psychophysiological measures was attributable to between-person differences. Specifically, the ICC1s, which denote the proportion of the total variance explained by between-person rather than within-person differences, ranged between .73 and .96 in Study 1 and between .67 and .95 in Study 2. Given that physiological parameters like EDA and CVA are influenced by a multitude of stable, biological factors, such as a person's age, gender, ethnicity, health status, and lifestyle factors, like physical activity and alcohol and tobacco consumption (Bari et al., 2018; Fatissou et al., 2016; Sammito et al., 2024; Venables & Mitchell, 1996), it is not surprising that most of their variance is driven by between-person differences rather than within-person fluctuations over the course of a single, relatively short achievement situation. Generally, a large proportion of between-person variance is not problematic per se; however, it highlights the presence of considerable differences between individuals in their physiological parameters. As outlined in Section 3.1, failing to adequately account for these substantial pre-existing differences between individuals in physiological parameters that are unrelated to the achievement situation may overshadow differences related to the achievement context, thereby undermining the finding of between-person associations between physiological states and psychological processes.

For the between-person analysis in EW1, we accounted for baseline differences in individuals' physiological parameters using a residualizing approach. Specifically, we predicted the values on the physiological measures from the baseline and extracted the residuals for each measure and time point. The resulting residualized values indicated how far the person's observed value deviated from the value predicted from their baseline assessment. By doing so, we not only controlled for between-person differences in their absolute level (similar to subtracting the baseline from the task): We additionally accounted for the possibility that the amplitude of changes between baseline and task might depend on the person's baseline value. Indeed, this approach considerably reduced between-person variance in the physiological measures, as indicated by lower ICC1s after baseline correction, ranging between .55 and .76 for Study 1, and between .42 and .72 for Study 2. The similar reduction in between-person variance in both studies suggests that the baseline duration (one minute in Study 1, three

minutes in Study 2) and its placement in the procedure (at the beginning in Study 1, after relaxation and validation phases in Study 2) did not affect the effectiveness of using the baseline assessment to reduce between-person variance in the physiological data.

Despite the decrease in the proportion of variance explained by between-person differences, a notable amount of between-person variance remained after the baseline correction. Again, this does not pose an issue *per se*, and in fact, a certain degree of between-person variation is a prerequisite for detecting between-person associations with other variables. However, the considerable remaining between-person variance raises the question of whether the employed baseline-correction approach was able to account for all the between-person differences in on-task physiological parameters driven by biological or environmental factors other than the achievement situation itself. If our approach had not been able to sufficiently correct for task-unrelated differences between individuals in their physiological parameters, this may explain the finding of few to no significant associations with affective states on the between-person level in EW1.

An aspect potentially missing from our approach to baseline correction may have been to not only quantify but also qualify the amplitude of individuals' physiological responses to the task. As introduced in Section 3.1, between-person approaches in the present context seem to refer to the question of whether the intensity of one individual's physiological response compared to another's corresponds to differences in their emotional experiences. Our residualizing approach to baseline correction, as well as those that subtract the baseline from on-task values, operationalize the intensity of a physiological response in the unit of the respective physiological measure. In our case, for example, a residual of two on SCR meant that the person showed two peaks per minute more than predicted from their baseline. What this operationalization does not account for is that quantitatively similar deviations from a predicted or baseline value may have qualitatively different meanings across individuals: For example, for an individual with generally low within-person changes in SCR, showing two peaks per minute above the expected value might constitute a strong response, whereas for another person with generally large fluctuations in SCR, this might be a random fluctuation. If indeed similar residuals have different meanings across individuals, using residualization as baseline correction is insufficient to determine whether between-person differences in the intensity of their physiological response relate to differences in psychological processes.

Instead, approaches may be needed that additionally qualify the physiological response by adding information on its intensity for the specific person. Harley et al. (2019) demonstrated a way of doing so: During the task, they extracted the minimum and maximum of an

individual's EDA and standardized each EDA value according to the person's EDA range. Specifically, they subtracted the minimum EDA value from each observed value and divided it by the difference between the individual's minimum and maximum EDA. This resulted in EDA values ranging between zero and one, with a value close to zero indicating an EDA close to this person's minimum, and values closer to one indicating an EDA close to this person's maximum. Doing so, they found significant moderate correlations between average, standardized SCL with anxiety ( $r = .50$ ) and shame ( $r = .55$ ) on the between-person level, while the correlations between EDA with shame and hopelessness were non-significant. In a similar approach, Kiuru, Malmberg, et al. (2022) and Kiuru, Trög, et al. (2022) normalized participants' on-task physiological values by first subtracting the baseline mean and then dividing by the baseline standard deviation. Despite this approach, they hardly found any significant between-person correlations between physiological and self-report measures on the between-person level. Hence, the question remains whether such between-person associations are genuinely absent or if other, more effective methods for making response intensity comparable across individuals could uncover meaningful patterns at the between-person level.

---

### **Insight 2**

#### **Associations of physiological states with psychological processes vary across the between- and within-person level.**

---

The multilevel correlations reported in EW1 clearly underline that associations of physiological and psychological states can substantially differ on the between- and the within-person level: While on the between-person level, only few (none in Study 1, three in Study 2) of the 24 inspected correlations between self-reported achievement emotions and physiological states reached statistical significance, more significant correlations were found on the within-person level (14 in Study 1, 18 in Study 2). This resembles findings by Zawadzki et al. (2017) who implemented an anger recall task and detected significant associations between anger and blood pressure only on the within- but not the between-person level. Similarly, in an achievement context, Kiuru, Malmberg, et al. (2022) found HR to significantly correlate with six self-reported achievement emotions on the within-person level, whereas on the between-person level, it only correlated with only one. In addition to these empirical findings, our results also align with theoretical perspectives suggesting that emotion and motivation theories – especially those relating psychological and physiological states – primarily refer to the within-person level (Mauss et al., 2024; Pekrun, 2023b; Pekrun & Marsh, 2022).

The central question, already raised in the previous section, is whether the absence of significant associations at the between-person level reflects a true nonexistence or is merely a result of methodological constraints. One of these methodological constraints, the challenge of adequate baseline correction, has been outlined in the previous section. Another methodological issue that possibly prevents the detection of between-person associations between physiological states and the subjective emotional experience might be a lack of comparability of the subjective emotional experiences between individuals using self-report measures (Mauss et al., 2024; Zawadzki et al., 2017): While self-report measures might be able to distinguish between a person experiencing low anxiety versus a person with intense anxiety, accurately capturing more nuanced differences between individuals' emotional states may be more difficult. Inherently, we experience only the range of our own emotional intensities, and at best, have observed and inferred emotional intensity in others. Consequently, the same self-reported value on an emotion scale may reflect different emotional intensities across individuals, or, conversely, individuals may report different values despite experiencing comparable emotional intensities on a hypothetical objective scale. Hence, for both physiological and affective states, within-person changes may be captured more accurately than differences between individuals, leading to more consistent associations at the within- rather than between-person associations level. However, regarding achievement emotions, this proposition is contradicted by numerous substantial between-person findings using self-report measures, both in prior studies (e.g., associations between personality traits and achievement emotions in Pekrun et al., 2023) and in the present work (i.e., moderate to large correlations between most self-reported achievement emotions and task performance, reported in EW1). Alternatively, another methodological challenge explaining the absence of between-person correlations may be model selection. Specifically, the linear relationships assumed in both EW1 in the present work, as well as in prior studies, may not adequately describe between-person associations. Instead, more complex non-linear functions may be necessary to capture associations between physiological states and psychological variables. For example, using the data of Study 2, Eckerskorn et al. (under revision) found a significant quadratic relation between individuals' self-concept and SCR on the between-person level using the same residualized baseline correction approach as EW1, underlining the possible need for non-linear models on the between-person level. Overall, in addition to their possible non-existence, several methodological factors, such as baseline correction, comparability of self-reported affect, and model selection, may explain the absence of substantial patterns linking physiological and psychological states at the between-person level.

Few correlations also reached statistical significance on the between-person level in Study 2 of EW1. Given they were observed in only one sample and for certain combinations of physiological measures and achievement emotions, I advise against drawing conclusions about between-person patterns based on them. Nevertheless, they are interesting in the sense that they underline the possibility of finding different correlational patterns on the two levels of analysis, underlining the importance of distinguishing between them (Hoffman, 2015; Murayama et al., 2017; Schuurman, 2023). For example, in Study 2, anxiety correlated negatively with SCR ( $r = -.26$ ) and positively with HRV ( $r = .37$ ) on the between-person level. This means that individuals who reported higher levels of anxiety showed lower baseline-corrected SCR and higher baseline-corrected HRV than individuals reporting lower anxiety. In contrast, on the within-person level, anxiety correlated positively with SCR ( $r = .21$ ) and negatively with HRV ( $r = -.22$ ). This suggests that when a person experienced higher levels of anxiety in a task block compared to another block, their SCR was higher and their HRV lower during that block than during the other block. Thus, for anxiety, the findings suggest a pattern of deactivation on the between-person level, but an activating pattern on the within-person level. Both patterns can be meaningfully interpreted when considering the findings of EW2, which relate physiological activation to effort. The deactivating pattern for anxiety on the between-person level might suggest that individuals who are generally anxious in mathematical situations were less engaged with the task than those with lower anxiety. Possibly, they did not even attempt to engage with the task, or they generally had lower math abilities, making the entire task overchallenging for them, which resulted in lower effort and, consequently, lower SCR and higher HRV compared to individuals with lower anxiety and possibly higher abilities. In contrast, the activating pattern for anxiety on the within-person level aligns with Pekrun's (2006) the proposition that an increase in anxiety mobilises more resources within a person. The two distinct patterns found for anxiety underline that associations on the between- and the within-person level can truly and substantially differ from one another. Although this is an isolated finding rather than a substantial pattern at the between-person level, it highlights the importance of distinguishing between the two levels in analyses and interpretations, as the patterns they reveal may differ substantially while both reflecting true relationships.

### ***Summary: Level of Analysis***

The findings of EW1, showing substantial patterns on the within- but not the between-person level, match the fact that theories linking physiological and psychological states typically refer to these mechanisms operating at the within-person level (Pekrun, 2023b). While our findings are therefore consistent with theoretical considerations, as well as previous results

(e.g., Kiuru, Malmberg, et al., 2022), the question remains whether the lack of associations at the between-person level reflects a true absence of such relationships, or results from methodological constraints. Regarding physiological variables, the large intraclass correlations underscore the importance of an adequate baseline correction in between-person analyses, raising the question of how to perform such a correction to make the intensity of a physiological response to a task comparable across individuals. Regarding achievement emotions, it is debatable whether self-report measures can sufficiently capture the nuanced differences in the intensity of individuals' affective experiences. However, this concern is counteracted by previous work that has found many substantial between-person associations between other variables and self-reported achievement emotions. Finally, linear models may not be suitable for detecting existing associations between physiological and psychological states on the between-person level. Although the reasons for the absence of substantial between-person patterns remain somewhat open, our findings clearly underline the importance of distinguishing between the two levels of analysis. Not only were substantial patterns observed solely at the within-person level, but in Study 2, the few significant between-person correlations suggested opposing patterns to those on the within-person level – both of which can be interpreted meaningfully without contradicting each other. Thus, the present findings underline the importance of considering the level of analysis when working with physiological measures in an achievement context, as findings from one level cannot be transferred to the other (Murayama et al., 2017).

### ***9.1.2 Meaning of Autonomic Changes in an Achievement Situation***

Physiological measures of the autonomic nervous system have been linked to various psychological constructs relevant in achievement settings, including cognition (e.g., Duschek et al., 2009; Forte et al., 2019; Luque-Casado et al., 2016), motivation (e.g., Albinet et al., 2024; Richter et al., 2016; Seery et al., 2009), and emotion (e.g., Horvers et al., 2021; Kiuru, Malmberg, et al., 2022; Roos et al., 2021). Such one-to-many relationships between physiological measures and various psychological constructs may complicate the application of these measures in research and practice, raising uncertainty about whether an observed physiological change reflects the intended construct or a different one (Richter & Slade, 2017). Consequently, a key goal of the present work has been to clarify the meaning of physiological states of the autonomic nervous system in an achievement situation, thereby also enhancing the understanding of their relationship with, and ability to indicate, broader constructs.

Based on a summary of prior research, I have developed the conceptual model depicted in Figure 1, referring to within-person processes in achievement situations. Synthesizing the findings of the present work indeed supports the model's proposition that physiological activation lies at the intersection of emotion and motivation, where they overlap in their function to mobilize resources within the individual (Insight 3). This refined understanding of physiological changes in an achievement situation, along with findings from all three empirical works, suggests that physiological measures of autonomic activation alone are not able to provide a comprehensive picture of an individual's psychological state during an achievement task (Insight 4).

---

### **Insight 3**

#### **Autonomic arousal corresponds to the mobilization of resources.**

---

Firstly, the findings of EW3, investigating SCR in relation to task characteristics, support the proposition that sympathetic activation reflects resources mobilized for a task. The adaptive test presented items in a way that individuals had a 50% chance of solving them correctly, meaning that items throughout the entire test would have been challenging, yet solvable for participants. In contrast, in the non-adaptive test, only a few items would have been at this challenging but solvable level of difficulty for each individual. Instead, items were likely either rather easy for them, or too difficult. Based on the MIT, this suggests that overall, more effort would be spent on the adaptive rather than the non-adaptive test, due to the test being sufficiently difficult to require a constantly high level of effort without exceeding the point of perceived difficulty where investment of effort stops. Consequently, the finding of a within-person main effect of test type on sympathetic activation, with higher levels of sympathetic activation during the adaptive test, support the claim that physiological activation corresponds to the mobilization of resources for a task.

While the findings of EW3 are based on differences in task characteristics, EW1 and EW2 each related physiological states to self-reported psychological variables in the achievement situation. EW1 investigated the associations between physiological states and the subjective experience of achievement emotions, finding activating patterns on the within-person level for enjoyment, pride, and anxiety, as well as deactivating patterns for anger, frustration, and boredom. EW2 explored the emotionally and motivationally relevant perceptions of task difficulty and value as predictors of physiological states, concluding that the observed within-person patterns for physiological states closely correspond to the patterns

expected for the motivational construct of effort. The first step in synthesizing across these two sets of findings is to map the different achievement emotions and effort onto the difficulty-value spectrum.

Given sufficiently high perceived value, different achievement emotions occur most intensely at different ranges of perceived difficulty (Pekrun, 2006): Intense enjoyment, pride, and anxiety occur in the ranges of low and moderate difficulty when the task is sufficiently controllable for a successful performance to be possible but not guaranteed. Intense anger and frustration occur when perceived control is low, which is the case when a task is perceived as very difficult. Additionally, when perceived value is low, intense boredom occurs, in particular intense when perceived difficulty is very low (underchallenge) or very high (overchallenge; (Goetz et al., 2019; Pekrun, 2024a). However, as discussed in EW2, boredom due to under- or overchallenge may not only be limited to low perceived value, but might also occur at the extreme ends of the difficulty spectrum, even when value is rather high. Also, the motivational construct of effort can be mapped onto the difficulty spectrum: According to MIT, the exertion of effort follows an energy conservation principle, meaning that no more resources than necessary and justified should be spent on a task (Brehm & Self, 1989; Richter et al., 2016). Consequently, effort increases with increasing, yet feasible (controllable) task demands, but diminishes when the task becomes so difficult (uncontrollable) that the expense of resources outweighs potential benefits. Perceived value moderates this association, as it determines the point of perceived difficulty at which the drop in effort occurs (Brehm & Self, 1989; Richter et al., 2016).

By identifying where each achievement emotion is most intense along the spectrum of perceived difficulty and value, it becomes apparent that the occurrence of different achievement emotions corresponds directly to the necessity and potential benefit of investing resources into a task: An intense experience of enjoyment, pride, and anxiety implies that the current achievement situation is somewhat controllable and valuable, meaning that investing resources into the task would be beneficial. Instead, an intense experience of anger and frustration suggests that the investment of resources might be meaningless, as the task is valuable, yet uncontrollable. Finally, an intense experience of boredom during a task implies that either the task is not meaningful, that the investment of resources is not necessary for a successful completion (underchallenge), or that the investment of resources does not lead to successful completion (overchallenge). Hence, the peak intensities of different achievement emotions correspond to varying necessities and benefits of resource investment. Their occurrence is fully

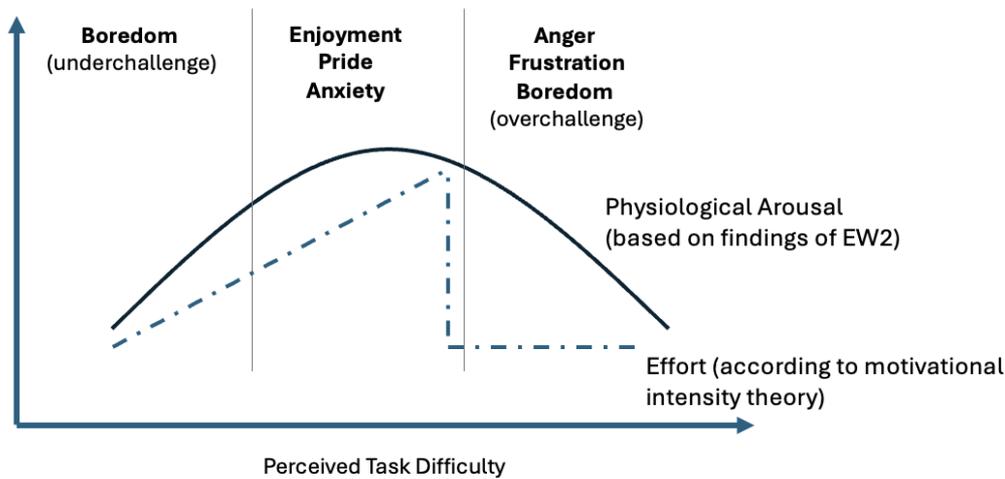
consistent with predictions on effort exertion according to the MIT, also suggesting low effort in both very high and very low difficulty, and a higher level of effort in between (Figure 6).

Having established a link between the occurrence of different achievement emotions and the exertion of effort through their relations to perceived difficulty and value, it becomes apparent that the patterns detected for physiological activation in EW1 and EW2 converge (Figure 6): In line with extensive research on MIT (Gendolla et al., 2019; Richter et al., 2016), EW2 reinforced a close link of sympathetic activation and parasympathetic deactivation with the motivational construct of effort by demonstrating high physiological activation at appraisals of moderate, compared to particularly high or low perceived difficulty, given sufficiently high perceived value. Meanwhile, the findings of EW1 showed physiologically activating patterns for joy, pride, and anxiety – those achievement emotions that are experienced most intensely when appraisal combinations suggest that the investment of resources into the task is meaningful. In contrast, EW1 found physiologically deactivating patterns for anger, frustration, and boredom – those achievement emotions that are most intense when the investment of resources into the task does not appear beneficial. Hence, physiological states are associated with the subjective experience of achievement emotions insofar as the occurrence of these emotions suggests the need for resource mobilization. This aligns with the claim that the physiological component of achievement emotions facilitates adaptive behaviour through engagement with or withdrawal from a task (Pekrun et al., 2023). Altogether, the synthesis of findings of EW1 and EW2 supports the proposed model of physiological states corresponding to the mobilization of resources for a task, a function that links emotions and motivation in the achievement context.

The finding of deactivating patterns for anger and frustration warrants a dedicated discussion, as it contrasts with their typical classification of anger and frustration as activating emotions (Pekrun, 2024a). Regarding anger, it may be important to consider that anger relates differently to perceived control, depending on its object focus: Anger as an activity emotion occurs when perceived control is high but the task is perceived as unpleasant, whereas anger as an outcome emotion is elicited by negative value and a lack of perceived control, typically due to someone else being in control (Pekrun, 2006). Possibly, in the former case, mobilizing resources might be beneficial to complete or change the task, whereas in the latter case, the mobilization of resources is inefficient as no influence over the task can be exerted anyway. In the present work, anger was significantly and negatively associated with performance in both samples, suggesting that the observed anger corresponded to the latter type, arising from a perceived lack of control.

**Figure 6**

*Mapping the Investigated Achievement Emotions, Effort, and Physiological Arousal onto Perceived Task Difficulty, Given Sufficiently High Value*



Therefore, in the present work, it may have been accompanied by physiological deactivation, whereas this may differ for anger driven by other appraisals of control. In line with this proposition, it has been reported in general emotional literature that although anger is typically physiologically activating (Kreibig, 2010), it can be deactivating when accompanied by a withdrawal-oriented motivational tendency (Stemmler et al., 2007). This further supports the idea that patterns of physiological arousal reflect the mobilization of effort associated with an emotion, rather than specific emotions themselves. Based on our findings for anger and frustration, the three-dimensional taxonomy of achievement emotions (Pekrun, 2024a) might require refinement by altering the classification of certain emotions (e.g., frustration as deactivating) or by specifying more precisely under which conditions certain emotions are activating or deactivating (e.g., anger in high vs. low control), possibly informed by the implications the experience of these emotions has for the investment of resources into a task.

It should be noted that the present reasoning for why certain correlations between achievement emotions and physiological activation are observed is based on the occurrence or non-occurrence of different achievement emotions within specific ranges of perceived difficulty. For example, we find a positive correlation between enjoyment and physiological activation, as physiological activation is higher in a range of perceived difficulty when enjoyment occurs compared to ranges where it does not occur (as depicted in Figure 6). An alternative approach to their association would be to focus on the occurrence of a particular achievement emotion and inspect how the intensity of this emotion relates to physiological

activation. This is, for example, demonstrated by Roos et al. (2023), who selected only instances where joy or anxiety was the dominant emotion, respectively, and related changes in control to sympathetic activation for those instances only. Such approaches focussing on certain parts of the difficulty spectrum may complement the one employed in the present work, providing a more fine-grained view of the physiological activation curve – for example, as suggested by Roos et al.'s (2023) exploratory results, we may find two peaks of activation in the moderate difficulty range instead of one: One peak may occur when joy/pride are most intense towards lower perceived difficulty, and another one when anxiety is more intense towards higher perceived difficulty. Future work should acknowledge the conceptual differences between these approaches and utilize them depending on the specific research question.

Altogether, the present findings linking physiological arousal to achievement emotions (EW1), task appraisals (EW2), and task characteristics (EW3) converge in suggesting that, in achievement situations, physiological states in the autonomic nervous system reflect the mobilization of resources for a task. This narrower meaning also allows for a better understanding of how physiological changes in an achievement situation relate to broader constructs, such as achievement emotions and motivation. Further, it has the following central implication for the use of autonomic measures as indicators of psychological processes.

---

### **Insight 4**

**Individual physiological measures need to be complemented by other measures to gain a comprehensive insight into individuals' emotional and motivational states.**

---

The understanding of physiological states as the mobilized resources during an achievement task comes with the limitation that they alone cannot provide a comprehensive insight into individuals' psychological states. The findings of EW1 and EW3 underline the lack of specificity of physiological states to a particular emotional state (Kreibig & Gendolla, 2014; Pekrun, 2023b), and the findings of EW2 and EW3 their inability to indicate specific task appraisals. In EW1, the statistically significant correlations found for each of the distinct emotions with physiological measures on the within-person level are low to moderate in size, ranging between  $|.09|$  and  $|.40|$ . This implies that physiological measures can hardly be seen as an indicator of one specific emotion, and further not even of a certain emotional valence, as both the positive emotions of enjoyment and pride, as well as the negative emotion of anxiety, were associated with physiological activation. Similarly, the findings of EW3 underline that physiological measures alone cannot provide information about a certain emotional valence, let

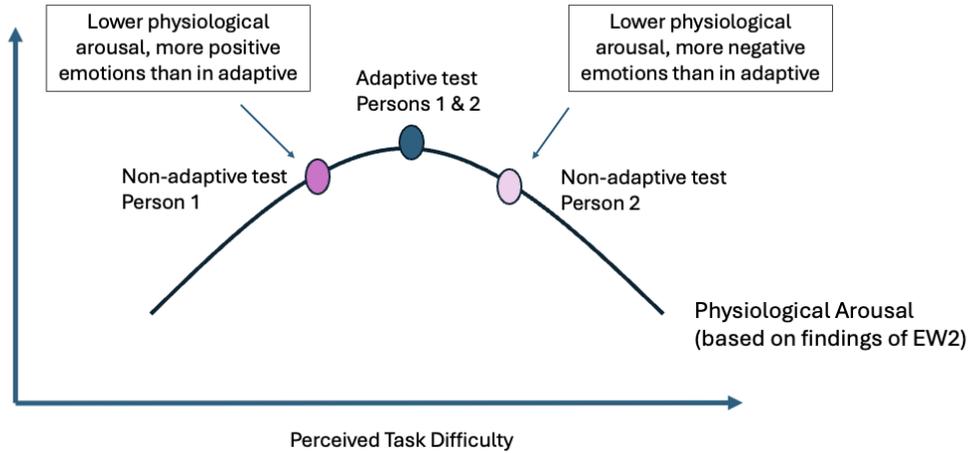
alone distinct emotional experiences. While the results showed generally higher sympathetic activation during the adaptive test, there was no main effect of test type on the subjective experience of achievement emotion, as this effect was moderated by perceived test difficulty. Specifically, individuals who perceived the adaptive test as more difficult than the non-adaptive test reported a more negative emotional experience on this test, whereas individuals who perceived the adaptive test as easier had a more positive emotional experience on this compared to the non-adaptive test (see Figure 7 for a visualization of findings). Consequently, the findings demonstrate that the physiological measure alone would not have provided a comprehensive understanding of how adaptive testing affects the test-takers' experience. While the measure of SCR was only able to capture the generally higher mobilization of resources in the adaptive test, the self-report measure of achievement emotions was essential in understanding that the adaptive test elicits more positive emotional experience in some individuals, but by a more negative emotional experience in others.

Further, the findings of EW3 and EW2 demonstrate that physiological measures cannot provide information on specific task perceptions. In EW3, physiological activation was higher for some individuals in the adaptive compared to the non-adaptive test, due to that test being perceived as easier, whereas for other individuals, the higher levels of activation were likely due to the adaptive test being perceived as more difficult (Figure 7). Similarly, the findings of EW2 also demonstrate that physiological states are not specific to task perceptions: The detected quadratic functions, as well as the moderating role of perceived value, imply that, for example, low sympathetic/high parasympathetic activation could be driven by either very low or very high perceived difficulty, and/or by low perceived value.

Taken together, the findings from all three empirical works underline that physiological measures are unable to distinguish between specific task perceptions of difficulty and value, or between positively and negatively valenced emotional states. Similarly, it has previously been found that both approach and avoidance behaviours due to perceptions of challenge or threat are accompanied by physiological activation (Chalabaev et al., 2009; Seery et al., 2009). This leads to the conclusion that physiological changes alone cannot provide information beyond the mobilization of resources. Instead, to gain a comprehensive picture of an individual's emotional or motivational state, individual physiological measures may need to be complemented by measures tapping into other modalities, which will be discussed in Section 9.3.2.

**Figure 7**

*Illustration of the Main Effect of Adaptive Testing on Physiological Activation and the Interaction Effect of Adaptive Testing with Perceived Difficulty on Achievement Emotions*



***Summary: Meaning of Autonomic Changes in an Achievement Situation***

A synthesis of findings supports the central assumption of the conceptual model in Figure 1, locating autonomic activation at the overlap between emotion and motivation in an achievement context, representing the mobilization of resources to deal with a task. This narrower definition of the meaning of physiological states in an achievement situation aligns with previous work on the MIT, commonly using autonomic activation to indicate effort (Albinet et al., 2024; Gendolla, 2025; Richter et al., 2016). The refined meaning also clarifies how physiological states are related to other constructs, such as achievement emotions. It further matches with findings from the field of cognition, showing higher activation in more difficult tasks (Luque-Casado et al., 2016; Mandrick et al., 2016) and in relation to better performance (Forte et al., 2019; Horvers et al., 2021), hence, in instances where more resources are required and mobilized. However, the narrower meaning of physiological states also implies that individual measures of autonomic activation can and should only be interpreted as the mobilization of resources and are therefore unable to provide a comprehensive insight into an individual's psychological state. This limitation complicates their application in research and practical applications: For example, we may not want to intervene if an individual shows high activation due to enjoyment or the perception of challenge, but we may want to intervene if the high arousal is driven by anxiety or the perception of threat. Hence, for many research and practical applications, physiological measures must be integrated with other information sources to yield a comprehensive understanding of an individual's state.

### ***9.1.3 Differences between Autonomic Measures***

Measures of activation in the autonomic nervous system are commonly used in educational psychology, as they are more affordable and easier to employ than brain imaging techniques, while allowing for continuous, unobtrusive data collection throughout an achievement situation (Ishikawa, 2023). The central branches of the ANS are the sympathetic and the parasympathetic nervous systems, which can show patterns of activation and deactivation independent of one another (Berntson et al., 1994). We recorded two indicators of EDA, reflecting sympathetic activation, as well as two measures of CVA, which is influenced by both sympathetic and parasympathetic activation (Wehrwein et al., 2016). This application of multiple measures follows the recommendation that using multiple indicators enables a more comprehensive picture of ANS activation and allows for cross-validation between them (Mauss et al., 2024; Pekrun, 2023b). Such recommendations seem to imply that different indicators may exhibit diverging patterns, raising the question of whether they carry different meanings in achievement situations. However, little is known about such potential differences between ANS indicators in their relation to psychological processes. Therefore, the present findings are used to compare patterns across indicators of sympathetic and parasympathetic activation, as well as between the two EDA measures. Indeed, the comparisons reveal that patterns can vary between different indicators of ANS activation, and that each ANS branch or even indicator may reflect a slightly different aspect within the broader function of resource mobilization (Insight 5). It should be noted that the following suggestions on possible differences are tentative and exploratory, grounded primarily in the patterns that emerged in the present data.

---

#### **Insight 5**

**Indicators of autonomic activation differ in their patterns and, therefore, possibly in their meaning.**

---

In EW1, at the within-person level, differences emerged in the relationships between different autonomic indicators and the subjective experience of achievement emotions. Specifically, SCR showed the most consistent correlations with self-reported emotions (five out of six correlations significant in Study 1, six out of six in Study 2). In Study 2, HR also showed significant correlations with all six self-reported emotions. For HRV, as well as for HR in Study 1, three out of six correlations were statistically significant. Furthermore, although all correlations were low to moderate in size, those between self-reported emotions and SCR and

HR tended to be larger than the correlations between self-reported emotions and HRV. This finding is partly in line with those at the within-person level of Kiuru, Malmberg, et al. (2022), who only found a consistent pattern of associations between HR, but not between HRV and self-reported emotions. However, unlike in the present work, where SCR seemed to be the measure most consistently linked with the subjective affective experience, they did not find substantial patterns of associations between SCR and self-reported emotions. Combining the findings from EW1 and Kiuru, Malmberg, et al. (2022), it appears that measures influenced by sympathetic activation (SCR and HR in our work, and HR in Kiuru, Malmberg, et al., 2022) show more consistent associations with the subjective experience of achievement emotions than HRV, a measure of only parasympathetic activation.

Furthermore, EW2 enables a comparison of the patterns exhibited by the different indicators in relation to perceived task difficulty. Given high perceived value, all three measures showed quadratic relationships with perceived difficulty, indicating reduced sympathetic and increased parasympathetic activation at both low and high perceived difficulty. For SCR, however, an additional linear effect of perceived difficulty (in interaction with perceived value) remained. These differences between the purely sympathetic and parasympathetic indicators are also visible in the respective figures. Given a high perceived value, HRV reached its minimum just below the midpoint of person-centred perceived difficulty, indicating a considerable increase in parasympathetic activation at both the lower and upper ends of the perceived difficulty scale. Instead, again given high perceived value, SCR also indicated a considerable decrease in sympathetic activation with high perceived difficulty, but only showed a slight decrease in low perceived difficulty.

The finding that perceived difficulty is more linearly related to sympathetic than to parasympathetic activation may explain why correlations, i.e., linear relationships, between SCR and self-reported anger and frustration in EW1 were larger and more consistent than those of HRV. Both anger and frustration are low at low perceived difficulty and high at high perceived difficulty (Pekrun, 2006; Shao et al., 2020). Since SCR showed not only a significant quadratic relationship but also a linear association with perceived difficulty in EW2, a significant negative within-person correlation between anger and frustration with SCR might have been observed in EW1 – their intensity increasing with perceived difficulty, and SCR decreasing with perceived difficulty. Instead, EW2 found HRV to be high at both low and high difficulty levels – corresponding to both low and high intensity in anger and frustration – possibly resulting in a lack of significant correlations between those emotions and in EW1. Instead, as both also have a quadratic relation with perceived difficulty, both SCR and HRV

showed rather consistent correlations with enjoyment, pride, anxiety, and boredom, which are also quadratically linked to perceived difficulty – enjoyment, pride, and anxiety being particularly intense at moderate difficulty, and boredom particularly low. As HR is influenced by both sympathetic and parasympathetic activation (Wehrwein et al., 2016), its patterns found in EW1 and EW2 share features of those for sympathetic and parasympathetic activation, with rather consistent linear associations with self-reported emotions in EW1, but a more quadratic relation with perceived difficulty in EW2.

Although the differences between the patterns of the three indicators are only small, based on the findings of EW2, I tentatively propose that sympathetic and parasympathetic activation may reflect different nuances of resource mobilization in an achievement situation: Sympathetic arousal (indicated by SCR) might be more closely related to the mobilization of resources in the sense of readiness for engagement with the task, whereas parasympathetic arousal (indicated by HRV) might be more reflecting the actual resources mobilized based on perceived task difficulty. Such a difference would become particularly apparent at the lower end of the difficulty spectrum, as observed in EW2: In rather low difficulty, one might be alert and willing to engage with the task – therefore already showing rather high sympathetic activation – while the task demands are low enough that they do not yet require the body to additionally reduce parasympathetic activation (i.e., uncoupled sympathetic activation; Berntson et al., 1994). Instead, at a moderate perceived difficulty, sympathetic activation remains high, while reciprocal parasympathetic deactivation frees up additional resources needed for the task. Finally, in high difficulty, when withdrawing from the task, sympathetic activation decreases, and parasympathetic activation increases reciprocally. If this were the case, sympathetic and parasympathetic activation might reflect different facets of resource mobilization in an achievement situation. However, as there is a lack of research employing multiple autonomic measures (Hsieh et al., 2011), this tentative proposition requires thorough replication and investigation before more definitive conclusions can be drawn. Nevertheless, the findings underscore the importance of considering activation in the two branches as separate, independent processes (Berntson et al., 1994).

The second notable discrepancy between measures is that between SCR and SCL in EW1. Despite commonly being used as indicators of sympathetic activation (e.g., Harley et al., 2019; Horvers et al., 2021; Lehikoinen et al., 2019), they did not correlate significantly with one another on a within-person level in either Study 1 or Study 2. Moreover, they even showed diverging within-person associations with self-reported achievement emotions: While SCR correlated positively with pride and negatively with anger in Study 1, as well as positively with

anxiety and negatively with anger and boredom in Study 2, the respective correlations with SCL were non-significant. This resembles findings by Harley et al. (2019) showing a positive association between diagnostic efficiency and SCR, but no such association with SCL. Similarly, Chang and Huang (2012) found no effect of attentional task demands on SCL but an effect on cardiovascular measures. Taken together, these findings support the notion that SCL may not be as reactive to environmental changes (Christopoulos et al., 2019). However, in the present results, SCL not only showed fewer or weaker relationships with self-reported achievement emotions, but even opposing patterns compared to SCR with joy, frustration, and boredom in Study 1, and joy, pride, and frustration in Study 2. Specifically, SCR correlated positively with joy and pride, while SCL showed negative correlations, and vice versa for frustration and boredom. Based on previous work linking the two measures with different neurological processes (Nagai et al., 2004), their diverging patterns might imply that they are reflective of different psychological processes (Harley et al., 2019). However, to my knowledge, there is no deeper understanding of what these different processes might be in the achievement context. Hence, the present work cannot make any claims about the differential meanings of SCR and SCL in achievement situations, beyond highlighting the observed differences, advising against using the two measures interchangeably, and emphasizing the need for more research to clarify their distinction.

***Summary: Differences between autonomic measures***

Altogether, the present work clearly demonstrates that while autonomic activation generally appears to reflect resource mobilization in an achievement situation (Insight 3), possible differences in the exact meaning of different indicators in this broader process remain yet to be explored. Based on the more pronounced linear association between SCR and task difficulty – in addition to its quadratic one – I tentatively proposed that while both serve the purpose of mobilizing resources for a given task, sympathetic activation may be more reflective of the readiness to engage, whereas parasympathetic activation may be more closely linked to the actual resources needed for the task. However, this is only one possible, tentative interpretation that requires thorough replications and investigations. Furthermore, a notable discrepancy emerged between SCR and SCL, the two indicators of EDA that supposedly reflect sympathetic activation. However, due to a lack of similar evidence to draw on and the use of SCL in only one of the empirical works, no specific propositions regarding differences in their meaning could be drawn from the findings. Overall, the findings clearly highlight the need for further exploration of differences in the meaning. Only once a more thorough understanding of potential differences is obtained will it be possible to meaningfully combine different

autonomic measures to gain a more nuanced insight into an individual's resource mobilization in an achievement situation.

## **9.2 Strengths and Limitations of the Present Work**

The present work provides several meaningful insights into the meaning and application of physiological measures of the autonomic nervous system during achievement situations. As in any empirical work, the methodological choices made have both strengths and limitations, which need to be considered for an adequate interpretation of findings. In the following sections, the central strengths and limitations of all three empirical works are outlined, providing perspective on the present findings and serving as a starting point for future research. The central strengths and limitations revolve around five aspects: the sample, the lab setting, the use of self-report measures, as well as the possibilities of inferring causality and within-person associations.

First, the collection of two separate samples, one with any voluntary participants and one restricted to female prospective primary school teachers, enabled a comparison of findings in EW1, which showed largely consistent patterns across the two groups. Nevertheless, both samples were selective and quite similar, including only adult, mostly female participants. Moreover, with most participants being university students (84% in Study 1, 100% in Study 2), both samples likely had higher cognitive abilities and more experience in academic settings than the general population. Recruiting in the university environment enabled us to obtain a considerable sample size, particularly compared to cognitive studies with typically much smaller samples. Further, the homogeneity of the sample and our familiarity with this population enabled us to create a testing procedure tailored to this population. For example, it allowed us to design an achievement task covering the full range of perceived difficulty, i.e., containing certain items that all or none of the participants could and that no one could solve, without requiring too many items. Besides these advantages, a selective and homogeneous sample limits the generalizability of findings if participants have certain characteristics that may cause their results to differ from the general population: On one hand, individuals who choose to be in a university environment and volunteer to participate in a study may have a particular affinity for challenging cognitive tasks, potentially leading them to respond differently – maybe more positively – to achievement situations than the general population. On the other hand, the population of primary school teachers sampled in Study 2 is known to have rather high levels of math anxiety (Artemenko et al., 2021), which may also impair the generalizability of their responses to a maths task. However, the descriptives for difficulty and

self-reported achievement emotions can alleviate these concerns. They showed that in both samples, perceived difficulty, as well as the intensity of positive and negative emotions, were distributed across the full range of the respective scale, suggesting that participants did not uniformly find the task either easy or hard, or only experienced particularly positive or negative emotions. Nevertheless, the homogeneous, selective sample, consisting primarily of female, adult student participants, potentially restricts the generalizability of the findings and calls for replications in other demographic groups, such as school-aged children and adults from different occupations and socioeconomic backgrounds.

Second, another aspect to be discussed regarding generalizability is the laboratory setting of the present work. On the one hand, the laboratory setting provided a controlled environment, keeping confounding influences to a minimum and thereby improving the quality of the physiological data by reducing noise in the signal. On the other hand, however, the laboratory is an artificial setting for achievement tasks, which are more likely to be encountered in educational environments, such as schools or universities, in the home through online learning or homework, or in organizational learning or assessment contexts. Hence, findings from a laboratory setting may lack generalizability to more natural achievement contexts. Furthermore, the present studies did not pose a high-stakes situation to participants. Based on the average perceived value of doing well around the mid-point of the scale in Study 2, it seems that participants cared about doing well on the task; however, it was not extremely important to them. While this may resemble other low-stakes situations, such as learning or homework, patterns in high-stakes situations, like final school exams or job assessment tests, might differ. In particular, responses to very low and very high task difficulty may vary in high-stakes situations – possibly, in high-stakes situations, we might observe a physiological mobilization of resources even at these extreme ends of the difficulty spectrum, as the individual might still want to increase their chances of a good performance even if it is either already certain or highly unlikely. Hence, in high-stakes situations, we might find different associations between physiological states and difficulty perceptions, as well as self-reported achievement emotions. However, such potentially divergent findings would not undermine the proposition that physiological states reflect the mobilization of resources, but only suggest that this resource mobilization may follow different patterns depending on the context. In order to establish which of the present findings are generalizable to real-life and high-stakes achievement settings, similar studies could employ experience sampling methods during everyday school life (e.g., Ketonen et al., 2023), or capture processes during actually relevant assessment situations (e.g., Graham et al., 2023).

Third, although the relevance of the present work was based on the goal of overcoming the limitations of self-report measures, it also utilized these very measures. The central approach of EW1 and EW2 has been to relate physiological changes to subjective emotional experiences and subjective task appraisals. Therefore, self-report appeared to be the most appropriate – and, in fact, the only – method for directly capturing these subjective experiences and perceptions. However, all the possible limitations of self-report outlined in Section 2.1 also apply to the present work: Memory biases may have affected responses when completing the self-report questionnaire after the respective block; social desirability biases may have occurred, for example, if participants felt obliged to indicate more frustration when noticing the task became more difficult; or observer effects may have taken place if the subjective emotional experience, perceptions of difficulty, or perceptions of value were altered by the repeated attention drawn to them. Although the present data do not provide specific evidence of systematic issues with the self-report measures, the possibility that biases have affected some participants' responses cannot be excluded. One way to avoid these biases would be to replace measures of task perceptions with indicators of task characteristics, for example, by relating physiological states to actual rather than self-reported task difficulty. However, this may overlook the importance of the individual's subjective perception of task characteristics. For other variables, such as the subjective affective experience, self-reports appear indispensable. However, future research might be able to design these assessments in a way that limits possible biases, such as avoiding memory bias by using a joystick during rather than a questionnaire following the task (e.g., Krieglmeier & Deutsch, 2010; Lizdek et al., 2012).

Fourth, the present correlational findings are limited in their ability to provide evidence on causal relationships. While one may consider the variation in average item difficulty across blocks as a manipulation of perceived task difficulty, and through that, possibly the emotional experience, there was no direct manipulation of achievement emotions or value. Hence, the findings are more correlational in nature, not automatically implying the existence of causal relationships. The absence of causal effects between physiological states and achievement emotions or task appraisals does not pose an issue in the investigations of the present work, which aimed to establish their meaning by investigating how the different states and appraisals coincided, rather than to establish the causal direction of these effects. However, what would undermine the presented interpretations are confounding effects of shared third variables. The main possible confounder to be discussed in this regard is temporal trends. In both samples, self-reported enjoyment, pride, and anxiety decreased over time; anger, frustration, and

boredom increased over time; SCR and HR decreased, and SCL and HRV increased over time<sup>3</sup>. If either one of both the self-reported and physiological variables were influenced by time independently of one another, such as by naturally decreasing physiological activation over time, the associations found between them in EW1 would be spurious. However, the correlations reported in EW2 also show that in Study 2, value decreased, and perceived difficulty increased over time. This suggests that what may appear as temporal trends in the self-reported and physiological measures could have also been a change over time driven by changes in task appraisals, which would fully support the proposed model and interpretations. Also for the findings of EW2, potential confounding effects of temporal order need to be considered. Specifically, individuals may have been most refreshed in the easiest two blocks – one being presented at the beginning of the procedure and one following break – possibly leading to higher activation coinciding with easier items due to temporal order. However, the quadratic effects of difficulty on all three physiological outcomes somewhat counteract this concern. For the more difficult blocks, it indeed is impossible to determine whether their lower activation levels were driven by higher difficulty or by fatigue from being further away from the start or the break. However, the quadratic effects showing lower physiological activation during blocks that were perceived as particularly easy do not support this concern, as we would expect highest activation in the easiest block if it was only driven by being placed at the beginning of the test. Altogether, the correlational data from the present work do not allow for the disentanglement of the effects of task-independent temporal trends and task perceptions on physiological states and their associations with other variables. Hence, while the present findings demonstrate the presence of correlational associations, their underlying mechanisms are yet to be established. Future work should aim to avoid an overlap between task-unrelated trends due to time or block order and trends in difficulty and value. For example, items could be presented with decreasing difficulty, opposing a potential natural decline in physiological activation. Alternatively, the order of blocks and the positioning of a break could be randomized across participants, or experimental manipulations, such as value inductions towards the end of a test, could be used to establish more secure causal associations.

Finally, all within-person analyses in the present work employed fixed within-person effects, meaning that the associations found represent aggregated within-person effects across all individuals. Such a fixed within-person effect can, as intended, point out differences between

---

<sup>3</sup> These associations were established with multilevel regression models with a random intercept, predicting the respective outcome from Time (Block 1 to 5/6). For all variables, the fixed effect of Time was significant with  $p < .05$ .

associations depending on the level of analysis and is an essential, valuable first step in gaining an understanding of within-person associations between variables, such as physiological states and affective experiences, task perceptions, or task characteristics. However, it is worth noting that a fixed within-person effect does not imply that this effect holds true for each and every individual (Algina & Swaminathan, 2011; Epskamp et al., 2022; Hoffman, 2015). Instead, within-person associations can substantially differ from one another. For example, two variables may correlate positively in some individuals and negatively in others, still leading to the finding of a positive fixed within-person effect if the positive associations are more prevalent or stronger. Hence, the present findings should always be interpreted with the caveat that they are fixed effects, rather than assuming that the found patterns are true for each individual person. Such possible heterogeneity in within-person associations can be explored in future research using random slopes models on larger samples with more measurement time points than are available in the present work (Pekrun & Marsh, 2022). If heterogeneity in the slopes is found, future work can explore possible predictors thereof, which will be crucial for future research and practical applications of physiological measures, ensuring that physiological changes within a specific person are interpreted correctly.

Overall, the methodological choices made in the present work come with several strengths in terms of study design and analysis. However, the findings also need to be interpreted within the limitations that accompany these choices, and future research is needed to establish the generalizability and robustness of the present results.

### **9.3 Considerations for Theory and Practice**

This final section comes back to the two central goals of educational psychology, namely to “enhance theoretical knowledge” and to “improve educational practice” (O’Donnell et al., 2025, p. 5). Addressing these goals, the following discusses the central insights gained in the present work from a theoretical and a practical angle, yielding meaningful implications and directions for future work. These considerations are organized around the three guiding themes of the present work, being the level of analysis, the meaning of physiological states in an achievement situation, and differences between indicators.

#### ***9.3.1 Considerations around the Level of Analysis***

The first set of insights (Insights 1 and 2) has demonstrated the importance of distinguishing between associations on the between- and the within-person level in analysis and interpretation. Associations of physiological states with the subjective emotional experience have primarily been found on the within- rather than the between-person level, supporting

related theoretical claims, such as by Mauss et al. (2024) stating that “Within-person associations more closely denote coherence as conceptualized by the theories of emotion” (p. 79), and by Pekrun (2023b) stating that “Physiological processes occur within persons, and theories on their functions for motivation, emotions, and engagement typically refer to within-person mechanisms” (p. 7). Consequently, applications of physiological measures in research and practice may be well advised to employ within-person approaches. In research, this may be implemented through within-person correlational (like in EW1 and EW2), or experimental designs (like in EW3). In practice, if physiological states are used for micro-adaptivity of learning tasks, adaptations of the task should be made based on within-person changes (e.g., a decline in physiological arousal relative to previous tasks), rather than based on absolute thresholds (e.g., SCR lower than three).

Certain research questions or practical applications may require a between-person comparison, such as when investigating which target group a learning environment is particularly (dis)engaging for, or when aiming to identify the most disengaged students to allocate the teacher’s resources to them. Firstly, it is necessary to consider theoretically whether we can even expect such associations. While relevant literature emphasizes a focus on the within-person level (e.g., Mauss et al., 2024; Pekrun, 2023b), it remains unclear whether associations on the between-person level are not assumed in the theory, or whether they are assumed to be present but deemed difficult to detect. Secondly, if such associations are deemed to be theoretically plausible, further work is needed to examine their existence. Although the present and prior findings show little evidence for substantial correlational patterns on the between-person level, the present work has discussed the possibility that this may be due to an inability to make the intensities of different individuals’ reactions comparable or due to a lack of suitable statistical modelling. Hence, while the present findings show more substantial correlational patterns between physiological states and subjective experiences on the within-person level, they do not discard the possibility of using physiological measures for between-person applications. However, to enable between-person approaches, research and practice may need to advance their methods of baseline correction and/or statistical modelling approaches. In order to explore the different alternatives, researchers may consider multiverse approaches, where different analysis methods or models are employed for the same research question, in order to compare findings depending on certain analytical decisions, like baseline correction and model selection (see Cantone & Tomaselli, 2024; Götz et al., 2024; Harder, 2020).

### ***9.3.2 Considerations around the Meaning of Physiological States***

The proposed narrower meaning of physiological states as the mobilization of resources, and the suggested conceptual model placing it at the overlap of emotion and motivation (Insight 3), give rise to several theoretical considerations. First, they raise the question of how physiological states are to be conceptualized in relation to the constructs of effort and achievement emotions. Clearly, there is a close association between physiological states and the effort construct: Similar to the meaning of physiological states derived from the present work, effort has been defined as the “mobilization of resources to execute action” (Silvestrini & Gendolla, 2019, p. 116). This invites debate over whether physiological states should be considered as either equivalent to effort, or as indicators of effort, like often done in research around MIT. Alternatively, physiological states could be considered as a parallel construct that reflects the mobilization of resources on a bodily level, whereas effort refers to the psychological process. The latter would be more in line with the proposed integrated approach, linking physiological states to both emotion and motivation through their function of resource mobilization, while maintaining the conceptualization of effort as a motivational construct associated with, but not being a component of, achievement emotions. To make matters even more complex, in the theory of achievement emotions, physiological states have been considered as a component of the emotion construct, with the function of *facilitating* effort (Pekrun et al., 2023), presumably implying a more predictive relationship between physiological changes and a rather behaviour-based conceptualization of effort. For practical applications, such nuanced conceptual distinctions are largely irrelevant, as, in any case, the implication remains the use of physiological measures to assess resource mobilization. However, theoretically disentangling these different processes and constructs is essential to advance theories, to foster a common terminology across fields, and to support the much-needed integration of emotional and motivational theoretical frameworks (Pekrun, 2024b).

Second, the proposed narrower meaning of physiological states in an achievement context also has implications for the possibility of using physiological measures in research and practice to assess individuals’ states. Specifically, it implies that physiological measures alone cannot provide a comprehensive insight into emotional or motivational states, lacking important context information such as emotional valence or goal orientation (Insight 4). Hence, when aiming to obtain insights beyond mere resource mobilization, physiological measures need to be complemented by other modalities. One possible modality to complement physiological measures may be self-report: Despite their previously outlined disadvantages (Section 2.1),

self-report measures are able to provide nuanced information about many processes during achievement situations (e.g., affective experiences or task perceptions; Pekrun, 2020), which can provide context to changes in autonomic activation. Alternatively, or additionally, task-related information may provide context, such as task difficulty, performance, or reaction time, which could support a distinction between autonomic deactivation occurring due to under- or over-challenge. Finally, other bodily measures may be able to complement the information gained from measures of the autonomic nervous system. For example, modalities like facial expressions, body movements, voice, or the central nervous system (e.g., EEG) have been used in affect detection systems, with these systems showing a better performance when combining information from different modalities (D’Mello & Kory, 2015; Harley et al., 2015). Consequently, while research and practice may benefit from a more specific definition of the meaning of physiological states, this also implies the need to develop and evaluate methods for triangulating across modalities to contextualize the information gained from autonomic measures. A promising, possibly essential next step in leveraging information from multiple modalities to infer psychological states may be the application of machine learning algorithms, which can detect complex patterns and integrate data from different sources to infer a certain psychological state – for example, in the context of emotion recognition, as demonstrated in a tutorial by Can et al. (2023).

### ***9.3.3 Considerations around the Differences between Indicators***

Even within the modality of the autonomic nervous system, the present findings demonstrate that different indicators exhibit variations in their patterns of associations with other variables, potentially reflecting different nuances of meaning. For practical applications, these observed differences suggest that indicators should not be used interchangeably. For theory and research, they raise the need to better understand the differences in the specific meaning of different autonomic indicators in achievement situations. However, detecting and understanding these differences poses a challenge, especially when focusing on bivariate, correlational associations and having only a limited understanding of the biological processes underlying these indicators. Advances may be made through experimental designs attempting to separate potentially different facets of resource mobilization. For example, to test the proposition that sympathetic activation is more linked to the readiness to engage with a task, and parasympathetic activation to the resources needed according to the task demands, research may focus on the low difficulty range where differences might be most salient, independently manipulating value (should mainly affect readiness to engage, i.e., sympathetic activation) and

difficulty (should mainly affect resources needed for task completion, i.e., parasympathetic activation). Further, interdisciplinary collaborations with experts in medicine and biology may be essential in understanding the underlying biological processes and functions behind different autonomic indicators, such as the differences between the SCR and SCL signals. Finally, differences between indicators may even be utilized through machine learning algorithms that might be able to detect how specific configurations of autonomic responses relate to psychological processes.

### 9.4 Outlook

The present work makes a meaningful contribution towards an enhanced understanding of the meaning and application of physiological measures of the autonomic nervous system in research and practice. This enhanced understanding is certainly a first step toward a meaningful application of physiological measures in both educational research and real-life settings, such as evaluating learning environments or developing micro-adaptive learning technologies. The next step, however, is equally important: Not only do we need to understand the meaning of physiological states, but we also need to identify what constitutes desirable and undesirable states in terms of learning and performance outcomes.

First, it is essential to explore which specific processes the resources that are mobilized through physiological activation are allocated. The proposed conceptual model (Figure 1) suggests that the mobilized resources are allocated to enhance task-relevant cognitive processes, which aligns with findings from cognitive research showing positive associations between physiological activation and task demands/performance (e.g., Forte et al., 2019; Luque-Casado et al., 2016). However, the exact nature of this relationship is yet to be established. A commonly referenced model in this context is the Yerkes-Dodson Law (Yerkes & Dodson, 1908). Although the original version described the relation between stimulus intensity and habit formation, it has been transferred to various other fields (Teigen, 1994). Regarding cognition, it is interpreted as performance being maximised at a medium level of activation; however, support for this claim is sparse (Duschek et al., 2009). Therefore, remains yet to be explored what the optimal level of physiological activation is in achievement situations to maximize learning and performance outcomes. Further, the effect of activation on achievement outcomes may also depend on emotional valence: It has been suggested that while a positively valenced activated state may enhance performance, a negatively valenced activated state leads to fewer resources being allocated to task-relevant cognitive processes and more resources being diverted into task-irrelevant processes such as worries (Pekrun et al., 2017).

Therefore, as this final part of the proposed model, linking the mobilization of and cognitive processes, has not been covered by the present work, it warrants further refinement. Empirical work can evaluate and inform such refinement, for example, by eliciting or observing changes in the mobilization of resources in different circumstances and capturing the processes to which these resources are allocated, as well as how this affects performance. For example, using a similar setup to the present work, participants could additionally report the occurrence of non-task-related cognitive processes, such as worries, allowing an exploration of the interplay between changes in resource mobilization, emotional valence, cognitive processes, and task performance. Such an advanced understanding of how the mobilization of resources affects cognition under what conditions is essential for identifying the desirable levels of activation in achievement situations to aim for in practical learning and assessment settings.

Second, beyond the allocation of resources during a specific achievement situation, also more long-term effects of activation need to be examined to determine optimal levels of activation for certain settings and contexts. For example, during a test, it may be beneficial to foster higher levels of activation in test-takers, enabling them to demonstrate their full potential. Instead, it may not be beneficial for students if high levels of resource mobilization are maintained throughout an entire school day. Such high levels of activation may not be sustainable, with students being highly engaged in earlier classes, but drained in lessons later in the day. Furthermore, consistently high levels of physiological activation may have negative effects on both mental and physical health (e.g., Thorp & Schlaich, 2015; Won & Kim, 2016). Hence, identifying desirable levels of activation should not only happen in the context of each individual situation, but needs to take on a broader, more comprehensive perspective on the learners' or test-takers' overall context. With such an enhanced understanding of the optimal activation levels in different contexts, we can move beyond using physiological measures merely to describe responses. Instead, we will be able to tell whether those responses are beneficial or undesirable and, where appropriate, intervene to improve learning and assessment outcomes.

---

# 10

---

## CONCLUSION

---

Miriam Wunsch



The starting point of the present work has been the idea that physiological measures may bring methodological advances in the field of educational psychology, as they are more objective and less obtrusive than self-report measures, and provide real-time information throughout an achievement situation. Indeed, the present empirical studies demonstrated that physiological measures of the autonomic nervous system can be used to obtain repeated-measures data from individuals during an achievement task. An important prerequisite for effectively utilizing physiological measures in research and practice is a clear understanding of their meaning and application in achievement situations. The present work contributed to this enhanced understanding by highlighting the central importance of considering the level of analysis when investigating or leveraging associations between physiological and psychological states. Furthermore, the present work proposed a conceptual model that integrates the constructs of cognition, motivation, and emotion at the within-person level, placing physiological activation at the overlap between emotion and motivation, where both share the function of mobilizing resources for an achievement task. Findings from three empirical works supported this model, showing relations of physiological activation with achievement emotions, emotionally and motivationally relevant task appraisals, and task characteristics. Consequently, the present work suggests that physiological states in an achievement situation are best understood as the mobilization of resources, and in this function, are linked with other constructs such as cognition, emotion, and motivation. Finally, the findings provided some insights into possible differences in the meaning of different autonomic indicators in an achievement situation within their broader function of resource mobilization. Overall, the present work highlighted the potential of using physiological measures of the autonomic nervous system in achievement situations and advanced our understanding of their meaning and application. At the same time, it also raised important open questions that need to be addressed in future work to fully realize the potential of physiological measures as a methodological advancement in educational psychology research and practice.



# 11

---

## REFERENCES

---

Miriam Wunsch



- Albinet, C. T., De Faria, C., & Causse, M. (2024). Cardiac pre-ejection period to index motivation and effort mobilization in cognitive studies: A critical narrative review. *Journal of Psychophysiology*, *38*(2), 81–101. <https://doi.org/10.1027/0269-8803/a000332>
- Algina, J., & Swaminathan, H. (2011). Centering in two-level nested designs. In J. Hox & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis* (pp. 285–312). Routledge. <https://doi.org/10.4324/9780203848852>
- Allen, J., Zheng, D., Kyriacou, P. A., & Elgendi, M. (2021). Photoplethysmography (PPG): State-of-the-art methods and applications. *Physiological Measurement*, *42*(10), 100301. <https://doi.org/10.1088/1361-6579/ac2d82>
- Anvari, F., Moeck, E. K., Franco, V. R., Elson, M., & Schneider, I. K. (2024). The “memory-experience gap” for affect does not reflect a general memory bias to overestimate past affect. *Emotion*, *24*(8), 1950–1961. <https://doi.org/10.1037/emo0001404>
- Artemenko, C., Masson, N., Georges, C., Nuerk, H.-C., & Cipora, K. (2021). Not all elementary school teachers are scared of math. *Journal of Numerical Cognition*, *7*(3), 275–294. <https://doi.org/10.5964/jnc.6063>
- Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (2nd ed., pp. 254–270). Routledge. <https://doi.org/10.4324/9781315697048-17>
- Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*(2), 182–200. <https://doi.org/10.1007/s11162-011-9251-2>
- Bari, D. S., Aldosky, H. Y. Y., Tronstad, C., Kalvøy, H., & Martinsen, Ø. G. (2018). Electrodermal responses to discrete stimuli measured by skin conductance, skin potential, and skin susceptance. *Skin Research and Technology*, *24*(1), 108–116. <https://doi.org/10.1111/srt.12397>
- Berntson, G. G., Cacioppo, J. T., Quigley, K. S., & Fabro, V. T. (1994). Autonomic space and psychophysiological response. *Psychophysiology*, *31*(1), 44–61. <https://doi.org/10.1111/j.1469-8986.1994.tb01024.x>
- Betz, K., Van Der Velden, R., Gawalko, M., Hermans, A., Pluymaekers, N., Hillmann, H. A. K., Hendriks, J., Duncker, D., & Linz, D. (2021). Interpretation der Photoplethysmographie: Schritt für Schritt. *Herzschrittmachertherapie + Elektrophysiologie*, *32*(3), 406–411. <https://doi.org/10.1007/s00399-021-00795-y>
- Biopac Systems, Inc. (2025a). BioNomadix Wireless PPG and EDA Transmitter. <https://www.biopac.com/product/bionomadix-wireless-ppg-and-eda-transmitter/>
- Biopac Systems, Inc. (2025b). MP160 System with AcqKnowledge. <https://www.biopac.com/product/mp160-system-with-acqknowledge/>
- Biopac Systems, Inc. (2025c). BioNomadix Wireless PPG and EDA Amplifier. <https://www.biopac.com/product/bionomadix-ppg-and-eda-amplifier/>
- Boucsein, W. (2012). *Electrodermal activity*. Springer US. <https://doi.org/10.1007/978-1-4614-1126-0>

- Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for psychological experiments*. <https://www.birmingham.ac.uk/documents/colleges/psych/saal/guide-electrodermal-activity.pdf>
- Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J., Adcock, R. A., Barch, D. M., Botvinick, M. M., Carver, C. S., Cools, R., Custers, R., Dickinson, A., Dweck, C. S., Fishbach, A., Gollwitzer, P. M., Hess, T. M., Isaacowitz, D. M., Mather, M., ... for the MOMCAI group. (2014). Mechanisms of motivation–cognition interaction: Challenges and opportunities. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 443–472. <https://doi.org/10.3758/s13415-014-0300-0>
- Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1), 109–131. <https://doi.org/10.1146/annurev.ps.40.020189.000545>
- Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review*, *33*(3), 1051–1095. <https://doi.org/10.1007/s10648-020-09585-3>
- Can, Y. S., Mahesh, B., & André, E. (2023). Approaches, applications, and challenges in physiological emotion recognition—A tutorial overview. *Proceedings of the IEEE*, *111*(10), 1287–1313. <https://doi.org/10.1109/JPROC.2023.3286445>
- Cantone, G. G., & Tomaselli, V. (2024). Theory and methods of the multiverse: An application for panel-based models. *Quality & Quantity*, *58*(2), 1447–1480. <https://doi.org/10.1007/s11135-023-01698-5>
- Capa, R. L., & Audiffren, M. (2009). How does achievement motivation influence mental effort mobilization? Physiological evidence of deteriorative effects of negative affects on the level of engagement. *International Journal of Psychophysiology*, *74*(3), 236–242. <https://doi.org/10.1016/j.ijpsycho.2009.09.007>
- Chalabaev, A., Major, B., Cury, F., & Sarrazin, P. (2009). Physiological markers of challenge and threat mediate the effects of performance-based goals on performance. *Journal of Experimental Social Psychology*, *45*(4), 991–994. <https://doi.org/10.1016/j.jesp.2009.04.009>
- Chang, Y.-C., & Huang, S.-L. (2012). The influence of attention levels on psychophysiological responses. *International Journal of Psychophysiology*, *86*(1), 39–47. <https://doi.org/10.1016/j.ijpsycho.2012.09.001>
- Chapleau, M. W., & Sabharwal, R. (2011). Methods of assessing vagus nerve activity and reflexes. *Heart Failure Reviews*, *16*(2), 109–127. <https://doi.org/10.1007/s10741-010-9174-6>

- Chen, Y.-Y., Fang, W.-N., Bao, H.-F., & Guo, B.-Y. (2024). The effect of task interruption on working memory performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *66*(4), 1132–1151. <https://doi.org/10.1177/00187208221139017>
- Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2019). The body and the brain: Measuring skin conductance responses to understand the emotional experience. *Organizational Research Methods*, *22*(1), 394–420. <https://doi.org/10.1177/1094428116681073>
- Constantinou, E., Vlemincx, E., & Panayiotou, G. (2023). Testing emotional response coherence assumptions: Comparing emotional versus non-emotional states. *Psychophysiology*, *60*(11), e14359. <https://doi.org/10.1111/psyp.14359>
- Dawson, M. E., Schell, A. M., Fillion, D. L., & Berntson, G. G. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology* (3rd ed., pp. 157–181). Cambridge University Press. <https://doi.org/10.1017/CBO9780511546396.007>
- Del Giudice, M., Buck, C. L., Chaby, L. E., Gormally, B. M., Taff, C. C., Thawley, C. J., Vitousek, M. N., & Wada, H. (2018). What is stress? A systems perspective. *Integrative and Comparative Biology*. <https://doi.org/10.1093/icb/icy114>
- D’Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys*, *47*(3), 1–36. <https://doi.org/10.1145/2682899>
- Duschek, S., Muckenthaler, M., Werner, N., & Reyes Del Paso, G. A. (2009). Relationships between features of autonomic cardiovascular control and cognitive performance. *Biological Psychology*, *81*(2), 110–117. <https://doi.org/10.1016/j.biopsycho.2009.03.003>
- Eatchel, K. A., Kramer, H., & Drews, F. (2012). The effects of interruption context on task performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 2118–2122. <https://doi.org/10.1177/1071181312561449>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 1–13. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Eckerskorn, F., Frenzel, A. C., Wünsch, M., Pekrun, R., Brings, L., & Kleine, M. (under revision). *Confronted with the “necessary evil”: The Impact of pre-service elementary school teachers’ mathematics self-concept on emotional and physiological responses to mathematics exposure*.
- Eidlin Levy, H., & Rubinsten, O. (2021). Numbers (but not words) make math anxious individuals sweat: Physiological evidence. *Biological Psychology*, *165*, 108187. <https://doi.org/10.1016/j.biopsycho.2021.108187>
- Epskamp, S., Hoekstra, R. H. A., Burger, J., & Waldorp, L. J. (2022). Longitudinal design choices: Relating data to analysis. In A.-M. Isvoranu, S. Epskamp, L. J. Waldorp, & D. Borsboom, *Network Psychometrics with R* (1st ed., pp. 157–168). Routledge. <https://doi.org/10.4324/9781003111238-12>
- Evers, C., Hopp, H., Gross, J. J., Fischer, A. H., Manstead, A. S. R., & Mauss, I. B. (2014). Emotion response coherence: A dual-process perspective. *Biological Psychology*, *98*, 43–49. <https://doi.org/10.1016/j.biopsycho.2013.11.003>

- Fairclough, S. H., & Ewing, K. (2017). The effect of task demand and incentive on neurophysiological and cardiovascular markers of effort. *International Journal of Psychophysiology*, *119*, 58–66. <https://doi.org/10.1016/j.ijpsycho.2017.01.007>
- Fatissou, J., Oswald, V., & Lalonde, F. (2016). Influence diagram of physiological and environmental factors affecting heart rate variability: An extended literature overview. *Heart International*, *11*(1), e32–e40. <https://doi.org/10.5301/heartint.5000232>
- Feiler, J. B., & Stabio, M. E. (2018). Three pillars of educational neuroscience from three decades of literature. *Trends in Neuroscience and Education*, *13*, 17–25. <https://doi.org/10.1016/j.tine.2018.11.001>
- Foroughi, C. K., Werner, N. E., Barragán, D., & Boehm-Davis, D. A. (2015). Interruptions disrupt reading comprehension. *Journal of Experimental Psychology: General*, *144*(3), 704–709. <https://doi.org/10.1037/xge0000074>
- Forte, G., Favieri, F., & Casagrande, M. (2019). Heart rate variability and cognitive function: A systematic review. *Frontiers in Neuroscience*, *13*, 710. <https://doi.org/10.3389/fnins.2019.00710>
- Frenzel, A. C., Fiedler, D., Marx, A. K. G., Reck, C., & Pekrun, R. (2020). Who Enjoys Teaching, and When? Between- and Within-Person Evidence on Teachers' Appraisal-Emotion Links. *Frontiers in Psychology*, *11*, 1092. <https://doi.org/10.3389/fpsyg.2020.01092>
- Gehlbach, H., & Barge, S. (2012). Anchoring and Adjusting in Questionnaire Responses. *Basic and Applied Social Psychology*, *34*(5), 417–433. <https://doi.org/10.1080/01973533.2012.711691>
- Gendolla, G. H. E. (2025). The intensity of mental effort: “The heart does not lie.” *Synthese*, *205*(6), 231. <https://doi.org/10.1007/s11229-025-05043-1>
- Gendolla, G. H. E., Wright, R. A., & Richter, M. (2019). Advancing issues in motivation intensity research: Updated insights from the cardiovascular system. In R. M. Ryan (Ed.), *The Oxford Handbook of Human Motivation* (pp. 372–392). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190666453.013.21>
- Gerritsen, R. J. S., & Band, G. P. H. (2018). Breath of life: The respiratory vagal stimulation model of contemplative activity. *Frontiers in Human Neuroscience*, *12*, 397. <https://doi.org/10.3389/fnhum.2018.00397>
- Goetz, T., Hall, N. C., & Krannich, M. (2019). Boredom. In K. A. Renninger & S. E. Hidi (Eds.), *The Cambridge Handbook of Motivation and Learning* (1st ed., pp. 465–489). Cambridge University Press. <https://doi.org/10.1017/9781316823279>
- Goldstein, D. S., & Kopin, I. J. (2007). Evolution of concepts of stress. *Stress*, *10*(2), 109–120. <https://doi.org/10.1080/10253890701288935>
- Götz, M., Sarma, A., & O'Boyle, E. H. (2024). The multiverse of universes: A tutorial to plan, execute and interpret multiverses analyses using the R package *multiverse*. *International Journal of Psychology*, *59*(6), 1003–1014. <https://doi.org/10.1002/ijop.13229>
- Graham, M. C., Husman, J., Pekrun, R., Villanueva, I., & Christensen, D. (2023). The dynamic experience of taking an examination: Ever changing cortisol and expectancy for success. *British Journal of Educational Psychology*, *93*(S1), 195–210. <https://doi.org/10.1111/bjep.12521>

- Hamill, R. W., Shapiro, R. E., & Vizzard, M. A. (2012). Peripheral autonomic nervous system. In *Primer on the Autonomic Nervous System* (pp. 17–26). Elsevier. <https://doi.org/10.1016/B978-0-12-386525-0.00004-4>
- Harder, J. A. (2020). The multiverse of methods: Extending the multiverse analysis to address data-collection decisions. *Perspectives on Psychological Science, 15*(5), 1158–1177. <https://doi.org/10.1177/1745691620917678>
- Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior, 48*, 615–625. <https://doi.org/10.1016/j.chb.2015.02.013>
- Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science, 47*(2), 151–180. <https://doi.org/10.1007/s11251-018-09480-z>
- Harrison, N. A., Kreibig, S. D., & Critchley, H. D. (2013). A two-way road: Efferent and afferent pathways of autonomic activity in emotion. In J. Armony & P. Vuilleumier (Eds.), *The Cambridge Handbook of Human Affective Neuroscience* (1st ed., pp. 82–106). Cambridge University Press. <https://doi.org/10.1017/CBO9780511843716.006>
- Heemskerck, C., Strand, S., & Malmberg, L. (2023). Physical activity predicts task-related behaviour, affect and tiredness in the primary school classroom: A within-person experiment. *British Journal of Educational Psychology, 93*(S1), 130–151. <https://doi.org/10.1111/bjep.12523>
- Hoffman, L. (2015). Introduction to the analysis of longitudinal data. In *Longitudinal Analysis* (1st ed., pp. 3–28). Routledge. <https://doi.org/10.4324/9781315744094>
- Hoffman, L., & Walters, R. W. (2022). Catching up on multilevel modeling. *Annual Review of Psychology, 73*(1), 659–689. <https://doi.org/10.1146/annurev-psych-020821-103525>
- Horvers, A., Tombeng, N., Bosse, T., Lazonder, A. W., & Molenaar, I. (2021). Detecting emotions through electrodermal activity in learning contexts: A systematic review. *Sensors, 21*(23), 7869. <https://doi.org/10.3390/s21237869>
- Hsieh, F., Ferrer, E., Chen, S., Mauss, I. B., John, O., & Gross, J. J. (2011). A network approach for evaluating coherence in multivariate systems: An application to psychophysiological emotion data. *Psychometrika, 76*(1), 124–152. <https://doi.org/10.1007/s11336-010-9194-0>
- Huber, K., & Bannert, M. (2023). What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements. *Journal of Computers in Education, 10*(1), 189–215. <https://doi.org/10.1007/s40692-022-00228-w>
- Ikezawa, S., Corbera, S., Liu, J., & Wexler, B. E. (2012). Empathy in electrodermal responsive and nonresponsive patients with schizophrenia. *Schizophrenia Research, 142*(1–3), 71–76. <https://doi.org/10.1016/j.schres.2012.09.011>
- iMotions. (2025). About Us: Ownership and Structure. <https://imotions.com/about-us/ownership-and-structure/>
- iMotions (Version 9.3). (2022). [Computer software].

- Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, 22(4), 337–349. <https://doi.org/10.1016/j.tics.2018.01.007>
- Ishikawa, M. (2023). Measuring the autonomic nervous system as a window into the mind and brain: A selective review. *European Psychologist*, 28(2), 67–82. <https://doi.org/10.1027/1016-9040/a000500>
- Jögi, A., Pakarinen, E., & Lerkkanen, M. (2023). Teachers' physiological and self-reported stress, teaching practices and students' learning outcomes in Grade 1. *British Journal of Educational Psychology*, 93(S1), 211–226. <https://doi.org/10.1111/bjep.12529>
- Johar, O., & Sackett, A. M. (2018). The self-contaminating nature of repeated reports of negative emotions. *Basic and Applied Social Psychology*, 40(5), 293–307. <https://doi.org/10.1080/01973533.2018.1496336>
- Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science*, 24(1), 10–16. <https://doi.org/10.1177/0963721414550708>
- Ketonen, E. E., Salonen, V., Lonka, K., & Salmela-Aro, K. (2023). Can you feel the excitement? Physiological correlates of students' self-reported emotions. *British Journal of Educational Psychology*, 93(S1), 113–129. <https://doi.org/10.1111/bjep.12534>
- Kiuru, N., Malmberg, L.-E., Eklund, K., Penttonen, M., Ahonen, T., & Hirvonen, R. (2022). How are learning experiences and task properties associated with adolescents' emotions and psychophysiological states? *Contemporary Educational Psychology*, 71, 1–16. <https://doi.org/10.1016/j.cedpsych.2022.102095>
- Kiuru, N., Trög, A.-S., Pasanen, M., Tourunen, A., Mikkonen, J., Ahonen, T., & Penttonen, M. (2022). Associations between adolescents' subjectively experienced emotions and psychophysiological reactions in achievement situations. *Merrill-Palmer Quarterly*, 68(1), 39–71. <https://doi.org/10.1353/mpq.2022.0003>
- Kleiger, R. E., Stein, P. K., & Bigger, J. T. (2005). Heart rate variability: Measurement and clinical utility. *Annals of Noninvasive Electrocardiology*, 10(1), 88–101. <https://doi.org/10.1111/j.1542-474X.2005.10101.x>
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Kreibig, S. D., & Gendolla, G. H. E. (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). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780203148211.ch31>
- Kreibig, S. D., Gendolla, G. H. E., & Scherer, K. R. (2012). Goal relevance and goal conduciveness appraisals lead to differential autonomic reactivity in emotional responding to performance feedback. *Biological Psychology*, 91(3), 365–375. <https://doi.org/10.1016/j.biopsycho.2012.08.007>
- Krieglmeyer, R., & Deutsch, R. (2010). Comparing measures of approach–avoidance behaviour: The manikin task vs. two versions of the joystick task. *Cognition & Emotion*, 24(5), 810–828. <https://doi.org/10.1080/02699930903047298>

- Kubios Oy. (2024). *Kubios HRV Scientific User's Guide 2024*.  
<https://www.kubios.com/downloads/HRV-Scientific-Users-Guide.pdf>
- Laborde, S., Allen, M. S., Borges, U., Dosseville, F., Hosang, T. J., Iskra, M., Mosley, E., Salvotti, C., Spolverato, L., Zammit, N., & Javelle, F. (2022). Effects of voluntary slow breathing on heart rate and heart rate variability: A systematic review and a meta-analysis. *Neuroscience & Biobehavioral Reviews*, *138*, 104711.  
<https://doi.org/10.1016/j.neubiorev.2022.104711>
- Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research – Recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, *08*.  
<https://doi.org/10.3389/fpsyg.2017.00213>
- Lang, P. J. (1988). What are the data of emotion? In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive Perspectives on Emotion and Motivation* (pp. 173–191). Springer Netherlands. [https://doi.org/10.1007/978-94-009-2792-6\\_7](https://doi.org/10.1007/978-94-009-2792-6_7)
- Lange, J., Dalege, J., Borsboom, D., van Kleef, G. A., & Fischer, A. H. (2020). Toward an integrative psychometric model of emotions. *Perspectives on Psychological Science*, *15*(2), 444–468. <https://doi.org/doi.org/10.1177/17456916198950>
- Lehikoinen, A., Ranta-Nilkku, E., Mikkonen, J., Kaartinen, J., Penttonen, M., Ahonen, T., & Kiuru, N. (2019). The role of adolescents' temperament in their positive and negative emotions as well as in psychophysiological reactions during achievement situations. *Learning and Individual Differences*, *69*, 116–128.  
<https://doi.org/10.1016/j.lindif.2018.11.004>
- Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emotion Review*, *6*(2), 100–112. <https://doi.org/10.1177/1754073913512003>
- Liu, Y., & Hou, S. (2018). Potential reciprocal relationship between motivation and achievement: A longitudinal study. *School Psychology International*, *39*(1), 38–55.  
<https://doi.org/10.1177/0143034317710574>
- Lizdek, I., Sadler, P., Woody, E., Ethier, N., & Malet, G. (2012). Capturing the stream of behavior: A computer-joystick method for coding interpersonal behavior continuously over time. *Social Science Computer Review*, *30*(4), 513–521.  
<https://doi.org/10.1177/0894439312436487>
- Loe, B., Sun, L., Simonfy, F., & Doeblner, P. (2018). Evaluating an automated number series item generator using linear logistic test models. *Journal of Intelligence*, *6*(2), 20.  
<https://doi.org/10.3390/jintelligence6020020>
- Lu, G., Yang, F., Taylor, J. A., & Stein, J. F. (2009). A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *Journal of Medical Engineering & Technology*, *33*(8), 634–641.  
<https://doi.org/10.3109/03091900903150998>
- Luk, G., & Christodoulou, J. A. (2024). Cognitive neuroscience and education. In P. A. Schutz & K. R. Muis (Eds.), *Handbook of educational psychology* (4th ed., pp. 383–404). Routledge.
- Luque-Casado, A., Perales, J. C., Cárdenas, D., & Sanabria, D. (2016). Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological Psychology*, *113*, 83–90. <https://doi.org/10.1016/j.biopsycho.2015.11.013>

- Mackersie, C. L., & Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: Heart rate variability and skin conductance. *Ear & Hearing, 37*(1), 118S-125S. <https://doi.org/10.1097/AUD.0000000000000305>
- Maier, K. J., Waldstein, S. R., & Synowski, S. J. (2003). Relation of cognitive appraisal to cardiovascular reactivity, affect, and task engagement. *Annals of Behavioral Medicine, 26*(1), 32–41. [https://doi.org/10.1207/S15324796ABM2601\\_05](https://doi.org/10.1207/S15324796ABM2601_05)
- Malim, T. (1994). *Cognitive Processes*. Macmillan Education UK. <https://doi.org/10.1007/978-1-349-13133-4>
- Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology, 121*, 62–73. <https://doi.org/10.1016/j.biopsycho.2016.10.002>
- Martin, A. J., & Burns, E. C. (2023). Gaps in contemporary motivation research: A biopsychological perspective. In M. Bong, J. Reeve, & S. Kim (Eds.), *Motivation Science* (1st ed., pp. 380–386). Oxford University Press New York. <https://doi.org/10.1093/oso/9780197662359.003.0062>
- Martin, A. J., Kennett, R., Pearson, J., Mansour, M., Papworth, B., & Malmberg, L.-E. (2021). Challenge and threat appraisals in high school science: Investigating the roles of psychological and physiological factors. *Educational Psychology, 41*(5), 618–639. <https://doi.org/10.1080/01443410.2021.1887456>
- Martin, A. J., Malmberg, L., Pakarinen, E., Mason, L., & Mainhard, T. (2023). The potential of biophysiology for understanding motivation, engagement and learning experiences. *British Journal of Educational Psychology, 93*(S1), 1–9. <https://doi.org/10.1111/bjep.12584>
- 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), 175–190. <https://doi.org/10.1037/1528-3542.5.2.175>
- Mauss, I. B., Zerwas, F. K., Wilhelm, F. H., & John, O. P. (2024). Coherence of emotional response systems: Theory, measurement, and benefits. In *Advances in Experimental Social Psychology* (Vol. 69, pp. 59–149). Elsevier. <https://doi.org/10.1016/bs.aesp.2023.11.002>
- Mendes, W. B. (2016). Emotion and the autonomic nervous system. In L. Feldman Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (4th edition, pp. 166–181). The Guilford Press.
- Miron, A. M., & Brehm, J. W. (2012). Emotional intensity theory and its cardiovascular implications for emotional states. In R. A. Wright & G. H. E. Gendolla (Eds.), *How motivation affects cardiovascular response: Mechanisms and applications*. (pp. 121–137). American Psychological Association. <https://doi.org/10.1037/13090-006>
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*(2), 112–117. <https://doi.org/10.1111/j.1467-8721.2009.01619.x>
- Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. In D. W. Putwain & K. Smart (Eds.), *BJEP Monograph Series II: Part 12 The Role of*

- Competence and Beliefs in Teaching and Learning*. British Psychological Society.  
<https://doi.org/10.53841/bpsmono.2017.cat2023.6>
- Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. R., & Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: A physiological account of a “default mode” of brain function. *NeuroImage*, *22*(1), 243–251. <https://doi.org/10.1016/j.neuroimage.2004.01.019>
- Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the “×” out of expectancy-value theory?: A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science*, *22*(8), 1058–1066. <https://doi.org/10.1177/0956797611415540>
- O’Donnell, A., Dobozy, E., Nagel, M. C., Barlett, B., Smala, S., Wormald, C., Yates, G., Reeve, J., Smith, J. K., Spooner-Lane, R., & Youssef-Shalala, A. (2025). *Educational psychology* (4th Australian Edition). John Wiley & Sons Australia, Ltd.
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R. (2020). Self-Report is indispensable to assess students’ learning. *Frontline Learning Research*, *8*(3), 185–193. <https://doi.org/10.14786/flr.v8i3.637>
- Pekrun, R. (2023a). Dissecting the elephant: Cognition, emotion, and motivation as distinct but intertwined entities. In M. Bong, J. Reeve, & S. Kim (Eds.), *Motivation Science* (1st ed., pp. 227–231). Oxford University Press New York.  
<https://doi.org/10.1093/oso/9780197662359.003.0038>
- Pekrun, R. (2023b). Mind and body in students’ and teachers’ engagement: New evidence, challenges, and guidelines for future research. *British Journal of Educational Psychology*, *93*(S1), 227–238. <https://doi.org/10.1111/bjep.12575>
- Pekrun, R. (2024a). Control-value theory: From achievement emotion to a general theory of human emotions. *Educational Psychology Review*, *36*(3), 1–36.  
<https://doi.org/10.1007/s10648-024-09909-7>
- Pekrun, R. (2024b). Overcoming fragmentation in motivation science: Why, when, and how should we integrate theories? *Educational Psychology Review*, *36*(1), 27.  
<https://doi.org/10.1007/s10648-024-09846-5>
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students’ learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48.  
<https://doi.org/10.1016/j.cedpsych.2010.10.002>
- Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, *88*(5), 1653–1670. <https://doi.org/10.1111/cdev.12704>
- Pekrun, R., & Marsh, H. W. (2022). Research on situated motivation and emotion: Progress and open problems. *Learning and Instruction*, *81*, 101664.  
<https://doi.org/10.1016/j.learninstruc.2022.101664>

- Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., Van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology, 124*(1), 145–178. <https://doi.org/10.1037/pspp0000448>
- Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education, 52*(3), 275–300. <https://doi.org/10.1080/15391523.2020.1719943>
- Portnova, G. V., Liaukovich, K. M., Vasilieva, L. N., & Alshanskaia, E. I. (2023). Autonomic and behavioral indicators on increased cognitive loading in healthy volunteers. *Neuroscience and Behavioral Physiology, 53*(1), 92–102. <https://doi.org/10.1007/s11055-023-01394-9>
- Putwain, D. W., Nicholson, L. J., & Pekrun, R. (2025). Test anxiety, emotion regulation, and achievement: Lagged and contemporaneous reciprocal relations. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000975>
- Quinio, A. E., & Lam, T. C. M. (2021). Methods and biases in measuring change with self-reports. In *Basic elements of survey research in education: Addressing the problems you advisor never told you about* (pp. 193–217). Information Age Publishing.
- Quintana, D. S., & Heathers, J. A. J. (2014). Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in Psychology, 5*. <https://doi.org/10.3389/fpsyg.2014.00805>
- Radtke, V., Martarelli, C., & Wolff, W. (2025). *Task-related effort – Electrodermal activity as an indicator of boredom vs. Difficulty related effort in cognitive-physical sequential tasks*. <https://doi.org/10.31219/osf.io/2t6p8>
- Radulescu, E., Nagai, Y., & Critchley, H. (2015). Mental effort: Brain and autonomic correlates in health and disease. In G. H. E. Gendolla, M. Tops, & S. L. Koole (Eds.), *Handbook of Biobehavioral Approaches to Self-Regulation* (pp. 237–253). Springer New York. [https://doi.org/10.1007/978-1-4939-1236-0\\_16](https://doi.org/10.1007/978-1-4939-1236-0_16)
- Ravazi, T. (2001). *Self-report measures: An overview of concerns and limitations of questionnaire use in occupational stress research*.
- Richter, M., Friedrich, A., & Gendolla, G. H. E. (2008). Task difficulty effects on cardiac activity. *Psychophysiology, 45*(5), 869–875. <https://doi.org/10.1111/j.1469-8986.2008.00688.x>
- Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). Three decades of research on motivational intensity theory. In *Advances in Motivation Science* (Vol. 3, pp. 149–186). Elsevier. <https://doi.org/10.1016/bs.adms.2016.02.001>
- Richter, M., & Slade, K. (2017). Interpretation of physiological indicators of motivation: Caveats and recommendations. *International Journal of Psychophysiology, 119*, 4–10. <https://doi.org/10.1016/j.ijpsycho.2017.04.007>
- Ringeisen, T., Lichtenfeld, S., Becker, S., & Minkley, N. (2019). Stress experience and performance during an oral exam: The role of self-efficacy, threat appraisals, anxiety, and cortisol. *Anxiety, Stress, & Coping, 32*(1), 50–66. <https://doi.org/10.1080/10615806.2018.1528528>
- Roos, A.-L., Goetz, T., Krannich, M., Donker, M., Bieleke, M., Caltabiano, A., & Mainhard, T. (2023). Control, anxiety and test performance: Self-reported and physiological

- indicators of anxiety as mediators. *British Journal of Educational Psychology*, 93(S1), 72–89. <https://doi.org/10.1111/bjep.12536>
- Roos, A.-L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review*, 33(2), 579–618. <https://doi.org/10.1007/s10648-020-09543-z>
- Salamone, J. D., & Correa, M. (2024). The neurobiology of activational aspects of motivation: Exertion of effort, effort-based decision making, and the role of dopamine. *Annual Review of Psychology*, 75(1), 1–32. <https://doi.org/10.1146/annurev-psych-020223-012208>
- Sammito, S., Thielmann, B., & Böckelmann, I. (2024). Update: Factors influencing heart rate variability—a narrative review. *Frontiers in Physiology*, 15, 1430458. <https://doi.org/10.3389/fphys.2024.1430458>
- Schuurman, N. K. (2023). *A “Within/Between Problem” Primer: About (Not) Separating Within-Person Variance and Between-Person Variance in Psychology*. <https://doi.org/10.31234/osf.io/7zgkx>
- Seery, M. D., Weisbuch, M., & Blascovich, J. (2009). Something to gain, something to lose: The cardiovascular consequences of outcome framing. *International Journal of Psychophysiology*, 73(3), 308–312. <https://doi.org/10.1016/j.ijpsycho.2009.05.006>
- 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. <https://doi.org/10.3389/fpsyg.2014.01040>
- Shao, K., Pekrun, R., Marsh, H. W., & Loderer, K. (2020). Control-value appraisals, achievement emotions, and foreign language performance: A latent interaction analysis. *Learning and Instruction*, 69, 101356. <https://doi.org/10.1016/j.learninstruc.2020.101356>
- Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40(1), 99–124. <https://doi.org/10.1146/annurev-neuro-072116-031526>
- Shenhav, A., Prater Fahey, M., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science*, 30(4), 307–314. <https://doi.org/10.1177/09637214211009510>
- Shimmer. (2025). EDA Electrodes. <https://www.shimmersensing.com/product/eda-electrodes/>
- Shuman, V., & Scherer, K. R. (2014). Concepts and structures of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 13–35). Routledge.
- Silvestrini, N., & Gendolla, G. H. E. (2019). Affect and cognitive control: Insights from research on effort mobilization. *International Journal of Psychophysiology*, 143, 116–125. <https://doi.org/10.1016/j.ijpsycho.2019.07.003>
- Solomon, P. (1995). The think aloud method: A practical guide to modelling cognitive processes. *Information Processing & Management*, 31(6), 906–907. [https://doi.org/10.1016/0306-4573\(95\)90031-4](https://doi.org/10.1016/0306-4573(95)90031-4)

- Špiljak, B., Vilibić, M., Glavina, A., Crnković, M., Šešerko, A., & Lugović-Mihić, L. (2022). A review of psychological stress among students and its assessment using salivary biomarkers. *Behavioral Sciences*, *12*(10), 400. <https://doi.org/10.3390/bs12100400>
- Stemmler, G., Aue, T., & Wacker, J. (2007). Anger and fear: Separable effects of emotion and motivational direction on somatovisceral responses. *International Journal of Psychophysiology*, *66*(2), 141–153. <https://doi.org/10.1016/j.ijpsycho.2007.03.019>
- Strohmaier, A. R., Schiepe-Tiska, A., & Reiss, K. M. (2020). A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety. *Frontline Learning Research*, *8*(1), 16–32. <https://doi.org/10.14786/flr.v8i1.427>
- Takeuchi, H., Taki, Y., Nouchi, R., Sekiguchi, A., Kotozaki, Y., Miyauchi, C. M., Yokoyama, R., Iizuka, K., Hashizume, H., Nakagawa, S., Kunitoki, K., Sassa, Y., & Kawashima, R. (2014). Regional gray matter density is associated with achievement motivation: Evidence from voxel-based morphometry. *Brain Structure and Function*, *219*(1), 71–83. <https://doi.org/10.1007/s00429-012-0485-3>
- Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. <https://doi.org/10.1016/j.cmpb.2013.07.024>
- Tatsumi, S., Kuratsune, D., & Kuratsune, H. (2025). Relationship between autonomic nervous function and brain functions such as memory and attention. *Physiology & Behavior*, *288*, 114721. <https://doi.org/10.1016/j.physbeh.2024.114721>
- Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, *4*(4), 525–547. <https://doi.org/10.1177/09593543940444004>
- Thorp, A. A., & Schlaich, M. P. (2015). Relevance of sympathetic nervous system Activation in obesity and metabolic syndrome. *Journal of Diabetes Research*, *2015*, 1–11. <https://doi.org/10.1155/2015/341583>
- Van Diest, I., Verstappen, K., Aubert, A. E., Widjaja, D., Vansteenwegen, D., & Vlemincx, E. (2014). Inhalation/exhalation ratio modulates the effect of slow breathing on heart rate variability and relaxation. *Applied Psychophysiology and Biofeedback*, *39*(3–4), 171–180. <https://doi.org/10.1007/s10484-014-9253-x>
- Van Doren, N., Dickens, C. N., Benson, L., Brick, T. R., Gatzke-Kopp, L., & Oravec, Z. (2021). Capturing emotion coherence in daily life: Using ambulatory physiology measures and ecological momentary assessments to examine within-person associations and individual differences. *Biological Psychology*, *162*, 108074. <https://doi.org/10.1016/j.biopsycho.2021.108074>
- Venables, P. H., & Mitchell, D. A. (1996). The effects of age, sex and time of testing on skin conductance activity. *Biological Psychology*, *43*(2), 87–101. [https://doi.org/10.1016/0301-0511\(96\)05183-6](https://doi.org/10.1016/0301-0511(96)05183-6)
- Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, *49*(3), 193–213. <https://doi.org/10.1080/00273171.2014.889593>
- Vu, T. V., Scharmer, A. L., Van Triest, E., Van Atteveldt, N., & Meeter, M. (2024). The reciprocity between various motivation constructs and academic achievement: A

- systematic review and multilevel meta-analysis of longitudinal studies. *Educational Psychology*, 44(2), 136–170. <https://doi.org/10.1080/01443410.2024.2307960>
- Waxenbaum, J. A., Reddy, V., & Varacallo, M. A. (2023). *Anatomy, Autonomic Nervous System*. StatPearls. <http://www.ncbi.nlm.nih.gov/books/NBK539845/>
- Wehrwein, E. A., Orer, H. S., & Barman, S. M. (2016). Overview of the anatomy, physiology, and pharmacology of the autonomic nervous system. *Comprehensive Physiology*, 1239–1278. <https://doi.org/10.1002/cphy.c150037>
- Wei, H., & Jin, Z. (2024). EEG correlates of trait test anxiety in the flanker task for adolescents. *Neuroscience Letters*, 826, 137725. <https://doi.org/10.1016/j.neulet.2024.137725>
- Weissman, D. G., & Mendes, W. B. (2021). Correlation of sympathetic and parasympathetic nervous system activity during rest and acute stress tasks. *International Journal of Psychophysiology*, 162, 60–68. <https://doi.org/10.1016/j.ijpsycho.2021.01.015>
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. <https://doi.org/10.1006/ceps.1999.1015>
- Won, E., & Kim, Y.-K. (2016). Stress, the autonomic nervous system, and the ommune-kynurenine pathway in the etiology of depression. *Current Neuropharmacology*, 14(7), 665–673. <https://doi.org/10.2174/1570159X14666151208113006>
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459–482. <https://doi.org/10.1002/cne.920180503>
- Zawadzki, M. J., Smyth, J. M., Sliwinski, M. J., Ruiz, J. M., & Gerin, W. (2017). Revisiting the lack of association between affect and physiology: Contrasting between-person and within-person analyses. *Health Psychology*, 36(8), 811–818. <https://doi.org/10.1037/hea0000466>
- Zimmer, H., & Richter, F. (2023). Novelty detection and orienting: Effects on skin conductance and heart rate. *Psychological Research*, 87(4), 1101–1113. <https://doi.org/10.1007/s00426-022-01735-2>