

# Staatsexamensarbeit

Lehramt Oberschule  
Sport und Geschichte

## **Reaktionen des Lehrkörpers auf verschiedene Arten von Unterrichtsstörungen: Messung des Pulses als Stressindikator innerhalb eines Laborsettings**

eingereicht von  
Tom Zitzmann

[tz25gase@studserv.uni-leipzig.de](mailto:tz25gase@studserv.uni-leipzig.de)  
Matrikelnummer: 3701729

Universität Leipzig

Erziehungswissenschaftliche Fakultät  
Institut für Bildungswissenschaften

Professur für Empirische Schul- und Unterrichtsforschung

Erstgutachten: Mandy Klatt  
Zweitgutachten: Dr. Christin Lotz

Abgabedatum: 26.10.2023

## **Abstract**

Der Lehrberuf zählt zu den Berufen mit der höchsten Stressbelastung, weshalb sich Lehrkräfte häufig beruflich umorientieren oder sich mit den gesundheitlichen Folgen konfrontiert sehen. Vielfältige Studien zeigen, dass vor allem der Unterricht sowie die damit einhergehenden Unterrichtsstörungen zu den Hauptstressoren zählen. Moderne und preiswerte Smartwatch Wearables bieten mit der Herzratenmessung potentiell die Möglichkeit, die physischen Auswirkungen von Stress während des Unterrichts aufzuzeichnen. Das Ziel der vorliegenden Arbeit ist es, zu untersuchen, inwiefern sich Smartwatch Wearables eignen, um im Rahmen einer simulierten Unterrichtssituation anhand der Herzratenaufzeichnung den Stress der Lehrkraft abzubilden. Zu diesem Zweck wurden die Herzratenmittelwerte in fünf jeweils zehnminütigen Zeitintervallen unterschiedlicher (psychischer) Belastung miteinander verglichen – darunter auch die simulierte Unterrichtseinheit – und die Unterschiede zwischen den Mittelwerten der Zeitintervalle mit einer einfaktoriellen Varianzanalyse mit Messwiederholung mit Kontrasten auf Signifikanz untersucht. Die Stichprobe der quasi-experimentellen Querschnittstudie bestand aus 76 Proband:innen im Alter zwischen 19 und 59 Jahren, die mindestens das dritte Semester der Lehramtsausbildung erreicht haben oder dem Lehrberuf bereits nachgehen. Es konnte gezeigt werden, dass sich die mittlere Herzrate der simulierten Unterrichtsphase statistisch signifikant von den anderen vier Belastungsphasen unterscheidet und somit der Stress während des Unterrichts anhand der Herzrate abgebildet werden kann. Die Herzratenmessung moderner und kostengünstiger Smartwatch Wearables hat somit das Potenzial ein wichtiger Anhaltspunkt im individuellen Stressmanagement von Lehrkräften zu werden.

# Inhaltsverzeichnis

|  |    |
|--|----|
| Abbildungsverzeichnis.....   | IV |
| Tabellenverzeichnis.....   | V  |
| Abkürzungs- & Symbolverzeichnis .....                              | VI |
| 1 Einleitung .....   | 1  |
| 2 Theoretischer Hintergrund .....                                  | 3  |
| 2.1 Definitionen und Grundlagen von Stress .....                   | 3  |
| 2.2 Gesundheitliche Risiken langanhaltender Stressexposition ..... | 7  |
| 2.3 Stress und Belastungen im Lehrberuf .....                      | 8  |
| 2.3.1 Grundlagen und Modelle .....                                 | 9  |
| 2.3.2 Stressforschung im Unterrichtskontext.....                   | 12 |
| 2.4 Herzratenunterschiede .....                                    | 16 |
| 2.5 Herzfrequenzüberwachung mittels Smartwatch Wearables .....     | 17 |
| 2.6 Fragestellung und Hypothese .....                              | 22 |
| 3 Methodisches Vorgehen.....                                       | 25 |
| 3.1 Stichprobe.....  | 25 |
| 3.2 Durchführung und Aufbau der Studie .....                       | 26 |
| 3.3 Variablen .....  | 29 |
| 3.4 Methodiken der statistischen Datenauswertung .....             | 31 |
| 3.4.1 Mittelwertbildung und z-Standardisierung der HR-Daten.....   | 31 |
| 3.4.2 ANOVA mit Messwiederholung .....                             | 32 |
| 3.4.3 Kontraste .....  | 35 |
| 3.4.4 Effektstärke $d$ .....                                       | 36 |
| 4 Ergebnisse .....   | 38 |
| 5 Diskussion .....   | 45 |
| 6 Fazit und Implikationen für die Praxis.....                      | 54 |
| Literaturverzeichnis.....  | 57 |
| Anhang .....   | 64 |
| A Setup der Unterrichtseinheit.....                                | 64 |
| B Syntax und Ergebnis-Output .....                                 | 65 |
| C Übersicht über digitalen Anhang .....                            | 68 |
| Selbstständigkeitserklärung.....                                   | 69 |

## Abbildungsverzeichnis

|   |    |
|---|----|
| <b>Abbildung 1:</b> Darstellung der verschiedenen Prozesse und deren Verknüpfungen innerhalb des transaktionalen Stressmodells nach Lazarus und Folkman (Heinrichs et al., 2015, S. 25) ..... | 6  |
| <b>Abbildung 2:</b> Anzahl der Proband:innen verteilt auf die studierten Schulformen .....  | 26 |
| <b>Abbildung 3:</b> HR-Mittelwerte mit SD der fünf betrachteten Phasen a) in <i>bpm</i> und b) z-standardisiert .....   | 40 |
| <b>Abbildung 4:</b> Aufbau der Unterrichtseinheit mit Positionen aller beteiligten Personen und des Equipments .....  | 64 |
| <b>Abbildung 5:</b> Boxplot zur Bestimmung von Ausreißern innerhalb der fünf Abstufungen der unbestimmten Variable .....  | 66 |

## **Tabellenverzeichnis**

|   |    |
|---|----|
| <b>Tabelle 1:</b> Stufen der Belastungsintensität in Abhängigkeit der prozentualen maximalen HR nach Hottenrott (2007).....                   | 17 |
| <b>Tabelle 2:</b> Klassifizierung der Effektstärke $d$ nach Cohen (1988).....   | 37 |
| <b>Tabelle 3:</b> Deskriptive Statistiken der HR-Messungen in <i>bpm</i> für alle fünf Phasen und für den Gesamtverlauf.....                  | 38 |
| <b>Tabelle 4:</b> Deskriptive Statistiken der z-standardisierten HR-Messungen in Einheit der Standardabweichung für alle fünf Phasen.....     | 39 |
| <b>Tabelle 5:</b> Kontrastergebnisse der Lehrphase im Vergleich zu Pre-, Postlehr-, Interview- und Endphase.....                              | 42 |
| <b>Tabelle 6:</b> Effektstärke Cohen's $d$ der HR-Mittelwertunterschiede zwischen Lehr-Phase und den verbleibenden Phasen.....                | 43 |
| <b>Tabelle 7:</b> SPSS-Output der Tests auf Normalverteilung .....  | 65 |
| <b>Tabelle 8:</b> SPSS-Output des Mauchly-Tests auf Sphärizität.....  | 66 |
| <b>Tabelle 9:</b> SPSS-Output der Tests der Innersubjekteffekte der einfaktoriellen Varianzanalyse mit Messwiederholung .....                 | 67 |
| <b>Tabelle 10:</b> Effektstärke $d$ der betrachteten Mittelwertunterschiede und zugehörige Werte für $U_3$ nach Cohen (1988) in Prozent ..... | 67 |

## Abkürzungs- & Symbolverzeichnis

|            |   |
|------------|---|
| ANOVA      | Varianzanalyse (engl. analysis of variance)           |
| <i>bpm</i> | beats per minute (Schläge pro Minute, Einheit der HR) |
| EKG        | Elektrokardiogramm                                    |
| EMU        | Evidenzbasierte Methoden der Unterrichtsdiagnostik    |
| <i>d</i>   | Cohen´s <i>d</i> , Maß der Effektstärke               |
| HR         | Herzrate  |
| HRV        | Herzratenvariabilität                                 |
| <i>M</i>   | Mittelwert  |
| <i>N</i>   | Stichprobenumfang                                     |
| <i>p</i>   | Signifikanzniveau                                     |
| PPG        | Photoplethysmographie                                 |
| <i>SD</i>  | Standardabweichung (engl. standard deviation)         |
| <i>SE</i>  | Standardfehler (engl. standard error)                 |
| SJT        | Situational Judgement Test                            |
| SPSS       | Software für statistische Datenanalyse                |

# 1 Einleitung

*„Die Abwesenheit von Stress ist Tod – nur Tote haben keinen Stress“*

Hans Selye

Auch wenn dieses Zitat von Hans Selye, dem Urvater der Stressforschung, bereits aus der Mitte des 20. Jahrhunderts stammt, könnte es den aktuellen Zeitgeist nicht treffender beschreiben. Im Zuge des technischen Fortschritts, der sowohl neue Arbeitsbedingungen als auch neue Arbeitsanforderungen mit sich bringt, entwickelte sich Stress zu einem Phänomen, welches aus unserer modernen Gesellschaft nicht mehr wegzudenken ist (Günthner, 2022). Immer mehr Menschen fühlen sich subjektiv gestresst, so geht es auch mehr als 80% der Personen, die im deutschsprachigen Raum leben (Heinrichs et al., 2015). Auch die Weltgesundheitsorganisation WHO sieht in Stress eine der größten Gefahren für die Gesundheit der Menschen im 21. Jahrhundert (Gangl, 2009). Stress am Arbeitsplatz spielt dabei eine wichtige Rolle, weshalb bestimmte Berufsgruppen stärker von Stress gefährdet sind als andere. Insbesondere Lehrkräfte sind im schulischen Umfeld einer Vielzahl an Stressoren ausgesetzt.

Das wissenschaftliche Interesse im Rahmen der Stressforschung liegt seit gewisser Zeit auf der Verwendbarkeit und Messbarkeit physiologischer Parameter als Indikator oder Prädiktor von Stress. In Studien kommen dafür oft teure, medizinische Geräte zum Einsatz, welche sich nicht für den alltäglichen Gebrauch eignen. Doch die Messung diverser Vitalparameter ist aufgrund des technischen Fortschritts auch für jedermann mittels Smartwatches oder Fitnesstrackern möglich. Dabei zeichnen sich Smartwatches sowohl durch ihre geringen Anschaffungskosten als auch ihren Tragekomfort aus und bieten viele verschiedene Funktionen wie beispielsweise die Überwachung der Herzrate (HR). Im Rahmen der vorliegenden Arbeit soll überprüft werden, inwiefern sich eine Smartwatch der Marke Fitbit dazu eignet, statistisch signifikante Mittelwertunterschiede der HR von Lehrkräften abzubilden. Zu diesem Zweck durchlaufen Lehrkräfte in einer Laborstudie fünf Phasen unterschiedlicher Beanspruchung, die jeweils zehn Minuten andauern. Die sogenannte Lehr-Phase stellt dabei eine simulierte Unterrichtssituation mit Störungen dar und ist objektiv betrachtet die Phase mit der höchsten Stressbelastung für die Lehrpersonen. Die weiteren vier Phasen sind vor und nach der Lehr-Phase zu verorten. Anhand der Ergebnisse über die Abbildbarkeit von HR-Mittelwertunterschieden kann zudem resümiert werden, ob

Smartwatches auch als Teil des individuellen Stressmanagements von Lehrpersonen im Arbeitsalltag Anwendung finden können.

Im folgenden Kapitel dieser Arbeit wird auf die theoretischen Grundlagen von Stress und Stressforschung im Allgemeinen eingegangen und außerdem Stress im Lehrberuf detailliert betrachtet. Des Weiteren wird in diesem Rahmen die Bedeutung von Unterschieden der HR dargelegt sowie die Möglichkeit erörtert, Smartwatches für die Messung der HR zu verwenden. Dabei werden bisherige Untersuchungen beleuchtet, welche Smartwatches zur Erhebung physiologischer Parameter verwendet haben. Das darauffolgende dritte Kapitel befasst sich mit der Beschreibung der Stichprobe, dem Ablauf und der Durchführung der Studie sowie der Methodik der Datenauswertung. Die Ergebnisse der statistischen Datenauswertung werden in Kapitel 4 dargelegt, welche im fünften Kapitel vor dem Hintergrund bestehender Forschungsergebnisse diskutiert werden. In diesem Zug findet auch eine Einordnung der Studie hinsichtlich ihrer Vorzüge und Limitationen statt. Im abschließenden Kapitel werden auf Grundlage der Diskussion Möglichkeiten vorgestellt, wie die gewonnenen Erkenntnisse in die Ausbildung von Lehrkräften einfließen können, und ein abschließendes Fazit gezogen.



## **2 Theoretischer Hintergrund**

Im folgenden Kapitel werden die theoretischen Grundlagen dargelegt, wobei auch auf den aktuellen Stand der Forschung eingegangen wird. Zunächst werden verschiedene Definitionen von Stress erörtert und diverse Stressmodelle beschrieben. Im Anschluss werden die physischen und psychischen Folgen thematisiert, die langanhaltende Stressexposition mit sich bringen kann. Außerdem wird die Bedeutung von Stress und Belastung im Lehrberuf sowohl theorie- als auch forschungsbasiert beleuchtet und die Möglichkeit der HR-Messung mittels moderner Smartwatches ausgeführt. Den Abschluss des Kapitels bildet die Herleitung der Fragestellung und der Hypothese vor dem Hintergrund der vorherigen Unterkapitel.

### **2.1 Definitionen und Grundlagen von Stress**

Stress ist in der heutigen Lebenskultur ein viel verwendeter Begriff, der laut van Dick und Stegmann (2007) dabei oft als Synonym für andere Begriffe, aber auch inkorrekt gebraucht wird. Anforderungen, welche von der Umwelt an eine Person gestellt werden, werden vom Standpunkt der Wissenschaft als Stressoren oder Belastungen bezeichnet. Führt diese hohe Belastung zu körperlichen Symptomen, werden diese Symptome von der Wissenschaft als Stressreaktionen definiert, gleichzeitig handelt es sich um Beanspruchung, wenn die Umweltanforderungen als belastend wahrgenommen werden, da die individuellen körperlichen und geistigen Ressourcen nicht für deren Bewältigung ausreichen (Van Dick & Stegmann, 2007). Im englischsprachigen Raum handelt es sich bei Stress um einen physikalischen Begriff, der die auf einen Gegenstand einwirkende mechanische Spannung beschreibt (Heinrichs et al., 2015).

Insbesondere aufgrund der vielfältigen, uneindeutigen und sich überschneidenden Verwendung des Begriffs *Stress* fällt es der Forschung schwer, eine eindeutige Definition festzulegen. Nach Heinrichs et al. (2015) muss ein multidimensionaler Ansatz verwendet werden, um alle Verknüpfungen zwischen Stressoren und deren subjektiven Einordnungen einer Person, den objektiv feststellbaren Merkmalen des Stressors sowie den physiologischen Reaktionsmustern zu definieren und zu analysieren. Demzufolge existiert eine Vielzahl von Definitionen, die der jeweiligen Zeit und dem gültigen Stand der Wissenschaft Rechnung tragen und damit vor allem eine Leitlinie für das diagnostische Verständnis von Stress bilden. Heinrichs et al. (2015) legen Stress eine

operationale Definition zu Grunde, welche davon ausgeht, dass Stress das Resultat der Bedrohung der psychologischen, aber auch der physiologischen Integrität eines Individuums darstellt. Dieser zieht verschiedene Reaktionen nach sich, die kognitiver, behavioraler, physiologischer oder auch emotionaler Natur sein können. Hierbei betonen Heinrichs et al. (2015) insbesondere, dass es irrelevant ist, ob die Einordnung des Stressors als potentielle Bedrohung auf subjektiven oder objektiven Maßstäben beruht. Dabei beeinflusst die Einschätzung der verfügbaren Bewältigungsressourcen im Vergleich zur Einschätzung des Ausmaßes der Bedrohung auch die Intensität der individuellen Stressregulation durch oben genannte Reaktionen (Heinrichs et al., 2015). Dies beschreibt ebenso auch Günthner (2022) als ein Ungleichgewicht zwischen Anforderungen und Bewältigungsmöglichkeiten, welches eine Einschränkung der Funktionalität und Stabilität des Individuums darstellt.

Aufgrund der weitläufigen Definition und den vielfältig auftretenden Formen von Stress, existieren auch viele Stresstheorien und dazugehörige Erklärungsmodelle, welche sich mit unterschiedlichen Phänomenen und Teilaspekten von Stress befassen. Im Folgenden soll eine Auswahl an Stressmodellen vorgestellt und diese in ihren Grundzügen dargelegt werden. Dabei soll diese Auswahl als Überblick dienen und erhebt keinen Anspruch auf Vollständigkeit.

Das *Allgemeine Anpassungssyndrom* nach Selye fokussiert sich auf eine Reihe von Prozessen, die in Situationen von Stress ablaufen. Er betrachtet Stress aus einer biologischen Perspektive und versucht damit, die mit Stress oftmals einhergehende körperliche Erschöpfung näher zu beschreiben (Günthner, 2022; Heinrichs et al., 2015). Anhand des Modells lässt sich eine Stresssituation respektive ein Stressprozess in drei aufeinanderfolgende Phasen unterteilen, namentlich die Alarmreaktion, die Widerstandsphase und die Erschöpfungsphase (Günthner, 2022; Heinrichs et al., 2015). Während der Alarmreaktion reagiert der Organismus direkt auf den Stressor und versucht bereits durch biologische Mechanismen, wie der Ausschüttung von Glukose in verschiedenen Gewebsschichten, auf diese einzugehen (Heinrichs et al., 2015). Die eigentliche Regulation findet in der Widerstandsphase statt, in der insbesondere metabolische und endokrine Mechanismen eingesetzt werden, um die vorangegangenen Reaktionen auf den Stressor zu kompensieren, was auch die vermehrte Freisetzung von Energie mit sich zieht (Heinrichs et al., 2015). Die finale Erschöpfungsphase beschreibt die Beeinträchtigung von Teilen des Organismus, welche auch zum Erliegen grundlegender Funktionen führen kann (Heinrichs et al., 2015). Dieses Modell eignet sich

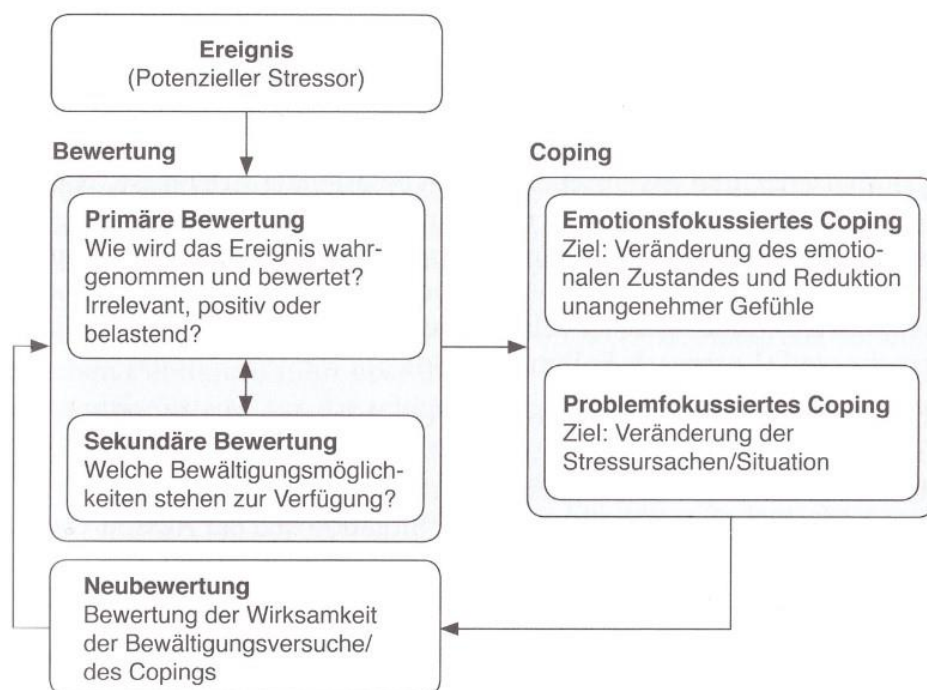
zur heuristischen Erklärung von Reaktionen des Körpers als Anpassung an Stress, ist allerdings in der aktuellen Stressforschung von geringerer Bedeutung (Günthner, 2022; Heinrichs et al., 2015).

Bei dem *Allostatic Load-Modell* nach McEwen handelt es sich ebenfalls um ein heuristisches Erklärungsmodell, welches anhand eines psychobiologischen Ansatzes versucht, den Entstehungsprozess von Erkrankungen zu beschreiben, die auf Stress zurückzuführen sind (Heinrichs et al., 2015). Es setzt dabei am zuvor erläuterten Modell von Selye an. Bei der Allostase handelt es sich sowohl um behaviorale als auch physiologische Mechanismen, welche das Anpassungsvermögen des Organismus an Belastungsfaktoren der Umwelt stärken (Heinrichs et al., 2015). So schildern Heinrichs et al. (2015), dass unter anderem auch das Ausschütten von Stresshormonen, was durch immensen psychischen Stress geschehen kann, zur Allostase zählt. Wird das allostatic System in einem unangebrachten Maß in Anspruch genommen, führt dies zu allostatic Belastung, welche sich in die vier Kategorien wiederholte akute Stressoren, zu geringe Reaktion auf Belastung, verlängerte Stressreaktion und fehlende Anpassung an wiederholt auftretende gleiche Stressoren einteilen lässt (Heinrichs et al., 2015). Dies zeigt, dass sowohl eine Überbeanspruchung als auch eine Unterbeanspruchung von Regulationsmechanismen zu nachhaltigen Schäden des Organismus führen können.

Eine Grundlage vieler moderner Stresstheorien bildet der psychologische Ansatz des *transaktionalen Stressmodells* nach Lazarus und Folkman (Günthner, 2022). Im Gegensatz zu vielen anderen Modellen berücksichtigt dieses bereits das Auftreten von Gefühlen und den Umgang mit diesen sowie Prozesse zur Stressbewältigung – dem Coping (Günthner, 2022; Seefeldt, 2000). Im Vordergrund stehen insbesondere die kognitiven Vorgänge bei der Bewertung von Stresssituationen, da sowohl die Situation als solche als auch die persönlichen Fähigkeiten zur Bewältigung dieser einer Bewertung unterliegen (Günthner, 2022; Heinrichs et al., 2015). Lazarus und Folkman unterscheiden dabei zwischen einer primären und sekundären Bewertung. Ein auftretender Stressor wird einer primären Bewertung hinsichtlich seiner Bedrohlichkeit und Relevanz unterzogen und die Situation dabei bezüglich ihrer Merkmale eingeschätzt (Heinrichs et al., 2015, Seefeldt, 2000). Die sekundäre Bewertung evaluiert die individuellen Möglichkeiten, Ressourcen und die Selbstwirksamkeitserwartung, welche zur Bewältigung zur Verfügung stehen (Seefeldt, 2000). Die Bezeichnung der beiden Bewertungsebenen entspricht nicht einer Priorisierung der Wichtigkeit oder einer zeitlichen Ablauffolge, beide sind von äquivalenter Bedeutung (Heinrichs et al., 2015). Neben den Bewertungen

spielt auch die Bewältigung eine wichtige Rolle im transaktionalen Stressmodell, wobei sich diese in problemfokussierte und emotionsfokussierte Bewältigung unterteilen lässt (Heinrichs et al., 2015). Zweitere fokussiert sich vor allem darauf, die negativen Folgen psychologischer und physiologischer Natur zu reduzieren, wohingegen Erstere nach einer aktiven Veränderung der auslösenden Stresssituation strebt (Heinrichs et al., 2015). Anschließend an die Bewältigungsversuche findet eine Neubewertung der Situation statt, welche entweder eine Erhaltung des Stresses mit sich bringt oder zu einer Entlastung führt, wenn die Bewältigungsversuche in Bezug auf die Situation als positiv und wirksam betrachtet werden (Heinrichs et al., 2015; Seefeldt, 2000). Eine Darstellung der beschriebenen Zusammenhänge der verschiedenen Bewertungs- und Bewältigungsschritte ist Abbildung 1 zu entnehmen.

Weitere Theorien beziehen sich auf die zur Stressbewältigung zur Verfügung stehenden Ressourcen, wie die *Theorie der Ressourcenerhaltung* nach Hobfoll oder auch das *Modell der Widerstandsfähigkeit* nach Kabasa (Günthner, 2022). Andere Modelle fokussieren sich wiederum auf Belastung im Arbeitskontext, so das *Effort-Reward-Imbalance Modell* oder auch das *Anforderungs-Kontroll-Modell* nach Karasek (Heinrichs et al., 2015).



**Abbildung 1:** Darstellung der verschiedenen Prozesse und deren Verknüpfungen innerhalb des transaktionalen Stressmodells nach Lazarus und Folkman (Heinrichs et al., 2015, S. 25).

## **2.2 Gesundheitliche Risiken langanhaltender Stressexposition**

In der Entwicklungsgeschichte des Menschen dienten physiologische Reaktionen auf Stress zur schnellen körperlichen Antwort auf potenziell lebensbedrohliche Situationen. Dabei sollte durch die Ausschüttung von Hormonen, die Aktivierung von Energiereserven, die Steigerung der Konzentration und Herzfrequenz sowie die zeitgleich stattfindende Reduktion anderer Körperfunktionen das Überleben gesichert werden (Kypta, 2006). Auf diese Weise sollten Kampf oder Flucht ermöglicht werden. In der modernen Welt sind derartige körperliche Reaktionen nicht von Nöten, dennoch laufen die genannten physiologischen Mechanismen ab, welche auf Dauer zu verschiedensten physischen und psychischen Symptomen und in Folge dessen zu gesundheitlichen Beeinträchtigungen führen können (Kypta, 2006).

Insbesondere die dauerhaft hohe Ausschüttung diverser Stresshormone resultiert in vielen körperlichen aber auch psychischen Symptomen, welche sich zu schwerwiegenden Krankheiten entwickeln können. Da die Bandbreite an Symptomen sehr groß ist, wird nur eine Auswahl dieser wiedergegeben.

Zu den körperlichen Symptomen gehören unter anderem Magenschmerzen, Kopfschmerzen, vermehrtes Auftreten von Infektionskrankheiten wie Erkältungen, Hautausschläge oder verschiedenste kardiale Beschwerden (Kypata, 2006). Studien belegen außerdem den Zusammenhang zwischen chronischem Stress und einem erhöhten Risiko an bestimmten Krankheiten zu erkranken oder diese in ihrem Verlauf zu beschleunigen. So bestätigten Thaker et al. (2007), dass langanhaltende Stressexposition zu einem gesteigerten Tumorwachstum bei Eierstockkrebs beiträgt. Vergleichbare Erkenntnisse konnten auch in Bezug auf Bauchspeicheldrüsenkarzinome festgestellt werden, welche unter chronischem Stress sowohl schneller wachsen als auch invasiver sind (Kim-Fuchs et al., 2014). Ebenso stellt das Reizdarmsyndrom eine der Erkrankungen dar, die auf Stress zurückzuführen ist, und sich aufgrund der psychischen Stressbelastung körperlich manifestiert, weshalb dieses Syndrom als psychosomatische Erkrankung eingeordnet wird (Talley & Spiller, 2002).

Zu den psychischen Symptomen von Stress zählen unter anderem vermehrtes Grübeln, Stimmungsschwankungen, Panikattacken, Reizbarkeit oder auch der Verlust des Selbstvertrauens (Kypta, 2006). Es ist bekannt, dass dauerhafter Stress begünstigend auf die Entstehung diverser psychischer Erkrankungen wirkt, so auch bei Depressionen, was anhand einer Vielzahl von Studien belegt ist (Hammen, 2005). In den meisten Fällen

ist neben der medizinischen Behandlung der potenziellen körperlichen Folgen sowohl bei Auftreten physischer als auch psychischer Stresssymptome ein verhaltenstherapeutischer Ansatz indiziert, welcher auf einen verbesserten Umgang mit Stress abzielt (Kypta, 2006).

Um ein eigenständiges Krankheitsbild handelt es sich bei Burnout, dessen primärer Auslöser anhaltender Stress ist und wörtlich das *Ausbrennen* in folgedessen beschreibt. Dieses Ausgebrannt sein spiegelt vordergründig die subjektive Wahrnehmung Betroffener wider (Koch et al., 2015). Ähnlich wie bei Stress existiert auch für Burnout keine eindeutige Definition, da über 120 verschiedene Symptome mit diesem Syndrom in Verbindung stehen, jedoch ist es notwendig, dass sowohl verringerte Leistungszufriedenheit als auch Erschöpfung und Depersonalisierung als Symptome geschlossen vorliegen müssen, um eine Burnout Diagnose stellen zu können (Koch et al., 2015; Kypta, 2006). Es ist anzumerken, dass alle bekannten Stresssymptome auch im Rahmen einer Burnout-Erkrankung auftreten können, wobei sich die für Burnout charakteristische Erschöpfung erst als letztes Symptom der Belastung zeigt (Kypta, 2006). Ausschlaggebend sind laut Kypta (2006) die Anzahl der Stressoren sowie der zeitliche Rahmen und die Intensität, mit denen eine Person konfrontiert ist. Da die Belastungsgrenzen individuell und oftmals nur schwer wahrnehmbar sind, wird das Übermaß an Stressoren vermehrt ignoriert oder heruntergespielt, was Betroffenen zusätzlich Energie und Willen nimmt, die Situation verschlimmert und folglich zur vollständigen Kraftlosigkeit und Erschöpfung des gesamten Körpers führt (Kypta, 2006). Burnout wird insbesondere als Überlastungsreaktion betrachtet, bei der die beruflichen Anforderungen eine zentrale Rolle spielen und die das Risiko für psychische Erkrankungen steigert (Koch et al., 2015). Psychologische und psychotherapeutische Maßnahmen beziehungsweise Rehabilitationen, welche meist im Kontext einer Gruppe vorgenommen werden, bilden die Grundlage der Behandlung des Burnout-Syndroms, indem der Umgang mit Stress erlernt wird und bereits existierende Bewältigungsmechanismen gestärkt werden (Koch et al., 2015).

## **2.3 Stress und Belastungen im Lehrberuf**

Im Folgenden wird spezifisch auf Stress und die damit verbundenen Belastungen im Lehrberuf eingegangen. Dabei werden insbesondere die Ursachen für Stress im Lehrberuf,

Definitionen und Modelle von Lehrer:innenstress sowie berufsscharakteristische Stressoren thematisiert. Außerdem werden neben den Richtungen auch der aktuelle Stand der Stressforschung im schulischen Kontext erörtert.

### **2.3.1 Grundlagen und Modelle**

Im Laufe der 1970er Jahre rückte der Stress, den Lehrkräfte bei der Ausübung ihres Berufes erfahren, vermehrt in das Zentrum des Forschungsinteresses (Kyriacou, 2001). Eine der ersten Definitionen von Lehrer:innenstress nach Kyriacou (2001) beschreibt diesen als die Erfahrung von unangenehmen, negativen Emotionen, die aus diversen Aspekten des Berufs als Lehrperson resultieren. Bei den genannten Gefühlen handelt es sich vordergründig um Nervosität, Depression, Frustration sowie Angst, aber auch Wut (Kyriacou, 2001). Weitere Definitionen von Lehrer:innenstress interpretieren diesen als das Niveau an Druck und Ansprüchen, die der einzelnen Person auferlegt werden oder fokussieren sich auf die physische und emotionale Erschöpfung (Kyriacou, 2001).

Kyriacou (2001) zufolge lässt sich der Lehrberuf anhand des Forschungsstandes zu Beginn des 21. Jahrhunderts, der hauptsächlich auf selbsteinschätzenden Umfragen basiert, als Hochstressberufsstand einordnen. Dies lässt sich auf die Grundlagen beziehungsweise strukturellen Charakteristika zurückführen, auf welchen der Lehrberuf aufbaut und dabei das Ziel des schlichten Einkommenserwerbs dem Ziel des ideell pädagogischen Handelns gegenübersteht (Rothland & Terhart, 2007). Zu diesen strukturellen Bedingungen, die eine Beanspruchung und Belastung darstellen, gehört der Expertise von Rothland und Terhart (2007) zufolge insbesondere auch der zweigeteilte Arbeitsplatz zwischen Schule und dem privaten Heim, welcher von der Öffentlichkeit wenig Berücksichtigung findet. Neben dem großen Maß an freier Einteilung der Arbeitszeit wird allerdings zugleich auch die Trennung von Arbeits- und Freizeit erschwert (Rothland & Terhart, 2007). Eine direkte Folge daraus zeigt sich in den undefinierten Grenzen des Aufgaben- und Zeiteinsatzes einer Lehrkraft, welches nach den Schilderungen von Rothland und Terhart (2007) im individuellen Ermessen der Lehrperson liegt. Außerdem stellt die Konkurrenz zwischen einer Reglementierung durch Lehrpläne und der pädagogischen Ermessensfreiheit der Lehrkraft ein weiteres strukturelles Charakteristikum dar, welches zur Belastung im Lehrberuf führt (Rothland & Terhart, 2007). Die eigene geringe Erfolgswahrnehmung, der Mangel an Anerkennung durch Beförderungen sowie durch die Gesellschaft sind nach Rothland und Terhart

(2007) ebenfalls zu diesen Charakteristika zu zählen. Auch Boyle et al. (1995) konnten belegen, dass der Mangel an Anerkennung und Wertschätzung in der Gesellschaft entscheidend zum Stressempfinden der Lehrkräfte beitragen.

Eine besondere Belastung erfahren Lehrkräfte zudem anhand verschiedener Erwartungen, die von unterschiedlichen Personengruppen, wie Schüler:innen, Eltern oder auch Vorgesetzten, an die Rolle der Lehrperson gestellt werden und folglich auch mit verschiedenen Aufgaben verbunden sind, wobei die Erwartungen teilweise sehr konträr und widersprüchlich sind (Rothland & Terhart, 2007). Hinzu kommen außerdem die von Helsper (1996) angeführten Antinomien des Lehrer:innenhandelns, welche Spannungsverhältnisse darstellen, die aus Gegensätzlichkeiten innerhalb verschiedener Bereiche des Lehrberufs resultieren, mit denen Lehrkräfte im Zusammenhang mit Entscheidungsfindungen konfrontiert sind und so eine Form von Unsicherheit mit sich bringen. Dazu zählen unter anderem die Gegenpole Einheitlichkeit und Differenz oder auch Autonomie und Heteronomie (Helsper, 1996). Rothland und Terhart (2007) gehen davon aus, dass die geschilderten strukturellen Schwierigkeiten und Problemstellungen bereits unabhängig vom tatsächlichen Arbeitsalltag ein breites Spektrum an potentiellen Belastungen mit sich bringen, die spezifisch für den Lehrberuf sind.

Zu vier Modellen und Theorien, die sich im Zusammenhang mit der Beschreibung und Untersuchung von Stressbelastung im Lehrberuf bewährt haben, gehören das *Job Characteristics Model* nach Hackman und Oldham, die Anwendung des transaktionalen Stressmodells auf den Lehrberuf nach Kyriacou und Sutcliffe, die *Handlungsregulationstheorie* nach Oesterreich, sowie das *Rahmenmodell der Belastung und Beanspruchung* nach Rudow (Van Dick & Stegmann, 2007). Es ist anzumerken, dass diese Aufzählung nur einen Teil der gängigen Modelle für Lehrer:innenstress beinhaltet und keinen Anspruch auf Vollständigkeit erhebt. Auch die folgenden Erläuterungen dienen ausschließlich als grobe Einordnung dieser.

Bei letztgenanntem Modell nach Rudow wird zwischen subjektiver und objektiver Belastung unterschieden, wobei Arbeitsaufgaben und -bedingungen als objektive Belastung klassifiziert werden, die „[...] durch den Prozess der Widerspiegelung [...] in die subjektive Belastung überführt [werden]“ (Van Dick & Stegmann, 2007, S. 35). Dieser Prozess wird insbesondere durch die Persönlichkeit und die individuellen Fähigkeiten der Lehrperson beeinflusst (Van Dick & Stegmann, 2007). Beanspruchungsreaktionen und die daran anschließenden Folgen bilden das Resultat der subjektiven Belastung (Van Dick & Stegmann, 2007). Dementsprechend wird jede



Lehrperson anders auf dieselben objektiven Belastungen reagieren, da sie sich aufgrund ihrer Persönlichkeit und individuellen Handlungsvoraussetzungen auch mit unterschiedlichen Beanspruchungsreaktionen konfrontiert sieht. Befinden sich Arbeitsaufgaben und -bedingungen im Einklang mit den Handlungskompetenzen und Ressourcen, die einer Lehrkraft zur Verfügung stehen, führt dies zu einer langfristigen Steigerung des Wohlbefindens sowie der psychischen und physischen Gesundheit, wohingegen ein Ungleichgewicht in chronischen Erkrankungen resultieren kann (Van Dick & Stegmann, 2007). Nach der Einschätzung von van Dick (2006) hat das Rahmenmodell nach Rudow mit der Vereinigung von Lehrer:innenmerkmalen und Tätigkeitsmerkmalen die Vorstellung von Kyriacou und Sutcliff umfassend erweitert. Diese gingen vom transaktionalen Stressmodell nach Lazarus aus und modifizierten es für die Anwendung auf den Lehrberuf (Van Dick & Stegmann, 2007). Dabei übernehmen objektive Charakteristika des Lehrberufs die Rolle von potentiellen Stressoren, die in physische und psychische Stressoren unterteilt werden und durch die Phasen der Bewertung und Bewältigung in Form von Stress Ausdruck finden können (Van Dick, 2006). Ist dieser Stress langanhaltend kann sich dieser, wie es auch Rudow in seinem Rahmenmodell übernommen hat, durch chronische Symptome manifestieren, welche psychischer, behavioraler oder physischer Natur sein können (Van Dick, 2006).

Mit dem *Job Characteristics Model* nach Hackman und Oldham wird außerdem ein Modell der Arbeitsmotivation zur näheren Betrachtung von Stress im Lehrberuf verwendet (Van Dick & Stegmann, 2007). Es basiert auf den folgenden drei Säulen Tätigkeitsmerkmale, psychologische Erlebniszustände und Auswirkungen der Arbeit, anhand derer Zusammenspiel eine Aussage über die Wirksamkeit von Veränderungen im Bereich der Tätigkeiten auf die beiden anderen Merkmale getroffen werden kann und soll (Van Dick & Stegmann, 2007). Dabei wird zusätzlich von Moderatorvariablen, wie individuelle Fähigkeiten und das Verlangen nach persönlicher Entfaltung, ausgegangen, die die Tätigkeitsmerkmale stark mit der Arbeitsmotivation im Beruf verknüpfen (Van Dick & Stegmann, 2007).

Bei Lehrer:innen handelt es sich um eine von Burnout besonders gefährdete Berufsgruppe, weshalb das Interesse an der Erforschung der Lehrer:innenbelastung groß ist (Krause & Dorsemagen, 2007). Die vermeintlichen Hauptquellen von Lehrer:innenstress respektive Belastung sind Kyriacou (2001) zu Folge unter anderem im Zeitdruck und im Arbeitspensum, dem Umgang mit Kolleg:innen, den vermeintlich schlechten Arbeitsbedingungen, der mangelnden Motivation der Schüler:innen und im

Aufrechterhalten der Disziplin zu verorten. Ein umfassenderes Konzept entwickelte Rudow (2000), welches Belastungsfaktoren den vier verschiedenen Kategorien kulturelle und soziale Bedingungen sowie Arbeitsaufgaben und Arbeitsumweltbedingungen zuordnet. Anhand dieser beiden Konzepte von Lehrer:innenstress wird exemplarisch verdeutlicht, dass dieser eine Vielzahl unterschiedlicher Quellen hat. Losgelöst von den genannten Belastungsquellen, beschäftigt sich ein weiterer Teil der Lehrer:innenstressforschung mit den Beanspruchungen von Lehrpersonen, zu denen beispielsweise emotionale Erschöpfung zählt (Krause & Dorsemagen, 2007).

Ein strukturelles Modell, welches verschiedene Bereiche der Forschung zu Lehrer:innenbelastung mit berücksichtigt und sich durch empirische Überprüfungen als sehr treffend erwiesen hat, wurde von Boyle et al. (1995) aufgestellt. Dabei erfolgt eine Unterteilung des Lehrer:innenstress in fünf verschiedene Dimensionen: Fehlverhalten von Schüler:innen, Arbeitspensum, schwierige zwischenmenschliche Verhältnisse zu Personengruppen innerhalb des Lehrberufs<sup>1</sup>, strukturelle Schwierigkeiten im Lehrberuf und die Stellung des Lehrberufs in der Gesellschaft (Boyle et al., 1995). Insbesondere das Fehlverhalten von Schüler:innen in Form von Disziplinlosigkeit, Unaufmerksamkeit sowie Unterrichtsstörungen wird von Lehrkräften als einer der bestimmenden Stressoren wahrgenommen (Aldrup et al., 2018; Aloe et al., 2014). Es lässt sich nach der Auffassung von Lohmann (2014) zwischen den folgenden vier Kategorien von störendem Schüler:innenverhalten unterscheiden: mangelnder Lerneifer, der sich beispielsweise durch Unaufmerksamkeit zeigt, verbale Störungen, wie Zwischenrufe oder Schwatzen, motorische Unruhe, welche sich zum Beispiel durch Zappeln auszeichnet, und aggressives Verhalten. Welche dieser Arten von Unterrichtsstörungen am stärksten zur Belastung von Lehrkräften führt, unterliegt dabei dem subjektiven Empfinden jeder einzelnen Lehrperson (Lohmann, 2014).

### **2.3.2 Stressforschung im Unterrichtskontext**

Das Anstellen neuer Untersuchungen ist nach Kyriacou (2001) wichtig, um die Datenlage in Bezug auf Lehrer:innenstress zu aktualisieren und Veränderungen in den Belastungen festzustellen, vor allem auch im Zusammenhang mit und in Folge von Schul- und

---

<sup>1</sup> Personengruppen innerhalb des Lehrberufs beziehen sich dabei auf Schüler:innen, Eltern, Vorgesetzte und Kolleg:innen.

Bildungsreformen. Dabei steht auch im Forschungsinteresse, inwiefern das Lehrpersonal mit derartigen strukturellen Veränderungen umgeht und in welchem Maß die Anpassung an diese eine weitere Belastungsquelle darstellt (Kyriacou, 2001). Außerdem sieht Kyriacou (2001) weiteren Bedarf in der detaillierteren Untersuchung beziehungsweise Unterscheidung der Stressprozesse, insbesondere zwischen den überhöhten Anforderungen und den Zweifeln am Selbstbild. Einen weiteren bedeutsamen Anhaltspunkt stellt dabei die Untersuchung dar, welchen Einfluss das zwischenmenschliche Klima im Klassenzimmer sowie das Lehrer:innen-Schüler:innen Verhältnis auf das individuelle Empfinden von Lehrer:innenstress nehmen (Kyriacou, 2001).

Diesen Ansatz verfolgten Aldrup et al. (2018) indem sie überprüften, welche Rolle die Lehrer:innen-Schüler:innen Beziehung für das Wohlbefinden von Lehrpersonen einnimmt und stellten dabei die Hypothese auf, dass das Verhältnis als Mediator fungiert. In diesem Kontext ergab die Befragung von Lehrkräften zum Fehlverhalten der Schüler:innen während des Unterrichts sowie zum eigenen Wohlbefinden in Abhängigkeit von Arbeitsenthusiasmus und emotionaler Erschöpfung, dass sinkender Arbeitsenthusiasmus und steigende emotionale Erschöpfung signifikant mit dem wahrgenommenen beziehungsweise bewerteten Fehlverhalten der Schüler:innen zusammenhängt (Aldrup et al., 2018). Außerdem verifizierten Aldrup et al. (2018), dass die Lehrer:innen-Schüler:innen Beziehung positiv mit dem Wohlbefinden von Lehrpersonen verknüpft ist und zudem auch zwischen Schüler:innenfehlverhalten und Arbeitsenthusiasmus vermittelt. Folglich resultiert ein negatives Lehrer:innen-Schüler:innen Verhältnis in einem niedrigeren Arbeitsenthusiasmus und gesteigerter emotionaler Erschöpfung und umgekehrt. Insbesondere vermittelt die Lehrer:innen-Schüler:innen Beziehung zwischen geringem Arbeitsenthusiasmus und dem Fehlverhalten der Schüler:innen (Aldrup et al., 2018). Zudem stellten Aldrup et al. (2018) im Rahmen ihrer Untersuchungen fest, dass Lehrkräfte, die mehr Fehlverhalten wahrnehmen, auch von einem verringerten Arbeitsenthusiasmus und gesteigerter emotionaler Erschöpfung berichteten. Dies lässt den Schluss zu, dass ein positives Verhältnis zwischen Lehrer:innen und Schüler:innen maßgeblich für das berufliche Wohlbefinden von Lehrpersonen ist und damit auch das Empfinden von Stress und Belastung beeinflusst (Aldrup et al., 2018).

In der Stressforschung in Bezug auf den Lehrberuf liegt das Interesse insbesondere darauf, adäquate Indikatoren und quantitativ messbare Parameter für Lehrer:innenstress und Burnout zu finden (Fisher, 2010; Junker et al., 2021; Runge et al.,

2020). Vor diesem Hintergrund untersuchte Fisher (2010), inwiefern das Alter respektive die Lehrerfahrung sowie die Unzufriedenheit im Beruf bei Lehrpersonen mit einem gesteigerten Burnout und Stressrisiko einhergeht. Dabei wurden in der Studie drei Tests als Instrumente verwendet<sup>2</sup>, welche auf der Selbstauskunft der Lehrkräfte beruhten und einen Stress- und Zufriedenheitswert, einen Burnoutwert und eine Einschätzung der individuellen Bewältigungskompetenzen lieferten (Fisher, 2010). Anhand der Testergebnisse stellte sich heraus, dass der Stresswert keinen signifikanten Zusammenhang mit dem Alter beziehungsweise der Lehrerfahrung aufweist, allerdings zeigte sich auch, dass jüngere Lehrkräfte respektive Lehrkräfte mit einer geringeren Berufserfahrung einen erhöhten Burnoutwert aufweisen (Fisher, 2010). Basierend auf den Ergebnissen einer multiplen Regressionsanalyse, kommt Fisher (2010) zu dem Schluss, dass die Jahre an Berufserfahrung, Burnout und die Zufriedenheit im Lehrberuf Lehrer:innenstress statistisch signifikant vorhersagen und auch Stress im Lehrberuf und Burnout statistisch signifikante Prädiktoren für die berufliche Zufriedenheit von Lehrkräften sind.

Einen anderen Ansatz verfolgen sowohl Junker et al. (2021) als auch Runge et al. (2020), die in ihren jeweiligen Studien die HR als potentiellen physischen Indikator für Lehrer:innenstress wählten. Um das Verständnis über Stressoren innerhalb des Klassenzimmers zu erweitern und deren Auswirkungen auf Lehrpersonen und deren physischen Stress in Form der HR zu untersuchen, zeichneten Junker et al. (2021) die HR von Lehrpersonen innerhalb einer realen Unterrichtseinheit mit Hilfe eines 7-Kanal-EKG auf. Dabei war es auch Ziel der Untersuchungen, herauszufinden, inwiefern Hauptstressoren innerhalb des Klassenzimmers, die durch eine Vielzahl an Studien bereits als solche bestätigt waren, auch anhand der HR abbildbar sind. Im Rahmen der Auswertung fokussierte sich das Interesse auf Perioden, in den die HR mindestens zwei Standardabweichungen über der durchschnittlichen HR der Proband:innen lag. Junker et al. (2021) konnten anhand ihrer Studie nachweisen, dass Lehrer:innenstress, der durch Stressoren innerhalb des Unterrichts hervorgerufen wird, anhand eines Anstieges der HR quantitativ messbar ist und sich insbesondere Lehrer:innen bezogene Aktivitäten und

---

<sup>2</sup> Bei den Tests beziehungsweise Umfragen handelte es sich um die *Classroom appraisal of resources and demands* (CARD) Umfrage zur Bestimmung eines Stress- und Zufriedenheitswerts, das *Maslach burnout inventory* (MBI) zur Bestimmung eines Burnoutwerts und das *Preventive resources inventory* (PRI) zur Einschätzung der individuellen Bewältigungskompetenzen der Lehrkräfte.

mangelndes Engagement und Motivation der Schüler:innen durch einen Anstieg in der HR auszeichnen.

In der Untersuchung von Runge et al. (2020) bildete ebenfalls die HR einen von mehreren Parametern, die innerhalb eines Mixed Methods Forschungsdesigns verwendet wurden, welches eine Kombination aus verschiedenen quantitativen und qualitativen Forschungsmethoden beinhaltet, um den Stress bei Lehrpersonen auch anhand physischer Merkmale kenntlich zu machen. Neben der HR wurden auch der Schlaf und die Schrittzahl von Lehrkräften über einen Zeitraum von zwölf Wochen mittels eines Fitnesstrackers der Marke Fitbit aufgezeichnet. Zusätzlich wurden vor Beginn der Messungen anhand der gleichen Tests (CARD, PRI und MBI) wie auch in der Studie von Fisher (2010) der Stress- und Burnoutwert sowie die individuellen Bewältigungskompetenzen bestimmt (Runge et al., 2020). Während der zwölfwöchigen Studiendauer wurden die Proband:innen wöchentlich im Rahmen eines Interviews zu den Tagen befragt, die sich anhand der aufgezeichneten Daten als stressigster beziehungsweise entspanntester Tag ausmachen ließen. Neben der Erfassung der häufigsten Stressoren der Proband:innen, die in Verbindung mit ihrem Beruf als Lehrkraft standen, war das Hauptziel der Studie, herauszufinden, in welchem Maß erschwingliche Fitnesstracker und die auf diesem Wege bereitgestellten Parameter verwendet beziehungsweise herangezogen werden können, um Lehrer:innenstress zu messen (Runge et al., 2020). Die Untersuchungen von Runge et al. (2020) kamen zu dem Ergebnis, dass sich im Setting ihrer Studie Stress im Lehrberuf anhand der Vitalparameter, die Fitnesstracker messen, abbilden lässt. Insbesondere stellte sich heraus, dass die Kombination aus einer hohen Schrittzahl, einer hohen HR und kurzem Schlaf ein Indikator für Stress ist und schlechtes Schüler:innenverhalten der Stressor, der am häufigsten auftritt beziehungsweise wahrgenommen wird (Runge et al., 2020). Es ist anzumerken, dass die Ergebnisse der Studie von Runge et al. (2020) aufgrund der geringen Stichprobengröße von vier Proband:innen nicht auf den gesamten Lehrberuf übertragen werden können.

Das Forschungsinteresse an der HR als physiologischer, quantitativer Parameter für die Messung von Stress erstreckt sich neben den Lehrkräften auch auf die Schüler:innen. Im Kontext einer Studie von Kärner und Warwas (2018) wurde die HR von Berufsschüler:innen über einen Zeitraum von sechs Unterrichtsstunden als Indikator für physischen Stress mit Hilfe eines Brustgurtes erfasst. Zusätzlich wurde auch das situative Stressempfinden der Berufsschüler:innen während der Einheiten erhoben. Dabei stand unter anderem im Fokus der Untersuchung, festzustellen, ob sich in Situationen, in

denen das individuelle Stresserleben erhöht ist, auch körperliche Stresssymptome akut verstärken (Kärner & Warwas, 2018). Es zeigte sich, dass sowohl die HR als auch das Stressempfinden der Proband:innen signifikant ansteigen, wenn diese im Unterrichtsetting mit einer gesteigerten Anforderung konfrontiert werden. Somit hat sich die HR im gewählten Setting als adäquater Indikator für Stress erwiesen (Kärner & Warwas, 2018).

## 2.4 Herzratenunterschiede

Bei Stress beziehungsweise Belastung mentaler und physischer Natur handelt es sich um Faktoren, die die HR direkt beeinflussen und zu einem Anstieg dieser führen (Custodis et al., 2014). Die HR oder auch Herzfrequenz gibt die Anzahl an Herzschlägen innerhalb eines bestimmten Zeitintervalls an, dabei handelt es sich meist um Herzschläge pro Minute und folglich wird die HR in  $\text{min}^{-1}$  beziehungsweise *bpm* (beats per minute) angegeben (Hottenrott, 2007). Dabei hängt die menschliche HR stark von der Intensität der Belastung ab. In Ruhe liegt die durchschnittliche HR eines Menschen bei etwa 70 Schlägen pro Minute, die maximal abrufbare HR eines Erwachsenen beträgt circa 180 *bpm* und bei alltäglicher Belastung liegt die HR zwischen 50 und 100 *bpm* (Noble et. al, 2017). Die HR des Menschen unterliegt dabei einer Vielzahl an Einflussfaktoren zu denen unter anderem Geschlecht, Gewicht, Körpergröße, Alter und auch die körperliche Fitness zählen (Hottenrott, 2007; Noble et. al, 2017). In Anlehnung an die Trainings- und Belastungssteuerung im Sport, lassen sich verschiedene Stufen der Belastungsintensität definieren. Diese werden anhand der individuellen maximalen HR bestimmt, die eine Person erreichen kann (Hottenrott, 2007). Der Körper ist nach Definition nur einer sehr geringen Belastung ausgesetzt, wenn die HR unter momentaner Belastung zwischen 50 – 60 % der maximalen HR liegt. Erreicht die HR den Bereich zwischen 80 – 90 % der maximalen HR definiert dies hohe Belastung (Hottenrott, 2007). Liegt die HR über diesem Bereich, also über 90 % der maximalen HR, ist die Belastungsintensität sehr hoch, was einem Bereich entspricht, der zumeist durch physische Belastung im Rahmen von Hochleistungssport erreicht wird (Hottenrott, 2007). Eine Auflistung aller fünf Belastungsstufen ist Tabelle 1 zu entnehmen.

**Tabelle 1**

*Stufen der Belastungsintensität in Abhängigkeit der prozentualen maximalen HR nach Hottenrott (2007).*

| % der maximalen HR | Belastungsintensität |
|--------------------|----------------------|
| 50 – 60            | sehr gering          |
| 60 – 70            | gering               |
| 70 – 80            | mittel               |
| 80 – 90            | hoch                 |
| > 90               | sehr hoch            |

Die nominellen HR-Werte in *bpm*, die die einzelnen Belastungsstufen festlegen, sind dabei individuell, weil sie von der maximalen HR eines jeden Menschen abhängen, die unter anderem durch die bereits erwähnten Faktoren beeinflusst wird (Hottenrott, 2007). Folglich variiert die HR zwischen der Ruhe-HR und maximaler HR, wobei der genaue Wert oder Bereich von der Intensität der Belastung abhängt und auch mentale Belastung zu einer Veränderung in der HR-Charakteristik beiträgt (Taelman et al., 2010).

## **2.5 Herzfrequenzüberwachung mittels Smartwatch Wearables**

Seit den 1970er Jahren wurden vermehrt Bestrebungen angestellt, eine präzise Messung der HR, wie sie zu diesem Zeitpunkt nur durch das EKG gewährleistet werden konnte, in mobiler Form zu realisieren (Hottenrott, 2007). Hottenrott (2007) berichtet, dass diese Bemühungen insbesondere vom Leistungssport angetrieben wurden und die ersten portablen HR-Messgeräte in Form von Brustgurten und Armbändern ausschließlich in diesem Bereich zum Einsatz kamen. Mittlerweile sind HR-Messungen und die damit einhergehende Kontrolle der individuellen Belastung sowohl im Fitnessbereich als auch im Leistungssport etabliert (Hottenrott, 2007). Bereits im Jahr 2007 vermutete Hottenrott (2007), dass der technische Fortschritt und die Weiterentwicklung im Bereich portabler HR-Messung langanhaltend seien wird.

Diese Entwicklung führte soweit, dass im letzten Jahrzehnt kommerzielle Smartwatch Wearables neben dem Gesundheits- und Fitnessbereich immer mehr Anklang in der Allgemeinbevölkerung fanden und seitdem auch häufiger zum Aufzeichnen physischer Aktivität im Alltag verwendet werden (Feehan et al., 2018;

Fuller et al., 2020; Jo et al., 2016; Hajj-Boustros et al., 2022). Zudem finden diese auch im Zusammenhang mit gesundheitsfördernder Forschung vermehrt Anwendung (Feehan et al., 2018). Dies lässt sich anhand der geringen Kosten im Vergleich zu aufwendigen Labormethoden zur Bestimmung von Vitalparametern und dem nicht-invasiven Charakter von Smartwatch Wearables begründen (Hajj-Boustros et al., 2022). Aufgrund dessen existiert eine Vielzahl wissenschaftlicher Veröffentlichungen, die sich mit der Gültigkeit und Zuverlässigkeit der Parameter auseinandersetzen, die von Smartwatch Wearables gemessen werden (Fuller et al., 2020; Hajj-Boustros et al., 2022; Jo et al., 2016).

In einem umfassenden Review betrachteten Fuller et al. (2020) insgesamt 158 Publikationen um eine Aussage über Zuverlässigkeit und Gültigkeit der Messungen, wie der der HR, treffen zu können. Es wurden neun Hersteller von Smartwatch Wearables, darunter unter anderem Apple, Garmin, Samsung und Fitbit, betrachtet, wobei nur Uhren und Armbänder in dieses Review mit eingeflossen sind. Dabei berücksichtigten Fuller et al. (2020) ausschließlich Studien in denen von mindestens zehn Proband:innen Daten mittels Smartwatch Wearables aufgezeichnet wurden. Es stellte sich heraus, dass es bei der Aufzeichnung der HR große Abweichungen zwischen den einzelnen Modellen gibt, da vor allem nur für Geräte von Apple, Fitbit und Garmin ausreichend Studien für einen Vergleich vorlagen (Fuller et al., 2020). Im Laborsetting detektierten Geräte der Marken Garmin und Apple die HR am präzisesten im Vergleich zu verifizierten Standards, während Fitbit Modelle insbesondere im alltäglichen Gebrauch dazu tendierten die HR zu unterschätzen, was jedoch vom Aktivitätslevel abhängt (Fuller et al., 2020). Zudem stellten Fuller et al. (2020) fest, dass insbesondere die älteren Modelle die HR ungenau aufzeichneten. Dennoch lagen die Abweichungen innerhalb des festgelegten Rahmens, in welchem die Messungen als genau und zuverlässig angekommen werden können. Neben der HR zeigte die Untersuchung von Fuller et al. (2020) auch eine gute Präzision bei der Messung der Schrittzahl innerhalb eines Laborsettings über die Mehrheit der Modelle hinweg. Zudem ist anzumerken, dass eine große Anzahl der betrachteten Studien auf kostengünstige Modelle entfiel, da diese aufgrund ihres Preises auch vermehrt in experimentellen Studien zur Erhebung von Parametern Anwendung fanden (Fuller et al., 2020). Ein Großteil der betrachteten Modelle sowie der Geräte auf dem Markt ermitteln die HR mittels Photoplethysmographie (PPG) (Allen, 2007; Fuller et al., 2020; Jo et al., 2016). Bei dieser optischen Methode handelt es sich um eine kostengünstige und einfache Technik, die Veränderungen des Blutvolumens misst (Allen, 2007).



Aufgrund der weiten Verbreitung dieser Technologie in modernen Fitnesstrackern und Smartwatches, untersuchten Jo et al. (2016) die Gültigkeit der HR-Messungen zweier Smartwatch Wearables, der Fitbit Charge HR und der Basis Peak, bei denen PPG verwendet wird, im Vergleich zu einer zeitgleich durchgeführten Messung der HR mit Hilfe eines 12-Kanal-EKG. Dieses gilt als Goldstandard bezüglich HR-Messungen (Jo et al., 2016). Dafür durchliefen Proband:innen abwechselnd Intervalle verschieden starker körperlicher Aktivität und Ruhe- beziehungsweise Regenerationsphasen innerhalb eines Laborsettings (Jo et al. 2016). Beide Geräte zeigten eine hohe Übereinstimmung mit den EKG-Messungen während der Regenerationsphasen. Außerdem lieferte die Basis Peak laut Jo et al. (2016) auch im Zusammenhang mit Aktivitäten geringer Intensität eine gute Übereinstimmung mit den Vergleichsdaten. Die Fitbit Charge HR zeigte in dessen eine sehr gute Genauigkeit in den Aktivitäten Laufen, Joggen und Rennen, wohingegen die Übereinstimmung mit den EKG-Daten für die restlichen Aktivitätsformen, die Rad fahren oder auch Kraftsportübungen umfassten, nicht gegeben war (Jo et al., 2016). Dies spiegelt ebenfalls die Erkenntnisse der Review-Studie von Fuller et al. (2020) wider, die bei der Genauigkeit der HR-Messungen von Fitbit Smartwatch Wearables eine deutliche Abhängigkeit von der Aktivität, bei welcher die Messung stattfand, feststellten. Insgesamt erfüllte die Basis Peak Smartwatch die Validierungskriterien trotz des geringfügigen Rückgangs in der Genauigkeit bei steigender physischer Belastung (Jo et al., 2016). Bereits Parak & Korhonen (2014) dokumentierten bei der Untersuchung von Pulsarmbändern, die PPG verwenden und am Handgelenk oder Unterarm getragen werden, den Trend einer Reduktion der Messgenauigkeit im Zusammenhang mit zunehmender physischer Aktivität. Außerdem bestätigten sie die Abhängigkeit der Genauigkeit von der Art der Aktivität und der auszuführenden Übung, welche sich auch in den Untersuchungen von Jo et al. (2016) für die Fitbit Charge HR abbilden ließ. Diese erfüllte die Validierungskriterien nicht und wies eine deutliche Verminderung der Genauigkeit mit steigender Intensität der Übungen auf (Jo et al., 2016). Es wird deutlich, dass die Genauigkeit der HR-Messung stark mit Modell, Intensität und ausgeübter Sportart variiert, wobei es keinen eindeutig erkennbaren Trend gibt, sondern jedes Modell individuell reagiert (Jo et al., 2016).

In Bezug auf die Genauigkeit der Messung von Vitalparametern durch Smartwatch Wearables wurde stetig die schnell voranschreitende Weiterentwicklung der Technologie wie PPG oder auch der Algorithmen, die der Berechnung der Parameter zu Grunde liegen, betont (Allen, 2007; Jo et al., 2016). Doch auch aktuelle Untersuchungen,

deren Untersuchungsgegenstand Smartwatch Wearables modernerer Generationen sind, decken sich mehrheitlich mit vorherigen Forschungsergebnissen. Im Rahmen einer Validierungsstudie untersuchten Jachymek et al. (2021) die Genauigkeit der HR-Messung mittels PPG von ausgewählten Smartwatches, die am Handgelenk getragen werden, darunter auch die Fitbit Charge 4. Auch in diesem Fall wurde die HR-Messung durch ein EKG als Referenz für die Genauigkeit gewählt. Zur Überprüfung der HR absolvierten die Proband:innen einen Stress-Test auf dem Laufband nach dem Bruce-Protokoll, bei welchem sowohl Geschwindigkeit als auch Steigung des Laufbandes kontinuierlich gesteigert werden (Jachymek et al., 2021). Die Fitbit Charge 4 schaffte es nach Jachymek et al. (2021) in keiner der Belastungsstufen die Validierungskriterien zu erfüllen. Außerdem konnte auch gezeigt werden, dass dieses Modell die HR tendenziell unterschätzt (Jachymek et al., 2021), wie zuvor bereits Fuller et al. (2020) für Smartwatch Wearables der Marke Fitbit im Allgemeinen feststellten. Des Weiteren konnten Jachymek et al. (2021) ebenfalls den Rückgang der Genauigkeit der HR-Messung mit steigender physischer Beanspruchung nachweisen, was die Erkenntnisse von Jo et al. (2016) und Fuller et al. (2020) widerspiegelt. Jachymek et al. (2021) konnten keinen Zusammenhang zwischen Alter, Geschlecht, Körpergröße, Gewicht beziehungsweise BMI und der Genauigkeit der HR-Messung innerhalb ihrer Studie feststellen und weisen explizit auf ein korrektes Tragen der Smartwatch hin, um extreme Abweichungen in den Messwerten zu vermeiden.

Vergleichbare Resultate bezüglich der Genauigkeit bei der Messung der HR erzielten Muggeridge et al. (2021) für das Vorgängermodell Fitbit Charge 3, wobei die Proband:innen ebenfalls unterschiedliche sportliche Aktivitäten in verschiedenen Intensitätsstufen ausübten und dabei die HR mittels Fitbit Charge 3 und dem Polar H10 Brustgurt aufgezeichnet wurde, welcher als Referenzkriterium diente. Während die Übereinstimmung zwischen Referenzwert und den HR-Daten der Fitbit Smartwatch insgesamt gut war, zeigte sich auch hier die Tendenz, dass die HR im Vergleich zum Referenzwert des Brustgurtes unterschätzt wird um durchschnittlich 7 *bpm* (Muggeridge et al., 2021). Auch beim Vorgängermodell der Fitbit Charge 4 zeigte sich bereits eine eindeutige Abhängigkeit der Genauigkeit in der HR-Messung von Art und Intensität der ausgeübten physischen Aktivität (Muggeridge et al., 2021).

In einer weiteren Validierungsstudie aus dem Jahr 2022 verglichen Hajj-Boustros et al. (2022) Smartwatch Wearables verschiedener Marken, darunter auch Fitbit, bezüglich ihrer Genauigkeit bei der HR-Messung. Dabei wurde ebenfalls ein Polar H10

Brustgurt als Referenz verwendet, dessen Genauigkeit im Vorfeld vielfach verifiziert wurde (Hajj-Boustros et al., 2022). Junge, gesunde Proband:innen trugen die Geräte (die Smartwatches und den Brustgurt) ähnlich wie bei den Untersuchungen von Muggeridge et al. (2021) und Jachymek et al. (2021) während fünf verschiedenen Aktivitäten, die jeweils zehn Minuten andauerten und dabei Sitzen, Laufen auf einem Laufband in verschiedenen Intensitäten, Krafttraining und Rad fahren auf einem Ergometer beinhalteten (Hajj-Boustros et al., 2022). Auch im Rahmen der Untersuchungen von Hajj-Boustros et al. (2022) zeigte sich, wie bereits in den zuvor betrachteten Studien, dass die Genauigkeit der HR-Messung stark durch die Intensität der ausgeübten Aktivität beeinflusst wird. Das Modell der Marke Fitbit wies für das Rad fahren und intensives Laufen auf dem Laufband eine hohe Übereinstimmung mit den HR-Messungen des Polar H10 Brustgurtes auf (Hajj-Boustros et al., 2022). Im Gegensatz dazu stellten Hajj-Boustros et al. (2022) fest, dass insbesondere im Sitzen die Genauigkeit der HR-Messung sehr gering ist. Auch hier kommen die Autoren zu dem Schluss, dass die Genauigkeit stark von der Marke beziehungsweise dem jeweiligen Modell abhängt, wobei im vorliegenden Fall die höherpreisigen Modelle verglichen mit dem Polar H10 Brustgurt die HR präziser bestimmten (Hajj-Boustros et al., 2022). Auf Grundlage der Ergebnisse der Validierungsstudien ist es angezeigt, die Genauigkeit der Vitalparameter, die mit Hilfe von Smartwatch Wearables gemessen werden, zu hinterfragen bevor sie weiteren Analysen und Interpretationen unterzogen werden (Hajj-Boustros et al., 2022; Muggeridge et al., 2021).

Insbesondere aufgrund des geringen Kostenfaktors und des aus dem Alltag gewohnten, unauffälligen Designs von Smartwatch Wearables, verglichen mit kostenintensiven, medizinischen Geräten, die durch ihren großen und daher auch ablenkenden Aufbau im Verdacht stehen, die Messdaten zu beeinflussen, werden diese immer häufiger in Pilot- und Proof-of-Concept-Studien eingesetzt (Shachter, 2022).

In einer Pilotstudie untersuchten Pakhomov et al. (2020) die HR-Daten von Studierenden der University of Minnesota, welche ein Fitbit Tracking Device besaßen und sich als Proband:innen für die Studie meldeten. Neben den HR-Daten wurde auch die subjektive Selbsteinschätzung alltäglicher Stressoren mehrmals täglich erhoben und ausgewertet. Dies erfolgte in einem digitalen Rahmenkonzept, welches die Fitbit App und die Web Applikationen für die Selbsteinschätzung sowie künstlich erzeugte Stress-situationen in Form von verschiedenen Tests beinhaltete. Auf diesem Weg konnten Pakhomov et al. (2020) am Beispiel der Marke Fitbit zeigen, dass sich kostengünstige

Smartwatch Wearables eignen, um sowohl standardisierte Stressoren als auch natürlich auftretenden psychosozialen Stress zu erkennen. Auch wenn der Untersuchung keine Betrachtung der Genauigkeit der HR-Messungen zu Grunde lag, da dies aufgrund der Vielfalt an Modellen nicht möglich war, zeigte sich, dass die HR um 5 – 10 *bpm* im Vergleich zum Ruhepuls stieg und dabei konsistent die Reaktion auf natürliche Stressoren, die nicht in Verbindung mit physischen Anstrengungen standen, widerspiegelte (Pakhomov et al., 2020). Es ist dabei hervorzuheben, dass die Messungen im natürlichen Lebensumfeld der Proband:innen stattfanden und nicht innerhalb eines Laborsettings, weshalb die Untersuchungen und Ergebnisse von Pakhomov et al. (2020) eine wichtige Grundlage für weitere Forschung im Bereich der Verwendung von Smartwatch Wearables zur Erfassung von alltäglichem Stress sind.

Des Weiteren verwendeten Shachter et al. (2022) Fitbit Charge 3 Smartwatches, um zu untersuchen, inwiefern sich Emotionen auf die HR auswirken, die in Sprachklassen vor, während und nach Sprachaktivitäten individuell auftreten und im Rahmen der Studie anhand einer Selbstauskunft protokolliert wurden. Diese wurde mittels des Fitbit Trackers erfasst. Anhand der HR-Daten konnten unter Berücksichtigung der Selbstauskünfte der Proband:innen physische Reaktionen auf die negativ konnotierten Emotionen Verlegenheit und Leiden innerhalb des Sprachunterrichts abgebildet werden (Shachter et al., 2022). Mit ihrer Studie zeigten Shachter et al. (2022), dass sich Smartwatch Wearables im von ihnen gewählten Setting zur HR-Messung eignen.

## **2.6 Fragestellung und Hypothese**

Der wissenschaftliche Fokus der Studie *ProVisioNET* liegt auf der Betrachtung von Lern- und Lehrprozessen im schulischen Zusammenhang, um diese gezielt mit Hilfe moderner Medien sichtbar zu machen und zu erforschen. Dabei ist zum einen von Interesse, worauf Lehrkräfte während des Unterrichtsgeschehens gezielt ihre Aufmerksamkeit richten, zum anderen wie sich die Interaktion von Lehrpersonen und Lernenden im Unterricht darstellt, aber auch inwiefern sich Lehrpersonen aufgrund ihres Grades an Lehrerfahrung unterscheiden. Festzustellen, ob und in welchem Maß die Unterrichtserfahrung der Lehrkräfte die Reaktion auf und die Wahrnehmung von Unterrichtsgeschehnissen beeinflusst, ist dabei das Hauptziel des Projektes. Im Rahmen dieser quasi-experimentellen Querschnittstudie (Döring & Bortz, 2016), die unter Laborbedingen

durchgeführt wurde, fand die Betrachtung des Unterrichts durch die Augen der Lehrperson statt, indem umfassende Daten durch die Verwendung einer Eyetracking-Brille, Audiorekordern und Kameras generiert wurden. Weitere Daten werden anhand eines Fitnesstrackers, der primär zur HR-Messung verwendet wurde, sowie durch Fragebögen zur Selbsteinschätzung des eigenen Verhaltens im Unterricht und zum strategischen Wissen der Lehrpersonen in Bezug auf Klassenführung und anhand der Reflektion des eigenen Eyetracking-Videos erhoben. Die Erkenntnisse, die auf diesem Weg gewonnen werden, sollen ihre Anwendung in einer breiter aufgestellten Ausbildung und Fortbildung von Lehrpersonen finden.

In der vorliegenden Arbeit steht die Untersuchung der Verwendbarkeit eines Fitnesstrackers beziehungsweise einer Smartwatch als Methode zur Messung von Unterschieden mittlerer HR von Lehrpersonen innerhalb eines Laborsettings im Vordergrund. Folgende grundlegende Forschungsfrage soll hierbei beantwortet werden: *Lassen sich bei Lehrkräften signifikante Unterschiede in der mittleren Herzrate zwischen verschiedenen Zeitphasen vor, während und nach einer simulierten Unterrichtssituation mittels einer Fitbit Smartwatch messen?*

Zur Beantwortung dieser Frage wird die HR von Lehrkräften und Lehramtsstudierenden mittels einer Fitbit Charge 4 Smartwatch im Zuge einer Laborstudie gemessen. Während der Aufzeichnung der HR halten die Versuchspersonen unter anderem eine 15-minütige Unterrichtseinheit ab, in der drei studentische Mitarbeitende mit typischem Verhalten eine Schulklasse simulieren. Außerdem werden die Proband:innen zu dieser Lehreinheit befragt und füllen Fragebögen bezüglich ihres Wissens zum Thema Klassenführung aus und geben eine Selbsteinschätzung zu ihrem Auftreten während der Unterrichtseinheit ab. In diesen Phasen sehen sich die Versuchspersonen jeweils mit einem unterschiedlichen Anspruch konfrontiert. Auf dieser Grundlage werden die mittleren HR dieser Phasen miteinander verglichen. Die bisherige Forschung bestätigt, dass Lehrpersonen während des Unterrichts, aber auch bei anderen Aktivitäten mit Bezug zu ihrem Beruf, einem erhöhten Stressniveau ausgesetzt sind (Aldrup et al., 2018; Fisher, 2010). Dabei beruhen diese Erkenntnisse meist auf einer Selbstauskunft der Lehrkräfte. Die HR kann in Bezug auf den Lehrberuf als physischer Stressindikator verwendet werden, wie Junker et al. (2021) im Rahmen einer Feldstudie durch HR-Messungen während des Unterrichts mittels eines 7-Kanal-EKG belegen konnten. In diesem Bezug wird auch auf die potentiell negativen physiologischen Folgen verwiesen, die dauerhafter beziehungsweise langanhaltender physischer Stress mit sich

bringt, der sich im betreffenden Fall durch die gesteigerte HR zeigt (Junker et al., 2021). Während Junker et al. (2021) mit dem 7-Kanal-EKG eine medizinische und auch kostenintensive Methode zur HR-Messung verwendeten, dienten Runge et al. (2020) ein Fitbit Fitnesstracker, um die HR sowie andere Vitalparameter von Lehrkräften sowohl während des Unterrichts, aber auch in ihrer Freizeit zu bestimmen. Die Untersuchung kam zu dem Schluss, dass eine erhöhte HR in Kombination mit der bestimmten Ausprägung weiterer Vitalparameter ein physischer Indikator für gesteigerten Gesamtstress bei Lehrpersonen ist (Runge et al., 2020). Basierend auf der dargelegten Studienlage wurde die folgende Hypothese aufgestellt:

*Der HR-Mittelwert in der Lehrphase ist am größten, wobei der HR-Mittelwertunterschied zwischen der Lehrphase und allen anderen Phasen jeweils signifikant ist.*

Die Hypothese wurde so formuliert, dass sie anhand einer einfaktoriellen Varianzanalyse mit Messwiederholung mit gerichteten Kontrasten beantwortbar ist.

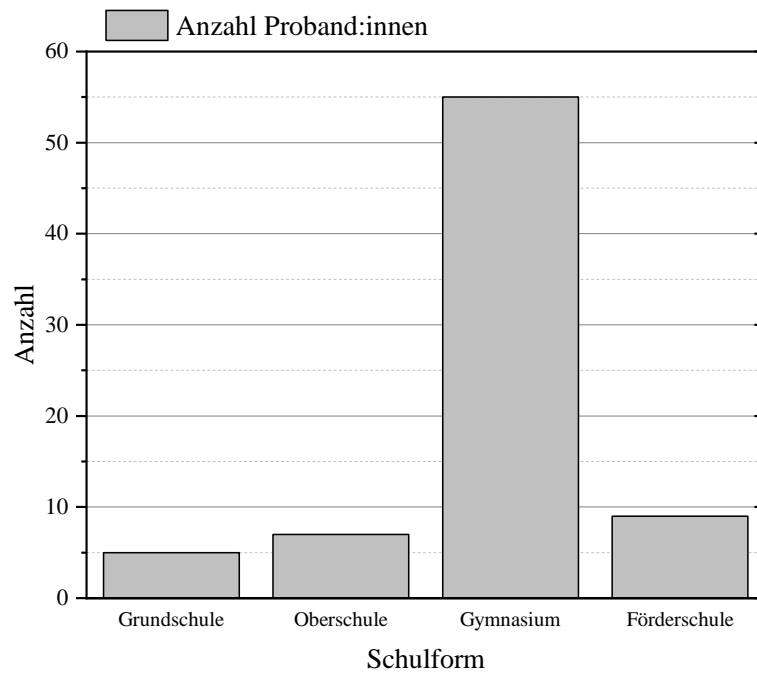
### 3 Methodisches Vorgehen

Im Verlauf dieses Kapitels sollen alle für diese Arbeit wichtigen Themen in Bezug auf die zugrunde liegende Studie erörtert werden. Dafür wird im Folgenden auf die Stichprobe sowie Ablauf und Durchführung der Studie eingegangen. Zusätzlich werden auch die zur Erhebung verwendeten Instrumente beschrieben, die Variablen definiert und die für die Datenauswertung relevanten statistischen Methoden ausführlich erklärt.

#### 3.1 Stichprobe

An der Studien haben insgesamt 79 Proband:innen teilgenommen, allerdings mussten drei Personen aufgrund fehlender beziehungsweise fehlerhafter HR-Daten der Fitbit Charge 4 von der Stichprobe exkludiert werden. Bei den verbleibenden  $N = 76$  Proband:innen, respektive Datensätzen, handelte es sich um 48 Frauen und 28 Männer. Zum Zeitpunkt der jeweiligen Studienteilnahme lag das durchschnittliche Alter der Teilnehmenden bei  $M = 30.26$  Jahren,  $SD = 10.52$  Jahren. Dabei wies die jüngste Person ein Alter von 19 Jahren auf, im Gegensatz dazu betrug das maximale Alter 59 Jahre. Die Teilnahme an der Studie war ab dem dritten Fachsemester des Lehramtsstudiums sowie für Lehrpersonen (inklusive Referendariat) möglich. Dabei wurde Lehramtsstudierenden eine Berufserfahrung von 0 Jahren zugeschrieben. Auf Grundlage dessen besaßen die Teilnehmenden eine durchschnittliche Berufserfahrung im Lehrberuf von  $M = 5.17$  Jahren,  $SD = 9.12$  Jahren. Folglich lag innerhalb der Stichprobe die minimale Lehrererfahrung bei 0 Jahren und die maximale Berufserfahrung bei 37 Jahren. Von den teilnehmenden Lehramtsstudierenden und Lehrpersonen haben fünf Personen ein Lehramtsstudium für die Grundschule absolviert, sieben haben Lehramt für Oberschule studiert, 55 für das Gymnasium und neun für die Förderschule. Die Verteilung der Proband:innen auf die verschiedenen Schulformen ist in Abbildung 2 graphisch dargestellt.

Alle Proband:innen meldeten sich freiwillig für die Teilnahme an der Studie. Dabei erfolgte die Akquise der Teilnehmenden mit Hilfe von Aushängen innerhalb der Universität Leipzig sowie in digitaler Form auf der Website der Universität Leipzig und der Professur für Empirische Schul- und Unterrichtsforschung des Instituts für Bildungswissenschaften. Des Weiteren wurden Schulen in Leipzig und im



**Abbildung 2:** Anzahl der Proband:innen verteilt auf die studierten Schulformen.

weitergefassten Leipziger Umland sowohl telefonisch als auch via E-Mail kontaktiert und auf die Studie hingewiesen. Die in dieser Arbeit diskutierten Daten der Stichprobe wurden im Zeitraum zwischen Juli 2021 und April 2023 erhoben.

### 3.2 Durchführung und Aufbau der Studie

Die Erhebungen im Rahmen der Studie *ProVisioNET* wurden alle in den Studienräumen der Professur für Empirische Schul- und Unterrichtsforschung der Universität Leipzig durchgeführt. Dabei fand die Durchführung für jede Versuchsperson einzeln statt.

Direkt nach dem Eintreffen wurde den Proband:innen jeweils eine Fitbit Charge 4 Smartwatch angelegt. Hierbei wurde insbesondere auf eine möglichst akkurate Positionierung des Gerätes am Handgelenk der Versuchsperson geachtet. Um die HR genau aufzuzeichnen, ist es notwendig, die Fitbit mit einem Finger breit Abstand hinter dem Handgelenksknochen zu tragen. Zusätzlich ist auf einen festen aber für die Proband:innen noch komfortablen Sitz der Smartwatch zu achten. Der Zeitpunkt des Anlegens der Uhr wurde jeweils notiert und markiert den Beginn der Studie beziehungsweise der Erhebung sowie den Beginn der kontinuierlichen HR-Auf-



zeichnung.<sup>3</sup> Neben der HR wurde mit Hilfe der Fitbit auch die Schrittzahl der Versuchsperson während der gesamten Studie dokumentiert.

Im Anschluss wurden die Proband:innen jeweils von der Versuchsleitung mit den drei studentischen Mitarbeitenden bekannt gemacht, die im Rahmen der Studie eine Schulklasse simulierten und dabei für das Unterrichtsgeschehen typische Ereignisse im Klassenraum nachahmten. Daraufhin erhielt die Versuchsperson durch die Versuchsleitung einen kurzen Überblick über den Hintergrund der Studie sowie über den Ablauf der Erhebung. Ist dieser Teil abgeschlossen, richteten sich die jeweiligen Proband:innen am Lehrer:innentisch für die simulierte Unterrichtsstunde ein und bauten dafür benötigtes Equipment und Materialien auf. Aus Studienzwecken sind für die Unterrichtseinheit folgende Aspekte erwünscht: bei den Handlungsmustern soll die Lektion nach Möglichkeit aus einem Vortrag, aus einem Lehrer:innengespräch oder Tafelarbeit bestehen. Arbeitsblätter, Power-Point-Präsentationen, Pinnwände und Overheadprojektoren dürfen als Medien eingesetzt werden. In Bezug auf die Sozialform soll es sich um Frontal- und Einzelunterricht handeln. Als Verlaufsform ist ein Unterrichtseinstieg angedacht. Als nächstes wurde eine Eyetracking-Brille aufgesetzt, auf die Bedürfnisse der Versuchsperson angepasst (insbesondere Sehstärke, wenn notwendig), ein erstes Mal kalibriert und durch die Versuchsleitung erklärt. Die Eyetracking-Brille erfasst mit einer Kamera das Gesehene der Versuchsperson sowie die Pupillenbewegung mittels Infrarotsensoren und nimmt über ein Mikrofon die Umgebungsgeräusche auf. Sobald die Eyetracking-Brille einsatzbereit war, wurden im Studienraum die insgesamt vier Kameras und der Audiorekorder gestartet. Mit diesem Setup wurde auch die Erwärmungsphase beziehungsweise das Kennenlernspiel zwischen den jeweiligen Proband:innen und den drei studentischen Mitarbeitenden aufgezeichnet. Mit Hilfe von gegenseitigem Ball zuwerfen, machten sich die Beteiligten mit den jeweiligen Namen vertraut. Dabei ist anzumerken, dass den studentischen Mitarbeitenden aus Studienzwecken ein Name zugewiesen wurde, der jeweils mit A, B oder C beginnt. Währenddessen trug die Versuchsperson bereits die Eyetracking-Brille. Danach wurde die Eyetracking-Brille mit der Versuchsleitung im Flur ein weiteres Mal kalibriert.

Mit Wiederbetreten des Studienraums durch die jeweiligen Proband:innen startete direkt die 15-minütige simulierte Unterrichtseinheit. Der genaue Zeitpunkt wurde auch hier dokumentiert. Die Lehrperson war dabei frei in der Wahl des Faches, der

---

<sup>3</sup> Die Funktionsweise und HR-Datenaufzeichnung der Fitbit Charge 4 wird in Kapitel 3.3 detailliert erläutert.

Klassenstufe sowie des Themas der Unterrichtsstunde. Die Versuchsleitung befand sich währenddessen hinter einer Trennwand innerhalb des Studienraums. Die drei studentischen Mitarbeitenden, die eine Schulklasse simulierten, führten während der Einheit verschiedene Arten von Unterrichtsstörungen aus, welche sich den Kategorien verbale Störungen, Mangel an Interesse und nervöse Störungen zuordnen lassen. Dazu zählen unter anderem am Handy spielen, störendes, lautes Hereinrufen, Tuscheln mit der Person am Nebentisch, lautes Fingerschnipsen oder auch gedanklich abwesend sein, was sich durch Malen auf dem Block kennzeichnet. Die entsprechenden Anweisungen zu den auszuführenden Störungen wurden der Schulklasse über einen Monitor mitgeteilt, der für die Versuchsperson nicht einsehbar war. Eine Abbildung, die die gesamte Anordnung der beteiligten Personen und das verwendete technische Equipment während der simulierten Unterrichtseinheit zeigt, ist in Anhang A beigelegt. Nach Ablauf der 15 Minuten erhielt die Versuchsperson in Form einer gelben Karte von der Versuchsleitung den Hinweis, dass noch eine Minute verbleibt. Verstrich auch diese Minute, so erhielten die jeweiligen Proband:innen eine rote Karte, die damit die Unterrichtseinheit beendete. Die betreffende Uhrzeit wurde wiederum dokumentiert.

Im Anschluss fand im Flur eine finale Kalibrierung mit der Versuchsleitung statt. Anschließend wurde die Brille abgesetzt und die Proband:innen füllten an einem Laptop im Studienraum einen Fragebogen aus, in dem die Einschätzung des eigenen Auftretens während der Lehreinheit abgefragt wurde.<sup>4</sup> Auch die studentischen Mitarbeitenden füllten den Fragebogen aus und gaben damit ihre Einschätzung bezüglich des Verhaltens der Versuchsperson ab. Ist dieser beantwortet, folgte das *Stimulated Recall Interview*, welches per Video und Audiorekorder aufgezeichnet wurde. Dafür schauten sich die Versuchsperson und die Versuchsleitung das Eyetracking-Video gemeinsam nochmals in voller Länge an. Die Proband:innen wurden dazu aufgefordert ihr Verhalten oder auch Gedanken, die zuvor während bestimmter Handlungen präsent waren, zu kommentieren und zu äußern. Für diese Zwecke bestand auch die Möglichkeit das Eyetracking-Video zu pausieren. Ist die Sichtung des Videos abgeschlossen, beantwortete die Versuchsperson einen weiteren Fragebogen am Laptop, der das strategische Wissen bezüglich Klassenführung betraf (SJT). Die Versuchsleitung wartete währenddessen bis der Fragebogen ausgefüllt war. Im Anschluss wurde den Proband:innen die Fitbit Smartwatch abgenommen und die genaue Zeit notiert, welche das Ende der

---

<sup>4</sup> Dabei handelt es sich um einen EMU Fragebogen (Evidenzbasierte Methoden der Unterrichtsdagnostik).

Datenerhebung markierte. Dabei beschränkte sich die Maximaldauer der Studie auf zwei Stunden.

### **3.3 Variablen**

Eine empirische Untersuchung ist aus mehreren relevanten Variablen aufgebaut, welche mindestens zwei Abstufungen aufweisen, wodurch sie sich von Konstanten unterscheiden (Bortz & Schuster, 2010). Grundsätzlich wird zwischen zwei Arten von Variablen unterschieden, den abhängigen und unabhängigen Variablen (Bortz & Schuster, 2010; Döring & Bortz, 2016; Schäfer, 2016). Dabei bezeichnet die abhängige Variable diejenige Größe innerhalb des Erhebungsprozesses, auf deren genauer Beschreibung und Betrachtung der Fokus der Untersuchung liegt (Schäfer, 2016). Die Ausprägung sowie Merkmale der abhängigen Variablen werden auf die Merkmale der unabhängigen Variablen zurückgeführt (Bortz & Schuster, 2010; Schäfer, 2016). Anhand der abhängigen Variablen werden die Auswirkungen der unabhängigen Variablen auf andere Merkmale im Forschungsprozess überprüft, insbesondere auch auf die abhängige Variable selbst (Bortz & Schuster, 2010; Schäfer, 2016). Folglich ist die abhängige Variable in ihrer Ausprägung stark durch den Einfluss der unabhängigen Variable geprägt (Schäfer, 2016). Im Gegensatz dazu ist die unabhängige Variable bereits durch Fragestellung und Untersuchungsdesign vorgegeben, dabei ist sie von anderen Variablen und deren Effekten unabhängig (Schäfer, 2016). Während des Experimentes unterliegt die unabhängige Variable einer systematischen Variation oder Beobachtung (Schäfer, 2016). Wenn es sich bei der zu überprüfenden Hypothese, wie im vorliegenden Fall, um eine Unterschiedshypothese handelt, definiert die unabhängige Variable die Gruppierungsvariable, wodurch die inhaltlichen Unterschiede der einzelnen Gruppen zu unterschiedlichen Effekten in der abhängigen Variable führen (Döring & Bortz, 2016).

In dieser Arbeit handelt es sich bei der unabhängigen Variable um die Phase, welche fünf Abstufungen aufweist, namentlich der Pre-, Lehr-, Postlehr-, Interview- und End-Phase. Diese fünf Phasen definieren jeweils zehnminütige Zeitintervalle innerhalb der Gesamterhebungsdauer von maximal zwei Stunden. Die einheitliche Länge der verschiedenen Phasen gewährleistet die Vergleichbarkeit ihres Einflusses auf die abhängige Variable. Die fünf Intervalle respektive Phasen zeichnen sich durch ihr unterschiedliches physisches und psychisches Anforderungsniveau an die Proband:innen

aus. Die Pre-Phase umfasst die ersten zehn Minuten der Gesamterhebung, in denen die Versuchsperson in die Studie und die Durchführung eingewiesen wird, sich mit dem Setting vertraut macht und sich für die Unterrichtseinheit vorbereitet. Es gibt keine gesteigerten Anforderungen an die Proband:innen. Die Lehr-Phase beschreibt die ersten zehn Minuten der simulierten Unterrichtseinheit, in denen die Versuchsperson selbst unterrichtet und mit Störungen seitens der Schulklasse konfrontiert wird. Die Mehrheit der Proband:innen absolvierte diese im Stehen. Folglich handelt es sich um eine Phase gesteigerter Belastung. Das zehnminütige Intervall direkt im Anschluss an die Unterrichtseinheit markiert die Postlehr-Phase. Dabei handelt es sich um die Übergangszeit zwischen der Lehreinheit und der anschließenden, weiterführenden Datenerhebung, sowie der Beantwortung des ersten Fragebogens (EMU), welche alleine im Sitzen stattfand. Daher kann von einem geringen Anforderungsniveau ausgegangen werden. Die Interview-Phase bildet die mittleren zehn Minuten des *Stimulated Recall Interviews* ab, in denen die Proband:innen gemeinsam mit der Versuchsleitung das Eyetracking-Video sichten und kommentieren. Dies fand vollständig im Sitzen statt. Bei der fünften Stufe handelt es sich um die End-Phase, welche die letzten zehn Minuten der maximal zweistündigen Erhebung beschreibt. Während dieses Intervalls beantwortet die Versuchsperson in Beisein der Versuchsleitung einen weiteren Fragebogen (SJT) im Sitzen. Anschließend wird die Datenerhebung beendet und die jeweiligen Proband:innen entlassen. Innerhalb der fünf festgelegten Intervalle beziehungsweise Phasen wird die Ausprägung der abhängigen Variable durch die variierende Intensität der Anforderung unterschiedlich stark beeinflusst.

Bei der abhängigen Variable handelt es sich um die HR der Proband:innen, wobei die mittlere HR je Proband:in und je Phase von Interesse ist. Die HR-Daten wurden während der Durchführung mit Hilfe einer Fitbit Charge 4 aufgezeichnet. Diese kam im April 2020 mit einer unverbindlichen Preisempfehlung von 149.99 € auf den Markt (Jackymek et al., 2021). Neben der aktuellen HR misst diese auch die Ruheherzfrequenz und legt Herzfrequenzzonen fest, welche durch den Algorithmus des Fitnesstrackers definiert werden oder auch individuell festgelegt werden können (Fitbit, 2023). Die Messung der HR ist dabei eine Grundeinstellung, welche automatisch und durchgängig in Echtzeit stattfindet (Fitbit Help, o. D.-a; Fitbit, 2023). Die Fitbit Charge 4 verwendet, wie die Mehrheit moderner Smartwatches und Fitnesstracker, PPG zur Ermittlung der HR (Jachymek, 2021). Dies findet durch einen optischen HR-Tracker statt, der mittels einer LED mehrmals pro Sekunde grünes Licht aussendet, welches durch eine

lichtempfindliche Fotodiode wieder detektiert wird (Fitbit Help, o. D.-b). Auf diese Weise wird die Veränderung des Blutvolumens in den Kapillaren am Handgelenk bestimmt, da die Kontraktion und Entspannung des Herzens bei jedem Schlag auch zu einer Erweiterung und folglich einem Zusammenziehen der Kapillaren führt (Fitbit Help, o. D.-b). Aus diesen Informationen errechnet der Fitnessstracker kontinuierlich die HR pro Minute (Fitbit Help, o. D.-b). Eine Maximierung des reflektierten Lichtes von hautoberflächennahen Kapillaren wird durch die Verwendung von grünem Licht erreicht (Fitbit Help, o. D.-b). Um die Genauigkeit der HR zusätzlich zu verbessern, misst ein Infrarotsensor, ob die Fitbit Charge 4 Hautkontakt am Handgelenk hat (Fitbit Help, o. D.-b). Die Daten werden sieben Tage lang auf der Uhr gespeichert, sind aber mit Hilfe der Fitbit App dauerhaft zugänglich und archivierbar (Fitbit, 2023). In der Fitbit App können neben den bisher erwähnten Parametern auch weitere Daten eingesehen werden, die eine von Fitbit konzipierte Gesundheitsmetrik bilden und dabei die Sauerstoffsättigung, Atemfrequenz, Ruheherzrate, Schwankungen der Hauttemperatur und die Herzratenvariabilität (HRV) berücksichtigen (Fitbit, 2023). Anhand dieser Parameter soll eine verbesserte Einschätzung des eigenen Gesundheitszustandes zum aktuellen Zeitpunkt möglich sein, der als eine Art Überblick und Anhaltspunkt dienen soll (Fitbit, 2023).

### **3.4 Methodiken der statistischen Datenauswertung**

In diesem Abschnitt werden die Methodiken der statistischen Datenauswertung erläutert, die verwendet wurden, um die Fragestellung und die im Zuge dessen aufgestellte Hypothese zu beantworten. Es wird auf die einfaktorielle Varianzanalyse (ANOVA) mit Messwiederholung mit Kontrasten und die Effektstärke  $d$  nach Cohen eingegangen. Zudem werden zuvor notwendig Auswertungsschritte, wie die Bildung von Mittelwerten und die z-Standardisierung, thematisiert.

#### **3.4.1 Mittelwertbildung und z-Standardisierung der HR-Daten**

Um die Mittelwertunterschiede zwischen der Lehr-Phase und jeweils der Pre-, Postlehr-, Interview- und End-Phase zu untersuchen, ist es notwendig, die durchschnittliche HR für jede:n Proband:in in den betrachteten fünf Phasen sowie über die gesamte Messzeit mit einer zeitlichen Obergrenze von zwei Stunden zu bestimmen.

Da die HR aller Proband:innen durch individuelle körperliche Faktoren, wie Vorerkrankungen, Alter oder auch Aktivitätslevel, beeinflusst wird, ist ein direkter Vergleich der HR-Daten nicht möglich. Um diese Levelunterschiede beziehungsweise individuellen Skalen auszugleichen, werden die HR-Daten durch eine z-Standardisierung umgerechnet, sodass sie anhand einer einheitlichen Skala besser vergleichbar sind (Schäfer, 2016). Jeder einzelne HR-Datenpunkt wird für jede Person der Stichprobe mit dem individuellen HR-Mittelwert und der *SD*, welche über die Gesamtdauer der Studie bestimmt wurden, auf folgende Weise standardisiert:

$$z_i(t) = \frac{HR_i(t) - \mu_i}{\sigma_i}$$

Die z-standardisierte HR beziehungsweise der z-Wert der Person *i* zum Zeitpunkt *t* (innerhalb der Studiendauer von zwei Stunden) wird anhand der HR der Person *i* zum Zeitpunkt *t* errechnet, indem von dieser HR der HR-Mittelwert  $\mu$  der Person *i* subtrahiert wird und anschließend durch die zugehörige Standardabweichung  $\sigma_i$  dividiert wird. Der z-Wert einer bestimmten HR gibt daher an, um welchen Faktor der Standardabweichung der HR-Messwert in *bpm* vom Gesamt-HR-Mittelwert abweicht. Meist werden z-Werte ohne Einheit angegeben, formal handelt es sich bei der Einheit um Standardabweichungen. Der Mittelwert aller z-Werte der HR einer Person, die an der Studie teilnahm, ist aufgrund der z-Standardisierung durch 0 gegeben und die zugehörige *SD* liegt daher bei 1 (Schäfer, 2016). Aus den z-standardisierten HR-Werten wird für jede der fünf Phasen und alle Proband:innen ein z-standardisierter HR-Mittelwert inklusive *SD* ermittelt. Dafür wurde, wie bereits für die Bestimmung der HR-Mittelwerte in *bpm*, Microsoft Office Excel verwendet. Die z-standardisierten HR-Mittelwerte bilden aufgrund ihrer gesteigerten Vergleichbarkeit im Gegensatz zu den Rohdaten in *bpm* den Ausgangspunkt der statistischen Auswertungen, welche im weiteren Verlauf dieses Kapitels erörtert werden. Die Bedeutung der Unterschiede der Phasenmittelwerte, welche über die gesamte Stichprobe gebildet werden, entsprechen der Diskussionsgrundlage dieser Arbeit.

### 3.4.2 ANOVA mit Messwiederholung

Generell handelt es sich bei einfaktoriellen Varianzanalysen um statistische Methoden, um die Differenzen zwischen mindestens zwei Mittelwerten zu vergleichen (Schäfer,

2016). Bei einer einfaktoriellen Varianzanalyse mit Messwiederholung durchläuft ein:e Proband:in oder im Allgemeinen die abhängige Variable verschiedene Bedingungen, was sich als Reihenfolgeeffekt in den erhobenen Daten der Messgröße zeigt (Schäfer, 2016). Anderenfalls kann es sich auch um die wiederholte Messung der abhängigen Variable zu verschiedenen, aufeinander folgenden Zeitpunkten handeln. In diesem Fall findet die Erhebung für die selbe Versuchsperson mehrmals in einem gewissen zeitlichen Abstand statt, somit trägt jede:r Proband:in mehrere Datenpunkte zur Untersuchung bei (Döring & Bortz, 2016; Janczyk & Pfister, 2013). Die Bedingungen werden dabei je nach Studiendesign variiert oder bleiben unverändert. In der vorliegenden Arbeit bilden die fünf verschiedenen Phasen, namentlich die Pre-, Lehr-, Postlehr-, Interview- und End-Phase, somit fünf Messwiederholungen. Folglich trägt jede Versuchsperson an fünf unterschiedlichen Zeitpunkten innerhalb der Gesamtstudiendauer je einen HR-Mittelwert pro Phase bei.

Im Rahmen einer Datenerhebung, die anhand einer ANOVA mit Messwiederholung ausgewertet wird, lassen sich verschiedene Varianzen definieren, welche sich jeweils auf bestimmte Teile oder Bedingungen innerhalb des zu betrachtenden Datensatzes beziehen (Janczyk & Pfister, 2013; Schäfer, 2016). Bei der Varianz, welche von Interesse ist, handelt es sich um die aufgeklärte Varianz, welche für Messwiederholungen aussagt, „dass sich über die Messpunkte hinweg [...] für alle [Proband:innen] Unterschiede ergeben [,welche] im Mittel von 0 verschieden sind.“ (Schäfer, 2016, S. 232) Diese wird durch die unabhängige Variable, welche hier durch die Phase gegeben ist, erzeugt und folglich als systematische Varianz bezeichnet (Schäfer, 2016). Dabei besitzt die Varianz zwischen den einzelnen Versuchspersonen, bei der es sich um die Intersubjektvarianz handelt, in diesem Zusammenhang keine Relevanz, da das Niveau der einzelnen Proband:innen im Bezug mit Messwiederholungen nicht von Interesse ist (Janczyk & Pfister, 2013; Schäfer, 2016). Neben der aufgeklärten Varianz ist auch die Fehlervarianz, die sich aufgrund von Messfehlern sowie der individuellen Reaktion einer Versuchsperson auf die verschiedenen Bedingungen beziehungsweise Phasen akkumuliert (Schäfer, 2016), von Bedeutung für die Varianzanalyse mit Messwiederholung. Aus dem Quotienten der aufgeklärten und der Fehlervarianz wird der *F*-Wert bestimmt:

$$F = \frac{\sigma_{UV}^2}{\sigma_{\text{Fehler}}^2}$$

Dabei entspricht der  $F$ -Wert der Prüfgröße der zugrunde liegenden  $F$ -Verteilung. Diese stellt eine Wahrscheinlichkeitsverteilung aller möglichen Varianz-Verhältnisse dar. Unter zur Hilfenahme der Freiheitsgrade der Fehlervarianz  $df_{\text{Fehler}} = (k - 1) \cdot (n - 1)$  und der aufgeklärten Varianz  $df_{\text{UV}} = k - 1$  wird die Wahrscheinlichkeit des auf Grundlage einer Hypothese ermittelten  $F$ -Werts überprüft, wobei  $k$  der Anzahl an Stufen der unabhängigen Variable entspricht und  $n$  der Anzahl an Proband:innen. Bei einem signifikanten Ergebnis, welches dann vorliegt, wenn die aufgeklärte Varianz groß gegenüber der Fehlervarianz ist, kann die Hypothese angenommen werden. In der Regel wird ein Signifikanzniveau von 5% zu Grunde gelegt. Auf diese Weise wird durch die Varianzanalyse geprüft, ob die festgestellten Differenzen in den Mittelwerten bedeutsam genug sind, um eine Verallgemeinerung auf die Grundgesamtheit zu rechtfertigen. In der mit SPSS ausgeführten Varianzanalyse werden zur Berechnung Quadratsummen verwendet, welche sich aus dem Unterschied zwischen Messwert und dem aus allen Datenpunkten gebildeten Mittelwert ergeben, die anschließend quadriert und für alle Messwerte aufsummiert werden (Schäfer, 2016).

Um eine einfaktorielle Varianzanalyse durchführen zu können, deren Ergebnisse auf statistischer Grundlage uneingeschränkt interpretierbar sind, müssen gewisse Voraussetzungen erfüllt sein (Janczyk & Pfister, 2013). Denn nur in diesem Fall ist es möglich, eine  $F$ -Verteilung als Prüfverteilung zu verwenden (Schäfer, 2016). Es ist notwendig, dass die abhängige Variable, im vorliegenden Fall die HR, intervallskaliert ist, da im Zuge der Berechnungen Operationen angewendet werden, welche nur unter dieser Bedingung durchgeführt werden können beziehungsweise dürfen (Schäfer, 2016). Zudem müssen die Messungen korreliert sein, was aufgrund der wiederholten Messung derselben Person gewährleistet ist (Janczyk & Pfister, 2013). Des Weiteren besteht die Notwendigkeit, dass die Messwerte der abhängigen Variable in jeder untersuchten Gruppe normalverteilt sind (Fahrmeir et. al, 2016; Schäfer, 2016). So wird sichergestellt, dass auch die Grundgesamtheit normalverteilt ist (Schäfer, 2016). Um zu überprüfen, ob dies erfüllt ist, kann der Shapiro-Wilk-Test oder der Kolmogorov-Smirnov-Test auf die Daten der einzelnen Gruppen angewendet werden. Die Normalverteilung kann allerdings auch als gegeben betrachtet werden, wenn die Stichprobe innerhalb einer Gruppe ausreichend groß ist (Bortz & Schuster, 2010; Döring & Bortz, 2016). Dabei wird meist eine Stichprobengröße von  $N \geq 30$  als Grenzwert betrachtet, ab welchem die Normalverteilung auch ohne weitere Testungen angenommen werden darf (Bortz & Schuster, 2010; Döring & Bortz, 2016). Eine weitere Bedingung an die Gruppen ist dabei,



dass sich die zugehörigen Varianzen nicht zu sehr voneinander unterscheiden beziehungsweise Sphärizität gegeben sein soll (Janczyk & Pfister, 2013; Schäfer, 2016). Diese lässt sich anhand des Mauchly-Tests nachprüfen, welcher signifikant ist, wenn die Bedingung nicht erfüllt wird (Janczyk & Pfister, 2013). Folglich sollten die Messwerte auch keine Ausreißer – Datenpunkte, die deutlich von der Mehrheit der anderen Messwerte abweichen – innerhalb der Stichprobe enthalten (Schäfer, 2016). Anhand eines Boxplots können Messwerte auf mögliche Ausreißer untersucht werden und daraufhin über den Umgang mit diesen entschieden werden.

Insgesamt ist die Varianzanalyse als robust gegenüber den hier genannten Voraussetzungen beziehungsweise deren Verletzung einzuordnen, sodass die Ergebnisse dennoch von einem statistischen Standpunkt aus sinnvoll interpretierbar sind (Schäfer, 2016).

### 3.4.3 Kontraste

Um zu bestimmen, zwischen welchen Gruppen, im vorliegenden Fall zwischen welchen Phasen, signifikante Mittelwertunterschiede in der HR vorliegen, wurden Einzelvergleiche durchgeführt. Da die zu untersuchende Hypothese davon ausgeht, dass die Lehr-Phase im Vergleich mit jeder übrigen Phase einen höheren HR-Mittelwert aufweist, und daher gerichtet ist, wurden Kontraste berechnet (Schäfer, 2016). In diesem Fall handelt es sich um a priori Kontraste oder auch geplante Vergleiche, da nur relevante Unterschiede von gesteigertem Interesse getestet wurden, welches auf theoretischen Erwartungen beruht (Janczyk & Pfister, 2013). Im Allgemeinen sind Kontraste „[...] eine bestimmte Kombination von Populationsparametern“ (Janczyk & Pfister, 2013, S. 108), welche sich als  $\psi$  mit den Erwartungswerten beziehungsweise Phasenmittelwerten  $\mu$  der einzelnen Gruppen und den jeweiligen Koeffizient  $k$  folgendermaßen darstellen lassen:

$$\psi = \sum_{i=1}^I k_i \mu_i = k_1 \mu_1 + \dots + k_I \mu_I$$

Daraus resultiert, dass die Anzahl der Koeffizienten der Anzahl an Gruppen (hier: Phasen) entspricht. Zudem muss die Summe der Koeffizienten 0 ergeben (Janczyk & Pfister, 2013). Den zu vergleichenden Gruppen werden im Fall von Einzelkontrasten und kombinierten Kontrasten verschiedene Vorzeichen zugewiesen. Die Koeffizienten sind für jeden betrachteten Kontrast neu zu wählen. Im Rahmen dieser Arbeit wurden

Einzelkontraste je zweier Gruppen zwischen der Lehr-Phase und jeweils der Pre-, Postlehr-, Interview- und End-Phase betrachtet, wobei der Lehr-Phase der Koeffizient „1“ und der zu vergleichenden Phase der Koeffizient „-1“ zugewiesen wurde. Die verbleibenden drei Phasen erhielten jeweils einen Koeffizienten von „0“. Die genaue Syntax der SPSS Rechnungen, welche auch die Koeffizienten der jeweils berechneten Kontraste enthält, ist in Anhang B.1 dokumentiert. Die notwendige Bedingung, dass die Summe der Koeffizienten 0 ergeben muss, war damit erfüllt. Die Gesamtzahl an Gruppen nimmt dabei keinen Einfluss auf den Einzelkontrast.

### 3.4.4 Effektstärke $d$

Die gängigste Methode zur Überprüfung der Bedeutsamkeit von statistisch signifikanten Mittelwertunterschieden bieten Effektstärkenmaße. Dabei ist Cohen's  $d$  das am weitesten verbreitete standardisierte Maß der Effektgröße bezüglich Mittelwertunterschieden zwischen zwei Gruppen (Döring & Bortz, 2016). Im Allgemeinen beschreiben Gruppendifferenz-Effektgrößenmaße, welche die Effektgröße in standardisierten Einheiten angeben, die Mittelwertdifferenz der beiden betrachteten Gruppen, die anhand der Standardabweichung relativiert wird. Dementsprechend berechnet sich Cohen's  $d$  wie folgt:

$$d = \frac{\mu_A - \mu_B}{\sigma}$$

Dabei ist der Mittelwertunterschied durch den Zähler definiert, der sich aus  $\mu_A$ , dem Mittelwert der Gruppe A und  $\mu_B$ , dem Mittelwert der Gruppe B ergibt. Cohen (1988) ging davon aus, dass beide Gruppen die gleiche Standardabweichung  $\sigma$  besitzen. Liegt ein Unterschied in der Standardabweichung der beiden Gruppen vor, ist die gepoolte Standardabweichung  $\sigma_{pool}$  zur Ermittlung von Cohen's  $d$  zu verwenden. Zusätzlich ist zu beachten, dass die gepoolte Standardabweichung durch

$$\sigma_{pool} = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}, n_A = n_B$$

gegeben ist, wenn der Stichprobenumfang  $n_A$  der Gruppe dem Stichprobenumfang  $n_B$  entspricht. Variiert die Stichprobengröße zwischen den betrachteten Gruppen wird die

folgende Gleichung zur Bestimmung der gepoolten Standardabweichung verwendet (Döring & Bortz, 2016):

$$\sigma_{pool} = \sqrt{\frac{(n_A - 1) \cdot \sigma_A^2 + (n_B - 1) \cdot \sigma_B^2}{(n_A - 1) + (n_B - 1)}}, n_A \neq n_B$$

Diese Formel gibt den allgemeinen Fall für  $\sigma_{pool}$  an. Die vereinfachte Variante, die bei identischer Stichprobengröße  $n_A = n_B$  Anwendung findet, lässt sich daraus ableiten. Da für die Daten, die im Rahmen dieser Arbeit ausgewertet wurden, der Stichprobenumfang für alle Gruppen gleich ist, wurde die reduzierte Formel der gepoolten Standardabweichung verwendet. Der Wertebereich der Effektstärke  $d$  ist sowohl für positive als auch negative Werte unbegrenzt, ein Wert von  $d = 0$  sagt hingegen aus, dass kein Effekt vorhanden ist (Döring & Bortz, 2016).

Um eine Aussage darüber zu treffen, ob die Notwendigkeit besteht, zusätzlich zu statistischen Signifikanzen auch praktische sowie theoretische Bedeutungen der Ergebnisse einer Studie zu erörtern, ist es erforderlich, Effektgrößen zu klassifizieren. Im Fall der Effektstärke  $d$  legt Cohen (1988) die Maßgabe zugrunde, dass es sich um geringe Effekte handelt, wenn  $|d| > 0.20$  ist. Bei einem Wert von  $|d| > 0.50$  beziehungsweise  $|d| > 0.80$  werden mittlere beziehungsweise große Effekte angenommen (Döring & Bortz, 2016). Die Klassifizierung der Effektstärke nach Cohen (1988) ist in Tabelle 2 aufgeführt.

**Tabelle 2**

*Klassifizierung der Effektstärke  $d$  nach Cohen (1988).*

| $ d $    | Klassifizierung  |
|----------|------------------|
| $> 0.20$ | geringer Effekt  |
| $> 0.50$ | mittlerer Effekt |
| $> 0.80$ | großer Effekt    |

## 4 Ergebnisse

Im folgenden Kapitel werden die Ergebnisse der zugrunde liegenden Untersuchung wiedergegeben. Unter zur Hilfenahme einer einfaktoriellen Varianzanalyse (ANOVA) mit Messwiederholung und gerichteten Kontrasten sowie der Effektstärke  $d$  soll die zuvor aufgestellte Hypothese in der anschließenden Diskussion falsifiziert oder verifiziert werden. Anhand der Ergebnisse der HR-Messung der Fitbit Charge 4 wird überprüft, ob der HR-Mittelwert in der Lehrphase am größten ist, wobei sich der HR-Mittelwert zwischen der Lehrphase und den restlichen vier Phasen jeweils statistisch signifikant unterscheidet.

Da drei Proband:innen aufgrund fehlender HR-Daten von der Stichprobe exkludiert werden mussten, wurde für 76 Proband:innen jeweils der HR-Mittelwert und die zugehörige Standardabweichung über den Gesamtverlauf der Studiendauer mit einem zeitlichen Maximum von zwei Stunden bestimmt. Zusätzlich wurden auch die Mittelwerte der fünf betrachteten Phasen, die jeweils zehn Minuten andauern, berechnet.<sup>5</sup>

**Tabelle 3**

*Deskriptive Statistiken der HR-Messungen in bpm unter Angabe des Stichprobenumfangs  $N$ , des Mittelwerts  $M$ , der Standardabweichung  $SD$  sowie des Maximums und Minimums für alle fünf Phasen und für den Gesamtverlauf.*

|                 | <i>N</i> | <i>M</i> | <i>SD</i> | <b>Min</b> | <b>Max</b> |
|-----------------|----------|----------|-----------|------------|------------|
| Pre-Phase       | 76       | 96.02    | 10.82     | 70.05      | 122.44     |
| Lehr-Phase      | 76       | 100.41   | 14.33     | 75.37      | 143.74     |
| Postlehr-Phase  | 76       | 93.55    | 12.35     | 70.17      | 136.53     |
| Interview-Phase | 76       | 81.85    | 9.66      | 62.36      | 111.04     |
| End-Phase       | 76       | 77.65    | 10.38     | 53.43      | 107.92     |
| Gesamtverlauf   | 76       | 90.13    | 10.42     | 68.28      | 123.66     |

Die in Tabelle 3 dargestellten Werte zeigen, dass der HR-Mittelwert in der Lehrphase ( $M = 100.41 \text{ bpm}$ ,  $SD = 14.33 \text{ bpm}$ ) nominell am größten ist. Dabei liegt der geringste HR-Mittelwert in dieser Phase bei  $75.37 \text{ bpm}$ , wohingegen das Maximum  $143.74 \text{ bpm}$  entspricht. Die verbleibenden Phasen weisen im Vergleich einen

<sup>5</sup> Eine ausführliche Aufstellung aller HR-Mittelwerte sowie der z-standardisierten Mittelwerte je Proband:in sind dem digitalen Anhang zu entnehmen.

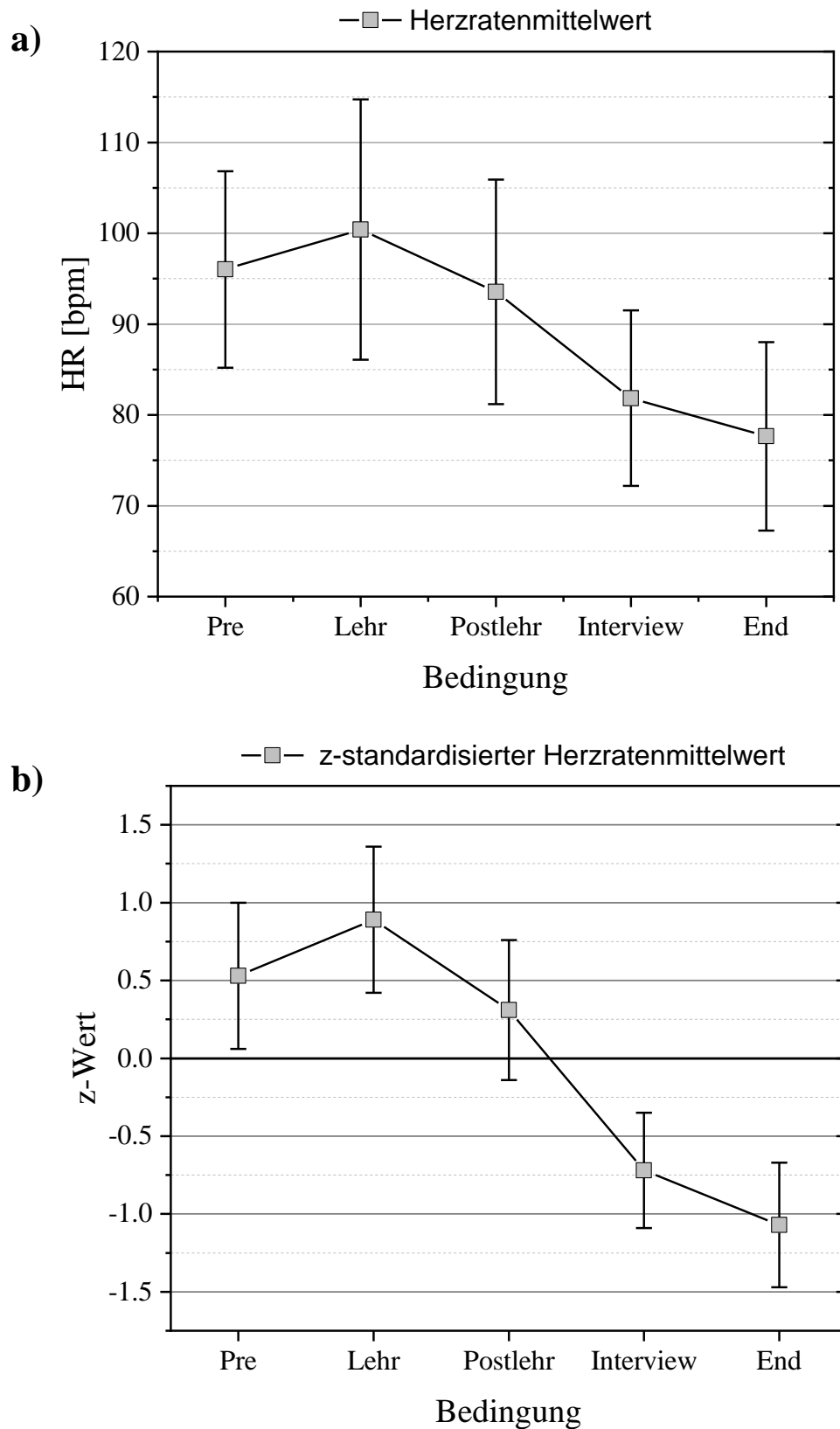
geringeren Mittelwert in der HR auf. Im weiteren Verlauf werden für die Durchführung der Varianzanalyse mit Messwiederholung die z-standardisierten HR-Daten als abhängige Variable verwendet, um individuelle Levelunterschiede zwischen den einzelnen Proband:innen auszugleichen und eine bessere Vergleichbarkeit der erhobenen HR-Daten zu gewährleisten, wie bereits in Kapitel 3.4.1 detailliert beschrieben. Die z-standardisierten HR-Daten spiegeln bezüglich der Mittelwerte für die betrachteten Phasen das gleiche Verhalten wider, wie die HR-Daten in *bpm*. Dies lässt sich anhand Abbildung 3 bestätigen, welche oben die HR-Phasenmittelwerte in *bpm* und unten die z-standardisierten Mittelwerte der Phasen abbildet. Die Lehr-Phase weist numerisch den höchsten z-standardisierten Mittelwert aller Phasen auf, wobei die Mittelwerte der übrigen Phasen in folgender Reihenfolge abnehmen: Pre-Phase, Postlehr-Phase, Interview-Phase und End-Phase (siehe Tabelle 4). Zudem ist zu erkennen, dass sich die Werte der Standardabweichung der ersten drei Phasen (namentlich Pre-, Lehr- und Postlehr-Phase) sehr ähnlich sind (0.47 bzw. 0.45), sich dabei allerdings von den letzten beiden Phasen (Interview-Phase: 0.37, End-Phase: 0.40) unterscheiden.

**Tabelle 4**

*Deskriptive Statistiken der z-standardisierten HR-Messungen in Einheit der Standardabweichung unter Angabe des Stichprobenumfangs  $N$ , des Mittelwerts  $M$ , der Standardabweichung  $SD$  sowie des Maximums und Minimums für alle fünf Phasen.*

|                 | <i>N</i> | <i>M</i> | <i>SD</i> | <b>Min</b> | <b>Max</b> |
|-----------------|----------|----------|-----------|------------|------------|
| Pre-Phase       | 76       | 0.53     | 0.47      | -0.91      | 1.37       |
| Lehr-Phase      | 76       | 0.89     | 0.47      | -0.80      | 1.81       |
| Postlehr-Phase  | 76       | 0.31     | 0.45      | -0.75      | 1.27       |
| Interview-Phase | 76       | -0.72    | 0.37      | -1.54      | 0.54       |
| End-Phase       | 76       | -1.07    | 0.40      | -2.27      | 0.16       |

Die Lehrphase weist sowohl für die HR-Daten als auch die z-standardisierten HR-Daten ein Maximum in ihrem Mittelwert, verglichen mit den übrigen Phasen, auf. Graphisch lässt sich dies auch anhand von Abbildung 3 belegen. Die z-Werte der HR-Mittelwerte mit *SD* weisen in Abhängigkeit der Bedingung beziehungsweise der Phase ein Maximum in der Lehrphase auf. Zudem wird auf diesem Weg nochmals verdeutlicht, dass die z-Werte der verbleibenden Phasen in der bereits beschriebenen Reihenfolge abnehmen.



**Abbildung 3:** HR-Mittelwerte mit *SD* der fünf betrachteten Phasen a) in *bpm* und b) z-standardisiert.

Zunächst ist es notwendig, die Daten auf das Vorhandensein verschiedener Voraussetzungen zu prüfen, die für eine ANOVA mit Messwiederholung einzuhalten

sind. Dafür wurden die Daten für die fünf Bedingungen, respektive die fünf Phasen, anhand eines Shapiro-Wilk-Test auf Normalverteilung untersucht, wobei ein Signifikanzniveau von 5% festgelegt wurde. Dieser war signifikant für Pre-Phase ( $p < .05$ ), Lehr-Phase ( $p < .05$ ) und Interview-Phase ( $p < .05$ ). Folglich sind die Daten dieser drei Phasen nicht normalverteilt. Im vorliegenden Fall ergibt sich für alle Gruppen beziehungsweise Phasen mit  $N = 76$  eine Stichprobengröße, die  $N \geq 30$  entspricht. Aufgrund dessen können die Daten als normalverteilt angenommen werden, da die ANOVA mit Messwiederholung in diesem Fall robust gegenüber der Normalverteilungsverletzung ist. Die vollständigen Ergebnisse des Shapiro-Wilk-Tests sind Anhang B.2 zu entnehmen.

Zudem sind die Daten auf Ausreißer zu untersuchen. Dies wurde anhand eines Boxplots durchgeführt. Die Post-Phase weist als einzige der fünf Phasen keine(n) Ausreißer auf. In der Pre-Phase weist ein Wert einen leichten Ausreißer mit einem niedrigen Wert auf, in der Lehr-Phase zeichnen sich zwei Werte durch ihren verhältnismäßig geringen Wert aus. Auch die End-Phase weist zwei leichte Ausreißer auf, wobei jeweils ein Wert auffallend niedrig und ein Wert auffallend hoch ist. Die meisten Ausreißer zeigen sich in der Interview-Phase, wobei es sich um fünf leichte und drei extreme Ausreißer handelt. Da es sich bei der Mehrheit der Ausreißer um leichte Ausreißer handelt und die Proband:innen, welche die extremen Ausreißer in der Interview-Phase darstellen, nur innerhalb der betreffenden Phase eine derartige Abweichung vom Rest der Stichprobe aufzeigen, wird keine Stichprobe ausgeschlossen und mit  $N = 76$  Proband:innen fortgefahren. Der Boxplot, auf dessen Grundlage diese Einschätzung getroffen wurde, ist in Anhang B.3 dargestellt.

Als dritte Voraussetzung sind die Daten auf Sphärizität zu überprüfen. Dafür wird der Mauchly Test herangezogen. Dieser war nicht signifikant ( $p > .05$ ), woraufhin Sphärizität angenommen werden kann. Es ist keine Korrektur des Freiheitsgrades im Zuge des Tests der Innersubjekteffekte notwendig. Die exakten Ergebnisse des Mauchly Test sind in Anhang B.4 aufgeführt.

Auf Grundlage dessen wurde eine ANOVA mit Messwiederholung mit gerichteten Kontrasten durchgeführt, wobei ein Signifikanzniveau von  $p \leq .05$  beziehungsweise 5% festgelegt wurde. Da nach dem Mauchly Test eine ausreichende Sphärizität angenommen werden kann, ist für die Betrachtung der Innersubjekteffekte keine Korrektur notwendig. Der Freiheitsgrad der Fehlervarianz lag bei  $df_{\text{Fehler}} = 300$ . Analog folgte für den Freiheitsgrad der aufgeklärten Varianz  $df_{\text{Fehler}} = 4$ . Der  $F$ -Wert

war gegeben durch  $F = 262.69$ . Auf Grundlage dessen zeigte die ANOVA mit Messwiederholung, dass sich die Bedingungen, respektive die fünf untersuchten Phasen, statistisch signifikant voneinander unterscheiden,  $F(4,300) = 262.69, p < .05$ .<sup>6</sup> Die Signifikanz indiziert, dass sich mindestens zwei der fünf Gruppen statistisch signifikant voneinander unterscheiden.

Um zu untersuchen, inwiefern sich die Lehrphase statistisch signifikant jeweils von Pre-, Postlehr-, Interview- und End-Phase unterscheidet, wurden gerichtete Kontraste berechnet, unter der Annahme, dass der Mittelwert der Lehrphase jeweils höher ist. Die Kontrastergebnisse sind in Tabelle 5 aufgeführt. Dabei werden für alle betrachteten Kontraste Kontrastschätzer, der zugehörige Standardfehler (*SE*) sowie der *p*-Wert, der eine Aussage über die Signifikanz der Mittelwertunterschiede trifft, berichtet.

**Tabelle 5**

*Kontrastergebnisse der Lehrphase im Vergleich zu Pre-, Postlehr-, Interview- und Endphase unter Angabe des Kontrastschätzers, des zugehörigen Standardfehlers *SE* und der Signifikanz.*

| Kontrast                  | Kontrastschätzer | <i>SE</i> | Sig.  |
|---------------------------|------------------|-----------|-------|
| Pre- vs. Lehr-Phase       | 0.37             | 0.08      | < .05 |
| Postlehr- vs. Lehr-Phase  | 0.58             | 0.08      | < .05 |
| Interview- vs. Lehr-Phase | 1.62             | 0.07      | < .05 |
| End- vs. Lehr-Phase       | 1.96             | 0.07      | < .05 |

Anhand der Kontrastergebnisse wird deutlich, dass es einen statistisch signifikanten HR-Mittelwertunterschied zwischen Lehr-Phase ( $M = 0.89, SD = 0.47$ ) und Pre-Phase ( $M = 0.53, SD = 0.47$ ) gab, welcher bei 0.37 ( $SE = 0.08$ ),  $p < .05$  lag. Des Weiteren waren auch die verbleibenden Mittelwertunterschiede statistisch signifikant: Dabei zeigte sich zur Postlehr-Phase ( $M = 0.31, SD = 0.45$ ) ein Unterschied von 0.58 ( $SE = 0.08$ ),  $p < .05$ , zur Interview-Phase ( $M = -0.72, SD = 0.37$ ) von 1.62 ( $SE = 0.07$ ),  $p < .05$ , sowie ein Unterschied zur End-Phase ( $M = -1.07, SD = 0.40$ ) von 1.96 ( $SE = 0.07$ ),  $p < .05$ . Folglich sind alle HR-Mittelwertunterschied zwischen der Lehr-Phase und den vier weiteren Phasen der Untersuchung statistisch signifikant.

<sup>6</sup> Der vollständige Test der Innersubjekteffekte ist in Anhang B.5 dokumentiert.



Um die Bedeutsamkeit der signifikanten Mittelwertunterschiede zwischen der Lehr-Phase und den restlichen Phasen einzuordnen, wird Cohen's  $d$  als Maß der Effektstärke berechnet. Dafür ist es notwendig den z-standardisierten HR-Mittelwert der Lehrphase sowie die zugehörige Standardabweichung jeweils mit den anderen vier Phasen zu vergleichen, wie bereits in Kapitel 3.4.4 ausführlich dargelegt wurde.<sup>7</sup> Anhand dessen ergeben sich folgende Effektstärken  $d$ , welche in Tabelle 6 aufgeführt sind.

**Tabelle 6**

*Effektstärke Cohen's  $d$  der HR-Mittelwertunterschiede zwischen Lehr-Phase und den verbleibenden Phasen unter Angabe der Effektstärke  $d$  sowie des dazugehörigen Betrags  $|d|$  und einer Einordnung der Größe des Effekts.*

| Vergleichsphase | $d$   | $ d $ | Effektstärke     |
|-----------------|-------|-------|------------------|
| Pre-Phase       | -0.78 | 0.78  | mittlerer Effekt |
| Postlehr-Phase  | -1.27 | 1.27  | großer Effekt    |
| Interview-Phase | -3.85 | 3.85  | großer Effekt    |
| End-Phase       | -4.51 | 4.51  | großer Effekt    |

Mit Ausnahme des Vergleichs mit der Pre-Phase entspricht das Maß der Effektstärke (Cohen's  $d$ ) einem großen Effekt. Mit einem Wert von  $d = -0.78$  beziehungsweise  $|d| = 0.78$  handelt es sich bei dem signifikanten HR-Mittelwertunterschied zwischen Lehr- und Pre-Phase um einen mittleren Effekt. Der negative Wert für die Effektstärke  $d$  verdeutlicht in allen vier Vergleichen, die durchgeführt wurden, dass die HR-Mittelwerte (z-standardisiert) der Pre-, Postlehr-, Interview- sowie Endphase jeweils geringer sind verglichen mit dem z-standardisierten HR-Mittelwert der Lehrphase. Im Fall der End-Phase liegt der HR-Mittelwert 4.51 Standardabweichungen unterhalb des HR-Mittelwerts der Lehrphase, der z-standardisierte HR-Mittelwert der Pre-Phase ist im Vergleich nur 0.78 Standardabweichungen geringer.

Zusammenfassend konnte anhand der deskriptiven Statistiken gezeigt werden, dass die Lehrphase mit  $M = 100.41 \text{ bpm}$ ,  $SD = 14.33 \text{ bpm}$  beziehungsweise für die z-standardisierten Werte  $M = 0.89$ ,  $SD = 0.47$  den größten HR-Mittelwert im Vergleich zu den anderen Phasen der Betrachtung aufweist. Die Existenz von statistisch signifikanten HR-Mittelwertunterschieden zwischen den Gruppen, respektive Phasen,

<sup>7</sup> Eine Einordnung der betreffenden Werte und die korrespondierende Aussagekraft über die Stärke des Effekts wurde auch in Kapitel 3.4.4 vorgenommen.

wurde durch die anschließende ANOVA mit Messwiederholung nachgewiesen,  $F(4,300) = 262.69, p < .05$ . Die Berechnung von paarweise gerichteten Kontrasten zwischen der Lehr-Phase und jeweils der Pre-, Postlehr-, Interview- und End-Phase zeigte auf, dass die betreffenden HR-Mittelwertunterschiede statistisch signifikant sind, wobei die vier Kontraste einen  $p$ -Wert von  $p < .05$  haben. Durch die Effektstärke Cohens  $d$  wurde deutlich, dass es sich bei den statistisch signifikanten HR-Mittelwertunterschieden zwischen Pre- und Lehr-Phase um einen mittleren Effekt und für die verbleibenden drei Unterschiede um große Effekte handelt.

## 5 Diskussion

Im Zuge der vorliegenden Arbeit wurde im Rahmen einer quasi-experimentellen Querschnittstudie untersucht, inwiefern sich der Puls respektive die HR von Lehrkräften als Indikator für Stress während des Unterrichts eignet. Zu diesem Zweck wurde die HR von 76 Proband:innen über einen Zeitraum von maximal zwei Stunden aufgezeichnet und fünf zehnminütige Phasen unterschiedlicher mentaler Anforderung miteinander verglichen. Dabei zeichnet sich eine dieser Phasen durch eine simulierte Unterrichtseinheit mit Unterrichtsstörungen aus. Diese sollte exemplarisch für den realen Unterricht und den damit einhergehenden Stress stehen. In einer Studie verdeutlichten Junker et al. (2021), dass verschiedene Arten von Unterrichtsstörungen innerhalb einer realen Unterrichtsstunde zu einer signifikanten Erhöhung der HR führen. Dabei wurde ein 7-Kanal-EKG für die Aufzeichnung der HR verwendet. Des Weiteren zeigten Runge et al. (2020), dass sich eine Kombination verschiedener Vitalparameter, unter anderem auch eine erhöhte HR, die mit einem Fitbit Fitnessstracker erhoben wurden, als Indikator für Stress im Lehrberuf eignen. Es ist allerdings anzumerken, dass die Studie von Runge et al. (2020) nur mit vier Proband:innen durchgeführt wurde und damit nur einen sehr geringen Stichprobenumfang aufweist. Daher kann sie nicht als repräsentativ für den gesamten Lehrberuf erachtet werden.

Auf Grundlage dieser Studienergebnisse wurde die Hypothese aufgestellt, dass der Mittelwert der Lehrphase, in der die simulierte Unterrichtseinheit stattfand, am größten ist. Zusätzlich wurde angenommen, dass sich der HR-Mittelwert der Lehrphase dabei statistisch signifikant von den weiteren Phasen unterschiedlicher Belastung – namentlich der Pre-Phase, Postlehr-Phase, Interview-Phase und der End-Phase – unterscheidet. Insbesondere lag der Fokus darauf, inwiefern dies mit einer Fitbit Charge 4 Smartwatch nachweisbar ist, welche durch ihren kompakten Aufbau, ihr alltägliches Design und den geringen Kostenfaktor einen Vorteil gegenüber medizinischen Messgeräten mit sich bringt.

Bereits durch die Mittelwertberechnung je Phase über die gesamte Stichprobe konnte sowohl für die z-standardisierten als auch die HR-Daten in *bpm* bestätigt werden, dass die Lehr-Phase mit  $M = 100.41 \text{ bpm}$ ,  $SD = 14.33 \text{ bpm}$  beziehungsweise  $M = 0.89$ ,  $SD = 0.47$  den höchsten Mittelwert aller fünf Phasen aufweist. Auf Grundlage dessen lässt sich der erste Teil der aufgestellten Hypothese verifizieren. Das anhand einer einfaktoriellen Varianzanalyse mit Messwiederholung festgestellte Vorliegen statistisch

signifikanter Mittelwertunterschiede zwischen den fünf Phasen, wurde unter der Berechnung von gerichteten Kontrasten bestätigt. Diese zeigten, dass die Unterschiede im HR-Mittelwert zwischen der Lehr-Phase und den übrigen Phasen jeweils statistisch signifikant sind. Somit sind die festgestellten HR-Mittelwertunterschiede systematischer Natur und beruhen nicht auf zufälligen Gegebenheiten. Die Effektstärke  $d$  nach Cohen belegt zusätzlich, dass es sich bei den vorliegenden statistisch signifikanten Mittelwertunterschieden zwischen der Lehr-Phase und der Postlehr-, Interview- und End-Phase um große Effekte handelt. Der Mittelwertunterschied zwischen Lehr- und Pre-Phase stellt einen mittleren Effekt dar. Auf Grundlage dieser Ergebnisse der durchgeführten Varianzanalyse mit Messwiederholung mit Kontrasten kann die zu Beginn aufgestellte Hypothese, dass der HR-Mittelwert in der Lehr-Phase am größten ist, wobei der HR-Mittelwertunterschied zwischen der Lehr-Phase und allen anderen Phasen jeweils signifikant ist, vollumfänglich verifiziert werden.

Damit entsprechen die Ergebnisse den Erwartungen, die sich anhand der bereits bestehenden empirischen Befunde aus den Untersuchungen von Runge et al. (2020) und Junker et al. (2021) ergaben und als Ausgangspunkte für die untersuchte Forschungsfrage dienten. Die Lehr-Phase, welche aufgrund der simulierten Lehreinheit mit gezielten Unterrichtsstörungen im Vergleich zum Beantworten von Fragebögen oder Interviews die Phase mit der objektiv größten Beanspruchung darstellt, weist den höchsten (z-standardisierten) HR-Mittelwert auf. Auch die verbleibenden Phasen bilden einen Trend ab, der zu erwarten war. Im Gegensatz dazu betrachteten Junker et al. (2021) keine Mittelwerte der HR über einen bestimmten Zeitraum, sondern gezielt Messzeitpunkte innerhalb der Gesamterhebung, in welchen die HR mindestens zwei Standardabweichungen oberhalb der durchschnittlichen HR lag und verknüpften diese mit Ereignissen innerhalb des Unterrichtsgeschehens. In diesem Zuge ist zu betonen, dass bei der Auswertung der HR-Mittelwerte der fünf Phasen sowie der Unterschiede zueinander, keine absoluten Werte der HR betrachtet oder berichtet werden, sondern relative Abweichungen zur mittleren HR eines:r jeden Proband:in untersucht werden, welche dann den Phasenmittelwert bilden. Dementsprechend lag die HR für die untersuchte Stichprobe in der Lehr-Phase 0.89 Standardabweichungen ( $SD = 0.47$ ) über der durchschnittlichen HR für den Zeitraum der Erhebung, wohingegen die HR in der End-Phase 1.07 Standardabweichungen ( $SD = 0.40$ ) unter der durchschnittlichen HR lag und damit auf eine deutlich entspanntere respektive weniger stressende Situation hindeutet. Da die End-Phase die letzten zehn Minuten der Erhebung umfasst, scheint der

deutliche Rückgang in der HR offensichtlich. Der Unterschied der HR-Mittelwerte wird nochmals durch die Betrachtung der Effektstärke  $d$  verdeutlicht. Diese entspricht für den Mittelwertunterschied zwischen Lehr- und End-Phase einem Wert von 4.51 und damit einem großen Effekt. Dies lässt sich nach Cohen (1988) auch folgendermaßen interpretieren, dass nahezu alle HR-Mittelwerte der Proband:innen aus der End-Phase einen geringeren Wert aufweisen, als der (z-standardisierte) HR-Mittelwert der Lehr-Phase.<sup>8</sup> Dies belegt das gesteigerte Anspannungs- und Stressniveau der Versuchspersonen während der Lehr-Phase, aber auch den im Vergleich dazu deutlich ruhigeren Zustand der Proband:innen während der End-Phase. Ähnliches gilt auch für den Mittelwert-Unterschied zwischen Interview- und Lehr-Phase mit einer Effektstärke von  $d = 3.85$ . Dies entspricht wiederum einem großen Effekt und sagt zudem aus, dass 99.99% der HR-Mittelwerte der Proband:innen in der Interview-Phase unter dem HR-Mittelwert der Lehr-Phase liegen. Analog haben 89.80% der HR-Mittelwerte der Studienteilnehmenden, welche der Postlehr-Phase zuzuordnen sind, einen geringeren Wert im Vergleich zum (z-standardisierten) HR-Mittelwert der Lehr-Phase. Dadurch wird für beide Fälle sichtbar, wie deutlich die statistisch signifikanten HR-Mittelwertunterschiede tatsächlich sind und große Effekte darstellen. Der Mittelwertunterschied zwischen Pre- und Lehr-Phase entspricht im Gegensatz dazu lediglich einem mittleren Effekt, so dass statistisch betrachtet nur 78.23% der Werte der Pre-Phase kleiner sind als der HR-Mittelwert der Lehr-Phase. Ein möglicher Grund dafür mag darin gegeben sein, dass ein großer Teil der Proband:innen aufgrund von Aufregung und der Ungewissheit bezüglich der Studienteilnahme bereits mit einer erhöhten HR in die Erhebung starteten, denn in der Pre-Phase liegt der HR-Mittelwert bereits 0.53 Standardabweichungen ( $SD = 0.49$ ) über der durchschnittlichen HR. Daher unterscheidet sich die Pre-Phase im Mittel weniger deutlich vom Mittel der Lehr-Phase, verglichen mit den verbleibenden drei Phasen nach der simulierten Unterrichtseinheit.

In Bezug auf die Gütekriterien der quantitativen Forschung kann das Vorliegen von Reliabilität und Objektivität bestätigt werden (Döring & Bortz, 2016). Dadurch, dass die Erhebungen alle unter gleichen Bedingungen in denselben Räumlichkeiten unter demselben Ablauf stattfanden, ist ein hohes Maß an Reliabilität gewährleistet. Zudem ist auch ein hohes Maß an Objektivität gegeben, welches zum einen die physiologische

---

<sup>8</sup> Bei Umrechnung von Cohen's  $d$  in Cohen's  $U_3$  ergibt sich bei  $d = 4.51$  ein Wert von  $U_3 = 99.997\%$ . Auf zwei Dezimalstellen gerundet gilt  $U_3 = 100.00\%$ . Eine Umrechnung aller Effektstärkenwerte  $d$  in Cohen's  $U_3$  ist Anhang B.6 zu entnehmen.

Messung als Methode der Datenerhebung mit sich bringt und zum anderen auch durch die teilautomatisierte, computerbasierte Auswertung unterstrichen wird. Lediglich die Validität der Erhebung ist eingeschränkt, da die Messung absoluter HR-Werte durch die Fitbit Charge 4 einer Unsicherheit unterliegt, welche im Folgenden noch erörtert wird. Des Weiteren ist anzumerken, dass bei der vorliegenden Stichprobe Lehrkräfte und Lehramtsauszubildende für das Gymnasium überrepräsentiert sind, da sie im Deutschen Bundesdurchschnitt für das Schuljahr 2021/2022 circa 25.5% aller Lehrkräfte ausmachten, wobei es in der untersuchten Stichprobe 72.37% sind (Statistisches Bundesamt, 2022). Um repräsentativ für Deutschland zu sein, müsste die Mehrheit der Proband:innen aus dem Grundschullehramt stammen.

Die empirischen Befunde der vorliegenden Arbeit gehen mit gewissen Limitationen einher, welche primär im Kontext mit der verwendeten Methode stehen. Dabei steht das Messinstrument, die Fitbit Charge 4 Smartwatch, im Vordergrund. Wie bereits in Kapitel 2.5 anhand einer Auswahl von Studien dargelegt, tendieren Smartwatches der Marke Fitbit im Allgemeinen, aber auch das Modell Fitbit Charge 4 im Speziellen dazu, die HR zu unterschätzen (Fuller et al., 2020; Jachymek et al., 2021). Daher sind die absoluten HR-Werte, welche von der Smartwatch in *bpm* aufgezeichnet wurden, vor der weiteren Verwendung kritisch zu betrachten. Für die vorliegende Erhebung stehen die relativen Unterschiede im Fokus des Interesses, welche gleichermaßen von der Unterschätzung der Messwerte beeinflusst werden. Somit ist davon auszugehen, dass die untersuchten Unterschiede mit geringen Einschränkungen den tatsächlich vorliegenden Trend abbilden. Um dennoch eine genauere respektive präzisere Messung der HR vor dem Hintergrund einer beispielsweise detaillierteren Auswertung gewisser Zeitintervalle oder Ereignisse zu gewährleisten, bietet ein Wechsel der Smartwatch zu einem höherpreisigen Modell der nächst höheren Preisklasse eine adäquate Möglichkeit. Es ist davon auszugehen, dass mit dem Preis auch die Präzision der HR-Messung steigt (Hajj-Boustros et al., 2022). Von einer zunehmenden Ungenauigkeit der HR-Messung mit steigender Belastung, wie sie unter anderem in den Untersuchungen von Jo et al. (2016), Fuller et al. (2020) und Jachymek et al. (2021) nachgewiesen wurden, kann im Rahmen der vorliegenden Studie nicht ausgegangen werden. Die exemplarisch angeführten Untersuchungen setzten die Verminderung der Messgenauigkeit bei hoher physischer Belastung in Zusammenhang mit bestimmten Muskelkontraktionen und extremer Schweißbildung. Diese beeinflussen die HR-Messung mittels PPG-Technologie, gehen im notwendigen Maße allerdings aus-

schließlich mit bestimmten Formen körperlicher Ertüchtigung einher. Die Messungen der maximalen HR-Werte, welche für die betrachtete Stichprobe im Bereich von 145 *bpm* liegen, besitzen dem entsprechend keine gesteigerte Ungenauigkeit im Vergleich zu niedrigeren HR-Messwerten. Allerdings unterliegen die erhobenen HR-Daten gewissen Fehlern, die auf menschlichen Handlungen basieren. Bei genauerer Betrachtung der Rohdaten fällt auf, dass sich die Anzahl an gemessenen Datenpunkten je Phase je Proband:in deutlich unterscheiden, obwohl jede der fünf betrachteten Phasen auf exakt zehn Minuten beschränkt ist, um eine größtmögliche Vergleichbarkeit der Daten zu gewährleisten. Außerdem liegen für einzelne Proband:innen Intervalle von bis zu 74 Sekunden innerhalb der zehnminütigen Phasen vor, in denen keine HR-Messung stattfand, obwohl die Messung in Abständen von einer beziehungsweise fünf Sekunden ablaufen sollte (Jachymek et al., 2021). Es ist anzunehmen, dass die Proband:innen durch unbewusstes und wohl stressbedingtes, wiederholtes Berühren und Bewegen der Fitbit Charge 4 Smartwatch den ursprünglichen Sitz des Geräts verändert haben, welcher allerdings essenziell für eine präzise Messung ist, wie in Kapitel 3.2 bereits erläutert wurde. Da die HR-Mittelwerte der Phasen je Proband:in auf einer unterschiedlichen Anzahl an Einzelmesswerten basieren, führt dies zu einer geringfügigen Reduktion der Vergleichbarkeit, welcher aufgrund der Stichprobengröße und dem verhältnismäßig geringen Anteil an Proband:innen mit Intervallen fehlender Messdaten kein großer Einfluss auf die Ergebnisse der Studie beigemessen wird. Derartige Fehler sind in Humanstudien jedoch kaum vermeidbar und sind im vorliegenden Fall vergleichbar mit den Fehlern in den Daten, welche Runge et al. (2020) im Rahmen ihrer Feldstudie unter der Verwendung von Fitbit Smartwatches berichteten. Nach Döring und Bortz (2016) lassen sich solche Verzerrungen der Daten auch als Bewegungsartefakte einordnen, die als mögliche Probleme bei physiologischen Messungen auftreten können.

Außerdem ist es wichtig anzumerken, dass es neben der HR mit der HRV einen weiteren physischen Indikator oder eine Messgröße gibt, die sich nach dem aktuellen Forschungsstand besser eignet als die HR, um physischen und mentalen Stress nachzuweisen und zu untersuchen. Die HRV beschreibt dabei die Änderungen in den zeitlichen Abständen zwischen zwei aufeinanderfolgenden Herzschlägen über einen gewissen Betrachtungszeitraum (Löllgen, 1999; Taelmann et al., 2010). Zudem handelt es sich bei der HRV um eine Messgröße der autonomen Funktion sowie der neurovegetativen Aktivität des Herzens. Damit bringt die HRV den Vorteil gegenüber der HR mit sich, dass sie nicht nur eine Aussage über die Intensität der Beanspruchung

gibt, sondern zugleich auch Informationen über die Qualität der Regulation des Herz-Kreislaufsystems beinhaltet (Hottenrott, 2007). Durch eine aufeinander abgestimmte Interaktion des parasympathischen und sympathischen Nervensystems, passt sich die HRV an Veränderungen der Umwelt an und reagiert somit fortlaufend auf Zeichen des eigenen Organismus (Hottenrott, 2007). Dementsprechend wird der sympathische Teil des autonomen Nervensystems vermehrt aktiviert, sobald eine Person einem Stressor ausgesetzt ist (Taelmann et al., 2010). Somit stellt die HRV einen quantitativen Parameter der autonomen Regulation dar und findet daher oftmals Anwendung um zu untersuchen, inwiefern das autonome Nervensystem auf körperliche und psychische Beanspruchung reagiert (Taelmann et al., 2010). Die Verwendbarkeit der HRV als Observable für mentalen Stress belegten Taelmann et al. (2010) im Rahmen einer Laborstudie, in der psychische und physische Beanspruchung in Anlehnung an Büroarbeit simuliert wurde. Außerdem legte diese Untersuchung dar, dass anhand der Charakteristika der HRV auch zwischen Ruhezustand und Situationen körperlicher respektive mentaler Belastung unterschieden werden kann. Die HRV verändert sich dabei folgendermaßen: die Variabilität der Dauer zwischen zwei Herzschlägen ist am größten, wenn sich eine Person in Ruhe befindet und reduziert sich mit zunehmender Belastung, wobei unter anderem auch chronischer Stress zu einer Einschränkung der HRV führen kann (Hottenrott, 2007). Viele moderne Smartwatch Wearables sind in der Lage neben der HR auch die HRV zu messen und aufzuzeichnen, allerdings trifft dies nicht auf alle Modelle zu. In einer Studie mit Fitbit Smartwatches untersuchten Chalmers et al. (2021), inwiefern sich die erhobenen Daten der HRV und HR eignen, um darin abzulesen, ob Proband:innen Stress ausgesetzt sind oder waren. Dabei kam die Untersuchung zu dem Ergebnis, dass insbesondere die HRV ein gutes Maß für die Erfassung von körperlichem und mentalem Stress ist, wenn man die HRV im Ruhezustand als Bezugspunkt mit einbezieht. Auch die Fitbit Charge 4, welche in der vorliegenden Arbeit für die Messung der HR als Stressindikator verwendet wurde, kann die HRV aufzeichnen, welche über die Fitbit App eingesehen werden kann (Fitbit, 2023). Allerdings ist es dafür notwendig, dass die Smartwatch für einen Zeitraum von mindestens 24 Stunden getragen wird, um die Parameter zu erfassen (Fitbit, 2023). Da die Erhebungen innerhalb der vorliegenden Studie einer zeitlichen Begrenzung von maximal zwei Stunden unterlagen, war es nicht möglich die HRV als potentiellen Stressindikator zu erheben und im Rahmen dieser Arbeit auszuwerten. Allgemein betrachtet ist es eher kompliziert einen derartig langen Erhebungszeitraum im Zusammenhang mit einer Laborstudie zu realisieren.



Gegebenenfalls bieten aktuellere Modelle im Zuge des technischen Fortschritts die Möglichkeit der HRV-Erfassung, ohne dass die Smartwatch mindestens einen Tag lang Daten aufzeichnen muss.

Die Vorzüge der Untersuchungen sowie der dazugehörigen Ergebnisse, die im Zuge dieser Arbeit dargelegt wurden, liegen insbesondere in der Art der Datenerhebung. Physiologische Messungen weisen nach Döring und Bortz (2016) insgesamt fünf Vorteile gegenüber anderen statistischen Datenerhebungsverfahren der Sozialforschung auf, wie beispielsweise Selbstauskünfte durch Persönlichkeitstests oder auch Interviews. Zum einen werden durch automatisierte, physiologische Messungen Gedächtnisfehler vermieden, die vermehrt bei Erhebungsmethoden auftreten, welche auf Selbstauskünften der Proband:innen basieren und oftmals die Daten verfälschen. Außerdem wird nach Döring und Bortz (2016) die Möglichkeit der direkten Einflussnahme auf die Ergebnisse der Datenerhebung seitens der Proband:innen auf ein Minimum reduziert. Die automatische Messung der HR durch die Fitbit Charge 4 kann im gewählten Setting kaum manipuliert werden, da es zudem auch keine Möglichkeit gibt, die HR künstlich zu erhöhen oder zu senken. Zum anderen bieten physiologische Messungen die Grundlage körperliche Veränderungsprozesse und Vorgänge abzubilden, welche im Rahmen einer Selbstauskunft nur schwer wiedergegeben werden können, sofern sie überhaupt durch die Versuchspersonen wahrgenommen werden. So werden im vorliegenden Fall auch geringe Veränderungen der HR aufgezeichnet, die die Proband:innen selbst nicht wahrnehmen. Des Weiteren können durch derartige Messungen physiologische Indikatoren in einem zeitlichen Verlauf dargestellt werden und damit zugleich Teilprozesse der Verarbeitung von äußeren und inneren Einflüssen betrachtet werden (Döring & Bortz, 2016). Dies ist im Rahmen anderer Datenerhebungsmethoden nicht realisierbar, auch weil bei diesen oft nur das Ergebnis von individuellen Verarbeitungsprozessen einer Versuchsperson erhoben wird.

Dadurch, dass als Messinstrument eine Fitbit Charge 4 Smartwatch verwendet wurde, konnten bestimmte Nachteile, welche auch physiologische Messungen mit sich bringen, deutlich reduziert werden. Die Verfügbarkeit und Beschaffbarkeit einer solchen Smartwatch stellt sowohl durch die immer weiter ansteigende Beliebtheit innerhalb der Allgemeinbevölkerung sowie den geringen Kostenfaktor im Vergleich zu medizinischen Gerätschaften der HR-Messung, wie bereits ausführlich thematisiert wurde, keinerlei Problem dar (Döring & Bortz, 2016). Auch das notwendige Wissen und die Expertise über die Verwendung des Messinstrumentes und über die Auswertung, die nach Döring

und Bortz (2016) einen gesteigerten Anspruch der physiologischen Datenerhebung gegenüber anderen Erhebungsverfahren darstellen, sind durch die Wahl des Messgeräts auf ein Minimum reduziert. Die Bedienoberfläche der Fitbit Charge 4 ist an bekannte Systeme von Smartphones angelehnt und daher auch sehr intuitiv oder kann sich anhand einer simpel gehaltenen Bedienungsanleitung angeeignet werden. Auch die Methodik zur Auswertung der erhobenen HR-Daten ist in Bezug auf den Aufwand nicht vergleichbar mit der Analyse von Daten, die beispielsweise ein EKG liefert. Dies lässt sich auch damit begründen, dass die Fitbit Charge 4 bereits direkt den für die weitere Auswertung relevanten Parameter in Abhängigkeit der Zeit ausgibt. Da sich die Smartwatch in ihrer Größe und der Art des Tragens nicht von einer gewöhnlich Armbanduhr unterscheidet, kann diese somit für Proband:innen als zumutbar und kaum einschränkend definiert werden (Döring & Bortz, 2016). Dennoch gibt es nach Döring und Bortz (2016) zwei Gesichtspunkte, unter denen physiologische Messungen kritisch zu betrachten sind. So werden einerseits auch im vorliegenden Fall lediglich die physiologischen Parameter betrachtet und analysiert, wobei die subjektiven Erfahrungen der Vorgänge aus der Sicht der Proband:innen nicht berücksichtigt werden. Andererseits merken Döring und Bortz (2016) an, dass physiologische Messungen die Reaktionen auf ein bestimmtes Szenario abbilden, welches zu untersuchen ist. Allerdings werden diese oft, wie auch im Falle dieser Arbeit, unter Laborbedingungen nachgestellt und den Teilnehmenden ist zudem bewusst, dass sie Teil einer Datenerhebung im Rahmen einer Studie sind. Daher ist auch hier zu hinterfragen, inwiefern sich die Ergebnisse auf den Alltag übertragen lassen und in welchem Maß gewisse physiologische Indikatoren allein durch das Wissen, dass es sich um eine Teilnahme an einer Studie handelt, unbewusst beeinflusst werden. Es ist angezeigt, sich letztere Frage auch in Bezug auf die hier betrachteten HR-Messungen zu stellen. Denn wie bereits erläutert, zeigt auch die Pre-Phase, in der die Proband:innen eine Einführung in das Studiengeschehen bekommen und noch keinen expliziten Stressoren ausgesetzt sind, eine erhöhte HR und auch im Vergleich mit den übrigen Phasen ist der HR-Mittelwertunterschied zur Lehr-Phase am geringsten. Außerdem bestätigt der Wert für die Effektstärke  $d$  nach Cohen, dass es sich bei dem Mittelwertunterschied zwischen Pre- und Lehr-Phase um einen mittleren Effekt handelt, wohingegen die weiteren HR-Mittelwertunterschiede der Betrachtung große Effekte darstellen. Dies lässt den Schluss zu, dass bereits das Studiensetting an sich und das Wissen darüber, dass eine ungewohnte Situation bevor steht, bei den Proband:innen zu einem stressbedingten Anstieg der HR führt. Dementsprechend bleibt auch die Frage offen, welcher Anteil des Stresses, der in

der Lehr-Phase zu einem signifikanten Anstieg der mittleren HR führte, tatsächlich auf die Unterrichtsstörungen und Stressoren während der simulierten Lehreinheit zurückzuführen ist. Ein Teil ist ohne Frage auch auf das gesamte Setting während der Lehr-Phase, inklusive Eyetracking-Brille, mehrerer Kameras und simulierter Schulklasse zurückzuführen. Es ist auch davon auszugehen, dass nach der Lehr-Phase der Einfluss der Studiensituation auf den Verlauf der HR nachlässt und im Laufe der maximal zweistündigen Erhebung immer weniger ins Gewicht fällt, da die unbekannte Situation bewältigt ist. Ungeachtet davon stellt dies die Richtigkeit und Aussagekraft der im Rahmen dieser Arbeit dargelegten Ergebnisse nicht in Frage. Trotz der belegten Ungenauigkeiten in der HR-Messung der Fitbit Charge 4 werden Trends in der HR richtig abgebildet. Daher ist es möglich anhand von HR-Mittelwertunterschieden Phasen zu unterscheiden, in denen Proband:innen jeweils unterschiedlich starker Beanspruchung in Form von Stress ausgesetzt sind.

## 6 Fazit und Implikationen für die Praxis

Im Rahmen der vorliegenden Arbeit wurde die Verwendbarkeit einer Fitbit Smartwatch zur Messung von HR-Mittelwertunterschieden zwischen verschiedenen Zeitphasen vor, während und nach einer simulierten Unterrichtseinheit bei Lehrkräften untersucht. Zu diesem Zweck wurde die HR innerhalb eines Laborsettings über einen Zeitraum von maximal zwei Stunden mittels einer Fitbit Charge 4 aufgezeichnet. Dabei wurden fünf jeweils zehnminütige Zeitintervalle bezüglich ihrer Mittelwertunterschiede miteinander verglichen, um festzustellen, inwiefern sich Unterricht inklusive Störungen und der damit verbundene Stress in der HR abbilden lassen. Daher wurde in dieser Lehr-Phase die höchste mittlere HR erwartet. Die weiteren vier Phasen unterscheiden sich untereinander und auch von der Lehr-Phase in ihrer psychischen und physischen Anforderung an die Proband:innen. Es konnte bestätigt werden, dass die Lehr-Phase den höchsten Mittelwert aller Phasen aufweist. Zudem zeigte sich, dass die somit paarweise entstehenden Mittelwertunterschiede zwischen der Lehr-Phase und jeder der übrigen vier Phasen statistisch signifikant sind. Auf Grundlage dessen lässt sich schlussfolgern, dass die Fitbit Charge 4 dazu verwendet werden kann, um im gewählten Setting Mittelwertunterschiede der HR zwischen verschiedenen Zeitphasen abzubilden.

Angesichts der Erkenntnis, dass sich Smartwatches dazu eignen, signifikante HR-Mittelwertunterschiede in Phasen unterschiedlich starker Belastung beziehungsweise Beanspruchung zu messen, bietet sich die Möglichkeit, diesen Wissenszuwachs in die Aus- und Weiterbildung von Lehrkräften einfließen zu lassen. Denn insbesondere vor dem Hintergrund der gesteigerten Jobaufgabe im Lehrberuf, welche von einer Vielzahl der Betroffenen mit einer hohen mentalen Belastung und dem Stressniveau begründet wird, kann die gezielte Verwendung von Smartwatches zur individuellen Überwachung physiologischer Anzeichen von Stress zu einer Verbesserung der Situation beitragen. Dafür ist es in erster Linie notwendig, Lehrkräfte und Lehramtsanwärter:innen zu sensibilisieren, dass sich das individuelle Stressempfinden anhand messbarer, körperlicher Parameter wie der HR bestätigen oder auch antizipieren lässt. Des Weiteren ist es wichtig, Lehrer:innen auch mit den Optionen und Vorteilen der individuellen Stressüberwachung mittels Smartwatches oder Fitnesstrackern vertraut zu machen, ihnen die Verwendung solcher Geräte naheulegen und ihnen auch das notwendige theoretische Wissen zu vermitteln. Dies kann in Form von Weiterbildungen für aktive Lehrkräfte oder auch in Seminaren während des Studiums realisiert werden. Dabei wäre es ratsam,

angehenden Lehrkräften spätestens in direkter Vorbereitung auf das Referendariat das Thema der individuellen Stressüberwachung durch Beobachtung der HR näher zu bringen. Der entscheidende Vorteil an dieser Art des Stressmanagements ist die bereits mehrfach erwähnte Individualität. Jede Lehrperson kann so überprüfen, welche Situationen, Aufgaben oder auch Tage sich aufgrund einer erhöhten HR als besonders stressig oder herausfordernd auszeichnen. Schließlich reagiert jede Person aufgrund von Erfahrungen und Prägung unterschiedlich auf dieselben Stressoren, wie bereits im Rahmen der Lehrer:innenstressmodelle in Kapitel 2.3.1 dargelegt. Auch anhand der HR-Mittelwerte, die in dieser Arbeit betrachtet wurden, zeigten sich bereits deutliche Unterschiede: Für eine:n exemplarische:n Proband:in stellte die Lehrsituation augenscheinlich keinen bedeutenden Stressor dar und die HR lag deutlich unter dem individuellen Durchschnitt, wohingegen die Interview-Phase eine merklich erhöhte HR aufwies. Auf dieser Grundlage hat jede Lehrkraft die Möglichkeit, sich den eigenen Stressoren bewusst zu werden und kann sich dementsprechend auch auf diese vorbereiten, wenn ihr Auftreten vorherzusehen ist. Um Stressoren erfolgreich zu antizipieren und zu bewältigen, können (angehende) Lehrkräfte ebenfalls im Rahmen von Weiterbildungen während und nach dem Studium Stressbewältigungsmethoden erlernen, welche in Akutsituationen Abhilfe schaffen, aber auch die individuellen Coping Mechanismen und die Resilienz stärken. Auch in diesem Zuge ist es wichtig, (angehende) Lehrer:innen mit einem großen Spektrum an Stressbewältigungsverfahren vertraut zu machen und sie vor allem dazu anzuregen, die Verfahren aktiv auszuprobieren. Denn wie bereits im transaktionalen Stressmodell verankert, sind Coping Mechanismen und damit auch die Stressbewältigungsverfahren, die zu einer effektiven Linderung von psychischen und physischen Stresssymptomen führen, individuell. In diversen Studien zeigten sich unter anderem Sport, Selbstfürsorge, ein gesunder Lebensstil, Ablenkung aber auch autogenes Training, progressive Muskelentspannung und emotionales Nacharbeiten von Stresssituationen als effektive Methoden zur Stressbewältigung (Runge et al., 2020; Seefeldt, 2000). Es sollte insbesondere darauf geachtet werden, dass auch Methoden vermittelt werden, die dazu beitragen, dass sich Lehrpersonen ihrer Selbstwirksamkeit im schulischen Kontext bewusstwerden und diese zusätzlich gestärkt wird. Auf diesem Wege werden Lehrer:innen im Umgang mit Stressoren gestärkt und sind damit weniger anfällig für Überforderung sowie den daraus resultierenden Stress und Burnout (Dicke et al., 2014; Friedman, 2003). Eine fast schon utopische Idealvorstellung, wie ein offener Umgang mit Stress im Rahmen des Lehralltags aussehen kann, beschreiben Runge et al.

(2020) anhand von Erfahrungen im Rahmen ihrer Studie: Dabei gehen Lehrer:innen, proaktiv auf Schüler:innen zu und informieren sie, wenn anhand der Vitalparameter zu erkennen ist, dass sie vermehrt gestresst sind, um auf diesem Wege die Schüler:innen um Rücksichtnahme zu bitten und im Voraus stressbedingtes Verhalten zu erklären.

Die aktuelle Forschung zeigt, dass neben der HR auch andere physiologische Parameter oder die Kombination mehrerer Parameter zur individuellen Stressüberwachung dienen können (Runge et al., 2020). Daher sollten Lehrpersonen alle Möglichkeiten und Indikatoren verwenden, die moderne Smartwatches auch mit individuellen Gesundheitsmetriken bieten, um sich ihrem persönlichen Stressniveau bewusst zu werden und möglichen Folgen wie Burnout durch frühzeitige Prävention zuvorzukommen.

## Literaturverzeichnis

- Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R. & Trautwein, U. (2018). Student Misbehavior and Teacher Wellbeing: Testing the mediating role of the teacher-student relationship. *Learning and Instruction*, 58, 126–136.  
<https://doi.org/10.1016/j.learninstruc.2018.05.006>
- Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), R1–R39.  
<https://doi.org/10.1088/0967-3334/28/3/r01>
- Aloe, A. M., Shisler, S., Norris, B., Nickerson, A. B. & Rinker, T. (2014). A multivariate meta-analysis of student misbehavior and teacher burnout. *Educational Research Review*, 12, 30–44.  
<https://doi.org/10.1016/j.edurev.2014.05.003>
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollständig überarbeitete und erweiterte Auflage). Springer.  
<https://doi.org/10.1007/978-3-642-12770-0>
- Boyle, G. J., Borg, M. G., Falzon, J. & Baglioni, A. J. (1995). A structural model of the dimensions of teacher stress. *British Journal of Educational Psychology*, 65(1), 49–67. <https://doi.org/10.1111/j.2044-8279.1995.tb01130.x>
- Chalmers, T. A., Hickey, B. A., Newton, P. J., Lin, C., Sibbritt, D., McLachlan, C. S., Clifton-Bligh, R., Morley, J. W. & Lal, S. (2021). Stress Watch: The use of heart rate and heart rate variability to detect stress: A pilot study using smart watch wearables. *Sensors*, 22(1), 151 – 160. <https://doi.org/10.3390/s22010151>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Lawrence Erlbaum Associates.
- Custodis, F., Reil, J., Schirmer, S. H., Adam, O., Möhlenkamp, S., Laufs, U. & Böhm, M. (2014). Herzfrequenz: klinische Variable und Risikomarker. *Deutsche Medizinische Wochenschrift*, 139(33), 1661–1672. <https://doi.org/10.1055/s-0034-1370223>

- Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A. & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and Emotional exhaustion: A Moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, 106(2), 569–583. <https://doi.org/10.1037/a0035504>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I. & Tutz, G. (2016). *Statistik: Der Weg zur Datenanalyse* (8. Auflage). Springer. <https://doi.org/10.1007/978-3-662-50372-0>
- Feehan, L. M., Geldman, J., Sayre, E. C., Park, C., Ezzat, A. M., Yoo, J. Y., Hamilton, C. B. & Li, L. (2018). Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. *JMIR Mhealth and Uhealth*, 6(8), e10527. <https://doi.org/10.2196/10527>
- Fitbit. (2023). *Fitbit Charge 4: Bedienungsanleitung Version 2.0*. Abgerufen am 08. August 2023, von [https://help.fitbit.com/manuals/manual\\_charge\\_4\\_de.pdf](https://help.fitbit.com/manuals/manual_charge_4_de.pdf)
- Fitbit Help. (o. D.-a). *What factors can affect my heart-rate reading in my Fitbit device?* Abgerufen am 01. September 2023 , von [https://help.fitbit.com/articles/en\\_US/Help\\_article/1582.htm](https://help.fitbit.com/articles/en_US/Help_article/1582.htm)
- Fitbit Help. (o. D.-b). *Wie tracke ich die Herzfrequenz mit meinem Fitbitgerät?* Abgerufen am 01. September 2023, von [https://help.fitbit.com/articles/de/Help\\_article/1565.htm](https://help.fitbit.com/articles/de/Help_article/1565.htm)
- Fisher, M. (2011). Factors Influencing Stress, Burnout, and Retention of Secondary Teachers. *Current Issues in Education*. 14(1).
- Friedman, I. A. (2003). Self-efficacy and burnout in teaching: the importance of interpersonal-relations efficacy. *Social Psychology of Education*, 6(3), 191–215. <https://doi.org/10.1023/a:1024723124467>



- Fuller, D., Colwell, E., Low, J. L., Orychock, K., Tobin, M., Simango, B., Buote, R., Van Heerden, D., Luan, H., Cullen, K., Slade, L. & Taylor, N. (2020). Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR Mhealth and Uhealth*, 8(9), e18694. <https://doi.org/10.2196/18694>
- Gangl, K. (2009). *Kundenkompass Stress: Aktuelle Bevölkerungsbefragung: Ausmaß, Ursachen und Auswirkungen von Stress in Deutschland*. FAZ-Inst. für Management-, Markt- und Medieninformation.
- Günthner, A. (2022). *Stress und Burnout: Ein verhaltenstherapeutisches Lehrbuch zu Stressmanagement und Burnout-Prävention* (1. Aufl.). Kohlhammer Verlag.
- Hajj-Boutros, G., Landry-Duval, M., Comtois, A. S., Gouspillou, G. & Karelis, A. D. (2022). Wrist-worn devices for the measurement of heart rate and energy expenditure: a validation study for the Apple Watch 6, Polar Vantage V and Fitbit Sense. *European Journal of Sport Science*, 23(2), 165–177. <https://doi.org/10.1080/17461391.2021.2023656>
- Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology*, 1(1), 293–319. <https://doi.org/10.1146/annurev.clinpsy.1.102803.143938>
- Heinrichs, M., Stächele, T. & Domes, G. (2015). *Stress und Stressbewältigung*. Hogrefe Verlag.
- Helsper, H. (1996). Antinomien des Lehrerhandelns in modernisierten pädagogischen Kulturen. Paradoxe Verwendungsweisen von Antinomie und Selbstverantwortlichkeit. In A. Combe/W. Helsper (Hrsg.), *Pädagogische Professionalität* (S. 521-569). Suhrkam.
- Hickey, B. A., Chalmers, T. A., Newton, P. J., Lin, C., Sibbritt, D., McLachlan, C. S., Clifton-Bligh, R., Morley, J. W. & Lal, S. (2021). Smart Devices and Wearable Technologies to Detect and Monitor Mental Health Conditions and Stress: A Systematic Review. *Sensors*, 21(10), 3461 – 3477. <https://doi.org/10.3390/s21103461>

- Hottenrott, K. (2007). *Trainingskontrolle: mit Herzfrequenz-Messgeräten* (2. Auflage). Meyer & Meyer Verlag.
- Jachymek, M., Jachymek, M. T., Kiedrowicz, R., Kaźmierczak, J. & Peregud-Pogorzelska, M. (2021). Validation of heart rate measurement of Fitbit Charge 4 and Xiaomi Mi Band 5 (Preprint). *JMIR Publications*.  
<https://doi.org/10.2196/preprints.30893>
- Janczyk, M. & Pfister, R. (2013). *Inferenzstatistik verstehen: Von A wie Signifikanztest bis Z wie Konfidenzintervall*. Springer. <https://doi.org/10.1007/978-3-642-34825-9>
- Jo, E., Lewis, K., Directo, D., Kim, M.J.Y. & Dolezal, B. (2016). Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking. *Journal of Sports Science & Medicine*. 15. 540-547.
- Junker, R., Donker, M. H. & Mainhard, T. (2021). Potential Classroom Stressors of Teachers: An Audiovisual and Physiological approach. *Learning and Instruction*, 75, 101495. <https://doi.org/10.1016/j.learninstruc.2021.101495>
- Kärner, T. & Warwas, J. (2018). Stress im Unterricht? Prozessanalysen zu Interaktionseffekten unterrichtlicher Anforderungen und individueller Ressourcenbewertungen auf physische und psychische Stresssymptome von Berufsschüler/innen. *Unterrichtswissenschaft*, 46(2), 185–214.  
<https://doi.org/10.1007/s42010-018-0014-z>
- Kim-Fuchs, C., Le, C. P., Pimentel, M. A., Shackelford, D. M., Ferrari, D., Angst, E., Hollande, F. & Sloan, E. K. (2014). Chronic stress accelerates pancreatic cancer growth and invasion: a critical role for beta-adrenergic signaling in the pancreatic microenvironment. *Brain Behavior and Immunity*, 40, 40–47.  
<https://doi.org/10.1016/j.bbi.2014.02.019>
- Koch, S., Lehr, D. & Hillert, A. (2015). *Burnout und chronischer beruflicher Stress*. Hogrefe Verlag.

- Krause, A. & Dorsemagen, C. (2007). Ergebnisse der Lehrerbelastungsforschung: Orientierung im Forschungsdschungel. In M. Rothland (Hrsg.), *Belastung und Beanspruchung im Lehrberuf* (1. Aufl., S. 11-31). VS Verlag für Sozialwissenschaften.
- Kypta, G. (2006). *Burnout erkennen, überwinden, vermeiden*. Carl-Auer Verlag.
- Kyriacou, C. (2001). Teacher stress: Directions for future research. *Educational review*, 53(1), 27–35. <https://doi.org/10.1080/00131910120033628>
- Löllgen, H. (1999). Herzfrequenzvariabilität. *Deutsches Ärzteblatt*, 96 (Heft 31-32), A-2029-2032.
- Lohmann, G. (2014). *Mit Schülern klarkommen: professioneller Umgang mit Unterrichtsstörungen und Disziplinkonflikten* (11. Aufl.). Cornelsen Schulverlag.
- Muggeridge, D. J., Hickson, K., Davies, A. V., Giggins, O. M., Megson, I. L., Gorely, T. & Crabtree, D. (2021). Measurement of heart rate using the polar OH1 and Fitbit Charge 3 wearable devices in healthy adults during Light, Moderate, Vigorous, and Sprint-Based exercise: Validation study. *JMIR Mhealth and Uhealth*, 9(3), e25313. <https://doi.org/10.2196/25313>
- Noble, A., Johnson, R., Thomas, A. & Bass, P. (2017). *Organsysteme verstehen - Herz-Kreislauf-System: Integrative Grundlagen und Fälle* (1. Aufl.). Elsevier.
- Pakhomov, S., Thuras, P., Finzel, R., Eppel, J. & Kotlyar, M. (2020). Using consumer-wearable technology for remote assessment of physiological response to stress in the naturalistic environment. *PLOS ONE*, 15(3), e0229942. <https://doi.org/10.1371/journal.pone.0229942>
- Parak, J., & Korhonen, I. (2014). Evaluation of wearable consumer heart rate monitors based on photoplethysmography. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2014*, 3670–3673. <https://doi.org/10.1109/EMBC.2014.6944419>

- Rudow, B. (2000). *Der Arbeits- und Gesundheitsschutz im Lehrberuf. Gefährdungsbeurteilung der Arbeit von Lehrerinnen und Lehrern*. Süddeutscher Pädagogischer Verlag.
- Rothland, M. & Terhart, E. (2007). Beruf: Lehrer – Arbeitsplatz: Schule. Charakteristika der Arbeitstätigkeit und Bedingungen der Berufssituation. In M. Rothland (Hrsg.), *Belastung und Beanspruchung im Lehrberuf* (1. Aufl., S. 11-31). VS Verlag für Sozialwissenschaften.
- Runge, N., Haarman, S. & Fisher, M. (2020) Using Fitbit fitness trackers to measure teacher stress and coping. *International Journal of Social Policy and Education*. 2(3), 56–70.
- Schäfer, T. (2016). *Methodenlehre und Statistik: Einführung in Datenerhebung, deskriptive Statistik und Inferenzstatistik*. Springer. <https://doi.org/10.1007/978-3-658-11936-2>
- Seefeldt, D. (2000). *Stress: Verstehen - Erkennen - Bewältigen. Mit praktischen Übungen für jeden*. (2., überarbeitete Auflage). Edition Wötzel.
- Shachter, J., Kangas, M., Sweller, N. & Stewart, J. (2022). Affective states and heart rate response: Measuring foreign language speaking performance reactions in a Japanese university classroom. *Journal for the psychology of language learning*, 4(2), 1–23. <https://doi.org/10.52598/jpll/4/2/2>
- Statistisches Bundesamt. (29.September 2022). Anzahl der Lehrkräfte an allgemeinbildenden Schulen in Deutschland im Schuljahr 2021/2022 nach Schulart [Graph]. In *Statista*. Abgerufen am 25. Oktober 2023, von <https://de.statista.com/statistik/daten/studie/162263/umfrage/anzahl-der-lehrkraefte-nach-schularten/>
- Taelman, J., Vandeput, S., Vlemincx, E., Spaepen, A. & Van Huffel, S. (2010). Instantaneous changes in heart rate regulation due to mental load in simulated office work. *European Journal of Applied Physiology*, 111(7), 1497–1505. <https://doi.org/10.1007/s00421-010-1776-0>

Talley, N. J. & Spiller, R. C. (2002). Irritable bowel syndrome: a little understood organic bowel disease? *The Lancet*, 360(9332), 555–564.

[https://doi.org/10.1016/s0140-6736\(02\)09712-x](https://doi.org/10.1016/s0140-6736(02)09712-x)

Thaker, P. H., Lutgendorf, S. K. & Sood, A. K. (2007). The neuroendocrine impact of chronic stress on cancer. *Cell Cycle*, 6(4), 430–433.

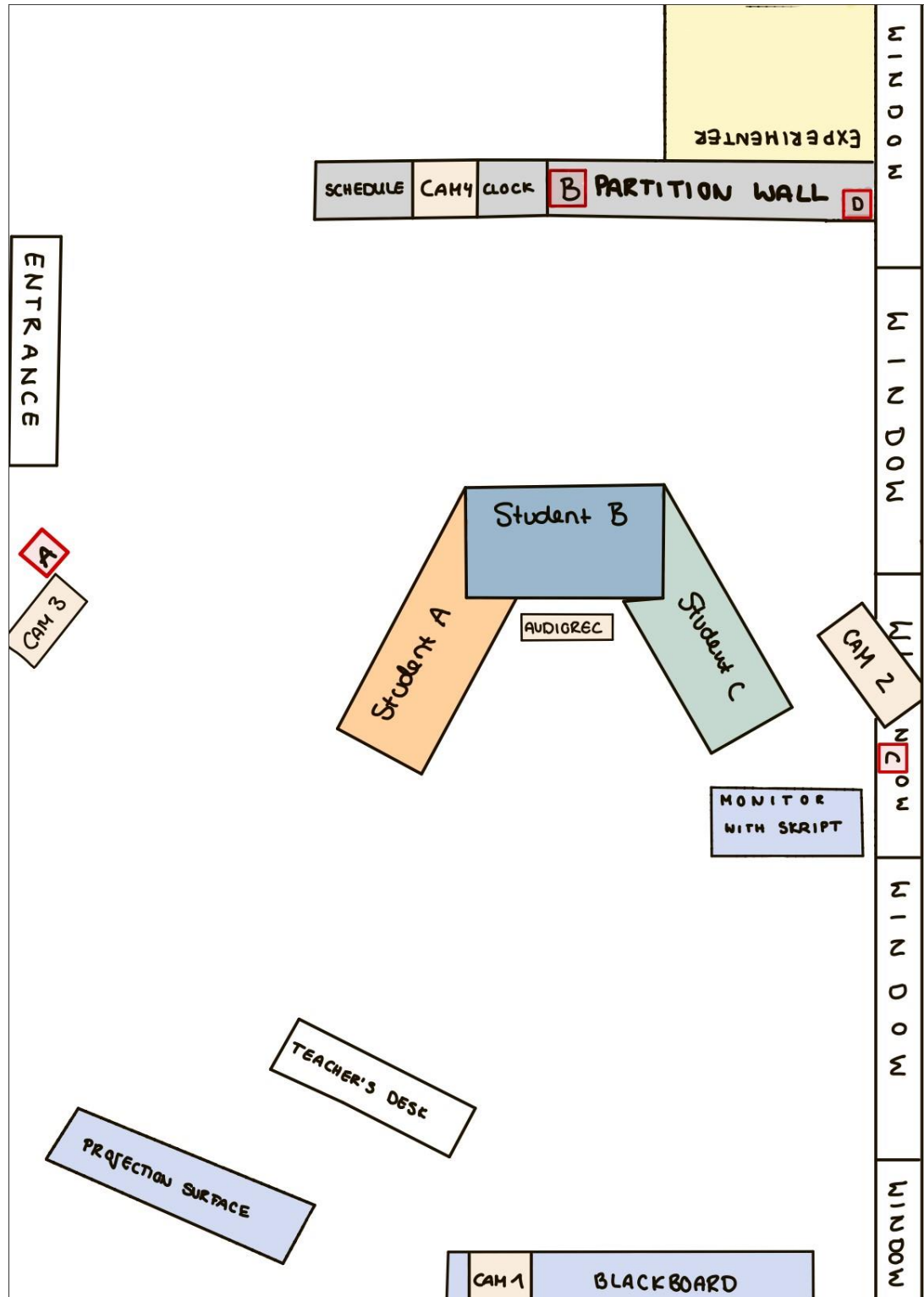
<https://doi.org/10.4161/cc.6.4.3829>

Van Dick, R. (2006). *Stress und Arbeitszufriedenheit bei Lehrerinnen und Lehrern: zwischen „Horrorjob“ und Erfüllung* (2. Aufl.). Tectum Verlag.

Van Dick, R. & Stegmann, S. (2007). Belastung, Beanspruchung und Stress im Lehrerberuf – Theorien und Modelle. In M. Rothland (Hrsg.), *Belastung und Beanspruchung im Lehrberuf* (1. Aufl., S. 34-51). VS Verlag für Sozialwissenschaften.

## Anhang

### A Setup der Unterrichtseinheit



**Abbildung 4:** Aufbau der Unterrichtseinheit mit Positionen aller beteiligten Personen und des Equipments.

## B Syntax und Ergebnis-Output

### Anhang B.1

SPSS-Syntax der einfaktoriellen Varianzanalyse mit Messwiederholung mit Kontrasten:

```
DATASET ACTIVATE DataSet2.  
GLM Pre Lehr Post Interview End  
  /WSFACTOR=Phasen 5 Polynomial  
  /METHOD=SSTYPE(3)  
  /PLOT=PROFILE(Phasen) TYPE=LINE ERRORBAR=NO  
MEANREFERENCE=NO YAXIS=AUTO  
  /PRINT=DESCRIPTIVE ETASQ  
  /MMATRIX "Lehrphase vs. Prelehrphase" ALL -1 1 0 0 0; "Lehrphase vs.  
Postlehrphase" ALL 0 1 -1 0 0; "Lehrphase vs. Interviewphase" ALL 0 1 0 -1 0;  
"Lehrphase vs. Endphase" ALL 0 1 0 0 -1  
  /CRITERIA=ALPHA(.05)  
  /WSDSIGN=Phasen.
```

### Anhang B.2

#### Tabelle 7

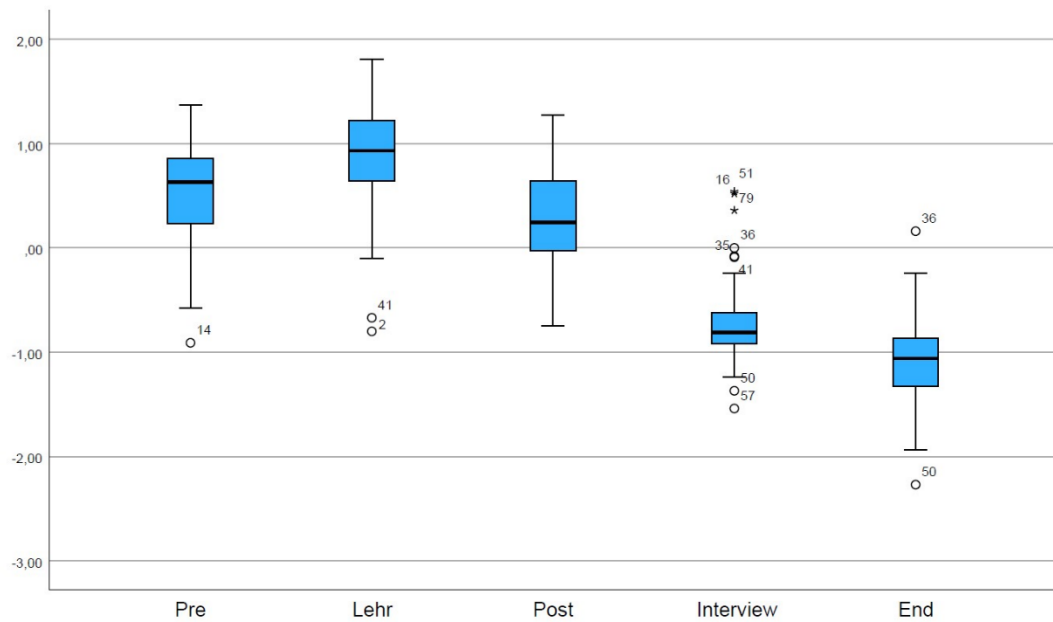
*SPSS-Output der Tests auf Normalverteilung (Kolmogorov-Smirnov-Test und Shapiro-Wilk-Test) innerhalb der fünf Abstufungen der unbestimmten Variable, welche durch die Pre-, Lehr-, Postlehr-, Interview- und End-Phase gegeben sind. Es wurden nur die Ergebnisse des Shapiro-Wilk-Tests in Rahmen der vorliegenden Arbeit ausgewertet.*

| Tests auf Normalverteilung |                                 |    |             |              |    |             |
|----------------------------|---------------------------------|----|-------------|--------------|----|-------------|
|                            | Kolmogorov-Smirnov <sup>a</sup> |    |             | Shapiro-Wilk |    |             |
|                            | Statistik                       | df | Signifikanz | Statistik    | df | Signifikanz |
| Pre                        | ,099                            | 76 | ,061        | ,961         | 76 | ,019        |
| Lehr                       | ,100                            | 76 | ,057        | ,928         | 76 | <,001       |
| Post                       | ,077                            | 76 | ,200*       | ,983         | 76 | ,379        |
| Interview                  | ,161                            | 76 | <,001       | ,891         | 76 | <,001       |
| End                        | ,078                            | 76 | ,200*       | ,987         | 76 | ,608        |

\*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

## Anhang B.3



**Abbildung 5:** Boxplot zur Bestimmung von Ausreißern innerhalb der fünf Abstufungen der unbestimmten Variable, welche durch die Pre-, Lehr-, Postlehr-, Interview- und End-Phase gegeben sind. Leichte Ausreißer werden durch Kreise und extreme Ausreißer durch Sterne gekennzeichnet.

## Anhang B.4

**Tabelle 8**

*SPSS-Output des Mauchly-Tests auf Sphärizität.*

| Mauchly-Test auf Sphärizität <sup>a</sup> |           |                           |    |      |                        |                     |             |
|---|-----------|---------------------------|----|------|------------------------|---------------------|-------------|
| Maß: MASS_1                               |           |                           |    |      |                        |                     |             |
| Innersubjekteffekt                        | Mauchly-W | Ungefähres<br>Chi-Quadrat | df | Sig. | Epsilon <sup>b</sup>   |                     |             |
|   |           |                           |    |      | Greenhouse-<br>Geisser | Huynh-Feldt<br>(HF) | Untergrenze |
| Phasen                                    | ,848      | 12,068                    | 9  | ,210 | ,925                   | ,979                | ,250        |

Prüft die Nullhypothese, dass sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Design: Konstanter Term  
Innersubjektdesign: Phasen

b. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.



## Anhang B.5

**Tabelle 9**

*SPSS-Output der Tests der Innersubjekteffekte der einfaktoriellen Varianzanalyse mit Messwiederholung.*

| Tests der Innersubjekteffekte |                        |                         |         |                        |         |       |
|-------------------------------|------------------------|-------------------------|---------|------------------------|---------|-------|
| Maß: MASS_1                   |                        |                         |         |                        |         |       |
| Quelle                        |                        | Typ III<br>Quadratsumme | df      | Mittel der<br>Quadrate | F       | Sig.  |
| Phasen                        | Sphärizität angenommen | 216,036                 | 4       | 54,009                 | 262,694 | <,001 |
|                               | Greenhouse-Geisser     | 216,036                 | 3,700   | 58,395                 | 262,694 | <,001 |
|                               | Huynh-Feldt (HF)       | 216,036                 | 3,915   | 55,177                 | 262,694 | <,001 |
|                               | Untergrenze            | 216,036                 | 1,000   | 216,036                | 262,694 | <,001 |
| Fehler(Phasen)                | Sphärizität angenommen | 61,679                  | 300     | ,206                   |         |       |
|                               | Greenhouse-Geisser     | 61,679                  | 277,465 | ,222                   |         |       |
|                               | Huynh-Feldt (HF)       | 61,679                  | 293,650 | ,210                   |         |       |
|                               | Untergrenze            | 61,679                  | 75,000  | ,822                   |         |       |

## Anhang B.6

**Tabelle 10:**

*Effektstärke  $d$  der betrachteten Mittelwertunterschiede und zugehörige Werte für  $U_3$  nach Cohen (1988) in Prozent.*

| Effektstärke $d$ | Cohen's $U_3$ |
|------------------|---------------|
| 0.78             | 78.23         |
| 1.27             | 89.80         |
| 3.85             | 99.99         |
| 4.51             | 100.00        |

## **C      Übersicht über digitalen Anhang**

Der digitale Anhang enthält folgende Dokumente:

### Microsoft Excel-Arbeitsblätter

- HR\_Daten\_bpm\_z\_std.xlsx
- Umrechnung\_d\_U3.xlsx

### PDF Dokumente

- ANOVA\_Messwdh\_Kontraste.pdf
- Deskriptive\_Stat\_HR\_bpm.pdf
- Staatexamensarbeit\_Zitzmann\_Tom.pdf

### SPSS Dokumente

- ANOVA\_Messwdh\_Kontraste.sav
- ANOVA\_Messwdh\_Kontraste.sps
- ANOVA\_Messwdh\_Kontraste.spv
- Deskriptive\_Stat\_Alter\_Erfahrung.sav
- Deskriptive\_Stat\_Alter\_Erfahrung.sps
- Deskriptive\_Stat\_Alter\_Erfahrung.spv
- Deskriptive\_Stat\_HR\_bpm.sav
- Deskriptive\_Stat\_HR\_bpm.spv
- Tests\_auf\_NVT\_Boxplot.sps
- Tests\_auf\_NVT\_Boxplot.spv

## Selbstständigkeitserklärung

Hiermit versichere ich, Tom Zitzmann, die vorliegende Arbeit selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt zu haben sowie alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, durch die Angabe der Quellen als Entlehnung kenntlich gemacht zu haben.

Leipzig, 26. Oktober 2023



---

Tom Zitzmann