

schungsarbeit von Bernay (2014: 62), ermutigt durch die Ausführungen über die Implementierung von Achtsamkeitsmeditation in die Methodik der Human- und Sozialwissenschaften von Konecki (vgl. 2016: 223; Konecki 2022) und basierend auf meiner eigenen Erfahrung mit der Achtsamkeitsmeditation, begleitet meine Meditationspraxis den Forschungsprozess. Die beschriebene phänomenologische Haltung (vgl. Finlay 2012: 24; Friesen et al. 2012: 1; van Manen 1990: 45) und die zu Beginn dieser Arbeit gezeigte attitudinale Komponente von Achtsamkeit (siehe Abschnitt 2.1 und vgl. Bishop 2004; Kabat-Zinn 2003: 145) weisen viele Ähnlichkeiten auf (vgl. Bentz/Shapiro 1998: 39). Zusätzlich verweist Konecki auf ein verbindendes Element zwischen der phänomenologischen Praxis und der Achtsamkeitsmeditation.

»This consciousness that appears [while meditating, WD] is prereflective, discursive, preverbal, and nonconceptual. We open our mind to all the things that could appear. In practical terms we do not refuse anything because we are not attached to any assumptions. So, Zen meditation could be understood as a practice of epoché.« (Konecki 2016: 223–224)

Dieser Forschungsstrang entwickelte sich somit zu einer achtsamkeitsbasierten hermeneutisch-phänomenologischen Untersuchung. Zusätzlich zu meiner täglichen Meditationspraxis wurden vor, zwischen und nach dem Lese-, Schreib und Reflexionsprozess Meditationszeiten eingewebt.

5.2 Zweiter Forschungsstrang – quantitative Analyse subjektiver Wirkungen mittels Prä-Post-Vergleich

Im zweiten quantitativen Forschungsstrang liegt der Fokus auf den möglichen subjektiven Wirkungen, die eine achtsamkeitsbasierte Lehrveranstaltung auf die Studierenden hat. In diesem Abschnitt wird die methodische Vorgehensweise zur Erhebung und Auswertung der quantitativen Daten beschrieben. Dabei wird konkret auf die Auswahl des verwendeten Fragebogeninventars und die Durchführung der Datenerhebung in den unterschiedlichen Lehrveranstaltungen an den unterschiedlichen Universitäten sowie die einzelnen Schritte der statistischen Analyse eingegangen.

5.2.1 Vorgehen bei der Rekrutierung der Teilnehmer*innen

Für diese Studie wurden Daten aus universitären Lehrveranstaltungen erhoben, um damit Zugang zu realen lebensweltlichen Situationen zu bekommen. Es handelt sich somit nicht um eine klassische experimentelle Interventionsstudie mit randomisiertem Design, wie sie in den meisten klinischen Studien der Achtsamkeitsforschung angewandt wird (vgl. Rosenkranz et al. 2019). Zwei Gründe können für diese Art der Teilnehmer*innenrekrutierung angeführt werden: (1) Das Interesse dieser Arbeit liegt in der Erforschung von achtsamkeitsbasierten Lehrveranstaltungen – im Sinne von Achtsamkeit als Pädagogik (*mindfulness as education*) (vgl. Ergas 2019c und siehe Abschnitt 6.1). Dabei wird Achtsamkeit nicht als punktuelle Intervention verstanden, sondern

als ein in die Lehr-Lern-Praxis eingewobenes Element aufgefasst. Somit unterscheidet sich dieses Verständnis von der klassischen Interventionslogik, denn im Fokus dieser Forschung stehen die realen lebensweltlichen Bedingungen der Lehrveranstaltung. Es wurden somit keine Interventionsgruppen bzw. Interventionsveranstaltungen im Sinne eines experimentellen Designs gebildet.

Die Achtsamkeitsforschung bemüht sich aktuell stark darum, in experimentellen Interventionsstudien das Training der Interventionsgruppe und das Training der aktiven Kontrollgruppe möglichst akkurat anzupassen, so dass sich die Trainings nur durch das Üben bzw. Nicht-Üben von Achtsamkeitsübungen unterscheiden (vgl. Krägeloh et al. 2019: 116). Aus methodischer Sicht ist ein solches Vorgehen nachvollziehbar, da darin die Hoffnung liegt, möglichst viele Störvariablen zu eliminieren. Der Nachteil liegt darin, dass es sich in der Regel um ein experimentelles und nicht um ein lebensweltliches Design handelt. (2) Der zweite Grund, warum für diese Studien kein randomisiertes experimentelles Design gewählt wurde, waren organisatorische und finanzielle Limitationen. Im regulären Lehrveranstaltungsbetrieb wäre es mit den vorhandenen Mitteln nicht möglich gewesen, ein solches Design aufzusetzen.

Es wurden reguläre achtsamkeitsbasierte Lehrveranstaltungen an den Universitäten Wien, München und Leipzig untersucht. Alle drei Lehrveranstaltungen sind ähnlich aufgebaut (siehe Abschnitt 6.2) und folgen – wie die meisten achtsamkeitsbasierten Programme (vgl. Ivtzan 2020) – in der Struktur dem MBSR-Programm (vgl. Crane et al. 2017; Crane et al. 2021 und siehe Abschnitt 2.3). Auch hinsichtlich der inhaltlichen Elemente (Übungen, Meditationen, erforschendes Gespräch) ähneln die drei Lehrveranstaltungen einander (siehe Abbildung 39).

An allen drei Universitäten werden die achtsamkeitsbasierten Lehrveranstaltungen als Wahllehrveranstaltung innerhalb eines Pflichtmoduls angeboten. Das bedeutet, dass es innerhalb der Module jeweils Parallellehrveranstaltungen ohne Bezug zum Thema Achtsamkeit gibt. An zwei Universitäten konnten aus diesen Parallellehrveranstaltungen Kontrollgruppen generiert werden. Für eine der beteiligten Universitäten war dies trotz erheblicher organisatorischer Bemühungen leider nicht möglich. Da das Hauptaugenmerk dieser Studie nicht auf dem Vergleich der drei Universitäten liegt, sondern die gepoolten Daten aller drei Universitäten analysiert wurden (siehe Abschnitt 7.5), konnten die Berechnungen dennoch durchgeführt werden.

Die Rekrutierung der Teilnehmer*innen für die Achtsamkeitsgruppe (Trainingsgruppe, TG) sowie die Kontrollgruppen (KG) erfolgte somit über die reguläre Lehrveranstaltungsanmeldung der jeweiligen Universität. Zu berücksichtigen ist, dass die Personen, die sich für die achtsamkeitsbasierte Lehrveranstaltung angemeldet hatten, dies mit einer gewissen Erwartung bzw. einem gewissen Interesse getan hatten. Dies kann zur Folge haben, dass diese Personen möglicherweise das Bedürfnis hatten, ihre Achtsamkeit zu fördern. Es könnte sein, dass diese sich dementsprechend vor der Lehrveranstaltung als weniger achtsam einschätzten oder auch besonders stark auf das Angebot ansprachen, und dass sie sich nach der Lehrveranstaltung folglich als achtsamer einschätzten (vgl. Rosenkranz et al. 2019; Krägeloh et al. 2019). Teilnehmer*innen aus der Kontrollgruppe haben wiederum keinen Bezug zum Konzept Achtsamkeit, da dieses in den Parallellehrveranstaltungen nicht thematisiert wird. Daraus ergibt sich mitunter ein methodologisches Problem, das in der Achtsamkeitsforschung be-

kannt ist. Personen, die Achtsamkeit nur begrifflich kennen, jedoch nicht im Kontext eines achtsamkeitsbasierten Programmes darüber informiert wurden, haben häufig ein anderes semantisches Verständnis entwickelt (vgl. Choi et al. 2021: 11). Ebenso dürfte sich das Praktizieren von Achtsamkeitsübungen bzw. Meditationen auf das konzeptionelle Verständnis der Personen auswirken. Auch dieses Phänomen ist in der Achtsamkeitsforschung bekannt (vgl. Krägeloh et al. 2018). Dem daraus resultierenden methodologischen Problem begegnet die Achtsamkeitsforschung mit der kontinuierlichen Weiterentwicklung bzw. Neuentwicklung von Achtsamkeitsfragebögen (vgl. Medvedev et al. 2019; Medvedev et al. 2022). Mit dieser methodologischen Problematik hinsichtlich des semantischen Verständnisses einzelner Items im Fragebogen haben jedoch auch randomisierte experimentelle Forschungsstudien zu kämpfen.

Zusammengefasst kann festgehalten werden, dass die Rekrutierung der Teilnehmer*innen aus organisatorischen und finanziellen Gründen nicht nach den Regeln des randomisierten experimentellen Designs erfolgen konnte. Daraus ergab sich jedoch die Gelegenheit, achtsamkeitsbasierte Lehrveranstaltungen unter möglichst realen lebensweltlichen Bedingungen zu beforschen. Somit wurden die Achtsamkeitsgruppe und die Kontrollgruppe aus regulären Lehrveranstaltungsgruppen generiert. Die Teilnehmer*innen hatten selbst gewählt, welche Lehrveranstaltung sie besuchten. Die daraus resultierenden methodologischen Limitationen wurden zugunsten der realen lebensweltlichen Orientierung in Kauf genommen, müssen jedoch in der Interpretation der Ergebnisse berücksichtigt werden und sind als methodische Limitation dieser Arbeit zu werten (siehe Abschnitt 8.3).

5.2.2 Erhebungssituation

Die Datenerhebung erfolgte an der Universität Wien, der Universität Leipzig und an der Ludwig-Maximilian-Universität München. In den achtsamkeitsbasierten Lehrveranstaltungen und den gewählten Parallellehrveranstaltungen in Wien, Leipzig und München wurden die Daten am Beginn der ersten und letzten Seminareinheit des Semesters erhoben. Aus raum-zeitlich-organisatorischen Gründen übernahmen neben mir auch die Lehrveranstaltungsleiter*innen der jeweiligen Lehrveranstaltung die Rolle der Testleiter*innen. Um einen einheitlichen Ablauf zu garantieren, wurden alle Testleiter*innen mit Einleitungstexten ausgestattet. Die Erhebung erfolgte in allen Lehrveranstaltungen digital mit dem Onlinetool Lime Survey. Den Zugang erhielten die Studierenden über einen Short-Link oder QR-Code. Als digitales Eingabegerät dienten Smartphone, Tablet oder Laptop. In einzelnen Fällen war kein digitales Gerät vorhanden, somit wurde auf einen Paper-Pencil-Test zurückgegriffen. Für die geplante Follow-Up-Datenerhebung wurden die Studierenden per E-Mail kontaktiert. Das Ausfüllen des gesamten Fragebogeninventars dauerte im Durchschnitt 30 Minuten.

5.2.3 Fragebogeninventar

Für die vorliegende Studie wurden standardisierte Skalen zur Erhebung unterschiedlicher psychologischer Variablen ausgesucht, die für die Studierenden als Zielgruppe passend waren und Anknüpfungen an das Konzept der pädagogischen Professionalität

ermöglichten. Das zusammengestellte Fragebogeninventar wurde einem Probedurchgang unterzogen. Studierende der Lehrveranstaltung Pädagogik der Achtsamkeit füllten das Fragebogeninventar im Sommersemester 2018 zu Testzwecken aus. Die positiven Rückmeldungen zur inhaltlichen Klarheit, Handhabbarkeit am Smartphone und Bearbeitungsdauer bestätigten die Eignung der Erhebungsmethode. In weiterer Folge werden die neun verwendeten Skalen einzeln bezüglich ihrer psychometrischen Eigenschaften beschrieben und auf Ergebnisse aus Validierungsstudien hingewiesen.

Achtsamkeit

Zur Selbsteinschätzung der Achtsamkeit wurde der Fragebogen *Comprehensive Inventory of Mindfulness Experiences (CHIME)* verwendet, um anhand von 37 Items die acht Komponenten des Achtsamkeitskonstrukts zu erfassen (vgl. Bergomi et al. 2014: 112). Die acht Unterskalen und dazugehörige Beispielimts sind:

- Gewährsein gegenüber inneren Erfahrungen
»Beim Sitzen oder Liegen nehme ich meine Körperempfindungen wahr.«
- Gewährsein gegenüber äußeren Erfahrungen
»Wenn ich Auto oder Zug fahre, bin ich mir meiner Umgebung, z.B. der Landschaft, bewusst.«
- Bewusstes Handeln
»Im Alltag werde ich durch viele Erinnerungen, Bilder oder Träumereien abgelenkt.«
- Annehmende Haltung
»Ich sehe meine Fehler und Schwierigkeiten, ohne mich zu verurteilen.«
- Dezentrierte Orientierung
»Ich kann meine Gedanken und Gefühle beobachten, ohne mich in ihnen zu verstricken.«
- Offene Haltung
»Ich versuche mich abzulenken, wenn ich unangenehme Gefühle erlebe.«
- Relativierung
»Im Alltag ist mir bewusst, dass viele Gedanken Interpretationen sind, die nicht unbedingt der Realität entsprechen.«
- Einsichtsvolles Verstehen
»Wenn ich mir unnötig das Leben schwer mache, wird mir das bald danach klar.«
(vgl. Bergomi et al. 2014: 119)

Neun der 37 CHIME-Items sind negativ formuliert (vgl. Bergomi et al. 2014: 119) und wurden später für die Datenanalyse umgepolt. Alle Items sollten von den Teilnehmer*innen auf einer 6-Punkte-Likert-Skala von 1= fast nie bis 6= fast immer eingeschätzt werden. Die zu bewertenden Aussagen bezogen sich dabei auf die vergangenen zwei Wochen (vgl. Bergomi et al. 2014: 113). Die interne Konsistenz des Gesamtkonstrukts Achtsamkeit wird mit einem Cronbachs Alpha ($\alpha = .90$) angegeben (vgl. Bergomi et al. 2014: 118). Die Retest-Reliabilität über einen Zeitraum von sieben Wochen beträgt ($r_{tt} = .89$) (Bergomi et al. 2014: 118). Bergomi et al. haben – basierend auf bereits bestehenden Achtsamkeitsskalen wie zum Beispiel dem *Five Facet Mindfulness Questionnaire (FFMQ)* (Baer et al. 2008) – mit dem CHIME den bislang umfassendsten Fragebogen in deutscher Sprache entwickelt.

Selbstmitgefühl

Zur Erfassung des Selbstmitgefühls wurde die deutsche Kurzversion (Hupfeld/Ruffieux 2011) der von Neff (2003) entwickelten *Self-Compassion Scale (SCS)* verwendet. Die deutsche Self-Compassion Scale Kurzform (SCS-D-Kurzform) besteht aus 12 Items. Diese wurden von den Teilnehmer*innen auf einer 5-Punkte-Likert-Skala von 1 = *sehr selten* bis 5 = *sehr oft* eingeschätzt. Das Gesamtkonstrukt Mitgefühl setzt sich aus den sechs Subskalen *Selbstbezogene Freundlichkeit* (self-kindness), *Selbstverurteilung* (self-judgment), *verbindende Humanität* (common humanity), *Isolation* (Isolation), *Achtsamkeit* (mindfulness) und *Überidentifizierung* (over-identification) zusammen (vgl. Hupfeld/Ruffieux 2011: 116). Die Items der negativorientierten Subskalen Selbstverurteilung, Isolation und Überidentifikation wurden vor der Berechnung des Self-Compassion-Gesamtwerts umgepolt. Für die deutsche Version der Self-Compassion-Gesamtskala wird von den Autor*innen eine sehr gute interne Konsistenz ($\alpha = .91$) und eine gute Retestreliabilität ($r_{tt} = .83$) berichtet. Die Korrelation zwischen der Kurzform und der Langform wird von den Autor*innen mit ($r = .91$) berichtet. Die Validität der einzelnen Faktoren ist wenig zufriedenstellend. Aus diesem Grund empfehlen die Autor*innen, die einzelnen Subskalen nicht zu interpretieren. Die SCS-D-Kurzform eignet sich jedoch zur Erfassung des Gesamtwerts des Selbstmitgefühls.

Emotionale Kompetenz

Die Skala zur Selbsteinschätzung emotionaler Kompetenzen (SEK-27) (Berking/Znoj 2008) fragt nach dem Umgang mit negativen Emotionen. Die 27 Items wurden von den Teilnehmer*innen auf einer 5-Punkte-Likert-Skala von 0 = *überhaupt nicht* bis 4 = *fast immer* bewertet. Das Konstrukt Emotionale Kompetenz teilt sich in folgende neun Subskalen:

- *Aufmerksamkeit* – eigene Emotionen bewusst wahrnehmen
- *Klarheit* – eigene Emotionen klar erkennen und benennen
- *Körperwahrnehmung* – emotionsbezogene Körperempfindungen wahrnehmen
- *Verstehen* – Ursachen des aktuellen emotionalen Erlebens verstehen
- *Regulation* – eigenes emotionales Erleben gezielt positiv beeinflussen
- *Akzeptanz* – eigene Gefühle akzeptieren
- *Resilienz* – belastende Gefühle tolerieren und aushalten
- *Selbstunterstützung* – sich selbst Mut machen
- *Konfrontationsbereitschaft* – sich in Situationen begeben, in denen negative Gefühle ausgelöst werden könnten (vgl. Berking/Znoj 2008: 145)

Die Autor*innen verweisen auf eine zufriedenstellende interne Konsistenz von ($\alpha = .90$) und einer Retestreliabilität von ($r_{tt} = .75$) für die Gesamtskala. Der SEK-27-Fragebogen wird des Weiteren als reliables, valides und veränderungssensitives Maß zur Selbsteinschätzung des konstruktiven Umgangs mit negativen Emotionen bezeichnet (vgl. Berking/Znoj 2008: 150).

Altruismus

Für die Selbsteinschätzung von Altruismus – als Haltung und Bereitschaft, anderen zu helfen, ohne direkten Nutzen zu erwarten – wurde die deutsche Version der *Generative Altruism Scale (GALS)* (Büssing et al. 2013) verwendet. Die 11 Items (z.B.: »Wenn ich Hilfsbedürftige sehe, spreche ich sie an, wie ich ihnen helfen kann.«) wurden auf einer 4-Punkte-Likert-Skala von 0 = *nie* bis 3 = *sehr oft* bewertet. Das eindimensionale Konstrukt weist mit einem Cronbachs Alpha von ($\alpha = .81$) eine gute interne Konsistenz auf (vgl. Büssing et al. 2013: 7). Laut der Autor*innen beinhaltet der Fragebogen keine Variablen zur Lebenszufriedenheit und ist auch für nicht-religiöse Personen anwendbar. Sie empfehlen dieses Instrument für empirische Studien, die den Zusammenhang zwischen Altruismus als Verhalten und der Entwicklung ethischer Werte untersuchen. Die Skala eignet sich demnach zur Auswertung von Bildungsprogrammen (vgl. Büssing et al. 2013: 1).

Selbstreflexion und Einsicht

Die Fähigkeit zur Selbstreflexion und Einsicht wurde durch Selbsteinschätzung anhand der *Self-Reflection and Insight Scale* (Grant et al. 2002) erhoben. Zum Zeitpunkt der Erhebung lag keine validierte deutsche Version vor. Aus diesem Grund wurde dieser Fragebogen einem Übersetzungsprozess unterzogen (vgl. Brislin 1986). Eine Kollegin am Institut für Anglistik übersetzte die Items ins Deutsche. Eine weitere Kollegin übersetzte die deutsche Version wieder zurück ins Englische. Diese rückübersetzte englische Version und die originale englische Version wurden von einer Kollegin mit englischer Muttersprache überprüft. Die 20 Items wurden von den Teilnehmer*innen auf einer 6-Punkte-Likert-Skala von 1 = *trifft überhaupt nicht zu* bis 6 = *trifft voll zu* bewertet. In der englischen Originalversion bilden 12 Items die Subskala Selbstreflexion und 8 Items die Subskala Einsicht.

- Selbstreflexion
 - »Ich habe ein klares Bedürfnis zu verstehen, wie mein Verstand funktioniert.«
- Einsicht
 - »Ich habe normalerweise eine sehr klare Vorstellung davon, warum ich mich auf eine bestimmte Art und Weise benommen habe.«

Die acht negativ formulierten Items wurden vor der Berechnung der Gesamtskala und den zwei Subskalen umgepolt. Die interne Konsistenz geben die Autor*innen mit einem Cronbachs Alpha von ($\alpha = .91$) für den Faktor Selbstreflexion und ($\alpha = .87$) für den Faktor Einsicht an (vgl. Grant et al. 2002: 826). Die Retestreliabilität über einen Zeitraum von sieben Wochen ist $r_{tt} = .77$ für Selbstreflexion und $r_{tt} = .78$ für Einsicht (vgl. Grant et al. 2002: 821). Die Autor*innen stufen dieses Maß als ein reliables und valides Instrument für die Testung der Fähigkeit zur Selbstreflexion und Einsicht ein.

Offenheit

Die deutsche Version der Offenheitsskala des *NEO-Fünf-Faktoren-Inventars (NEO-FFI)* (Borkenau/Ostendorf 1993, 2008) wurde in das Fragebogeninventar aufgenommen, um nach der Offenheit für Erfahrungen zu fragen. Zwölf Items fragen nach Neugier, Kreativität, Bereitschaft, sich auf neue und ungewöhnliche Erfahrungen einzulassen

und nach Interesse für vielfältige kulturelle Praktiken. Die Aussagen wurden von den Teilnehmer*innen auf einer 5-Punkte-Likert-Skala von 0 = *starke Ablehnung* bis 4 = *starke Zustimmung* bewertet. Für die Offenheitsskala berichten die Autor*innen eine interne Konsistenz von ($\alpha = .75$). Durch eine Testwiederholung nach zwei Jahren konnten die Autor*innen eine stabile Retestreliabilität ($r_{tt} = .76$) vorweisen (vgl. Borkenau/Ostendorf 2008). Zusätzlich zu den Selbstberichten führten die Autor*innen mit der Fremdbereichsversion des NEO-FFI auch eine Befragung mit Freunden und Bekannten der Befragten durch. Für die Offenheitsskala der beiden Versionen konnte dabei eine Korrelation ($r = .57$) festgestellt werden.

Selbstwirksamkeitserwartung

Die Skala zur Erhebung der *allgemeinen Selbstwirksamkeitserwartung* (SWE) (Schwarzer/Jerusalem 1999: 13) diente zur Selbsteinschätzung der eigenen Fähigkeiten und Fertigkeiten hinsichtlich der Bewältigung von herausfordernden Situationen. Die 10 Items (z.B. »Schwierigkeiten sehe ich gelassen entgegen, weil ich meinen Fähigkeiten immer vertrauen kann.«) wurden von den Teilnehmer*innen mit einer 4-Punkte-Likert-Skala von 1 = *stimmt nicht* bis 4 = *stimmt genau* bewertet. Bei dieser Skala handelt es sich um ein einfaktorielles Konstrukt mit einer internen Konsistenz von Cronbachs Alpha ($\alpha = .79$) (vgl. Schwarzer/Jerusalem 1999: 13). Diese Skala ist in zahlreichen Studien eingesetzt und in über 25 Sprachen übersetzt worden. Sie gilt als valides Instrument zur Messung der allgemeinen Selbstwirksamkeitserwartung.

Ungewissheitstoleranz

Mit Hilfe der *Ungewissheitstoleranzskala* (UGTS) (Dalbert 1999) wurde durch Selbstbeurteilung der Umgang mit ungewissen Situationen erhoben. Die Teilnehmer*innen bewerteten acht Items (z.B. »Ich weiß gerne, was auf mich zukommt.«) auf einer 6-Punkte-Likert-Skala von 1 = *stimmt überhaupt nicht* bis 6 = *stimmt genau*. Drei negativ formulierte Items wurden vor der Berechnung des Gesamtwerts umkodiert. Die acht Items der Skala bilden ein einfaktorielles Konstrukt. Mit einem Cronbachs Alpha von ($\alpha = .72$) verweist die Autorin auf eine gute interne Konsistenz (vgl. Dalbert 1999: 11). Mit einer Retestreliabilität von ($r_{tt} = .72$) nach einer Wiederholung nach 6 Monaten wird die Skala als stabil beschrieben.

Engagement für das Studium

Die kurze Version der deutschen Utrecht Work Engagement Scale (UWES-9) für Studierende (Schaufeli et al. 2006) wurde in das Fragebogeninventar aufgenommen. Die Skala setzt sich aus 9 Items (z.B. »Wenn ich morgens aufstehe, freue ich mich auf die Veranstaltungen.«) zusammen, die von den Teilnehmer*innen auf einer 7-Punkte-Likert-Skala von 0 = *nie* bis 6 = *immer* eingeschätzt wurden. Nach Validierungsstudien in zehn Ländern mit 14.521 Teilnehmer*innen attestieren die Autor*innen der allgemeinen UWES-Kurzform akzeptable psychometrische Eigenschaften. In der Langform der UWES konnten die vom Konstrukt Engagement theoretisch angenommen Dimensionen Elan, Hingabe und Vertiefung in der Faktorenanalyse bestätigt werden. Für die Kurzform empfehlen die Autor*innen die Interpretation des Gesamtwerts (Schaufeli et al. 2006: 712). Die interne Konsistenz wird von Schaufeli et al. mit einem Cronbachs Alpha von ($\alpha = .80$) als

zufriedenstellend angegeben (vgl. Schaufeli et al. 2006: 709). Auf Grund der Retestrelia-
bilität von ($r_{tt}=.73$) nach einer Testwiederholung nach einem Jahr wird die Skala von den
Autor*innen als stabil eingestuft (vgl. Schaufeli et al. 2006: 709). Für die spezielle Version
für Studierende konnten keine Validierungsstudie gefunden werden.

Tabelle 17: Übersicht über die verwendeten Selbstberichtsmaße

Erfassungsbereich	Skala	Uni 1	Uni 2	Uni 3
Achtsamkeit	Comprehensive Inventory of Mindfulness Experiences (CHIME)	x	x	x
Selbstmitgefühl	Self-Compassion Scale Short Form (SCS-DSF)	x	x	x
Emotionale Kompetenz	Selbsteinschätzung emotionaler Kompetenzen (SEK-27)	x	x	x
Altruismus	Generative Altruism Scale (GAIS)	x	x	x
Selbstreflexion und Einsicht	Self-Reflection and Insight Scale (SRIS)	x	x	x
Offenheit für Erfahrungen	Offenheitsskala des NEO-Fünf- Faktoren-Inventars (NEO-FFI)	x	x	x
Selbstwirksamkeitserwartung	Skala zur allgemeinen Selbstwirksamkeits- erwartung (SWE)	x	x	x
Ungewissheitstoleranz	Ungewissheitstoleranzskala (UGTS)	x	x	x
Engagement für das Studium	Utrecht Work Engagement Scale für Studie- rende (UWES)	x	x	x

5.2.4 Statistische Datenauswertung

Die Überprüfung der quantitativen Fragestellungen erfolgte durch die statistische Aus-
wertung der mittels standardisierter Selbstbeurteilungsinstrumente erhobenen Daten,
mit Hilfe von SPSS 25.

Gruppenvergleich

Um zu überprüfen, ob sich die Trainingsgruppe und Kontrollgruppe hinsichtlich der
neun Konstrukte unterscheiden, wurde zu jedem der neun Konstrukte eine zweifakto-
rielle Varianzanalyse mit Messwiederholung (Mixed ANOVA) (vgl. Bühner/Ziegler 2017:
530) gerechnet. Das jeweilige Konstrukt ist die abhängige Variable (AV); die unabhä-
ngigen Variablen (UV) werden durch den Faktor Gruppenzugehörigkeit (AG x KG) und den
Faktor Zeit (T1 x T2) gebildet. Mit Hilfe der zweifaktoriellen Varianzanalyse mit Mess-
wiederholung wird überprüft, ob der Zwischensubjektfaktor (Gruppenzugehörigkeit),
der Innersubjektfaktor (Messzeitpunkt) und/oder das Zusammenwirken beider Fakto-
ren einen Unterschied erzeugen (vgl. Bühner/Ziegler 2017: 535).

Überprüfung der Effekte innerhalb der Trainingsgruppe

Für die Überprüfung der Effekte innerhalb der Trainingsgruppe wurde der t-Test für gepaarte Stichproben gerechnet, um zu überprüfen, ob sich die Mittelwerte der Trainingsgruppe vom Zeitpunkt vor der achtsamkeitsbasierten Lehrveranstaltung zum Zeitpunkt danach signifikant voneinander unterscheiden.

Überprüfung der Voraussetzungen

Vor der Anwendung der zweifaktoriellen Varianzanalyse mit Messwiederholung wurden zunächst die notwendigen Voraussetzungen überprüft. Um die Annahmen hinsichtlich der Normalverteilung der abhängigen Variablen sowie die Homogenität der Varianzen zu überprüfen, wurden der Shapiro-Wilk-Test und der Levene-Test angewandt. Laut Eid et al. kann für die Varianzanalyse jedoch von einer gewissen Robustheit gegenüber Verletzungen der Normalverteilung und Varianzhomogenität ausgegangen werden (vgl. 2013: 390). Auf die Überprüfung der Voraussetzung der Sphärizität konnte verzichtet werden, da bei zweifaktorieller Varianzanalyse mit Messwiederholung bei zwei Stufen eine Sphärizität gegeben ist (vgl. Eid et al. 2013: 461). Um Gruppenunterschiede in den Ausgangswerten zu untersuchen, wurden t-Tests für ungepaarte Stichproben (für den Vergleich von Interventions- und Kontrollgruppe), eine einfaktorielle Varianzanalyse (für den Vergleich der drei Universitäten) sowie Chi-Quadrat-Tests und Mann-Whitney-U-Tests gerechnet.

Für die Berechnung des t-Tests für gepaarte Stichproben ist eine Normalverteilung der Differenzen erforderlich, die mittels Shapiro-Wilk-Test überprüft wurde. Bei hinreichend großen Stichproben ($n > 30$) darf auch bei Abweichung von dieser Voraussetzung von einer Robustheit dieses Tests ausgegangen werden.

Für die Berechnung des t-Tests für ungepaarte Stichproben ist eine Normalverteilung der Differenzen erforderlich, die mittels Shapiro-Wilk-Test überprüft wurde, sowie die Varianzhomogenität, die mittels des Levene-Tests überprüft wurde.

Berechnung der Effektstärken

Zusätzlich zu der statistischen Signifikanztestung wird auch die Effektstärke d berechnet. Mit Hilfe der Effektstärke als einem standardisierten Maß ist ein Vergleich mit anderen Studien möglich. Die Berechnungen basieren auf der Formel von Cohens d (1988). Für den t-Test bei gepaarten Stichproben wird Cohens d_z als Effektstärke berechnet, indem die Differenz der Mittelwerte aus der Prä-Post-Testung durch die Standardabweichung geteilt wird.

$$d = \frac{M}{SD}$$

Die Effektstärke für Mittelwertunterschiede zwischen zwei Gruppen ergibt sich aus der Differenz der Gruppenmittelwerte, die durch die gepoolte Standardabweichung geteilt wird.

$$d = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sum (x_1 - \mu_1)^2 + (x_2 - \mu_2)^2}{n_1 + n_2 - 2}}}$$

Diese beiden Berechnungsvarianten werden für die Angabe der Effektstärke im Zusammenhang mit dem t-Test angewandt. Nach Cohen (1992) kann ein Effekt von $d = 0.2$ als klein, $d = 0.5$ als mittel und $d = 0.8$ als groß eingestuft werden.

Für die zweifaktorielle Varianzanalyse mit Messwiederholung wird das von SPSS berechnete partielle Eta-Quadrat (η_p^2) als Effektstärke angegeben. Die Grenzen für die Effektgröße des partiellen Eta-Quadrats (η_p^2) liegen nach Cohen (1988) bei $d = 0.01$ (klein), $d = 0.06$ (mittel) und $d = 0.14$ (groß). Für eine bessere Vergleichbarkeit wird im Ergebnissteil neben dem partiellen Eta-Quadrat (η_p^2) auch d nach Morris (2008) eine Effektstärke aus der d -Familie angeführt.

Morris schlägt für die Effektstärkenberechnung vor, die Differenz aus der Prä-Post-Veränderung in der Interventionsgruppe und der Prä-Post-Veränderung in der Kontrollgruppe zu bilden und diese durch die gepoolte Standardabweichung der Prä-Messung zu teilen.

$$\Delta = \delta T - \delta C = \frac{(\mu T, post - \mu T, pre) - (\mu C, post - \mu C, pre)}{\sigma}$$

Vermeidung der Alphafehlerkumulierung

Wie bereits angeführt, werden in diesem Forschungsstrang zahlreiche statistische Tests durchgeführt. Die Anzahl der Tests beeinflusst jedoch die Wahrscheinlichkeit, einen p-Wert unterhalb von 0.05 zu erhalten und somit ein Ergebnis als statistisch signifikant zu interpretieren (Fehler erster Art). Dieser Zusammenhang ergibt sich logischerweise aus der angewendeten Methodik statistischer Testverfahren. Zunächst wird ein Signifikanzniveau von 0.05 festgelegt, und daraufhin werden die statistischen Tests ausgeführt. Dabei stellt das gewählte Signifikanzniveau stets einen Kompromiss zwischen der Möglichkeit einer Entdeckung und der Akzeptanz fehlerhafter Testergebnisse dar. In den Sozialwissenschaften wird üblicherweise ein Signifikanzniveau von 0,05 verwendet, während in der Physik zum Beispiel mit einem extrem niedrigen p-Wert von weniger als 0.0000057 gearbeitet wird (Chen et al. 2017: 1725). Bei einem p-Wert von 0,05 besteht die Wahrscheinlichkeit, bei einem von 20 durchgeführten Tests ein fehlerhaftes Ergebnis zu erhalten. Jeder zusätzlich durchgeführte Test erhöht die Wahrscheinlichkeit eines solchen Fehlers. Diese Problematik wird als Alphafehlerkumulierung bezeichnet. Es existieren verschiedene Methoden, um diesem Effekt entgegenzuwirken, wie beispielsweise die Bonferroni-Korrektur oder die Bonferroni-Holm Korrektur (Chen et al. 2017: 1727). In dieser vorliegenden Arbeit wird die Bonferroni-Holm Korrektur angewendet. Diese Methode basiert auf der Anordnung der p-Werte in aufsteigender Reihenfolge und dem anschließenden Vergleich dieser p-Werte mit sukzessiv weniger konservativen Alphaniveaus. Die Bonferroni-Holm Korrektur ordnet die p-Werte zuerst nach ihrer Größe und vergleicht sie anschließend mit zunehmenden Grenzen. Die niedrigste Grenze wird mithilfe der normalen Bonferroni-Korrektur berechnet. Die nächste Grenze entspricht der Bonferroni-Korrektur, die angewandt worden wäre, wenn ein Test weniger durchgeführt worden wäre. Dieses Vorgehen wird fortgesetzt, wobei der größte p-Wert unverändert bleibt (Holm 1979). Dank dieser Methode weist die Holm-Methode niemals weniger Hypothesen zurück als die Bonferroni-Methode, während die Gesamtwahrscheinlichkeit, einen Fehler vom Typ I zu begehen, stets bei oder unter 0,05 bleibt.