

Entwicklung und Förderung der statistischen Kompetenz in  
Statistikgrundlagenkursen

**Inauguraldissertation**

zur

Erlangung des akademischen Grades eines  
Doktors der Naturwissenschaften

(Dr. rer. Nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Universität Greifswald

vorgelegt von

Dipl.-Psych. Ramona Stock

Greifswald, 09.01.2023

Dekan\*in: Prof. Dr. Gerald Kerth

1. Gutachter\*in: PD Dr. Anette Hiemisch

2. Gutachter\*in: Prof. Dr. Stephan Dutke

Tag der Promotion: 23.05.2023

# INHALTSVERZEICHNIS

Inhaltsverzeichnis.....	3
Zusammenfassung.....	5
Abstract.....	6
A. Einleitung.....	7
B. statistische Kompetenz als Lernziel.....	9
B.1 Statistische Kompetenz.....	9
B.1.1 Kognitive Voraussetzungen statistischer Kompetenz.....	9
B.1.2 Motivationale Voraussetzungen statistischer Kompetenz.....	15
B.2 Messung von Kompetenz.....	20
B.2.1 Messung der motivationalen Komponente von Kompetenz.....	22
B.3 Zusammenfassung.....	24
C. Untersuchung I – Die Bedeutung von Kompetenzüberzeugungen für die Entwicklung statistischer Kompetenz.....	26
C.1 Einleitung.....	26
C.2 Theoretischer Hintergrund.....	27
C.2.1 Selbstreguliertes Lernen.....	27
C.2.2 Selbstwirksamkeitserwartung und Selbstreguliertes Lernen.....	29
C.2.3 Feedback und Selbstreguliertes Lernen.....	32
C.2.4 Implizite Theorien und selbstreguliertes Lernen.....	34
C.3 Hypothesen.....	36
C.4 Methode.....	39
C.4.1 Stichprobe.....	39
C.4.2 Durchführung.....	39
C.4.3 Messinstrumente.....	39
C.5 Ergebnisse.....	42
C.5.1 Aktuelle Selbstwirksamkeitserwartung.....	42
C.5.2 Überschätzung.....	44
C.6 Diskussion.....	48
C.6.1 Selbstwirksamkeitserwartung und Lernverhalten.....	48
C.6.2 Selbstwirksamkeitserwartung und Leistung.....	50
C.6.3 Implizite Theorien.....	51
C.6.4 Praktische Implikationen.....	53
D. Untersuchung II – Die Förderung der Entwicklung statistischer Kompetenz durch individualisiertes Lernen im Inverted Classroom.....	54
D.1 Einleitung.....	54
D.2 Theoretischer Hintergrund.....	55

D.2.1 Wissenserwerb durch Individualisierung im IC .....	55
D.2.2 Methoden zur individuellen Förderung des Wissenserwerbs im IC.....	60
D.3 Hypothesen.....	63
D.4 Methode.....	65
D.4.1 Untersuchungsdesign.....	65
D.4.2 Struktur der Statistikkurse .....	66
D.4.3 Durchführung .....	68
D.4.4 Stichprobe.....	68
D.4.5 Messinstrumente .....	69
D.5 Ergebnisse .....	73
D.5.1 Voranalysen: Unterschiede zwischen den Gruppen .....	73
D.5.1 Statistikleistung in IC und TL .....	74
D.5.2 Homogenisierung im IC und TL .....	77
D.5.3 Statistikmotivation im IC und TL.....	78
D.5.4 Explorative Datenanalyse: Unterschiede zur Kontrollgruppe .....	80
D.6 Diskussion .....	83
D.6.1 Förderung des statistischen Wissenserwerbs im IC.....	83
D.6.2 Förderung der Motivation im IC .....	85
D.6.3 Limitationen und Ausblick .....	87
E. Gesamtdiskussion .....	89
Literatur.....	92
Anhang Untersuchung I.....	110
Anhang Untersuchung II .....	112
Eigenständigkeitserklärung .....	<b>Fehler! Textmarke nicht definiert.</b>
Lebenslauf .....	<b>Fehler! Textmarke nicht definiert.</b>
Danksagung .....	114

# ZUSAMMENFASSUNG

Ziel von Statistikkursen an Universitäten ist es, den Studierenden statistische Kompetenz als Grundlage des wissenschaftlichen Arbeitens zu vermitteln. Trotzdem verlassen Studierende diese Kurse teilweise mit statistischen Fehlkonzepten und können das Gelernte in Studium und Praxis nicht anwenden. Zudem bereiten den Studierenden hohe Statistikangst und ein geringes Interesse an Statistik als Teil ihres Studienfachs Probleme.

In der vorliegenden Arbeit geht es um die Frage, wie Statistikkurse zu Studienbeginn gestaltet sein sollten, um die statistische Kompetenz der Studierenden besser zu fördern. Dabei wird davon ausgegangen, dass für statistische Kompetenz neben kognitiven Voraussetzungen in Form von gut vernetztem und in Schemata gespeichertem Wissen in Statistik und anderen Bereichen, auch motivationale Voraussetzungen notwendig sind. Als zentrale motivationale Voraussetzung werden dabei Kompetenzüberzeugungen in Statistik aufgefasst. Solche Kompetenzüberzeugungen zeigen nicht nur kurz- und langfristige Zusammenhänge zur Leistung, sondern stehen auch mit anderen motivationalen Konstrukten wie Angst oder Interesse in Verbindung. In der folgenden Arbeit wurden zwei Untersuchungen durchgeführt, welche auf die Bedeutung dieser Kompetenzüberzeugungen bei der Entwicklung und Förderung der statistischen Kompetenz in Statistikkursen fokussierten.

Die erste Untersuchung befasste sich damit, ob Kompetenzüberzeugungen zu Beginn eines Statistikkurses mit Lernverhalten und Leistung zum Ende des Kurses zusammenhängen. Dabei interessierte vor allem, ob a) eine besonders hohe oder eine besonders realistische Überzeugung von der eigenen statistischen Kompetenz zu Kursbeginn von Vorteil ist und b) ob eine Überzeugung von der Veränderbarkeit der statistischen Kompetenz die Einflüsse der anfänglichen Überzeugungen moderieren kann. In der Untersuchung wurden  $n = 88$  Psychologiestudierende in mehreren Statistikkursen zu Beginn und zum Ende eines Semesters befragt. Es zeigte sich, dass eine hohe Kompetenzüberzeugung mit besseren Leistungen einhergeht – während die Kompetenzüberzeugung für bessere Leistungen nicht unbedingt auch realistisch sein muss. Zudem ist eine Überzeugung von der Veränderbarkeit von Kompetenz von Vorteil: sie kann negative Effekte einer geringen Kompetenzüberzeugung zu Beginn des Semesters auf die Leistung kompensieren, führt aber auch bei einer unrealistisch hohen Kompetenzüberzeugung zu besseren Leistungen.

In der zweiten Untersuchung wurde analysiert, ob ein nach dem Lehrformat des „Inverted Classroom“ (IC) unterrichteter Statistikeinführungskurs die Kompetenzüberzeugungen und damit verbundene Konstrukte von Statistikangst und Interesse sowie die Leistung von Studierenden fördern kann. Letzteres sollte darauf zurückzuführen sein, dass im Gegensatz zu „traditionellen Vorlesungen“ (TL) im IC Möglichkeiten zur individuellen Anpassung des Kurses an die Voraussetzungen von Studierenden zur Verfügung stehen und die Studierenden so beim Lernen weniger überfordert und motivierter sind. In der Untersuchung wurden  $n = 27$  Studierende ein Semester lang in einem Statistikkurs im IC und  $n = 43$  Studierende in einem Kurs als TL unterrichtet. Eine zusätzliche Kontrollgruppe (KG) von  $n = 24$  Studierenden erhielt keinen Kurs. Die Ergebnisse zeigten, dass Studierende im IC zwar bessere und auch homogener Leistungen erreichen als im TL, dies kann jedoch nicht auf die Überforderung oder Motivation beim Lernen zurückgeführt werden. Auch die Kompetenzüberzeugungen und das Interesse waren im IC deutlich höher. Sowohl im IC als auch in der TL wurde die Statistikangst im Semesterverlauf geringer, während sich in der KG keine Veränderung zeigt.

Zu Beginn der Statistikausbildung sollten infolgedessen Kompetenzüberzeugungen – insbesondere auch die Überzeugung von der Veränderbarkeit statistische Kompetenz – gefördert werden. Dazu sind individualisierte Lehrformate wie der IC geeignet, welche die Entwicklung statistischer Kompetenz und statistischer Kompetenzüberzeugungen ermöglichen. Solche individualisierten Lehrformate können nicht nur in Statistikkursen zu Beginn, sondern auch in Statistikkursen im weiteren Studienverlauf eingesetzt werden.

## ABSTRACT

University statistics courses aim to teach statistical competence as a basis for the competence to conduct scientific research. However students leave these courses with statistical misconceptions and cannot apply their knowledge on statistical problems in studies or practice. In addition, students in statistics courses often have to deal with high statistics anxiety and low interest in statistics.

The main question of this thesis is how introductory statistics courses can better promote students' statistical competence. It is assumed that statistical competence requires cognitive preconditions, such as cognitive schemata of statistical concepts, as well as motivational preconditions. Amongst the most important motivational preconditions for competence are self-perceptions of competence. Research shows that competence self-perceptions correlate with short- and longterm achievement and are connected to other motivational constructs such as anxiety and interest.

For this thesis two studies were conducted. These studies focused on the development and advancement of statistical competence in introductory statistics courses especially the role of competence self-perceptions.

The first study investigated the connection between competence self-perceptions at the beginning of a statistics course with learning behavior and achievement at the end of the course. The focus of the study were two-fold: a) Understand if it is better to have a high or an accurate (e.g. high but unrealistic) self-perception of one's competence and b) whether the correlations of competence self-perception with learning behavior and achievement can be moderated by a growth mindset of statistical competence.  $N = 88$  psychology students in different statistics course filled out a questionnaire at the beginning and the end of the course. Results show that the higher the competence self-perception, the higher the achievement, while the accuracy of the competence self-perception doesn't matter. In addition a growth mindset is important for those correlations. If the students assume that statistics competence can be changed, this compensates for low competence self-perceptions at the beginning of the course and leads to better achievement when those perceptions are unrealistically high.

The second study investigated if an introductory statistics course taught as "inverted classroom" (IC) compared to a course taught as "traditional lecture" (TL) positively affected self-perceptions and related constructs such as anxiety and interest and leads to better achievement. It was assumed that these positive effects on achievement could be explained through more individualized learning in the IC compared to the TL, which leads to less cognitive overload and more motivation while learning. For the study,  $n = 27$  students from different subjects and semesters were taught a course as IC while  $n = 43$  students attended the same course as TL. An additional control group of  $n = 24$  students did not attend a statistics course. Results show that student's achievement in the IC is higher and more homogenous. However, these results cannot be explained due to cognitive overload and motivation in the IC. In addition, competence self-perceptions and interest were higher in the IC than in the TL. IC and TL both reduced anxiety while there was no such reduction in the control group.

Competence self-perceptions – especially the perception that competence can be changed ("growth mindset") should be promoted at the beginning of a student's statistics education. Individualized courses such as the "inverted classroom" can promote these competence self-perceptions. Courses with individualized learning may be not only suited for introductory statistics courses but also for statistics courses in later semesters.

## A. EINLEITUNG

“Few college students escape taking an introductory course in statistics. But for many of these students, the lessons don't seem to stick. They remember the pain but not the substance.“

(Peterson, 1991, p.56; zitiert nach Gal & Ginsburg, 1994)

Statistikurse sind in vielen Studienfächern ein zentraler Bestandteil der wissenschaftlichen Ausbildung (Onwuegbuzie, 2003; Chiesi & Primi, 2010). In diesen Kursen sollen den Studierenden die wichtigsten statistischen Grundlagen – in Form statistischer Kompetenz – vermittelt werden (Rumsey, 2002). Statistische Kompetenz ist eine der Voraussetzungen dafür, dass im weiteren Studienverlauf die Kompetenz zum wissenschaftlichen Arbeiten (Forschungskompetenz) erworben werden kann (Schladitz et al., 2015; Schweizer et al., 2011; Thiel & Böttcher, 2014). Diese Kompetenz zum wissenschaftlichen Arbeiten ist ein wichtiges Ziel der akademischen Ausbildung an Universitäten (Groß Ophoff et al., 2017). Die statistische Kompetenz ist jedoch nicht nur als Grundlage für wissenschaftliches Arbeiten in Studium und Berufspraxis zu sehen, sondern sie ist auch für die Interpretation statistischer Informationen im Alltag von Bedeutung (Gal, 2002; Gal & Ginsburg, 1994; Garfield, 2002; Galligan, 2013). Die Entwicklung statistischer Kompetenz ist infolgedessen auch ein allgemeines Bildungsziel.

Tatsächlich haben Studierende jedoch oft Schwierigkeiten, die in Statistikkursen gelernten Konzepte sowohl auf Fragestellungen ihres Fachbereichs als auch auf statistische Problemstellungen außerhalb des Klassenzimmers anzuwenden (Fong et al., 1986; Ciarocco et al., 2013). Selbst die besten Absolventen<sup>1</sup> von Statistikkursen haben oft nur ein geringes konzeptuelles Verständnis der gelernten Inhalte (Clark et al., 2007). Statistische Fehlkonzepte liegen deshalb sowohl bei Wissenschaftlern in der Ausbildung als auch in der Praxis vor (Haller & Kraus, 2002; Sotos et al., 2007; Huck, 2009; Harris et al., 2009). Ein erfolgreicher Abschluss von Statistikkursen geht also nicht zwingend mit statistischer Kompetenz einher.

Aufgrund der Relevanz für Studium, Beruf und Allgemeinbildung stellt sich die Frage, wie die Kompetenzentwicklung durch die Statistikausbildung im Studium gefördert werden kann. Eine solche Förderung ist besonders in Statistikgrundlagenkursen von Bedeutung. Diese Kurse sind nicht nur die erste, sondern oft auch die einzige Lehrveranstaltung, in der Studierende in Kontakt mit Statistik kommen (Giesbrecht et al., 1997). Dementsprechend müssen hier die notwendigen Grundlagen für die spätere Kompetenzentwicklung im Studium gelegt werden (Rumsey, 2002). Zu diesen Grundlagen zählt nicht nur statistisches Wissen, sondern vor allem auch die Motivation. Oft sind Studierende in Statistikkursen bereits von Beginn an wenig motiviert: hohe Statistikangst und eine geringe Wertschätzung von Statistik, weil der Bezug der statistischen Inhalte zum eigenen Studienfach unklar

---

<sup>1</sup> Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung der Sprachformen für unterschiedliche Geschlechter (männlich, weiblich und divers) verzichtet und das generische Maskulinum verwendet. Mit sämtlichen Personenbezeichnungen sind jedoch ausdrücklich alle Geschlechter gemeint.

ist, sind typische Ursachen dieser mangelnden Motivation (Benson & Blackman, 2003; Chiesi & Primi, 2010; Condrón et al., 2018; Connors et al., 1998; Dykeman, 2011; Griffith et al., 2012; Macher et al., 2012; Onwuegbuzie & Wilson, 2003; Sciotto, 1995; Vittengl et al., 2004; Waples, 2016). Gelingt es nicht, die geringe Motivation der Studierenden frühzeitig zu verändern, kann sich dies nicht nur negativ auf die Kompetenzentwicklung im Kurs, sondern im gesamten Studienverlauf auswirken (Gal, Ginsburg & Schau, 1997; Roberts, 2016; Stark & Mandl, 2000). Um zu verhindern, dass der Einführungskurs in Statistik die Grundlage für eine Ablehnung statistischer Inhalte legt, ist es notwendig die Motivation bei der Ausgestaltung dieser Kurse mit zu berücksichtigen.

In der folgenden Arbeit geht es um die Frage, wie statistische Kompetenz – insbesondere unter Berücksichtigung der Motivation von Studierenden – in Statistikgrundlagenkursen gefördert werden kann. Um die Kompetenzentwicklung fördern zu können, muss dabei zunächst bekannt sein, was unter „statistischer Kompetenz“ als Lernziel zu verstehen ist, d.h. was die Studierenden zum Ende eines Statistikkurses überhaupt können sollen. Deshalb wird zu Beginn eine theoretische Einführung gegeben, in der die kognitiven und vor allem auch motivationalen Voraussetzungen beschrieben werden, welche für statistische Kompetenz notwendig sind und infolgedessen die Lernziele eines Kurses darstellen. Zudem wird darauf eingegangen, wie die statistische Kompetenz als Lernziel gemessen werden kann. Danach werden zwei Untersuchungen zur Kompetenzentwicklung in Statistikkursen vorgestellt. Die erste Untersuchung versucht die Frage zu beantworten, welche Rolle motivationale Voraussetzungen zu Beginn von Statistikkursen für die Kompetenzentwicklung im Kurs spielen. Die zweite Untersuchung stellt einen Statistikkurs in einem speziellen Lehrformat vor, welches in besonderem Ausmaß Motivation und Kompetenzentwicklung bei den Studierenden fördern könnte.



## **B. STATISTISCHE KOMPETENZ ALS LERNZIEL**

### **B.1 Statistische Kompetenz**

Kompetenz wird häufig definiert als eine Verbindung aus Wissen, Fähigkeiten und Einstellungen, die eine erfolgreiche Problemlösung in einem bestimmten Fachbereich erlaubt (Baartman et al., 2007; Klieme & Leutner, 2006; Koeppen et al., 2008; Zlatkin-Troitschanskaia et al., 2012; Weinert, 2001).

Im Fachbereich Statistik kann sich erfolgreiche Problemlösung beispielsweise darin zeigen, dass aus statistischer Information (z.B. statistischen Ergebnissen eigener Untersuchungen, statistischen Ergebnissen, die in wissenschaftlichen Studien präsentiert werden oder statistischer Information im Alltag) die richtigen Schlussfolgerungen abgeleitet werden. Diese Form der statistischen Kompetenz wird als *Statistical Reasoning* bezeichnet (Garfield, 1998; Lovett, 2001). Erweitert man diesen Begriff der statistischen Kompetenz, kann Kompetenz sich natürlich auch darin zeigen, dass für die eigene Untersuchung die korrekten statistischen Verfahren zur Auswertung ausgewählt und angewendet werden. Auch in diesem Fall steht statistische Information zur Verfügung (z.B. Art des Untersuchungsdesigns) aus der durch Schlussfolgerung das korrekte Verfahren abgeleitet werden muss (Lovett, 2001). Diese statistische Kompetenz stellt das Lernziel von Statistikgrundlagenkursen dar. Um dieses Lernziel zu erreichen, d.h. um zum Ende des Kurses statistische Aufgaben in Studium, Beruf und Alltag lösen zu können, müssen bei den Studierenden bestimmte kognitive und motivationale Voraussetzungen gefördert werden.

#### **B.1.1 Kognitive Voraussetzungen statistischer Kompetenz**

Das Vorhandensein kognitiver Voraussetzungen in Form von Wissen auf einem bestimmten Niveau ist zentral für die erfolgreiche Kompetenzanwendung. Welches Wissen für statistische Kompetenz notwendig ist, kann dabei aus einer strukturellen (*models of competence levels*) und einer inhaltlichen Perspektive (*models of competence structure*) untersucht werden (Koeppen et al., 2008). Bei der strukturellen Perspektive geht es darum, auf welche Art statistisches Wissen im Gedächtnis gespeichert bzw. organisiert sein sollte, um eine erfolgreiche Problemlösung zu ermöglichen. Die strukturelle Perspektive ist vor allem wichtig um zu verstehen, was Statistikanfänger von erfolgreichen Problemlösern auf Ebene des Wissens unterscheidet. Aus einer inhaltlichen Perspektive heraus interessiert, welche Wissensbereiche (z.B. Wissen über Statistik, Wissen über Forschungsmethoden, Wissen über einen bestimmten Fachbereich, z.B. Biologie) für die Problemlösung notwendig sind. Die inhaltliche Perspektive beschäftigt sich damit vor allem mit dem Unterschied zwischen der statistischen Kompetenz und der Forschungskompetenz.

##### **B.1.1.1 Struktur statistischer Kompetenz**

Nach dem Ansatz von Broers (2001, 2002, 2009) ist statistisches Wissen in Form propositionaler Netzwerke im Gedächtnis organisiert (vgl. Tyroller, 2005). Ein solches propositionales Netzwerk ist aus einzelnen

Knotenpunkten (Propositionen) und Verbindungen zwischen diesen Knotenpunkten (Beziehungen zwischen den Propositionen) aufgebaut. Eine Proposition ist die kleinste Wissenseinheit, die wahr oder falsch sein kann (z.B. die Aussage „ $-1 \leq r_{xy} \leq 1$ “) (Broers, 2001). Propositionen können sowohl semantisches Wissen (z.B. „ $r_{xy}$  beschreibt die Größe und Richtung des linearen Zusammenhangs zwischen  $x$  und  $y$ “) als auch prozedurales Wissen enthalten (z.B. die Berechnung von  $r_{xy}$  in einem Statistikprogramm) (Huberty et al., 1993; Kintsch, 1998; Lovett, 2001).

Durch Lernen werden neue Propositionen im Netzwerk ergänzt und neue Verknüpfungen zwischen den Propositionen hergestellt. Erst durch diese Verknüpfung von Propositionen untereinander kommt es zum konzeptuellen Verständnis (Broers, 2002; 2009). Dies bedeutet, dass alleine Wissen in Form isolierter Propositionen, welche statistische Definitionen oder die Durchführung einer statistischen Berechnung enthalten, nicht automatisch mit dem Verständnis eines statistischen Konzepts einhergehen. Beispielsweise korrelierte bei Huberty et al. (1993) die Leistung von Studierenden in mehreren Aufgaben, welche alle die Definition eines Begriffs erforderten, höher untereinander, als mit der Leistung in Aufgaben, die konzeptuelles Verständnis erforderten. Demnach hatten einige Studierende zwar umfangreiches propositionales Wissen (d.h. konnten alle Aufgaben zur Definition von Begriffen lösen), die einzelnen Propositionen waren für das Verständnis jedoch noch nicht ausreichend miteinander vernetzt. In einer Studie von Clark et al. (2007) wurden die besten Absolventen eines Statistikkurses nach dem Kurs zu statistischen Konzepten wie z.B. Varianz befragt. Obwohl alle das Konzept der Varianz definieren und diese berechnen konnten, zeigte keiner der Befragten ausreichendes Verständnis der Varianz. Dies lässt darauf schließen, dass bei Studierenden keine ausreichende Verknüpfung von Propositionen stattfindet. Die unzureichende Verknüpfung könnte eine Erklärung dafür sein, dass Studierende auch bei der Lösung statistischer Probleme oft wenig erfolgreich sind.

Für das erfolgreiche Lösen von statistischen Problemen – d.h. statistische Kompetenz – ist es notwendig, dass statistisches Wissen in Form von Schemata im Gedächtnis gespeichert ist (Sweller et al., 1998). Im propositionalen Netzwerk können Schemata als hierarchisch übergeordnete Strukturen verstanden werden, die durch Verknüpfungen vieler Propositionen entstehen (Tyroller, 2005). Die für Schemata notwendigen Verknüpfungen werden dadurch gebildet, dass bei der Problemlösung ähnliche Merkmale zwischen Problemen erkannt werden. Diese Merkmale werden abstrahiert und als Problemschema im Gedächtnis gespeichert (Broers, 2002). Beispielsweise kann es ein Schema für „Korrelationsprobleme“ geben. In diesem ist festgelegt, dass die Korrelation angewendet wird, wenn der Zusammenhang zwischen zwei kontinuierlichen Variablen analysiert werden soll. Das Schema kann zudem auch weiteres Wissen enthalten, beispielsweise wie der Korrelationskoeffizient berechnet und interpretiert wird.

Befunde deuten darauf hin, dass Studierende bei der Lösung statistischer Probleme häufig deshalb Schwierigkeiten haben, weil ihre Problemschemata nicht oder nur unzureichend ausgebildet sind. Dies zeigt sich beispielsweise bei Aufgaben, in denen Studierende Statistikprobleme in Kategorien klassifizieren sollen (z.B. „Korrelationsprobleme“, „t-Test Probleme“ etc.). Zwar konzentrieren sich

Studierende mit mehr Statistikerfahrung bei der Klassifikation tendenziell weniger auf oberflächliche Ähnlichkeiten (z.B. Fragestellungen zum Zusammenhang zwischen Erschöpfung und mentaler Aufmerksamkeit) als auf strukturelle Ähnlichkeiten (z.B. Skalenniveau der Variablen in der Fragestellung) zwischen den Problemen (Quilici & Mayer, 2002). Jedoch orientieren sich selbst Studierende, die bereits mehrere Statistikkurse besucht haben, nicht konsistent an strukturellen Ähnlichkeiten (Rabinowitz & Hogan, 2008). In einer Untersuchung von Lavigne et al. (2008) sollten drei Studierende, welche in einem Statistikkurs relativ gut abgeschnitten hatten, Statistikprobleme in gleiche Kategorien (z.B. „Probleme für einen t-Test für unabhängige Stichproben“) sortieren. Obwohl die Studierenden wussten, dass Merkmale wie Skalenniveau und Variablen für die Kategorisierung wichtig sind und Wissen über unterschiedliche Testverfahren hatten, fiel es ihnen schwer dieses Wissen bei der Sortierungsaufgabe auf die Statistikprobleme anzuwenden. Demnach scheint das statistische Wissen der Studierenden auch nach einem Statistikkurs oft noch nicht ausreichend in Schemata integriert zu sein.

Auch die Befunde mehrerer Untersuchungen von Lovett (2001) deuten drauf hin, dass Studierenden auch wenn man den gesamten Problemlöseprozess betrachtet, vor allem solche Schritte Schwierigkeiten bereiten, welche die Anwendung schematischen Wissens erfordern. In der Untersuchung sollten Studierende, die einen SPSS-Kurs abgeschlossen hatten, beispielsweise eine Statistikaufgabe lösen, welche die Auswahl, Durchführung und Interpretation der statistischen Analysen für eine Forschungsfrage erforderte. Während der Problemlösung zeigte sich, dass die Studierenden vor allem die ersten Schritte des Problemlöseprozesses nicht oder nur unzureichend ausführten. Diese Schritte betreffen die Übersetzung der Forschungsfrage in ein bestimmtes Schema eines statistischen Problems. Hingegen bereiteten die Schritte der Durchführung der statistischen Analysen den Studierenden, welche nur den Abruf auswendig gelernter Rechenschritte erfordert, kaum Schwierigkeiten. Erst der Bezug der Ergebnisse auf die Forschungsfrage schien für die Studierenden wieder problematisch zu sein, da auch dieser ein Verständnis der statistischen Konzepte voraussetzt. Auch in einer Studie von Hong und O’Neil (1992) zeigte sich, dass Studierenden bei der Problemlösung schematisches Wissen fehlt. Dort sollten mehrere Professoren und Studierende in der Promotionsphase statistische Probleme lösen und ihre Problemlöseschritte grafisch veranschaulichen. Dadurch sollten Unterschiede in den mentalen Modellen der Problemlöser erfasst werden. Studierende, welche die Probleme am besten lösten, wiesen ähnliche mentale Modelle auf, wie die Professoren. Bei schlechteren Problemlösern waren die mentalen Modelle nicht vorhanden oder fehlerhaft.

Auch fehlerhafte Schemata können Schwierigkeiten bei der statistischen Problemlösung bereiten. Letzteres ist der Fall, wenn Propositionen im Wissensnetzwerk nicht unzureichend, sondern falsch miteinander verknüpft wurden (Broers, 2002). Beispielsweise kann die Verknüpfung von Propositionen über den korrelativen Zusammenhang  $r_{xy}$  mit der Proposition „-1 ist kleiner als 1“ dazu führen, dass eine Korrelation von  $r_{xy} = -1$  als kleiner eingestuft wird, als  $r_{xy} = 1$  (Liu, Lin & Kinshuk, 2010). Durch diese Verknüpfung entsteht ein Fehlkonzept der Korrelation. Es wurde bisher bereits eine ganze Reihe typischer Fehlkonzepte in Statistik identifiziert (Cooper & Shore, 2008; Garfield, 1998; Liu, Lin & Kinshuk,

2010; Sotos et al., 2007). Diese Fehlkonzepte führen ebenfalls zu falschen Schlussfolgerungen bei statistischen Problemen (Garfield, 2000).

In Statistikgrundlagenkursen sollte daher als kognitive Voraussetzung der statistischen Kompetenz Wert auf ein vollständiges und korrektes Verständnis der statistischen Konzepte gelegt werden. Dies bedeutet, dass Wissen nicht in Form isolierter Propositionen vorliegt, sondern dass diese Propositionen korrekt miteinander verknüpft und in übergeordneten Schemata gespeichert sein müssen (Broers, 2002). Ist dies nicht der Fall, können unvollständige und fehlerhafte Konzepte lange Zeit bestehen, was dazu führt, dass eine falsche Anwendung statistischer Konzepte sogar bei praktizierenden Wissenschaftlern auftritt (Haller & Kraus, 2002; Harris et al., 2009).

Dies bedeutet, dass das Lernziel „statistische Kompetenz“ nicht erreicht werden kann, wenn Studierende einen möglichst großen Umfang an auswendig gelerntem Wissen wiedergeben können. Dieser Unterschied zwischen Wissen auf der Ebene einzelner Propositionen und Wissen auf der Ebene statistischer Kompetenz wird nicht zuletzt auch in dem bekannten theoretischen Ansatz der statistischen Kompetenz (*Statistical Reasoning*) von Garfield et al. (2008) thematisiert: *Statistical literacy* wird dort als Basiswissen beschrieben, das notwendig ist, um ein statistisches Konzept zu definieren, zu beschreiben oder zu interpretieren (Rumsey, 2002; Ziegler, 2014). Demzufolge erfordert *Statistical Literacy* nur den Abruf einzelner Konzepte aus dem Gedächtnis, sodass der kognitive Aufwand relativ gering ist. Um statistische Probleme auf der Ebene des *Statistical Reasoning* zu lösen, hingegen ist es notwendig mehrere Konzepte miteinander zu kombinieren (Garfield et al., 2008; Sabbag, Garfield & Zieffler, 2018).

#### B.1.1.2 Inhalt von statistischer Kompetenz und Forschungskompetenz

Statistische Kompetenz umfasst im engeren Sinne die erfolgreiche Analyse und Evaluation statistischer Information (Schladitz et al., 2015). Diese statistische Kompetenz ist jedoch nur ein Teilbereich der Forschungskompetenz bzw. der Kompetenz zum wissenschaftlichen Arbeiten in einem Fach. Forschungskompetenz bedeutet, den gesamten Forschungsprozess zu einer Fragestellung im eigenen Fach im Rahmen des Studiums durchführen zu können (Böttcher & Thiel, 2017; Gess et al., 2018; Schweizer et al., 2011). Der Forschungsprozess umfasst neben der Datenanalyse und – evaluation beispielsweise auch Schritte wie z.B. die Generierung einer Forschungsfrage, Planung und Durchführung der Datenerhebung und Verschriftlichung der Forschungsarbeit, für die eigenständige Kompetenzen notwendig sind (Bortz & Döring, 2016; Eid et al., 2017). Theoretische Modelle der Forschungskompetenz enthalten neben der statistischen Kompetenz daher weitere Teilbereiche, die andere Schritte des Forschungsprozesses wie z.B. die Literaturrecherche zur Ableitung einer Forschungsfrage oder die kritische Reflexion von Ergebnissen betreffen (Schladitz et al., 2015; Thiel & Böttcher, 2014). Auch empirisch lassen sich Hinweise darauf finden, dass sich die Forschungskompetenz in solche Teilbereiche zerlegen lässt: Beispielsweise ergab eine Faktorenanalyse zur Ermittlung der Struktur von drei Instrumenten zur Forschungsselbstwirksamkeit, dass sich diese am besten in die vier Faktoren *Data Analysis Self Efficacy*, *Data Collection Self Efficacy*, *Research Integration Self-Efficacy* und *Technical Writing Self Efficacy*

zerlegen lässt (Forester, Khan & Hesson-McInnis, 2004). Demnach scheint auch bei der selbsteingeschätzten Forschungskompetenz die statistische Kompetenz (in der Studie die *Data Analysis Self Efficacy*) nur einen Teil auszumachen. Dies lässt zunächst darauf schließen, dass aus inhaltlicher Sicht zwischen den Lernzielen „statistischer Kompetenz“ und „Forschungskompetenz“ differenziert werden kann: Ziel von Statistikgrundlagenkursen ist es, das notwendige inhaltliche Wissen für den Umgang mit statistischer Information zu liefern. Ziel der gesamten Statistikausbildung ist es hingegen, dieses inhaltliche Wissen über Statistik mit weiteren Inhalten zu Forschungsmethoden und fachlichem Hintergrundwissen zu ergänzen, sodass diese Wissensinhalte zusammen zur Forschungskompetenz integriert werden können.

In der häufig anzutreffenden modularisierten Statistikausbildung im Studium wird davon ausgegangen, dass diese Integration automatisch im Verlauf des Studiums stattfindet: nach dem ersten Statistikgrundlagenkurs wird das dort erworbene Wissen mit Wissen aus anderen Veranstaltungen des Fachs verknüpft, sodass die Studierenden zum Ende des Studiums erfolgreich Forschungsprobleme in ihrem Fach lösen können. Befunde zur Struktur der Forschungskompetenz deuten jedoch darauf hin, dass die inhaltlichen Teilbereiche nur teilweise zu einer übergeordneten Forschungskompetenz integriert werden. Bei einer Prüfung der faktoriellen Struktur der selbsteingeschätzten Forschungskompetenz nach dem fachübergreifenden RMKR-W Modell (Thiel & Böttcher, 2014) zeigte sich beispielsweise, dass ein Faktorenmodell mit der Forschungskompetenz als übergeordneten Faktor und den angenommenen fünf Teilbereichen (*Recherchekompetenz, Methodenkompetenz, Reflexionskompetenz, Kompetenz zur Kommunikation und fachliches Wissen*) als untergeordnete Faktoren keine bessere Passung zur tatsächlichen Struktur der selbsteingeschätzten Kompetenz aufweist, als ein Modell ohne einen übergeordneten Faktor (Böttcher & Thiel, 2017). Dies lässt darauf schließen, dass die Studierenden stark zwischen den Inhalten der einzelnen Teilbereiche von Forschungskompetenz differenzieren – und sie die Inhalte so möglicherweise auch in ihrer Wissensstruktur nur unzureichend miteinander verknüpft haben. Besonders groß scheint der Unterschied zwischen den Inhalten von Statistikveranstaltungen und anderen Veranstaltungen mit forschungsmethodischen Inhalten wahrgenommen zu werden. Beispielsweise ergab die empirische Prüfung der Struktur des BFK Modells der selbsteingeschätzten Forschungskompetenz in den Bildungswissenschaften (Schladitz et al., 2015), dass sich die faktorielle Struktur der Forschungskompetenz statt in die angenommenen drei Teilbereiche in nur zwei Teilbereiche zerlegen lässt: einen Faktor für die Teilbereiche Informationskompetenz und forschungsbezogenem kritisches Denken und einen Faktor für den Teilbereich statistische Kompetenz (Schladitz et al., 2015). Zudem zeigte eine Untersuchung zur Forschungskompetenz bei Psychologiestudierenden, dass sich diese in drei separate untergeordnete Faktoren zerlegen lässt, welche genau den Inhalten der unterschiedlichen Lehrveranstaltungen im Fach entsprechen: je eine Veranstaltung zur Grundlagenstatistik, zur Weiterführenden Statistik und zur Forschungsmethodik (Schweizer et al., 2011). Hier findet also ebenfalls nur eine teilweise Verknüpfung der statistischen Wissensinhalte der einzelnen Lehrveranstaltungen statt.

Vor allem für die Weiterentwicklung der statistischen Kompetenz zur Forschungskompetenz scheint es demnach sinnvoll zu sein, in Statistikgrundlagenkursen als kognitive Voraussetzungen der Kompetenz nicht nur isoliert statistisches Wissen zu vermitteln, sondern auch Wissensinhalte aus anderen Bereichen (z.B. methodische oder fachspezifische Inhalte) in den Kurs zu integrieren. Es zeigt sich beispielsweise, dass Studierende der Psychologie, die zwei Semester eine integrative Lehrveranstaltung zu Statistik und Forschungsmethoden besuchten, langfristig besser abschnitten, als Studierende des traditionellen Curriculums, in dem Statistik und Forschungsmethoden als separate Veranstaltungen angeboten wurden (Barron & Apple, 2014). Entscheidend für den Lernerfolg ist dabei die Verzahnung der Lehrveranstaltungen: werden Statistik und Forschungsmethoden zwar im selben Semester, jedoch als separate Lehrveranstaltungen unterrichtet, schneiden die Studierenden nicht besser ab (Scepansky & Carkenord, 2004). Für die Kompetenzentwicklung aus kognitiver Sicht ist es also notwendig, dass Studierende nicht nur über strukturell, sondern auch über inhaltlich gut vernetztes Wissen verfügen.

## **B.1.2 Motivationale Voraussetzungen statistischer Kompetenz**

Viele Kompetenzdefinitionen umfassen motivationale Komponenten als Bestandteil von Kompetenz (Weinert, 2001; Stoof et al., 2002; Zlatkin-Troitschanskaia & Pant, 2016). Dies ist darauf zurückzuführen, dass motivationale Faktoren (z.B. Interesse, Zielorientierungen, Kompetenzüberzeugungen) neben dem Wissen ebenfalls Einfluss auf Lernen und Leistung – und somit die Kompetenzentwicklung – haben (z.B. Elliott et al., 2017; Pugh & Bergin, 2006; Robbins et al., 2004). Demnach kann die Motivation der Studierenden eine wichtige Rolle dabei spielen, ob diese zum Ende eines Statistikkurses ein statistisches Problem erfolgreich lösen können oder nicht. Beispielsweise werden sich Studierende, die der Überzeugung sind, sie wären sowieso „schlecht“ in Statistik, weniger Mühe bei der Bearbeitung statistischer Aufgaben geben und so bei deren Lösung weniger erfolgreich sein – seien es nun Übungsbeispiele, Prüfungsaufgaben oder die wissenschaftliche Abschlussarbeit. Statistische Kompetenz erfordert demnach nicht nur kognitive, sondern auch motivationale Voraussetzungen.

In Statistik wurden als motivationale Voraussetzungen bisher vor allem Statistikangst und Statistikeinstellungen untersucht (z.B. Onwuegbuzie, 2003; Emmioğlu & Capa-Aydin, 2011; Macher et al., 2012; 2013; Nolan et al., 2012). Im Gegensatz zur Statistikangst umfassen Einstellungen in Statistik neben affektiven Inhalten auch statistische Kompetenzüberzeugungen sowie den Wert von Statistik (z.B. Roberts & Saxe, 1982; Schau et al., 1995; Wise, 1985).

### **B.1.2.1 Statistikangst**

Eines der Hauptprobleme der Statistikausbildung an Universitäten ist, dass viele Studierende in Statistikkursen bereits von Beginn an Angst vor Statistik haben (Condron et al., 2018; Macher et al., 2012; Onwuegbuzie & Seaman, 1995; Onwuegbuzie & Wilson, 2003; Waples, 2016). Diese Statistikangst kann als eine emotionale Reaktion verstanden werden, die von leichtem Unwohlsein bis hin zu starker Nervosität oder Panik reichen kann (Onwuegbuzie, DaRos & Ryan, 1997).

Die Statistikangst zu Studienbeginn ist möglicherweise dadurch zu erklären, dass die Studierenden mit dem Fach Mathematik bereits während ihrer Schulzeit vorwiegend negative Erfahrungen gemacht haben (Waples, 2016). Diese Mathematikangst wird auf das Fach Statistik übertragen, sodass die Studierenden bereits mit einer negativen Haltung in den ersten Kurs starten: bisher hatten sie immer Probleme mit Mathematik, warum sollte es in Statistik anders sein? In vielen Untersuchungen zeigt sich tatsächlich ein Zusammenhang zwischen Statistikangst und Mathematikangst (Baloğlu, 2004).

Eine hohe Statistikangst korreliert negativ mit der Statistikleistung (Ali & Iqbal, 2012). Diese negative Auswirkung auf die Leistung (z.B. in einer Prüfung) kann sowohl direkt als auch indirekt erklärt werden: Einerseits geht eine höhere Statistikangst mit einem erhöhten Angstzustand vor und während der Prüfung einher und wirkt darüber auf das Prüfungsergebnis (Macher et al., 2013). Andererseits korreliert eine hohe Angst vor mathematischen und statistischen Inhalten mit negativen Einstellungen gegenüber Statistikkursen und einer geringen Motivation, sich mit Statistik zu beschäftigen und

infolgedessen auch mit der Leistung (Tremblay et al., 2000). Zudem neigen Studierende mit hoher Angst dazu zu prokrastinieren, können infolgedessen weniger Zeit und Aufmerksamkeit für das Lernen aufwenden und erzielen so schlechtere Prüfungsergebnisse (Macher et al., 2012; Onwuegbuzie, 2004). Besteht also bereits zu Beginn des Studiums Statistikangst, behindert diese den Erfolg der gesamten Statistikausbildung.

Statistikangst steht in engem Zusammenhang mit einer negativen Kompetenzüberzeugung in Statistik. Für die Entstehung von Angst in einer bedeutsamen Situation im akademischen Kontext (z.B. einer Statistikprüfung) ist es entscheidend, welche Möglichkeiten der subjektiven Kontrolle (z.B. das Ergebnis der Statistikprüfung durch Lernen beeinflussen zu können) der Situation bestehen (Frenzel et al., 2020). Beispielsweise erleben Studierende, die davon überzeugt sind, statistische Inhalte aufgrund ihrer eigenen Kompetenz meistern zu können deutlich weniger Angst vor einem Statistikkurs (Condrón et al., 2018). Auch in anderen Untersuchungen korrelieren Überzeugungen der statistischen aber auch der mathematischen Kompetenz mit der Statistikangst (Bandalos et al., 1995; González et al. 2016; Macher et al., 2012, 2013). Die ist ein weiterer Hinweis darauf, warum bereits Erstsemester ohne jegliche Vorerfahrung Angst vor Statistik haben: sie erwarten aufgrund ihrer geringen Kompetenzüberzeugung in Mathematik, dass sie auch mit Statistik Probleme haben werden. Weiterhin zeigt sich in Messinstrumenten eine hohe Korrelation des Faktors Statistikangst in Form negativen Affekts und des Faktors der Kompetenzüberzeugungen (Nolan et al., 2012; Zimprich, 2012). Auch dies deutet auf den engen inhaltlichen Zusammenhang zwischen den Konstrukten hin. Kompetenzüberzeugungen werden teilweise sogar direkt als eine Komponente von Statistikangst erhoben (Onwuegbuzie, 2003).

Die Kompetenzüberzeugung ist somit ein wichtiger Prädiktor, wenn nicht sogar ein Bestandteil der Statistikangst. Je weniger die Studierenden zu Beginn des Studiums also von ihrer statistischen Kompetenz überzeugt sind, desto höher ist ihre Statistikangst. Infolgedessen scheint es für eine Reduktion der Statistikangst sinnvoll zu sein, die Kompetenzüberzeugung der Studierenden in Statistik zu fördern.

#### B.1.2.2 Kompetenzüberzeugungen

Kompetenzüberzeugungen zählen zu den wichtigsten motivationalen Einflussfaktoren auf die Leistung in akademischen Bereichen (z.B. Richardson et al., 2012; Robbins et al., 2004). Unter Kompetenzüberzeugungen in einem akademischen Bereich (z.B. Statistik) lassen sich unterschiedliche theoretische Konstrukte zusammenfassen (Marsh et al., 2017; Möller & Trautwein, 2015). Ein bekanntes Konstrukt ist die akademische Selbstwirksamkeitserwartung (Bandura, 1977). Diese bezeichnet die Überzeugung eines Lernalters, eine bestimmte Aufgabe im akademischen Bereich erfolgreich ausführen zu können (Raufelder & Ringeisen, 2015). Dies kann beispielsweise die Überzeugung sein, eine gute Note in einem Statistikkurs zu erreichen oder eine statistische Aufgabe lösen zu können (z.B. „Bitte geben Sie an, wie überzeugt Sie davon sind, die folgende Aufgabe aufgrund Ihrer aktuellen Fähigkeiten lösen zu können: den p-value eines statistischen Testverfahrens interpretieren.“; *Current Statistics Self Efficacy*, Finney & Shraw, 2003). Ein



weiteres Konstrukt der Kompetenzüberzeugung, das akademische Selbstkonzept, ergibt sich aus der Wahrnehmung und Evaluation der eigenen Kompetenz in einem Bereich (Möller & Trautwein, 2015). Es umfasst also die Überzeugung in Statistik „gut“ oder „schlecht“ zu sein (z.B. „Ich denke nicht schnell genug für Statistik.“, *Computational Self Concept*, Cruise et al., 1985). Obwohl theoretisch Unterschiede zwischen der Selbstwirksamkeitserwartung, dem Selbstkonzept und weiteren Konstrukten der Kompetenzüberzeugung bestehen sollten (Bong & Skaalvic, 2003; Pajares, 1996; Möller & Trautwein, 2020), überlappen sich diese empirisch stark und zeigen alle ähnliche Zusammenhänge zur Leistung (Hughes et al., 2011; Valentine et al., 2004). Demnach scheint es für die Statistikleistung unerheblich zu sein, ob Studierende ein hohes statistisches Selbstkonzept oder eine hohe statistische Selbstwirksamkeitserwartung haben: eine hohe Kompetenzüberzeugung geht mit besserer statistischer Kompetenz einher (Onwuegbuzie, 2003; Emmioğlu & Capa-Aydin, 2011; Nolan et al., 2012).

Kompetenzüberzeugungen werden aus bisherigen Erfahrungen mit der eigenen Kompetenz in einem Bereich abgeleitet (Bandura, 1994; Bong & Skaalvik, 2003). Eine geringe Überzeugung von der statistischen Kompetenz kann sich beispielsweise durch die Erfahrung Übungsaufgaben nicht lösen zu können oder schlechte Noten ergeben (Waples, 2016; Zimprich, 2012). Dabei spielt für die Ausbildung von Kompetenzüberzeugungen weniger der Erfolg oder Misserfolg bei der Kompetenzanwendung an sich eine Rolle, als vielmehr dessen Interpretation durch den Lerner (Schunk, 1985): eine schlechte Note geht beispielsweise eher dann mit einer negativen Kompetenzüberzeugung einher, wenn sich beim Lernen eigentlich angestrengt wurde oder wenn alle anderen bessere Noten haben.

Kompetenzüberzeugungen ergeben sich nicht nur aus bisherigen Erfahrungen bei der Anwendung der eigenen Kompetenz, sondern beeinflussen wiederum die zukünftigen Kompetenzanwendungen (Marsh, et al. 2017; Richard et al., 2006). Dies ist damit zu erklären, dass Studierende mit einer hohen Kompetenzüberzeugung in Statistik ein positiveres Lernverhalten zeigen, als Studierende mit geringer Kompetenzüberzeugung. Sind Studierende von ihrer statistischen und mathematischen Kompetenz überzeugt, erleben sie nicht nur weniger Statistikangst (und vermeiden dadurch statistische Inhalte), sondern nutzen auch mehr tiefenorientierte Lernstrategien, haben eine bessere Selbstregulation und sind persistenter und konzentrierter beim Lernen (Bandalos et al., 2003; Diseth, 2011; Liem et al., 2008; González et al., 2016). Infolgedessen erzielen diese Lerner, wenn sie ihre statistische Kompetenz (z.B. in einer Prüfung) anwenden sollen, bessere Resultate.

Interessant ist dabei, dass Kompetenzüberzeugungen häufig auch dann einen Einfluss auf zukünftige Leistungen haben, wenn bisherige Leistungen kontrolliert werden (Valentine et al., 2004). Dies lässt die Annahme zu, dass positive Kompetenzüberzeugungen über einen längeren Zeitraum hinweg zu einer immer stärkeren positiven Entwicklung der Kompetenz führen könnten – motiviert durch ihre positiven Kompetenzüberzeugungen wachsen die Studierenden über sich hinaus. In einer korrelativen Untersuchung von Arens et al. (2017) zeigte sich beispielsweise ein solcher reziproker Zusammenhang zwischen Kompetenzüberzeugungen in Mathematik und der Mathematikleistung über mehrere

Messzeitpunkte. Demnach können positive Kompetenzüberzeugungen vor allem auch langfristig die Kompetenzentwicklung unterstützen.

Aufgrund dieser Wechselwirkung zwischen Kompetenzüberzeugungen und Leistung scheint es besonders sinnvoll zu sein, in Statistikkursen nicht nur die kognitive Komponente der Kompetenz (i.S. von Statistikwissen), sondern auch die Kompetenzüberzeugungen in Statistik als motivationale Komponente der Kompetenz als Lernziel zu fördern (Huang, 2011; Sproesser et al., 2016). Dies gilt insbesondere für Kurse zu Studienbeginn, da sich bereits vorhandene negative Kompetenzüberzeugungen über die Statistikausbildung hinweg auch chronifizieren können: starten die Studierenden bereits mit einer geringen Kompetenzüberzeugung aufgrund ihrer Erfahrungen mit Mathematik in der Schule in die Statistikausbildung und machen dort erneut schlechte Erfahrungen, bestätigt dies ihre negativen Überzeugungen und es kommt anstatt zu positiver Kompetenzentwicklung zu einem Teufelskreis (Waples, 2016). Zudem scheint der erste Statistikkurs für die Förderung von Kompetenzüberzeugungen auch besonders geeignet zu sein: letztere ändern sich bei Anfängern in Statistik nämlich stärker als bei Fortgeschrittenen (Walker & Brakke, 2017; Zimmermann & Goins, 2012). Später im Studium haben sich die Kompetenzüberzeugungen also möglicherweise so gefestigt, dass sie nur schwer veränderbar sind.

### B.1.2.3 Wert und Interesse

Die fehlende Relevanz bzw. der geringe Wert von Statistik wird neben der Statistikangst oft als weitere Ursache für die geringe Motivation in Statistikkursen aufgeführt (Benson & Blackman, 2003; Chiesi & Primi, 2010; Condrón et al., 2018; Connors et al., 1998; Griffith et al., 2012; Sciutto, 1995; Vittengl et al., 2004; Waples, 2016). Studierenden fehlt häufig die Verbindung zwischen den statistischen Inhalten und denen des eigenen Studienfachs – Statistik wird als irrelevant für das Studium wahrgenommen. Dieses Problem behindert vor allem die Interessensentwicklung in Statistik: sowohl kurz- als auch langfristiges Interesse an Statistik beinhalten die persönliche Relevanz des Themas (Harackiewicz & Knogler, 2017). Studierende, die Interesse für ein Thema entwickelt haben, strengen sich beim Lernen wiederum mehr an und nutzen vermehrt tiefenorientierte Lernstrategien (Krapp, 1999; Pintrich & De Groot, 1990; Schiefele et al., 1995). Infolgedessen geht ein höheres Interesse im akademischen Bereich mit besseren Leistungen einher – besonders im Bereich Mathematik (Jansen et al., 2016; Schiefele et al., 1992). Dies lässt darauf schließen, dass es für die Kompetenzentwicklung in Statistik wichtig ist, dass Studierende Interesse am Fach entwickeln.

Während viele Ansätze zur Interessensentwicklung darauf fokussieren, die persönliche Relevanz des Themas zu erhöhen (Harackiewicz & Knogler, 2017), könnte auch das Erleben positiver Kompetenzerfahrungen in Statistik das Interesse am Fach fördern. Die Interessensentwicklung steht nämlich in engem Zusammenhang mit der Entwicklung von Kompetenzüberzeugungen (Hidi & Renniger, 2006; Köller et al., 2001; Marsh et al., 2005): Lerner, die positive Erfahrungen mit der Kompetenz in einem Bereich machen (d.h. gute Leistungen haben), interessieren sich auch mehr für diesen Bereich (Bandura

& Schunk, 1981; Sproesser et al., 2016). Bereits kurzfristig (z.B. im Verlauf weniger Unterrichtseinheiten) zeigt sich, dass eine Zunahme der Kompetenzüberzeugungen auch mit einer Zunahme im Interesse einhergeht (Fryer et al., 2022; Niemvitra & Tapola, 2007; Sproesser, et al., 2016). Langfristige korrelative Studien deuten ebenfalls darauf hin, dass bisherige Kompetenzerfahrungen und die daraus resultierenden Kompetenzüberzeugungen einen positiven Einfluss auf das Interesse haben (Du et al., 2021; Köller et al., 2001; Marsh et al., 2005; Viljaranta et al., 2014). Dabei ist der Zusammenhang teilweise auch reziprok: höheres Interesse geht mit einem höheren Selbstkonzept zu einem späteren Zeitpunkt einher (Marsh et al., 2005). Kompetenz und Kompetenzüberzeugungen können aufgrund ihrer Erhöhung des Interesses sogar zukünftige berufsbezogene Entscheidungen beeinflussen (Lent et al., 1994): Beispielsweise kann die Kompetenzüberzeugung in einem Fach selbst über die Leistung hinaus sowohl Interesse als auch Berufswahl voraussagen (Lent et al., 1993). Dies zeigt sich auch bei Entscheidungen über wissenschaftliche Karrieren. Sowohl bei angehenden Medizinerinnen als auch bei Psychotherapeuten korreliert die Kompetenzüberzeugung im Bereich der Forschung mit der Intention zur Verfolgung einer Karriere im Forschungsbereich (Wright & Holttum, 2012; Bierer et al., 2014). Dabei mediiert die Kompetenzüberzeugung die Beziehung zwischen den Fähigkeiten bzw. Kompetenzerfahrungen und dem Anstreben einer Forschungskarriere (Blanco, 2011; Adedokun, 2013). Dies lässt darauf schließen, dass die Förderung statistischer Kompetenzüberzeugungen in Statistikkursen nicht nur während des Studiums, sondern auch im Anschluss daran die Interessens- und damit auch die Kompetenzentwicklung unterstützen kann.

Die Kompetenzentwicklung während der Statistikausbildung wird aus motivationaler Sicht also mitbestimmt durch die Statistikangst, die statistischen Kompetenzüberzeugungen und die persönliche Relevanz von Statistik für die Studierenden. Darunter sind die Kompetenzüberzeugungen besonders wichtig, da positive Kompetenzerfahrungen Ängste bei den Studierenden abbauen und das Interesse fördern können. Insofern ist es in Statistikkursen zu Beginn des Studiums zunächst wichtig, solche positiven Kompetenzerfahrungen zu fördern. Der Fokus auf Kompetenzüberzeugungen als zentrale motivationale Voraussetzung ist auch von praktischer Bedeutung, da sich Kompetenzüberzeugungen leichter ändern als andere motivationale Konstrukte wie z.B. Einstellungen (Sizemore & Lewandowski, 2009; Walker & Brakke, 2017). Die Kompetenzüberzeugungen stellen also das wichtigste Lernziel der statistischen Kompetenzentwicklung aus motivationaler Sicht dar.

## B.2 Messung von Kompetenz

Das Lernziel von Statistikgrundlagenkursen ist es, dass Studierende nach Abschluss des Kurses erfolgreich statistische Probleme lösen – d.h. die korrekten Schlussfolgerungen für eine „statistischen Problemsituation“ ableiten – können. Um zu prüfen, ob die Studierenden diese Lernziel tatsächlich erreicht haben, können unterschiedliche Formen des Assessments eingesetzt werden.

Am besten sollte die Messung von Kompetenz als Lernziel über authentische Aufgaben erfolgen. Authentische Aufgaben sind der späteren (praktischen) Anwendungssituation so ähnlich wie möglich und erfordern zur Lösung eine Kombination aus Wissen, Fähigkeiten und Einstellungen (Gulikers et al., 2007; Lichtenberg et al., 2007). Eine solche authentische Aufgabe zur Messung der Forschungskompetenz ist beispielsweise die Bachelor- oder Masterarbeit. Für den Statistikbereich gibt es mittlerweile eine ganze Reihe von authentischen Assessments zur Messung von Kompetenz wie Portfolios (Sciutto, 2002) oder die Analyse empirischer Artikel (Bachiochi et al., 2011). Solche authentischen Assessments sind jedoch oft zu aufwendig, um sie in großen Gruppen – wie einem ganzen Statistikgrundlagenkurs – einzusetzen (Garfield & Chance, 2000).

Eine alternative ökonomischere Form des Assessments von statistischer Kompetenz in der Anwendung sind standardisierte Aufgaben mit vorgegebenen Antwortformaten (z.B. Multiple Choice). Solche Aufgaben werden häufig dafür kritisiert, nur für die Messung von auswendig Gelerntem und nicht zur Kompetenzmessung geeignet zu sein (Lichtenberg, et al., 2007; Stratmann et al., 2009). Es ist jedoch anzunehmen, dass dies auch von der Konstruktion der Aufgaben abhängig ist. Werden die kognitiven Voraussetzungen der statistischen Kompetenz als propositionales Netzwerk betrachtet, misst eine Aufgabe ein umso höheres Level an Kompetenz, je stärker eine Integration der einzelnen Propositionen im Netzwerk zur erfolgreichen Aufgabenlösung notwendig ist. Beispielsweise erfordert die Aufgabe „Was ist die Standardabweichung?“ nur den Abruf einer isolierten Proposition im Netzwerk. Wird die Aufgabe umformuliert z.B. in „Wie verändert sich die Standardabweichung, wenn ein weiterer, extremer Wert hinzukommt?“, müssen bereits mehrere Propositionen verknüpft werden – z.B. die der Standardabweichung mit der des Mittelwerts. Letztere Aufgabe ist so konstruiert, dass sie nicht einfach durch auswendig gelerntes Wissen lösbar ist. Standardisierte Aufgaben mit vorgegebenen Antworten können sogar so komplex formuliert werden, dass sie die Ableitung richtiger Schlussfolgerungen im Sinne der statistischen Kompetenz erfordern (z.B. aus mehreren vorgegebenen statistischen Verfahren das richtige für eine Forschungsfrage auswählen oder die richtigen aus mehreren vorgegebenen Schlussfolgerungen über statistische Ergebnisse einer Forschungsfrage wählen). Solche Aufgaben finden sich beispielsweise im *Statistical Reasoning Assessment* (SRA; Garfield, 1998; Garfield & Chance, 2000). Das SRA umfasst 20 Items im multiple Choice Format, die statistisches Schlussfolgern z.B. im Bereich Zufall, Wahrscheinlichkeit, Stichprobenziehung, Daten und deren Repräsentation oder Zusammenhänge messen. In den Items wird dabei besonderer Wert darauf gelegt, auch typische falsche statistische Schlussfolgerungen aufgrund von Fehlkonzepten zu messen. Interessanterweise korreliert

die Leistung im SRA nur sehr gering mit der Leistung, die über die üblichen Tests in einem Statistikkurs gemessen wurde (Garfield & Chance, 2000). Dies spricht jedoch nicht zwingend für eine geringe Validität des SRA zur Messung statistischer Kompetenz. Stattdessen könnte dies ein Hinweis darauf sein, dass die üblichen Tests in Statistik (z.B. Klausuren) keine Kompetenz im Sinne des *Statistical Reasoning* messen. Weiterentwicklungen des SRA wie das *Comprehensive Assessment of Outcomes in a First Statistics course* (CAOS; DelMas et al., 2006), die von einer Gruppe von Statistikexperten spezifisch für die Messung von *Statistical Reasoning* entwickelt und validiert wurden, zeigen auch, dass Studierende nach Abschluss eines Statistikgrundlagenkurses nur teilweise bessere Testwerte beim statistischen Schlussfolgern erreichen (DelMas et al., 2007). Demnach scheint es zur Förderung statistischer Kompetenz nicht nur sinnvoll zu sein, Statistik anders zu unterrichten, sondern auch andere Assessments einzusetzen, um die Lernziele in Statistikgrundlagenkursen zu messen. Für die Messung von Kompetenz als Lernziel können dazu auch Aufgaben mit vorgegeben Antwortformaten eingesetzt werden – wichtig ist, dass diese anwendungsbezogenes Wissen messen (Blömeke, 2013).

Statistische Kompetenz als Lernziel kann auch durch Selbsteinschätzungen der Kompetenz gemessen werden. Dazu sollen Lerner in Fragebögen meist ihre aktuelle Kompetenz in einem Bereich selbst einschätzen bzw. beurteilen wie beispielsweise im Fragebogen zur Bildungswissenschaftlichen Forschungskompetenz (BFK; Schladitz et al., 2015) (z.B. „Ich kann Daten aus Diagrammen, Tabellen und Texten gut erfassen und sicher interpretieren.“). Solche Selbsteinschätzungen der Kompetenz geraten immer wieder in die Kritik, da diese nur im geringen bis mittleren Bereich mit objektiven Kriterien der Kompetenzanwendung korrelieren (Blanch-Hartigan, 2011; Dunning et al., 2004). Im BFK Modell korrelieren die selbsteingeschätzten Kompetenzfacetten beispielsweise kaum mit den durch Testaufgaben gemessenen Facetten der Forschungskompetenz (Schladitz et al., 2015). Auch in Statistik sind die Zusammenhänge teilweise gering: Beispielsweise korreliert die selbsteingeschätzte Kompetenz der *Survey of Attitudes Towards Statistics* (SATS; Schau et al., 1995) in verschiedenen Untersuchungen zwischen  $r = .12$  und  $r = .64$  mit unterschiedlichen objektiven Kriterien der statistischen Kompetenz (z.B. Noten und Tests). Demnach bildet die subjektive Kompetenzmessung nicht unbedingt dasselbe Konstrukt ab, wie die objektive Kompetenzmessung – nämlich das aktuelle Statistikwissen. Geringe Korrelationen von Selbsteinschätzungen zeigen sich vor allem im unteren Kompetenzbereich, in dem es häufig zur Überschätzung der Kompetenz kommt, ein Phänomen, das in der Literatur als „unskilled but unaware“ bezeichnet wird (Hall et al., 2010; Hiemisch, 2012; Schlösser et al., 2013). Selbsteinschätzungen der statistischen Kompetenz sind infolgedessen nur bedingt geeignet, die notwendigen kognitiven Voraussetzungen für die Lösung eines statistischen Problems genau zu messen. Hingegen sind Selbsteinschätzungen zentral, wenn es um die Messung der motivationalen Komponente der Kompetenz – die Kompetenzüberzeugungen – geht: wie gut oder schlecht sich Lerner bei der Lösung statistischer Probleme einschätzen entspricht einer positiven bzw. hohen oder negativen bzw. niedrigen Kompetenzüberzeugung in Statistik.

## **B.2.1 Messung der motivationalen Komponente von Kompetenz**

Kompetenzüberzeugungen stellen das wichtigste motivationale Konstrukt für die statistische Kompetenzentwicklung dar. Infolgedessen sollten Kompetenzüberzeugungen ebenfalls als Lernziel von Statistikgrundlagenkursen gemessen werden. Dies kann vor allem deshalb sinnvoll sein, weil bei Messungen erfolgreicher statistischer Problemlösung (z.B. in Testaufgaben oder Klausuren) die Unterschiede in der Motivation von Studierenden nur eine untergeordnete Rolle spielen: im Normalfall sind alle Studierenden motiviert, eine möglichst gute Note zu erreichen und strengen sich an, egal ob sie von ihrer statistischen Kompetenz überzeugt sind oder nicht.

Kompetenzüberzeugungen im akademischen Bereich lassen sich auf unterschiedlichen Ebenen der Spezifität messen. Beispielsweise kann die gemessene Überzeugung sich auf die Kompetenz bei einer bestimmte Aufgabe, die Kompetenz in einem bestimmten Fach oder sogar die allgemeine akademische Kompetenz im Studium beziehen (Bong, 2002; Choi, 2005; Finney & Shraw, 2003). Kompetenzüberzeugungen sollen Leistungen besonders dann gut vorhersagen können, wenn die im Messinstrument erfragte Kompetenzüberzeugung mit der Kompetenz übereinstimmt, die in der Leistungssituation angewendet werden soll (Pajares, 1996). Demnach sollten beispielsweise allgemeine Kompetenzüberzeugungen im Fach Statistik die Leistung in einer Statistikaufgabe weniger gut vorhersagen können, als spezifische Kompetenzüberzeugungen in Bezug auf die Aufgabe (Finney & Shraw, 2003). Bisherige Befunde sind allerdings eher heterogen (Choi, 2005; Richardson et al., 2004; Bong, 2000; Finney & Shraw, 2003; Valentine et al., 2004): oft liefern Kompetenzüberzeugungen bezogen auf ein bestimmtes Fach die besten Vorhersagewerte. Demnach sollten Kompetenzüberzeugungen als Lernziel direkt bezogen auf das Fach Statistik gemessen werden. Dies ist auch deswegen sinnvoll, weil die so gemessenen Kompetenzüberzeugungen nicht nur Indikator für die Kompetenz zum Kursende sein sollen, sondern auch für die Kompetenzentwicklung in der weiteren Statistikausbildung.

Zudem gibt es Hinweise darauf, dass der Zeitpunkt der Messung von Kompetenzüberzeugungen und Leistung bei der Vorhersage von Leistungen eine Rolle spielt (Valentine et al., 2004): zu Beginn eines Semesters gemessene Kompetenzüberzeugungen in Statistik zeigen beispielsweise geringere Korrelation zur letztendlichen Kursleistung, als zum Ende des Semesters gemessene Überzeugungen (Nolan et al., 2012). Dies deutet einerseits darauf hin, dass Kompetenzüberzeugungen kurzfristige Leistungen besser abbilden als langfristige Leistungen. Immerhin entwickeln sich Kompetenzüberzeugungen gemeinsam mit der Leistung: je besser Studierende werden, desto höher sollten auch ihre Kompetenzüberzeugungen werden (und umgekehrt). Andererseits ist das Ergebnis ein Hinweis darauf, dass es für die Vorhersage von Leistungen sinnvoll sein könnte, zwischen Überzeugungen bezogen auf die aktuelle Kompetenz und Überzeugungen bezogen auf die Erlernbarkeit von Kompetenz zu unterscheiden (Finney & Shraw, 2003; Schunk, 1996). Während aktuelle Kompetenzüberzeugungen möglicherweise besser sind, um kurzfristige Leistungen abzubilden (z.B. den aktuellen Stand der Kompetenz nach einem Kurs), könnten Überzeugungen von der Erlernbarkeit von Kompetenz für langfristige Leistungen eine Rolle spielen (z.B. die weitere Kompetenzentwicklung

nach Abschluss eines Kurses). Bei der Messung von Kompetenzüberzeugungen nach Abschluss eines Statistikkurses ist es möglicherweise sinnvoll, beide Aspekte mit aufzunehmen: für die weitere Statistikausbildung im Studium ist es wichtig zu wissen, ob Studierende durch den Kurs auch subjektiv einen Kompetenzzuwachs wahrnehmen und ob dieser Kompetenzzuwachs sie dazu motiviert, auch weiterhin an ihrer statistischen Kompetenz zu arbeiten und Statistik nicht zu vermeiden.

## B.3 Zusammenfassung

Statistikurse zu Beginn des Studiums sollen Studierenden statistische Kompetenz vermitteln, um sie auf den Umgang mit Statistik im weiteren Studium, Beruf und Alltag vorzubereiten. Statistische Kompetenz bedeutet eine erfolgreiche Lösung statistischer Probleme (Weinert, 2001). Unter statistischer Problemlösung i.e.S. ist dabei die Ableitung korrekter Schlussfolgerungen aus statistischer Information zu verstehen (z.B. statistische Ergebnisse richtig auswerten, interpretieren oder kritisch bewerten) (Garfield, 1998).

Für die Ableitung der richtigen Schlussfolgerungen aus statistischer Information ist Wissen über statistische Konzepte notwendig. Eine erfolgreiche Problemlösung ist dabei umso wahrscheinlicher, je besser das Wissen in einem propositionalen Netzwerk integriert ist. Auf Ebene der Wissensstrukturen erfordert statistische Kompetenz dabei, dass die Propositionen als einzelne Wissenseinheiten im Netzwerk korrekt untereinander verknüpft und zu übergeordneten Einheiten (z.B. Schemata) zusammengefasst wurden (Broers, 2002; Mayer, 2008; Sweller et al., 1998). Auf der Ebene der Wissensinhalte kann für statistische Kompetenz neben Wissen über statistische Grundlagen auch Wissen über fachliche Hintergründe oder Forschungsmethoden notwendig sein. Unzureichende oder fehlerhafte Verknüpfungen des Wissens gehen mit mangelhaftem Verständnis statistischer Konzepte einher und begünstigen falsche Schlussfolgerungen (Broers, 2002). Ein Beispiel dafür ist bei der Interpretation statistischer Ergebnisse allein von der Regel auszugehen, dass  $p < .05$  ein erwünschtes Ergebnis im Signifikanztest darstellt, welches für die angenommenen Forschungshypothesen spricht. Dies muss jedoch nicht immer der Fall sein. Beispielsweise kann die Forschungshypothese der statistischen Nullhypothese des durchgeführten Signifikanztests entsprechen (in welchem Fall  $p < .05$  gegen die Forschungshypothese spricht). Wurde das Konzept der Forschungshypothese nicht mit dem der statistischen Hypothesen des Signifikanztests integriert, kann es also zu Fehlern bei der Interpretation der statistischen Ergebnisse kommen.

Ebenso kann es für die erfolgreiche Problemlösung entscheidend sein, dass man sich den Umgang mit Statistik zutraut. Kompetenzüberzeugungen zählen sowohl im akademischen Kontext allgemein als auch speziell im Fach Statistik zu den besten Prädiktoren der Leistung (Nolan et al., 2012; Richardson et al., 2012; Robbins et al., 2004). Wer nicht davon überzeugt ist statistische Konzepte verstehen zu können, wird sich für die Interpretation statistischer Information einfache und oberflächliche Strategien suchen (Diseth, 2011). Ein Beispiel dafür ist, dass viele Leser die statistische Information im Ergebnisteil wissenschaftlicher Artikel überspringen und sich stattdessen auf die „einfacher“ aufbereitete Darstellung im Diskussionsteil verlassen. Dies kann ebenso zur Ableitung falscher Schlussfolgerungen führen. Beispielsweise weil der Leser ohne die nötige statistische Hintergrundinformation aus dem Ergebnisteil die Interpretation der Ergebnisse durch den Autor der Studie missversteht. Kompetenzüberzeugungen sind auch langfristig für die Weiterentwicklung von Kompetenz von Bedeutung. Einerseits stehen sie in engem Zusammenhang mit der Entwicklung der Leistung (Arens et



al., 2017; Marsh et al., 2017): positive Kompetenzüberzeugungen fördern gute Leistungen und gute Leistungen erhöhen wiederum die Überzeugung von der eigenen Kompetenz. Andererseits können Kompetenzüberzeugungen auch Ängste abbauen und das Interesse am Fach fördern (Adedokun, 2013; Bandalos et al., 1995; Bierer et al., 2014; Blanco, 2011; Condrón et al., 2018; González et al., 2016; Macher et al., 2012, 2013; Wright & Holtum, 2012).

Statistische Kompetenz erfordert demnach sowohl kognitive als auch motivationale Voraussetzungen. Dabei sind die motivationalen Voraussetzungen nicht nur für die Kompetenzanwendung, sondern auch für die Kompetenzentwicklung von besonderer Bedeutung. Statistikgrundlagenkurse müssen so unterrichtet werden, dass diese Voraussetzungen aufgebaut werden.

Ob das Lernziel „statistische Kompetenz“ in einem Statistikgrundlagenkurs erreicht wurde, sollte sowohl mit objektiven Messungen der Kompetenz als auch mit subjektiven Messungen über die Selbsteinschätzung der Kompetenz gemessen werden. Objektive Messungen sollten auf Verständnis und Anwendung statistischer Konzepte fokussieren (Blömeke, 2013). Subjektive Messungen über Selbsteinschätzungen sind wichtig für die separate Messung der Motivation in Form der Kompetenzüberzeugungen, da diese mit objektiven Messungen nur unzureichend erfasst werden (Lichtenberg et al., 2007). Dabei muss berücksichtigt werden, welche Kompetenzüberzeugung als Lernziel gefördert und gemessen werden soll (z.B. aktuelle oder lernbezogene Kompetenzüberzeugungen; Finney & Shraw, 2003; Schunk, 1996).

Im Folgenden werden zwei Untersuchungen vorgestellt, welche sich mit der Förderung des Lernziels „statistische Kompetenz“ in Statistikgrundlagenkursen auseinandersetzen. Diese Untersuchungen fokussieren darauf, wie die Kompetenzentwicklung sich vor dem Hintergrund einer problematischen motivationalen Ausgangslage in Statistikgrundlagenkursen gestalten lässt. Dabei werden die Kompetenzüberzeugungen als zentrale motivationale Voraussetzung der statistischen Kompetenz betrachtet. In der ersten Untersuchung geht es darum, welchen Einfluss Kompetenzüberzeugungen zu Beginn eines Statistikkurses auf das Lernen und die Leistung zum Ende eines Kurses haben. Dabei wird zwischen dem Einfluss einer möglichst hohen vs. einer möglichst realistischen Kompetenzüberzeugung unterschieden. In der zweiten Untersuchung wird ein Statistikgrundlagenkurs in einem stark individualisierten Lehrformat – dem Inverted Classroom – vorgestellt. Es soll die Frage beantwortet werden, ob die Individualisierung im Kurs positive Lernerfahrungen ermöglicht und damit in besonderem Ausmaß statistische Kompetenz und positive Kompetenzüberzeugungen fördert.

# C. UNTERSUCHUNG I – DIE BEDEUTUNG VON KOMPETENZÜBERZEUGUNGEN FÜR DIE ENTWICKLUNG STATISTISCHER KOMPETENZ

## C.1 Einleitung

In selbstregulierten Lernprozessen des Kompetenzerwerbs spielen motivationale Einflussfaktoren eine entscheidende Rolle (Pintrich, 2000; Winne & Hadwin, 1998; Zimmermann, 2000). Dies gilt insbesondere für den Kompetenzerwerb in Statistik: Viele Studierende starten bereits mit ungünstigen motivationalen Voraussetzungen in ihre Statistikausbildung. Vor allem die Statistikangst wird als Problem thematisiert, da diese zu einer Vermeidung statistischer Inhalte führt und somit den Kompetenzerwerb behindert (Macher et al., 2012; Onwuegbuzie, 2004; Onwuegbuzie & Wilson, 2003; Tremblay et al., 2000). Für eine erfolgreiche Statistikausbildung ist es wichtig, diese Angst möglichst früh im Studium abzubauen.

Eine Möglichkeit zur Reduktion der Statistikangst könnte die Stärkung der Überzeugung der Studierenden von der eigenen statistischen Kompetenz zu Beginn der Statistikausbildung sein. Zu diesen Überzeugungen von der eigenen Kompetenz zählt die Selbstwirksamkeitserwartung (Bandura, 1977). Die Selbstwirksamkeitserwartung bezeichnet die Überzeugung eines Lerners, eine bestimmte Aufgabe in einem Bereich (z.B. Statistik) erfolgreich ausführen zu können (Marsh et al., 2017; Raufelder & Ringeisen, 2015) (z.B. Kann ich die Statistikaufgabe lösen? Kann ich den Statistikkurs bestehen? Kann ich Statistik bei meiner wissenschaftlichen Abschlussarbeit richtig anwenden?). Die Selbstwirksamkeitserwartung ist einer der wichtigsten Einflussfaktoren des selbstregulierten Lernens und der Leistung und komplementär zur Statistikangst: je überzeugter man davon ist, eine Aufgabe erfolgreich ausführen zu können, desto weniger wird diese Aufgabe aus Angst vermieden (z.B. Bandura, 1977; Condrón et al., 2018; González et al. 2016; Marsh et al., 2017; Pajares, 1996, 2002; Waples, 2016; Zimmerman, et al., 2017). Bisher galt die Annahme, dass die Leistung von Lernern umso höher ist, je mehr sie von ihrer eigenen Kompetenz überzeugt sind. Demnach wäre es für die statistische Kompetenzentwicklung vor allem zentral, dass Lerner eine hohe Selbstwirksamkeitserwartung haben, die ihnen das Gefühl vermittelt, statistische Aufgaben erfolgreich meistern zu können. Seit einiger Zeit wird jedoch die Frage diskutiert, ob eine zu hohe Selbstwirksamkeitserwartung die Kompetenzentwicklung auch behindern kann – nämlich dann, wenn sie auf einer Überschätzung der eigenen Kompetenz beruht (Möller & Trautwein, 2015). Wird die Kompetenz falsch eingeschätzt, kann es dazu kommen, dass der Lernprozess frühzeitig abgebrochen wird (z.B. wird die Prüfungsvorbereitung beendet, obwohl der Stoff noch nicht richtig „sitzt“) (Bol & Hacker, 2012; Hadwin & Webster, 2013; Sitzman & Yeo, 2013). Infolgedessen können Lerner trotz oder sogar aufgrund ihrer hohen Selbstwirksamkeitserwartung schlechte Leistungen erzielen, welche wiederum die Angst vor Statistik begünstigen. In diesem Fall wäre es für die statistische Kompetenzentwicklung vor allem zentral, dass die Selbstwirksamkeitserwartung den Stand der Kompetenz korrekt abbildet. Vor allem die Überschätzung des eigenen Kompetenzstands wird im Gegensatz zu einer zu geringen Einschätzung des Kompetenzstands häufig problematisch gesehen (Hattie, 2013; Stone, 1994).

Um Maßnahmen zu entwickeln mit denen die Statistikangst abgebaut und die Entwicklung der statistischen Kompetenz unterstützt werden kann, muss zunächst die Frage beantwortet werden: Schneidet ein Lerner umso besser im Statistikkurs ab, je überzeugter er von seinen statistischen Fähigkeiten ist? Oder ist es für die Erreichung der Kursziele auch wichtig, dass die Überzeugung von den statistischen Fähigkeiten mit den tatsächlichen Fähigkeiten übereinstimmt?

Zudem muss für die Kompetenzentwicklung berücksichtigt werden, dass die Selbstwirksamkeitserwartung größtenteils auf eigenen (Miss-)Erfolgen beruht und sich somit in Folge von Rückmeldungen über die eigene Leistung (z.B. Noten) verändern kann (Bandura, 1994; Lent et al., 1991; Richard et al., 2006). Allerdings ist diese Veränderung der Selbstwirksamkeit von der Interpretation der Rückmeldung abhängig (Usher & Pajares, 2008). Beispielsweise kann eine negative Rückmeldung den Lerner motivieren den Lernaufwand in Zukunft zu erhöhen, sodass trotz anfänglicher Überschätzung bessere Leistungen erzielt werden (Bi et al., 2017). Es kann jedoch auch sein, dass negative Rückmeldungen demotivieren und mit weiteren schlechten Leistungen einhergehen (Sitzman & Johnson, 2011). Letzteres sollte vor allem dann zutreffen, wenn Lerner die Annahme haben, dass Kompetenz durch Lernen nicht veränderbar ist (Dweck & Leggett, 1988). Diese implizite Annahme von der Unveränderbarkeit von Kompetenz hat einen negativen Einfluss auf die Setzung von Lernzielen, die Anstrengung und die Auswahl von Lernstrategien (Dweck & Molden, 2017). Für die langfristige Steuerung selbstregulierter Lernprozesse und der Leistung ist daher möglicherweise nicht nur entscheidend wie hoch die Selbstwirksamkeitserwartung von Lernern ist oder ob diese Selbstwirksamkeitserwartung mit der tatsächlichen Kompetenz übereinstimmt. Die Vorstellung des Lerners von der Veränderbarkeit von Kompetenz (Implizite Theorie) könnte diesen Zusammenhang beeinflussen (Dweck & Leggett, 1988). In Maßnahmen der Kompetenzförderung wäre es in diesem Fall auch wichtig, die Überzeugung der Lerner von der Veränderbarkeit der eigenen Fähigkeiten zu stärken.

In der folgenden Untersuchung geht es um den Zusammenhang zwischen der Überzeugung von der eigenen statistischen Kompetenz – die statistische Selbstwirksamkeitserwartung – zu Beginn eines Statistikkurses mit dem Lernen und der Leistung im Kurs. Dazu wird betrachtet, ob eine hohe Selbstwirksamkeitserwartung und eine geringe Überschätzung positiv mit der Kompetenzentwicklung zusammenhängen und ob diese Zusammenhänge durch die implizite Theorie von der Veränderung der Kompetenz moderiert werden. Zudem wird untersucht, ob sich die statistische Selbstwirksamkeitserwartung im Verlauf eines Statistikkurses verändert.

## **C.2 Theoretischer Hintergrund**

### **C.2.1 Selbstreguliertes Lernen**

Der größte Teil der Lernprozesse von Studierenden an Hochschulen erfolgt selbstreguliert. Selbstregulierte Lernprozesse sind dadurch gekennzeichnet, dass „die Person in Abhängigkeit von der Art ihrer Lernmotivation selbstbestimmt eine oder mehrere Steuerungsmaßnahmen (...) ergreift und den

Fortgang des Lernprozesses selbst überwacht.“ (Schiefele & Pekrun, 1996, S. 258). Der Lerner an der Hochschule bestimmt also selbst, was er lernt, wann und wie er lernt und worauf er hin lernt (Kopp, 2020). Demnach ist es für das Lernen im Hochschulkontext besonders wichtig, dass Lerner selbst motiviert sind.

Einer der bekanntesten Ansätze zum selbstregulierten Lernen ist das zyklische Modell der Selbstregulation nach Zimmerman (Zimmermann, 2000; Zimmerman et al., 2017). In diesem Modell wird selbstreguliertes Lernen aus einer sozial-kognitiven Perspektive betrachtet und die Rolle der Selbstwirksamkeitserwartung in allen Phasen des selbstregulierten Lernens betont (Kopp, 2020; Pajares, 2002; Zimmerman et al., 2017). Das Modell bietet demnach einen angemessenen Rahmen um nachvollziehen zu können, wie selbstreguliertes Lernen generell funktioniert und wie die Selbstwirksamkeitserwartung das selbstregulierte Lernen beeinflussen könnte. Nach dem Modell läuft das selbstregulierte Lernen in drei Phasen ab. In der Vorbereitungsphase (forethought phase) werden die Lernziele festgelegt und Lernstrategien entsprechend geplant (task analysis). Diese Vorbereitung wird von der Motivation (self-beliefs/values) des Lerners – insbesondere der Selbstwirksamkeitserwartung – gesteuert. In der nächsten Phase folgt die Ausführung der Handlung (performance phase). Während der Ausführungsphase überwacht der Lerner kontinuierlich die Zielerreichung (self-observation) und verwendet Strategien, um die Ausführung zu optimieren (self-control). In der letzten Phase (self-reflection phase) wird das Lernergebnis bewertet (self-judgement). Je nachdem, wie auf die Bewertung dieses Lernergebnis reagiert wird (self-reaction) kann wiederum die Motivation und Aufgabenplanung zukünftiger Lernprozesse beeinflusst werden. Demnach können bisherige Lernergebnisse (z.B. in Form von Rückmeldungen) die Selbstwirksamkeitserwartung verändern und/oder dazu genutzt werden, das selbstregulierte Lernen im nächsten Zyklus anzupassen. Das zyklische Modell der Selbstregulation ist in Abbildung C.1 dargestellt.

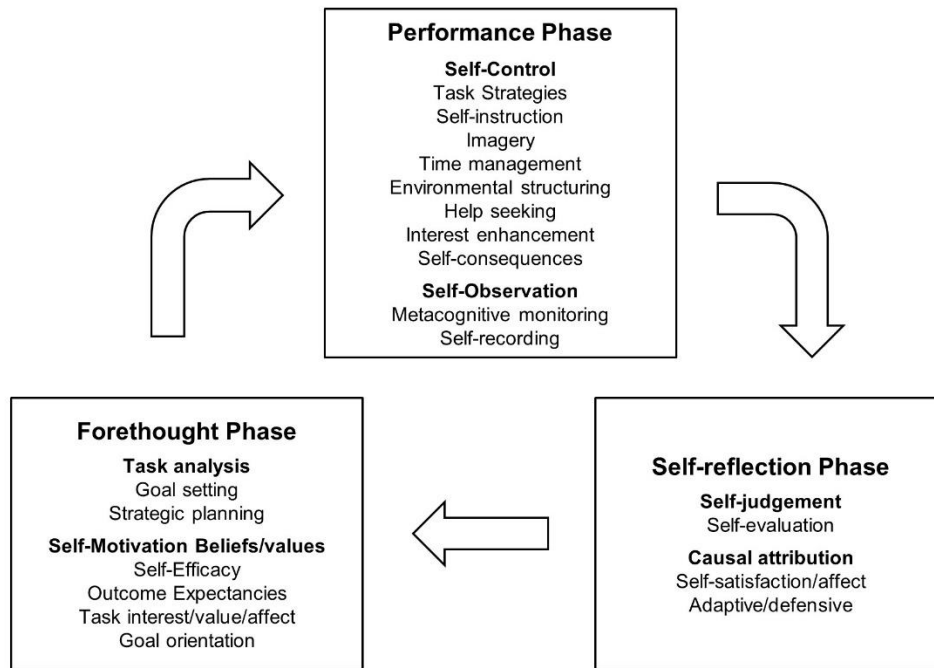


Abbildung C.1 Zyklisches Modell des Selbstregulierten Lernens (nach Zimmerman, Schunk & DiBenedetto, 2017)

## C.2.2 Selbstwirksamkeitserwartung und Selbstreguliertes Lernen

Die Selbstwirksamkeitserwartung kann im selbstregulierten Lernprozess zwei Funktionen erfüllen (Stone, 2000). Die erste Funktion ist die motivierende Unterstützung des selbstregulierten Lernens. Diese Motivation entsteht dadurch, dass die Selbstwirksamkeitserwartung dem Lerner Auskunft darüber gibt, welcher Zustand der Kompetenz in Zukunft erreichbar ist. Kann ich die statistischen Inhalte im Kurs verstehen? Kann ich den Kurs bestehen oder werde ich sogar mit der Bestnote abschließen?

Die Motivation aufgrund einer hohen Selbstwirksamkeitserwartung zeigt sich besonders in der Vorbereitungsphase (forethought phase) des selbstregulierten Lernprozesses. Beispielsweise korreliert die Selbstwirksamkeitserwartung mit den selbstgesetzten Lernzielen: Lerner mit einer höheren akademischen Selbstwirksamkeitserwartung streben höhere Lernziele in Form besserer Noten an und erreichen über die Ziele mediiert bessere Leistungen (Fan et al., 2008; Zimmerman et al., 1992). Zudem geht eine hohe Selbstwirksamkeitserwartung mit besseren Strategien bei der Erreichung von Lernzielen – d.h. günstigerem Lernverhalten – einher. Beispielsweise hat die Selbstwirksamkeitserwartung sowohl direkt als auch indirekt einen positiven Effekt auf den Einsatz tiefenorientierter Lernstrategien (Diseth, 2011, Liem et al., 2008; Walker et al., 2006). Lerner mit höherer Selbstwirksamkeitserwartung zeigen zudem eine höhere Anstrengungsbereitschaft bei schwierigen oder uninteressanten Aufgaben (Pintrich & DeGroot, 1990; Schunk, 1989; 1991). Die beiden letztgenannten Punkte sind in der Statistikausbildung besonders von Bedeutung: Einerseits ist für die statistische Kompetenz ein tiefergehendes Verständnis der Inhalte i.S. einer Verknüpfung von Wissensseinheiten notwendig (Broers, 2002). Diese Verknüpfung wird vor allem durch tiefenorientierte Lernstrategien erreicht (Alao & Guthrie, 1999). Andererseits ist das Interesse an

Statistik bei vielen Studierenden eher gering, da der obligatorische „Pflichtkurs“ in Statistik als nicht relevant für das eigene Studienfach wahrgenommen wird (Chiesi & Primi, 2010; Vittengl et al., 2004).

Andererseits kann die Selbstwirksamkeitserwartung auch eine metakognitive Funktion haben (Hadwin & Webster, 2013). Diese zeigt sich unter anderem in der Ausführungsphase (performance phase) bei der metakognitiven Überwachung des selbstregulierten Lernprozesses: sie gibt dem Lerner Auskunft über den aktuellen Stand der Kompetenz (Ist) und somit darüber, wie viel Lernaufwand noch investiert werden muss, um ein selbstgesetztes Lernziel (Soll) zu erreichen (Sitzman & Yeo, 2013; Vancouver et al., 2001; 2002; Vancouver & Kendall, 2006). Habe ich die statistischen Inhalte ausreichend verstanden? Habe ich genug gelernt, um die Prüfung im Kurs bestehen zu können? Dies wird deutlich, wenn man das selbstregulierte Lernen analog zu einem mechanischen Regelkreislauf betrachtet (Landmann et al., 2014). Das Erreichen eines gewünschten Zielzustands wird in diesem Kreislauf per Feedback über die aktuelle Diskrepanz zwischen aktuellem Zustand (Ist) und Zielzustand (Soll) geregelt. So lange eine Diskrepanz besteht, werden Handlungen initiiert, um diesen Zielzustand zu erreichen. Gibt es keine Abweichung mehr zwischen Ist und Soll, wird die Handlung abgebrochen. Um Lernhandlungen richtig zu planen ist es daher von Vorteil, wenn die aktuelle Selbstwirksamkeitserwartung (wahrgenommener Ist-Zustand) mit der aktuellen Kompetenz (tatsächlicher Ist-Zustand) übereinstimmt (Stone, 2000). Hingegen würde im Fall einer Überschätzung der wahrgenommene Ist-Zustand näher am Zielzustand (Soll) als der tatsächliche Ist-Zustand liegen. Infolgedessen kann es sein, dass Lernhandlungen zu früh abgebrochen oder gar nicht erst initiiert werden.

Die Übereinstimmung von wahrgenommenem und tatsächlichem Kompetenzzustand wird als Kalibrierung bezeichnet (Keren, 1991; Lichtenstein et al., 1982). Um die Kalibrierung zu ermitteln wird die Leistung bei einer Aufgabe mit der Überzeugung, diese Aufgabe lösen zu können (bzw. richtig gelöst zu haben), in Beziehung gesetzt (Hattie, 2013; Nietfeld et al., 2006; Stone, 2000). Die so erfasste Kompetenzüberzeugung ist demnach vergleichbar mit den spezifischen Selbstwirksamkeitserwartungen bezogen auf bestimmte Aufgaben, wie sie im akademischen Kontext häufig gemessen werden (Pajares, 1996): Kann ich die folgenden Testaufgaben lösen (z.B. Ferla et al., 2009; Finney & Shraw, 2003; Jansen et al., 2015; Lent et al., 1991)? Diese Überzeugung, eine bestimmte Aufgabe aufgrund des momentanen Kompetenzzustands lösen zu können, wird auch als *aktuelle Selbstwirksamkeitserwartung* bezeichnet (Finney & Shraw, 2003). Ein Lerner kalibriert also dann „richtig“, wenn die aktuelle Selbstwirksamkeitserwartung auf einer korrekten Wahrnehmung des momentanen Kompetenzstands basiert.

Je besser die Kalibrierung eines Lernalters, desto besser sollte dessen Selbstregulation im Lernprozess sein (Zimmerman & Moylan, 2009): anhand des wahrgenommenen aktuellen Kompetenzzustands werden so lange Lernhandlungen ausgeführt bis der angestrebte wahrgenommene Zielzustand der Kompetenz erreicht wird. Infolgedessen sollten Lerner mit guter Kalibrierung im Vergleich zu Lernern, die ihren tatsächlichen Kompetenzzustand überschätzen, bessere Leistungen erzielen. Beispielsweise haben Schüler, deren Kalibrierung in Mathematik in der fünften Jahrgangsstufe

besser ist, einen höheren Zuwachs an Leistungen von der fünften zur achten Klasse (Rinne & Mazzocco, 2014). Ein ähnlicher Zusammenhang zwischen Kalibrierung und Leistung zeigte sich auch bei Grundschulern über mehrere Quizaufgaben im Verlauf eines Schuljahrs: innerhalb eines Schülers geht eine bessere Kalibrierung in der vorherigen Aufgabe auch mit einer besseren Leistung in der nachfolgenden Aufgabe einher (Rutherford, 2017). Zudem kann ein Training von Psychologiestudierenden nicht nur deren Kalibrierung im Vergleich zu einer Kontrollgruppe verbessern, sondern die Verbesserung der Kalibrierung geht mit einem gleichzeitigen Zuwachs an Leistung einher (Nietfeld et al., 2006). Eine aktuelle Selbstwirksamkeitserwartung, die richtig kalibriert ist, scheint damit für die Kompetenzentwicklung von Vorteil zu sein.

Das Gegenteil könnte der Fall sein, wenn die aktuelle Selbstwirksamkeitserwartung auf einer starken Überschätzung der eigenen Kompetenz beruht. Eine höhere Kompetenzwahrnehmung würde in diesem Fall nach dem mechanischen Regelkreislauf zwar mit der Setzung höherer Lernziele einhergehen, gleichzeitig wird aber weniger Lernaufwand investiert, weil es zu einer Fehlsteuerung bei der Zielerreichung kommt (Hadwin & Webster, 2013; Vancouver et al., 2001). Infolgedessen erreichen die Überschätzer ihre angestrebten Lernziele nicht, d.h. sie erzielen schlechte Leistungen.

Solche negativen Zusammenhänge wurden bisher vor allem in Studien gefunden, in denen die Wirkung der Selbstwirksamkeitserwartung auf Leistungen nicht inter- sondern intraindividuell über mehrere Messzeitpunkte untersucht wurde. Beispielsweise zeigt sich, dass Studierende mit höherer Selbstwirksamkeitserwartung zu Beginn eines Kurses weniger Studienzeit fürs Lernen einplanen und schlechtere Leistungen erzielen – zumindest wenn die selbstgesetzten Lernziele konstant gehalten werden (Vancouver & Kendall, 2006). Auch in einer Untersuchung von Richard et al. (2006) in einem Statistikkurs, in der über mehrere Messzeitpunkte jeweils die Selbstwirksamkeitserwartung vor einem Test erfragt wurde, deuten die Befunde auf einen negativen Zusammenhang zwischen Selbstwirksamkeitserwartung und Leistung hin. Zudem wurde in einer Gruppe von Überschätzern auch interindividuell eine negative Korrelation zwischen der Selbstwirksamkeitserwartung beim ersten und der Leistung beim zweiten Test gefunden (Moores & Chang, 2009). Eine fehlkalibrierte Selbstwirksamkeitserwartung i.S. einer Überschätzung kann die Kompetenzentwicklung also behindern.

Eine Überschätzung der Kompetenz tritt insbesondere dann auf, wenn Lerner wenig Erfahrung in einem Bereich haben oder deren tatsächliche Kompetenz in dem Bereich eher gering ist (Hacker & Bol, 2019; Stone, 2000). Betrachtet man Studierende in ihrem ersten Statistikkurs ist es durchaus denkbar, dass diese beiden Punkte zutreffen. Die Studierenden haben einerseits keine Vorerfahrung in Statistik. Andererseits werden an einem Statistikkurs zu Beginn des Studiums häufig auch Studierende teilnehmen, deren statistische Kompetenz gering ist. Dies liegt daran, dass Statistikkurse oft Pflichtveranstaltungen in Fächern sind, die primär keine mathematische (oder gar keine naturwissenschaftliche Orientierung) aufweisen und eventuell sogar deshalb von den Studierenden gewählt wurden, weil diese schlecht im Mathematikunterricht waren und das Fach im Studium

vermeiden möchten. Demnach ist anzunehmen, dass eine falsche Kalibrierung bei Studierenden zu Beginn der Statistikausbildung ebenfalls ein Problem für Lernen und Leistung darstellen könnte.

### **C.2.3 Feedback und Selbstreguliertes Lernen**

Zusammenfassend scheint eine ungünstige Selbstregulation des Lernens also sowohl bei einer geringen statistischen Selbstwirksamkeitserwartung zu Studienbeginn (aufgrund der *motivationalen Funktion* der Selbstwirksamkeitserwartung) als auch bei einer Überschätzung der statistischen Kompetenz (aufgrund der *metakognitiven Funktion* der Selbstwirksamkeitserwartung) zu Studienbeginn gefördert werden zu können. Ist die Motivation zu gering, glauben die Lerner nicht daran, in Statistik eine gute Leistung erreichen zu können. Infolgedessen investieren sie nur wenig Aufwand beim Statistiklernen. Glauben die Lerner, in Statistik sehr viel besser zu sein, als sie tatsächlich sind, kann es zu Problemen bei der metakognitiven Steuerung des Lernprozesses kommen: Der zu investierende Lernaufwand wird falsch eingeschätzt. Beispielsweise korrelierte der selbst berichtete Lernaufwand bei einem Mathematikproblem mit der Genauigkeit der selbsteingeschätzten Kompetenz zur Lösung dieses Problems (Chen, 2003).

Infolgedessen werden Studierende in beiden Fällen schlechtere Leistungen erzielen. Dies äußert sich in Problemen beim Verständnis neuer Vorlesungsinhalte, falschen Lösungen von Hausaufgaben bis hin zu schlechten Prüfungsnoten. Diese Leistungen funktionieren als negatives Feedback (Stone, 2000). Dieses negative Feedback sollte nach dem Modell des selbstregulierten Lernens in der Selbstreflexionsphase (self-reflection phase) einen Einfluss darauf haben, wie Lerner ihre eigene Kompetenz einschätzen und somit auch auf deren Selbstwirksamkeitserwartung. Beispielsweise sollte eine Überschätzung der eigenen Kompetenz durch negatives Feedback korrigiert werden, sodass sich die Kalibrierung von Lernern mit der Zeit verbessert. Daraufhin sollten Lerner ihre Zielsetzung und Handlungsplanung im nächsten Lernzyklus anpassen (z.B. mehr Anstrengung investieren) und dadurch ihre Leistung verbessern.

Tatsächlich zeigen bisherige Befunde, dass negatives Feedback bei einer Überschätzung Lernprozesse verbessern kann. Beispielsweise sind Lerner, die sich zu Beginn eines Lernzyklus als überdurchschnittlich einschätzen, nach dem Lernzyklus unzufriedener mit ihren Leistungen. Diese Unzufriedenheit korreliert dann jedoch mit der Zunahme an Anstrengung im darauffolgenden Lernzyklus und diese wiederum mit der Leistung (Bi et al., 2017). Offenbar nutzen die Lerner das negative Feedback aus dem ersten Lernzyklus, um ihren Lernprozess anzupassen. Auch bei Labuhn, et al. (2010) konnten Schüler in der Gruppe der Überschätzer durch die Nutzung einer Kombination aus individuellem und sozialem Feedback über den Lernfortschritt ihre Kalibrierung und ihre Leistung verbessern.

Insbesondere negative Rückmeldungen über die eigenen Leistungen wie es bei einer Überschätzung der Fall ist, können Lernprozesse jedoch auch beeinträchtigen. Gibt man Probanden bei einer Entscheidungsaufgabe beispielsweise entweder positives Feedback, leicht negatives Feedback



oder stark negatives Feedback zeigt sich zwar, dass leicht negatives Feedback die Anstrengung am stärksten erhöht und die Leistung verbessert. Stark negatives Feedback geht im Vergleich zu den anderen Arten von Feedback allerdings mit weniger Anstrengung und Leistungsverbesserungen einher (Stone, 1994). In einer Untersuchung von Sitzman und Johnson (2012) zur beruflichen Weiterbildung mit mehreren Onlinemodulen zeigte sich, dass negative im Vergleich zu positiven Rückmeldungen über die eigene Leistung nach einem Modul schlechte Leistungen und hohe Dropoutraten begünstigen. Dies gilt sowohl, wenn die negativen Rückmeldungen die anfängliche Einschätzung der eigenen Kompetenz bestätigen (korrekte negative Selbsteinschätzung) als auch dann, wenn diese wiederlegt wird (Überschätzung). Möglicherweise führen negative Rückmeldungen bei Überschätzung sogar zu einer Art „Rebound“-Effekt, welcher die Selbstwirksamkeitserwartung sogar zu weit nach unten korrigiert. Beispielsweise geht eine fälschlicherweise zu hohe Einschätzung der bisherigen Leistung bei Schülern langfristig mit einer Abnahme des Selbstkonzepts einher, welche wiederum mit den Leistungen korreliert (Sticca et al., 2017). Zudem zeigt sich in einer Studie von Hadwin und Webster (2013), dass Studierende im Verlauf des Semesters die Erreichung ihrer selbstgesetzten Lernziele zwar nicht zuverlässiger vorhersagen können, d.h. die Studierenden wurden nicht besser in ihrer „Kalibrierung“. Allerdings nahmen ihre selbstgesetzten Lernziele im Verlauf des Semesters ab, d.h. die Überschätzung bzw. der Optimismus bei der Vorhersage der Erreichung der Lernziele reduzierte sich.

Neben den Befunden zu positiven und negativen Auswirkungen von Feedback auf Kalibrierung und Leistung gibt es zuletzt auch Untersuchungen, die darauf schließen lassen, dass Lerner Rückmeldungen über ihre Leistungen in einigen Fällen gar nicht zur Korrektur ihrer Überschätzung nutzen. Beispielsweise zeigt sich, dass sich die Übereinstimmung zwischen der Selbsteinschätzung der erreichten Testleistung und der tatsächlich erreichten Testleistung (Kalibrierung) auch nach mehreren Tests im Verlauf eines Semesters nicht systematisch verbessert. Vor allem der Anteil der Studierenden, die ihre Leistung überschätzten, ist auch zum Ende des Semesters noch hoch im Verhältnis zu den Unterschätzern (Hawker et al., 2016). Sogar wenn die Studierenden nach jedem Test Feedback erhalten zeigt sich, dass die Studierenden im Durchschnitt nicht nur in jedem Test ihre Leistung überschätzen – diese Überschätzung erhöht sich sogar mit jedem Test: die Studierenden schätzten trotz Feedback für jeden Test ein, dass sie besser abschneiden würden, als im vorherigen Test (Foster et al., 2017). Bei Hacker et al. (2000) zeigte sich zwar ein Anstieg des Zusammenhangs zwischen vorhergesagter und tatsächlicher Testleistung über mehrere Tests hinweg, nachdem Studierende gebeten wurden, Quellen ihrer Über- und Unterschätzung zu identifizieren und Fehler zu verbessern. Dieser Anstieg war aber vor allem in der Gruppe von Studierenden zu beobachten, die bereits im ersten Test am besten abschnitten. Diese Gruppe gehörte möglicherweise von Beginn an zu den guten „Kalibrierern“, da die Genauigkeit der selbsteingeschätzten Kompetenz umso besser ist, je höher die tatsächliche Kompetenz (Bol & Hacker, 2012; Hacker & Bol, 2019; Hattie, 2013). Bei schlechteren Studierenden zeigte sich über alle Tests hinweg kein Zusammenhang zwischen vorhergesagter und erreichter Testleistung. Studierende scheinen negatives Feedback also von selbst nicht notwendigerweise zu nutzen, um ihre Überschätzung zu korrigieren.

Insgesamt scheint es also unterschiedliche Möglichkeiten zu geben, wie sich negatives Feedback auf den Lernprozess auswirkt: Eine Korrektur einer überhöhten Selbstwirksamkeitserwartung durch negatives Feedback im Lernprozess kann einerseits dazu führen, dass Lerner ihre Anstrengung erhöhen, um die angestrebten Lernziele zu erreichen. Das Feedback kann aber auch dazu führen, dass Lerner aufgeben und gar keine Anstrengung mehr investieren. Zudem scheint negatives Feedback in einigen Fällen nicht zu einer besseren Übereinstimmung zwischen Selbstwirksamkeitserwartung und tatsächlichem Kompetenzzustand zu führen. Wie sich negatives Feedback auf die Selbstwirksamkeit und zukünftiges Lernen auswirkt, könnte durch die impliziten Theorien eines Lerners beeinflusst werden.

#### **C.2.4 Implizite Theorien und selbstreguliertes Lernen**

Implizite Theorien (Dweck & Leggett, 1988) oder Mindsets (Dweck, 2006; Dweck & Molden, 2017) sind die Überzeugungen eines Lerners bezüglich der Veränderbarkeit von Kompetenz. Personen mit einer *inkrementellen Theorie* bzw. einem *Growth Mindset* sind davon überzeugt, dass sich Kompetenz verändern lässt während Personen mit einer *Entitätstheorie* bzw. einem *Fixed Mindset* vom Gegenteil ausgehen.

Nach dem Sozialkognitiven Ansatz der Motivation (Dweck & Leggett, 1988) bestimmen implizite Theorien, wie sich Rückmeldungen – insbesondere negative Rückmeldungen – über die aktuelle Kompetenz auf das Lernen auswirken. Lerner mit einer Entitätstheorie neigen dazu, Leistungen ihrer eigenen Kompetenz zuzuschreiben und zwar vor allem dann, wenn sie diese als Misserfolg wahrnehmen (d.h. wenn es sich um schlechte Leistungen handelt). Hingegen attribuieren Lerner mit inkrementeller Theorie sowohl Erfolge als auch Misserfolge auf ihre eigene Anstrengung (Robbins & Pals, 2002). Infolgedessen sollten letztere auch negatives Feedback nutzen, um ihren Lernprozess besser steuern zu können (Ehrlinger et al., 2016). Beispielsweise attribuieren Lerner mit inkrementeller Theorie negative Rückmeldung in einem Test nicht nur eher auf einen Mangel an Anstrengung (Hong et al., 1999). Zudem erklären sich Schüler, die bisher negative Rückmeldung über ihre Leistungen in Englisch erhalten haben, bei inkrementeller Theorie eher bereit einen weiteren Kurs zu belegen um ihre Leistung zu verbessern, als Schüler mit Entitätstheorie (Hong et al., 1999). In einer zweijährigen Längsschnittstudie von Blackwell et al. (2007) zum Lernverhalten im Fach Mathematik zeigte sich, dass Schüler, die zu einer stabilen Theorie tendieren, schlechte Lernergebnisse eher auf ihre mangelnde Kompetenz attribuieren und danach weniger motiviert sind, ihre Kompetenz zu verbessern. Schüler mit stabiler Theorie erzielten bereits nach kurzer Zeit schlechtere Noten, als ihre Klassenkameraden mit inkrementeller Theorie obwohl sich die Leistungen zu Beginn der Erhebung nicht unterschieden. Taberero & Wood (1999) fanden, dass Lerner mit einer inkrementellen Theorie nach einem ersten Lerndurchlauf mit anschließendem Feedback eine höhere Selbstwirksamkeitserwartung (bezüglich ihrer Leistung im folgenden Durchlauf) aufweisen, weniger unzufrieden mit der eigenen Leistung sind und sich für den folgenden Durchlauf höhere Ziele setzen. Diese Unterschiede nach dem ersten Lerndurchlauf mediieren zudem die Wirkung der

impliziten Theorien auf die Leistung nach dem zweiten Durchlauf. Die Effekte sind unabhängig von der Leistung im ersten Durchlauf. In einer Studie von Dimotakis et al. (2017) wurde der Effekt von Feedback über die Leistung in einem Assessment Center auf die Selbstwirksamkeitserwartung untersucht. Dabei zeigte sich, dass die implizite Theorie den Zusammenhang zwischen negativem Feedback und der Überzeugung, die Leistung im Assessment Center in Zukunft verbessern zu können, moderiert. Während die Selbstwirksamkeit bei Personen mit Entitätstheorie mit steigendem negativem Feedback abnahm, blieb sie bei Personen mit Inkrementeller Theorie konstant. Hier hatte das Feedback also offenbar keine demotivierende Wirkung.

Ähnliche Effekte zum Feedback wurden auch für die Lernzielorientierung gefunden. In einer Untersuchung von Dahling & Ruppel (2016) wurde das Feedback der Teilnehmer in einem Wissenstest manipuliert. Teilnehmer, die negatives Feedback erhielten, hatten nach dem Feedback eine geringere Selbstwirksamkeitserwartung bezogen auf einen Wiederholungstest – allerdings nur dann, wenn sie eine geringe Lernzielorientierung aufwiesen. Da sich eine Lernzielorientierung aus einer inkrementellen Theorie ableiten lässt (Blackwell et al., 2007; Dweck & Leggett, 1988; Robbins & Pals, 2002), handelte es sich dabei vermutlich um Teilnehmer, die ihre Kompetenz als wenig veränderbar einschätzen.

Anscheinend interpretieren Lerner mit inkrementeller Theorie im Vergleich zu ihren Mitlernern also auch eine schlechte Leistung nicht negativ, erleiden dadurch keinen Verlust ihrer Selbstwirksamkeit und geben nicht auf. Im Gegenteil, die Leistung wird möglicherweise sogar als Information genutzt um sich in Zukunft mehr anzustrengen.

Demnach scheint die Interpretation von Rückmeldungen über den eigenen Kompetenzstand und deren Auswirkungen auf den Lernprozess von den impliziten Theorien abhängig zu sein. Dweck und Leggett (1988) gehen deshalb von einem Moderationseffekt zwischen impliziten Theorien und der aktuellen Überzeugung von der eigenen Kompetenz aus (Chen & Tutwiler, 2017; Cury, et al., 2006; Hong et al., 1999). Lerner mit inkrementeller Theorie gehen auch dann, wenn sie davon überzeugt sind, dass ihre aktuelle Kompetenz in einem Bereich gering ist (sie zuvor also negative Rückmeldungen über ihre Kompetenz erhalten haben) schwierige Aufgaben an und zeigen höhere Ausdauer. Erhalten Lerner mit Entitätstheorie hingegen negatives Feedback und schätzen daher ihre aktuelle Kompetenz als gering ein, vermeiden sie weitere Lernhandlungen, zeigen also „hilfloses Verhalten“ (Dweck & Leggett, 1988).

Die wenigen Untersuchungen, die einen solchen Moderationseffekt mit in ihre Analysen einbezogen, konnten diesen jedoch nicht eindeutig belegen (Chen & Tutwiler, 2017). Bei Tabernero und Wood (1999) zeigte sich keine Interaktion zwischen der Selbstwirksamkeit nach dem ersten Lerndurchlauf und impliziter Theorie auf die Zielsetzung für den nächsten Durchlauf. Bei Braten und Olausson (1998) verwendeten Lerner mit einer Kombination aus hoher akademischer Selbstwirksamkeitserwartung und inkrementeller Theorie signifikant mehr Lernstrategien, die anderen Gruppen unterschieden sich jedoch nicht.

Der Moderationseffekt impliziter Theorien bezüglich der Selbstwirksamkeitserwartung oder gar der Überschätzung von Kompetenz auf die Leistung – insbesondere auf langfristige Leistungen – wurde bisher jedoch noch nicht untersucht.

### C.3 Hypothesen

In dieser Studie soll überprüft werden, wie sich eine ungünstige Selbstwirksamkeitserwartung moderiert durch die implizite Theorie der Veränderbarkeit von Kompetenz auf Lernverhalten und Leistung in einem Statistikkurs auswirkt (Dweck & Leggett, 1988). Als ungünstig für selbstreguliertes Lernen werden sowohl eine geringe Selbstwirksamkeitserwartung als auch eine Selbstwirksamkeitserwartung, die auf einer starken Überschätzung beruht, aufgefasst. Zudem wird untersucht, welche Rolle die implizite Theorie bei der Veränderung der Selbstwirksamkeitserwartung spielt.

Je höher die Selbstwirksamkeit von Lernern, desto motivierter sind diese: je höher die Selbstwirksamkeitserwartung, desto mehr Anstrengung wird in eine Aufgabe investiert und desto höher ist die Persistenz wenn Probleme bei der Aufgabebearbeitung auftreten (Schunk, 1989; 1991). Umgekehrt sollten Lerner mit einer geringen Selbstwirksamkeitserwartung in Statistik zu Beginn eines Kurses wenig motiviert sein und sich beim Statistikkennen gar nicht erst anstrengen.

H1a: Es gibt einen positiven Zusammenhang zwischen der statistischen Selbstwirksamkeitserwartung und dem Lernverhalten.

In zahlreichen Untersuchungen wurden zudem positive Effekte einer hohen Selbstwirksamkeitserwartung auf die Leistung im akademischen Bereich gefunden (Richardson et al., 2012; Robbins et al., 2004; Valentine et al., 2004). Insbesondere in Statistik könnte die langfristige Leistung durch eine geringe Selbstwirksamkeitserwartung beeinträchtigt werden. Lerner mit einer geringen Selbstwirksamkeitserwartung verwenden weniger komplexe tiefenorientierte Lernstrategien (Diseth, 2011, Liem et al., 2008; Walker et al., 2006). Zu diesen komplexeren Lernstrategien zählen beispielsweise die Organisation und Elaboration (Duncan & McKeachie, 2005). Die Organisation und Integration von Lerninhalten ist eine notwendige Voraussetzung für das Verständnis und die Anwendung dieser Inhalte (Mayer, 2008). Eine geringe Selbstwirksamkeitserwartung in Statistik sollte daher zum Ende eines Kurses mit einer geringeren statistischen Kompetenz einhergehen.

H1b: Es gibt einen positiven Zusammenhang zwischen der statistischen Selbstwirksamkeitserwartung und der Leistung.

Die obigen Zusammenhänge der Selbstwirksamkeitserwartung mit Lernen und Leistung sollte von der impliziten Theorie des Lerners abhängen (Dweck & Leggett, 1988): nehmen Lerner statistische Kompetenz als veränderbar (inkrementelle bzw. flexible Theorie) oder unveränderbar (Entitäts- bzw. stabile Theorie) wahr. Aus dieser impliziten Theorie leitet sich die Zielorientierung eines Lerners ab (Dweck &

Leggett, 1988): Lerner, die von der Veränderbarkeit von Kompetenz ausgehen weisen tendenziell auch eine höhere Orientierung an Lernzielen auf (Blackwell et al., 2007). Solche Lerner sind vor allem daran interessiert, ihre Kompetenz in einem Bereich zu erweitern (Elliott & Dweck, 1998). Lerner mit dieser Einstellung reagieren auf negative Rückmeldungen über die eigene Kompetenz nicht mit einer Verringerung der Selbstwirksamkeit und brechen den Lernprozess ab, sondern erhöhen ihre Anstrengung (Dahling & Ruppel, 2016; Dimotakis et al., 2017; Tabernero & Wood, 1999). Infolgedessen ist anzunehmen, dass eine geringe statistische Selbstwirksamkeit zu Beginn des Kurses für Lerner mit inkrementeller Theorie kaum eine Rolle spielt – sie strengen sich unabhängig davon an und erreichen somit bessere Leistungen. Im Gegensatz dazu sollten sich Lerner mit stabiler Theorie weniger anstrengen und schlechtere Leistungen erzielen, welche nicht nur die zukünftige Selbstwirksamkeit beeinträchtigen, sondern auch die weitere Kompetenzentwicklung behindern.

H1c: Der Zusammenhang zwischen der Selbstwirksamkeitserwartung und dem Lernverhalten wird von impliziten Theorien der Kompetenz moderiert.

Bei einer flexibleren Theorie ist der Zusammenhang zwischen Selbstwirksamkeitserwartung und Lernverhalten schwächer, als bei einer stabilen Theorie.

H1d: Der Zusammenhang zwischen der Selbstwirksamkeitserwartung und der Leistung wird von impliziten Theorien der Kompetenz moderiert.

Bei einer flexibleren Theorie ist der Zusammenhang geringer, als bei einer stabilen Theorie.

H1e: Die Veränderung der Selbstwirksamkeit über das Semester ist von den impliziten Theorien abhängig.

Bei Lernern mit stabiler Theorie kommt es zu einer Abnahme der Selbstwirksamkeit.

Einige Untersuchungen deuten darauf hin, dass eine Überschätzung negative Auswirkungen auf das Lernverhalten und die Leistung haben kann (Moores & Chang, 2009; Sitzman & Johnson, 2012; Sticca et al., 2017). Überschätzer sollten weniger Anstrengung beim Lernen investieren, weil sie sich näher am eigenen Lernziel sehen, als ihre Kommilitonen, welche ihre Kompetenz korrekt einschätzen (Vancouver & Kendall, 2006; Sitzman & Yeo, 2013).

H2a: Es gibt einen negativen Zusammenhang zwischen Überschätzung und Lernverhalten.

H2b: Es gibt einen negativen Zusammenhang zwischen Überschätzung und Leistung.

Auch hier ist anzunehmen, dass sich eine kompensatorische Wirkung einer flexiblen Theorie der eigenen Kompetenz zeigt. Für Lerner mit einer flexiblen Theorie sollte es unabhängig von der Einschätzung der aktuellen Kompetenz wichtig sein, so viel wie möglich zu lernen. Infolgedessen sollten sie sich generell beim Lernen mehr anstrengen.

H2c: Der Zusammenhang zwischen Überschätzung und Anstrengung wird von den impliziten Theorien der Kompetenz moderiert.

Bei einer flexiblen Theorie zeigt sich ein geringerer negativer Zusammenhang als bei einer stabilen Theorie.

Eine anfängliche Überschätzung zu Beginn eines Statistikkurses sollte langfristig dazu führen, dass Lerner im Kurs zunächst negative Lernerfahrungen machen – d.h. negatives Feedback über den Stand ihrer aktuellen Kompetenz erhalten. Dieses Feedback kann zur Erhöhung der Anstrengung führen (Bi et al., 2017) aber auch dazu, dass der Lernprozess beendet wird (Sitzman & Johnson, 2012). Welche Reaktion gezeigt wird, sollte von der Interpretation der negativen Rückmeldung über die eigene Kompetenz abhängen. Nur Lerner mit einer stabilen Theorie sollten diese Information ihrem eigenen Mangel an Kompetenz zuschreiben, weitere Lernhandlungen unterlassen und so langfristig schlechtere Leistungen erzielen.

H2d: Der Zusammenhang zwischen der Überschätzung und der Leistung wird von impliziten Theorien der eigenen Kompetenz moderiert.

Bei einer flexiblen Theorie zeigt sich ein schwächerer negativer Zusammenhang zwischen Überschätzung und Leistung.

Überschätzer sollten negatives Feedback über ihre Leistung dazu nutzen, ihre Kalibrierung anzupassen (Foster et al., 2017). Untersuchungen zeigen jedoch, dass sich die Kalibrierung durch Feedback über die eigene Leistung nicht in jedem Fall verbessert (Foster et al., 2017; Hacker et al., 2000; Hadwin und Webster, 2013; Hawker et al., 2016). Eine Verbesserung der Kalibrierung und damit Verringerung der Überschätzung über den Semesterverlauf ist möglicherweise ebenfalls von den impliziten Theorien abhängig. Während Studierende mit flexibler Theorie ihre Kalibrierung immer weiter anpassen, könnten Studierende mit stabiler Theorie dazu neigen, die Überschätzung trotz negativer Rückmeldung aufrecht zu erhalten.

H2d: Die Veränderung der Kalibrierung bei Überschätzung ist von der impliziten Theorie der Lerner abhängig.

Bei inkrementeller Theorie verringert sich die Überschätzung im Verlauf des Semesters.

## C.4 Methode

### C.4.1 Stichprobe

Für die Untersuchung wurden im WiSe 2017/18, SoSe 2018 und WiSe 2018/19 insgesamt  $n = 88$  Studierende der Psychologie an der Universität Greifswald befragt. Um eine größere Stichprobe zu erhalten wurden die Befragungen in drei unterschiedlichen Lehrveranstaltungen im Fach Statistik durchgeführt. 72.7 % der Studierenden befanden sich im ersten Fachsemester, 12.5 % im dritten und 14.8 % im sechsten Fachsemester. Im Durchschnitt waren die Befragungsteilnehmer  $M = 20.67$  ( $SD = 2.83$ ) Jahre alt. 93.2 % der Teilnehmer waren weiblich.

### C.4.2 Durchführung

In jeder Lehrveranstaltung wurden die Studierenden gebeten, an einer freiwilligen Befragung zu Selbsteinschätzung im Fach Statistik teilzunehmen. Als Aufwandsentschädigung konnten die Teilnehmer Versuchspersonenstunden erhalten oder an einer Verlosung mehrerer Büchergutscheine teilnehmen. Die Befragungen wurden zum Teil als Paper-Pencil-Version und zum Teil online durchgeführt. Die Onlineversion wurde mit SoSciSurvey (Leiner, 2019) realisiert und den Studierenden unter [www.soscisurvey.de](http://www.soscisurvey.de) zur Verfügung gestellt.

Zu Beginn des Semesters wurden die Studierenden jeweils zu ihrer Selbstwirksamkeit, ihren impliziten Theorien zur Kompetenz und dem geplanten Lernaufwand für das kommende Semester befragt. Zudem wurde ein Test der aktuellen Kompetenz durchgeführt, um die Überschätzung berechnen zu können. Zum Ende des Semesters sollten die Studierenden erneut einen Fragebogen zu Selbstwirksamkeit und impliziten Theorien ausfüllen. Danach wurde die Kursleistung über einen Test gemessen.

### C.4.3 Messinstrumente

#### C.4.3.1 Leistung

Die Leistung der Studierenden wurde zu Beginn und zum Ende des Semesters jeweils mit einem Test im Multiple Choice Format erhoben. Jeder Test enthielt fünf bis neun praxisbezogene Problemlöseaufgaben, für deren Lösung die Anwendung statistischer Kompetenz (*statistical reasoning*) erforderlich ist. Dieses Aufgabenformat ist angelehnt an das *Statistical Reasoning Assessment* (SRA; Garfield, 1998). Die Aufgabeninhalte waren an den aktuellen Wissensstand der Teilnehmer angepasst. Der Test zu Beginn des Semesters (Pretest) bezog sich also auf statistische Inhalte des bisherigen Studiums, der Test zum Ende des Semesters (Nachtest) auf die Inhalte der Lehrveranstaltung. Im Gegensatz zu ihren Kommilitonen, welche bereits Vorwissen aus der bisherigen Statistikausbildung haben, wurden die Erstsemester deshalb erst fünf Wochen nach Beginn der Lehrveranstaltungen befragt. Dieses Vorgehen sollte gewährleisten, dass die Testaufgaben die aktuelle Kompetenz möglichst valide messen. Die Testergebnisse wurden in beiden Tests in Prozent (0 – 100) berechnet.

#### C.4.3.2 Aktuelle Selbstwirksamkeitserwartung

Die Selbstwirksamkeitserwartung wurde in dieser Untersuchung als Einschätzung des gegenwärtigen Status der Kompetenz im Fach Statistik erhoben. Es wurde somit die aktuelle Selbstwirksamkeitserwartung gemessen. Diese kann von der Selbstwirksamkeit bezogen auf das Erlernen von Statistik abgegrenzt werden, welche auch die Einschätzung des zukünftig erreichbaren Status an Kompetenz durch Lernen mit einbezieht (Finney & Shraw, 2003, Stone, 2000). Die aktuelle Selbstwirksamkeitserwartung wurde mit der Skala *Cognitive Competence* (CC) aus der *Survey of Attitudes towards Statistics* (SATS, Schau, Stevens, Dauphinee & Del Vecchio, 1995; deutsch: Zimprich, 2012) gemessen. In dieser Skala sollen sechs Aussagen zu den aktuellen Fähigkeiten im Fach Statistik beurteilt werden (z.B. „Ich kann Statistik-Gleichungen recht gut verstehen.“). Das Item „Ich werde Statistik sicher erlernen.“ aus der Originalskala wurde in dieser Untersuchung ausgeschlossen, da es Überzeugungen über die Veränderung von Kompetenz abbildet und keine aktuelle Einschätzung der statistischen Kompetenz. Im Gegensatz zur Originalskala, welche ein 7-stufiges Antwortformat verwendet, wurde in der Studie eine 5-stufige Ratingskala von 0 (*trifft gar nicht zu*) bis 100 (*trifft voll zu*) eingesetzt. Die interne Konsistenz liegt in der Stichprobe bei  $\alpha = .74$  und ist damit geringer als in anderen Untersuchungen (Nolan et al., 2012; Zimprich, 2012). Zum Ende des Semesters erreicht die interne Konsistenz einen Wert von  $\alpha = .81$ .

#### C.4.3.2 Überschätzung

Zur Messung der Überschätzung wurden die Teilnehmer im Test zu Beginn und zum Ende des Semesters vor jeder Problemlöseaufgabe nach ihrer Überzeugung gefragt, die folgende Aufgabe korrekt lösen zu können. Dazu wurde eine fünfstufige Skala verwendet von 0 (*Ich traue mir die Lösung der Aufgabe überhaupt nicht zu.*) bis 100 (*Ich traue mir die Lösung der Aufgabe voll zu.*). Solche spezifisch an das Kriterium angelehnten Methoden zur Einschätzung der eigenen Kompetenz sollten in der Regel besonders hoch mit dem tatsächlichen Kompetenzkriterium korrelieren (Bandura, 1986; Pajares, 1996). Dadurch sollte verhindert werden, dass das Ausmaß der gemessenen Überschätzung auf eine unpräzise Einschätzungsmethode zurückzuführen ist. Für die Berechnung der Überschätzung wurde in dieser Studie der *Bias* von Pajares und Graham (1999) verwendet. Dieser *Bias* erlaubt es, nicht nur die Abweichung der selbsteingeschätzten von der tatsächlichen Leistung sondern auch die Richtung dieser Abweichung zu quantifizieren (Boekharts & Rozendaal, 2011; Chen, 2003). Dazu wurde für jede Problemlöseaufgabe im Test eine richtige (vollständige) Antwort mit 100 bewertet und eine falsche (unvollständige) Antwort mit 0. Diese Bewertung wurde dann von der Überzeugung über die richtige Aufgabenlösung durch den Teilnehmer abgezogen. Für den *Bias* ergibt sich dadurch eine Skala von -100 bis 100, wobei Werte unter 0 für eine Unterschätzung der Kompetenz und Werte größer als 0 für eine Überschätzung der Kompetenz stehen. War ein Teilnehmer beispielsweise zu 75% überzeugt, die Aufgabe korrekt lösen zu können (selbsteingeschätzte Kompetenz = 75) und hat die Aufgabe falsch gelöst (tatsächliche Kompetenz = 0),



wird für die Aufgabe ein Wert von 75 berechnet, was für eine relativ starke Überschätzung steht. Für den *Bias* im Gesamttest wurde im Anschluss der Mittelwert gebildet.

#### C.4.3.3 Implizite Theorie

Die Skala zur Messung der Impliziten Theorien der eigenen Kompetenz wurde angelehnt an die *Implicit Theories of Intelligence Scale* (Dweck, 1999) entwickelt. Dazu wurden die Formulierungen der Items an den Fachbereich Statistik angepasst. Diese Anpassung wurde vorgenommen, weil sich implizite Theorien unterschiedlicher Kompetenzbereiche unterscheiden können (Dweck & Molden, 2017). Die allgemeine Annahme der Veränderbarkeit von Kompetenz muss daher nicht unbedingt der Auffassung über die Erlernbarkeit von Statistik entsprechen. Die Skala umfasst insgesamt sechs Aussagen zur statistischen Kompetenz, denen der Lerner auf einer Skala von 0 bis 100 mehr oder weniger zustimmen kann (z.B. „Jeder hat ein bestimmtes Ausmaß an statistischen Fähigkeiten, daran kann man selbst nicht viel ändern.“). Um die Teststärke zu erhöhen wurden die impliziten Theorien der Lerner nicht wie üblich als dichotome (Dweck et al., 1995), sondern als kontinuierliche Variable betrachtet (Lüftenegger & Chen, 2017; Robbins & Pals, 2002). Ein hoher Wert entspricht somit einer flexibleren Theorie über die statistische Kompetenz, d.h. die statistische Kompetenz wird eher als veränderbar wahrgenommen. Die interne Konsistenz der Skala liegt zum ersten Messzeitpunkt bei  $\alpha = .77$  und zum zweiten Messzeitpunkt bei  $\alpha = .83$ .

#### C.4.3.4 Anstrengung

Um einen Indikator für den geplanten Lernaufwand zu erhalten, wurden angelehnt an die Skala *Effort* aus der *Survey of Attitudes towards Statistics* (Schau et al., 1995; Schau, 2003) und die Skala *Effort Regulation* aus dem *Motivated Strategies of Learning Questionnaire* (MSLQ; Pintrich, Smith, García, & McKeachie, 1991) neun Items zum Lernaufwand in der besuchten Statistikveranstaltung formuliert (z.B. „Ich bin während der Statistikveranstaltung aufmerksam.“). Auch hier wurde das fünfstufige Antwortformat von 0 bis 100 verwendet. Die Items weisen eine zufriedenstellende interne Konsistenz auf ( $\alpha = .71$ ).

## C.5 Ergebnisse

### C.5.1 Aktuelle Selbstwirksamkeitserwartung

Die durchschnittliche Selbstwirksamkeitserwartung der Studierenden zu Semesterbeginn liegt auf einer Skala von 0 bis 100 bei  $M = 64.66$  ( $SD = 17.25$ ) (Tabelle C.1a). Demnach schätzen die Psychologiestudierenden ihre statistische Kompetenz bereits am Kursanfang insgesamt eher positiv ein. Am geringsten ist die durchschnittliche Selbstwirksamkeitserwartung dabei bei den Studierenden im ersten Semester ( $M = 62.27$ ,  $SD = 17.25$ ). Bei den Drittsemestern liegt die durchschnittliche Selbstwirksamkeitserwartung bereits höher bei  $M = 69.09$  ( $SD = 16.25$ ) und ist nochmals etwas höher im sechsten Semester ( $M = 72.69$ ,  $SD = 16.02$ ).

#### C.5.1.1 Zusammenhang der aktuellen Selbstwirksamkeitserwartung mit Lernverhalten und Leistung

Zum Test des in H1a und H1b angenommenen Zusammenhangs zwischen aktueller Selbstwirksamkeit und Anstrengung im Kurs bzw. Leistung zum Ende des Kurses wurde jeweils die bivariate Korrelation errechnet.

Zwischen der aktuellen Selbstwirksamkeit und der Anstrengung gibt es keinen Zusammenhang ( $r = .04$ ;  $t(86) = 0.33$ ,  $p = .74$ ). Damit muss H1a abgelehnt werden. Studierende, die ihre aktuelle statistische Kompetenz höher einschätzen, strengen sich damit nicht mehr an, als andere Studierende. Es zeigt sich, dass eine höhere aktuelle Selbstwirksamkeit zu Beginn des Semesters gemäß H1b mit besseren Ergebnissen im Nachtest einhergeht ( $r = .25$ ). Der Zusammenhang ist signifikant ( $t(86) = 2.43$ ,  $p = .01$ ).

Der in H1c und H1d angenommene Moderationseffekt der impliziten Theorien auf diesen Zusammenhang wurden mittels schrittweiser linearer Regressionsanalysen geprüft (Cohen et al., 2003). Im ersten Schritt wurden dazu die beiden Prädiktoren aktuelle Selbstwirksamkeit und implizite Theorie und im zweiten Schritt der Interaktionsterm zwischen beiden Prädiktoren für den Test der Moderation in das Regressionsmodell aufgenommen. Für die Regressionsanalysen wurden die Prädiktoren in zentrierter Form verwendet.

Für die Anstrengung im Kurs ergibt sich kein Interaktionseffekt zwischen der aktuellen Selbstwirksamkeitserwartung und der impliziten Theorie der Lerner ( $\Delta R^2 = 0.00$ ;  $F_{\text{inc}}(1/84) = 0.36$ ,  $p = .55$ ). Die Ergebnisse sprechen damit gegen die Annahmen in H1c.

Beim Zusammenhang mit der Leistung zeigt sich ein signifikanter Interaktionseffekt zwischen der aktuellen Selbstwirksamkeit und der impliziten Theorie der Studierenden ( $F_{\text{inc}}(1/84) = 5.53$ ,  $p = .02$ ;  $\Delta R^2 = 0.06$ ). Dies spricht für die Hypothese (H1d).

Die Richtung des Interaktionseffekts zwischen aktueller Selbstwirksamkeitserwartung wurde im Anschluss mit Hilfe des SPSS Makros PROCESS (Hayes, 2018) geprüft. Dabei wurden die bedingten Regressionsgewichte für den Zusammenhang zwischen aktueller Selbstwirksamkeit und Testleistung

für eine niedrige, mittlere und hohe Ausprägung des Moderators implizite Theorien berechnet. Da die Moderatorvariable eine linksschiefe Verteilung aufweist, wurden das 16., 50. und 84. Perzentil als Ausprägungen der impliziten Theorie verwendet (Baltes-Götz, 2018). Auch für diese Analysen wurden die Prädiktoren aktuelle Selbstwirksamkeitserwartung und implizite Theorie zentriert.

Bei einer eher flexiblen impliziten Theorie von der Veränderbarkeit statistischer Kompetenz ergibt sich kein Zusammenhang zwischen aktueller Selbstwirksamkeit und Leistung ( $b = -0.00$ ;  $t(84) = 0.04$ ,  $p = 0.96$ ). Auch bei einer mittleren Ausprägung des Moderators ist der Zusammenhang nicht signifikant von Null verschieden ( $b = 0.20$ ;  $t(84) = 1.65$ ,  $p = 0.10$ ). Bei einer impliziten Theorie, die sich eher an der Stabilität statistischer Kompetenz orientiert, zeigt sich jedoch ein signifikant positiver Zusammenhang zwischen Selbstwirksamkeit und Leistung ( $b = 0.51$ ;  $t(84) = 3.15$ ,  $p = 0.00$ ). Steigt die Selbstwirksamkeit um 10 Skaleneinheiten an, schneiden die Studierenden im Nachtest im Schnitt um etwa 5 Prozentpunkte besser ab. Nur unter Annahme der Unveränderbarkeit statistischer Kompetenz ist die Nachtestleistung also umso besser, je höher die aktuelle Selbstwirksamkeitserwartung ausgeprägt ist. Der Moderationseffekt ist in Abbildung C.2a grafisch dargestellt.

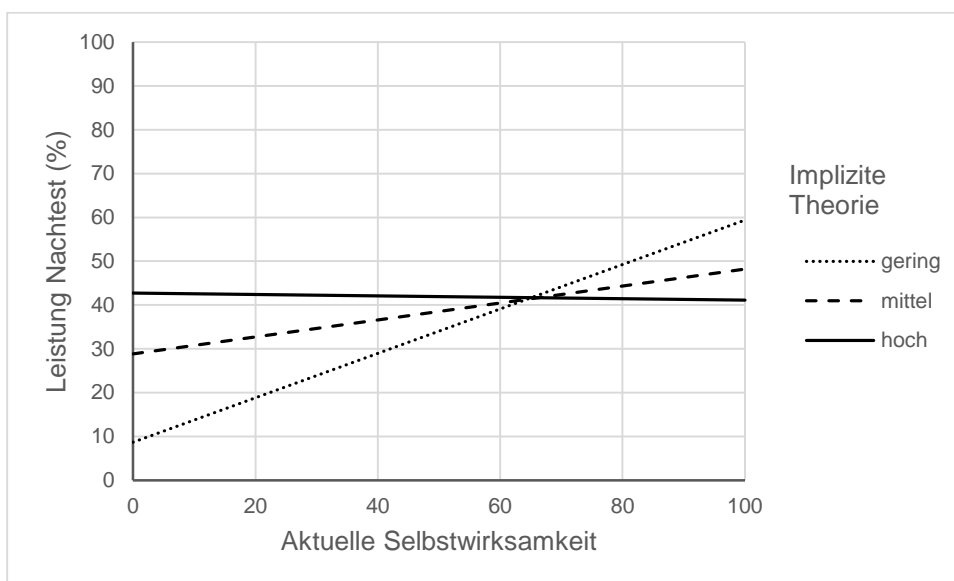


Abbildung C.2a. Moderation des Zusammenhangs zwischen aktueller statistischer Selbstwirksamkeitserwartung und Leistung durch die implizite Theorie. Eine hohe Ausprägung der Skala *Implizite Theorie* entspricht der Annahme der Veränderbarkeit statistischer Kompetenz.

### C.5.1.2 Veränderung der aktuellen Selbstwirksamkeitserwartung

Um beurteilen zu können, ob die Veränderung der aktuellen Selbstwirksamkeitserwartung im Verlauf des Semesters von der impliziten Theorie der Studierenden abhängt (H1e) wurde eine Varianzanalyse mit Messwiederholung gerechnet, in der die implizite Theorie als Kovariate eingeführt wurde.

Dabei zeigt sich zunächst, dass sich die aktuelle Selbstwirksamkeitserwartung im Semesterverlauf nur gering verändert. Die mittlere aktuelle Selbstwirksamkeitserwartung der Psychologiestudierenden liegt zu Beginn des Semesters bei  $M = 64.66$  ( $SD = 17.25$ ) und steigt bis zum Ende des Semesters nur auf einen durchschnittlichen Wert von  $M = 70.00$  ( $SD = 17.04$ ) an. Diese Zunahme ist nicht signifikant ( $F(1/86) = 0.036$ ,  $p = .55$ ;  $\eta^2 = .00$ ). Es ergibt sich auch kein signifikanter Interaktionseffekt zwischen der Veränderung in der aktuellen Selbstwirksamkeitserwartung und der impliziten Theorie ( $F(1/86) = 0.00$ ,  $p = .93$ ;  $\eta^2 = .00$ ). Der Unterschied in der aktuellen Selbstwirksamkeitserwartung zwischen dem Beginn und dem Ende des Semesters wird demnach nicht wie in H1e angenommen davon beeinflusst, ob Studierende von der Veränderbarkeit statistischer Kompetenz überzeugt sind oder nicht.

**TABELLE C.1a**

*Deskriptive Statistik und Skalenkennwerte in der Gesamtstichprobe (N = 88)*

	Semesterbeginn (T1)		Semesterende (T2)		$\alpha_1$	$\alpha_2$	n
	M	SD	M	SD			
Aktuelle Selbstwirksamkeit	64.66	17.25	70.00	17.04	.74	.81	5
Implizite Theorien	76.28	13.86	75.80	14.13	.77	.83	6
Bias <sup>1</sup>	6.36	22.57	12.97	22.14			
Bias Kalibrierung <sup>2</sup>	19.51	14.54	20.73	15.02			
Anstrengung	73.80	13.75			.71		9
Testleistung (Pretest bzw. Nachtest)	42.56	21.91	40.62	19.51			

*Anmerkungen.* n = Anzahl der Items.  $\alpha_1$  = Semesterbeginn.  $\alpha_2$  = Semesterende.

<sup>1</sup>. Die Biasvariable hat einen Wertebereich von -100 bis +100. Negative Werte stehen für eine Unterschätzung und positive Werte für eine Überschätzung. <sup>2</sup>. Hier wurde von den Werten der Biasvariablen der jeweilige Betrag gebildet, sodass diese von 0 bis 100 skaliert ist. Höhere Werte stehen für eine schlechtere Kalibrierung.

### C.5.2 Überschätzung

Insgesamt überschätzten 61.4% der Studierenden in der Gesamtstichprobe ihre aktuelle Kompetenz zu Kursbeginn. 35.2% der Studierenden neigen zur Unterschätzung (Tabelle C.1b). Die restlichen 3.4% schätzen ihre statistische Kompetenz korrekt ein. Diese Ergebnisse entsprechen auch den Ergebnissen anderer Autoren, nach denen Lerner eher dazu neigen, sich zu überschätzen (Chen, 2003; Foster et al., 2017; Pajares & Graham, 1999). Der mittlere Bias liegt bei einem Wert von  $M = 6.37$  ( $SD = 23.57$ ) (Tabelle C.1a). Berücksichtigt man, dass die Skala des Bias theoretisch Werte von - 100 (starke Unterschätzung) bis + 100 (starke Überschätzung) einschließt, handelt es sich jedoch nur um eine leichte Überschätzung.

**Tabelle C.1b***Häufigkeit von (Fehl-)Kalibrierung in der Gesamtstichprobe (N = 88)*

Semesterbeginn (T1)			Semesterende (T2)		
Unterschätzer	korrekte Kalibrierung	Überschätzer	Unterschätzer	korrekte Kalibrierung	Überschätzer
35.2%	3.4%	63.5%	20.9%	9.3%	69.8%

Um die Gruppe der Überschätzer zu untersuchen, wurden aus der Stichprobe die  $n = 54$  Studierenden ausgewählt, deren Bias größer als Null war, d.h. welche ihre Kompetenz überschätzten (vgl. Moores & Chang, 2009). Höhere Werte beim Bias entsprechen in dieser Gruppe inhaltlich einer stärkeren Überschätzung, während geringere Werte (gegen 0) für eine korrektere Einschätzung der Kompetenz sprechen.

### C.5.2.1 Zusammenhang der Überschätzung mit Lernverhalten und Leistung

In H2a und H2b sollte es einen negativen Zusammenhang der Überschätzung zu Beginn des Kurses mit der Anstrengung sowie mit der Kursleistung geben. Diese Zusammenhänge wurden ebenfalls mittels bivariater Korrelationen geprüft. Weder mit der Anstrengung ( $r = .08$ ;  $t(52) = 0.57$ ,  $p = .57$ ) noch mit der Leistung im Nachtest ( $r = .05$ ;  $t(52) = 0.13$ ,  $p = .72$ ) lässt sich ein Zusammenhang beobachten. Demnach strengen sich Studierende, die ihre statistische Kompetenz überschätzen weder mehr an noch erreichen sie schlechtere Leistungen zum Ende des Semesters. H2a und H2b müssen demnach verworfen werden.

Gemäß H2c und H2d sollte der Zusammenhang zwischen Überschätzung und Anstrengung bzw. Leistung durch die implizite Theorie moderiert werden. Diese Annahmen wurden ebenfalls mittels schrittweiser Regressionsanalysen getestet (Cohen, Cohen, West & Aiken, 2003).

Zwischen dem Ausmaß an Überschätzung und der impliziten Theorie hinsichtlich ihrer Wirkung auf die Anstrengung zeigt sich kein Interaktionseffekt ( $\Delta R^2 = 0.00$ ;  $F_{inc}(1/50) = 0.14$ ,  $p = .71$ ). Damit muss H2c ebenfalls abgelehnt werden.

Für die Leistung im Nachtest bei den Überschätzern ergibt sich ein Interaktionseffekt zwischen dem Bias und der impliziten Theorie ( $\Delta R^2 = 0.05$ ). Dieser ist in der Stichprobe jedoch nicht signifikant ( $F_{inc}(50) = 2.63$ ,  $p = .11$ ). Die Ergebnisse sprechen damit teilweise für H2d.

Da die Interaktion zwischen Überschätzung und impliziter Theorie auf die Leistung im Nachtest nach Cohen (1988) zumindest einem kleinen Effekt entspricht ( $\Delta R^2 > 0.01$ ) wurden auch hier analog zum obigen Vorgehen mit SPSS PROCESS die bedingten Regressionsgewichte für eine niedrige, mittlere und hohe Ausprägung des Moderators berechnet, um die Richtung der Moderation zu bestimmen. Für eine hohe Ausprägung, d.h. eine stark inkrementell orientierte implizite Theorie ergibt sich ein Regressionsgewicht von  $b = 0.55$  ( $t(50) = 1.64$ ,  $p = 0.10$ ). Nimmt die Überschätzung also um zehn Skalenpunkte zu, steigt die Leistung im Nachtest im Schnitt um 5.5 Prozentpunkte. Bei einer mittleren Ausprägung des Moderators reduziert sich dieser Anstieg auf weniger als 3 Prozentpunkte ( $b = 0.27$ ;

$t(50) = 1.26, p = 0.21$ ). Orientieren sich Lerner eher an einer stabilen impliziten Theorie, steigt die Leistung im Nachtest bei höherer Überschätzung nicht an ( $b = -.02; t(50) = 0.10, p = 0.93$ ). Eine höhere Überschätzung der eigenen Kompetenz zu Beginn des Kurses führt also nur dann zu einem höheren Zuwachs an Leistung im Kurs, wenn die Lerner an einer flexiblen Theorie der statistischen Kompetenz orientiert sind, d.h. die statistische Kompetenz als veränderbar ansehen. Es ist jedoch anzumerken, dass die bedingten Regressionsgewichte weder für eine geringe, mittlere noch hohe Ausprägung der impliziten Theorien signifikant von Null abweichen. Abbildung C.2b zeigt den oben beschriebenen Moderationseffekt.

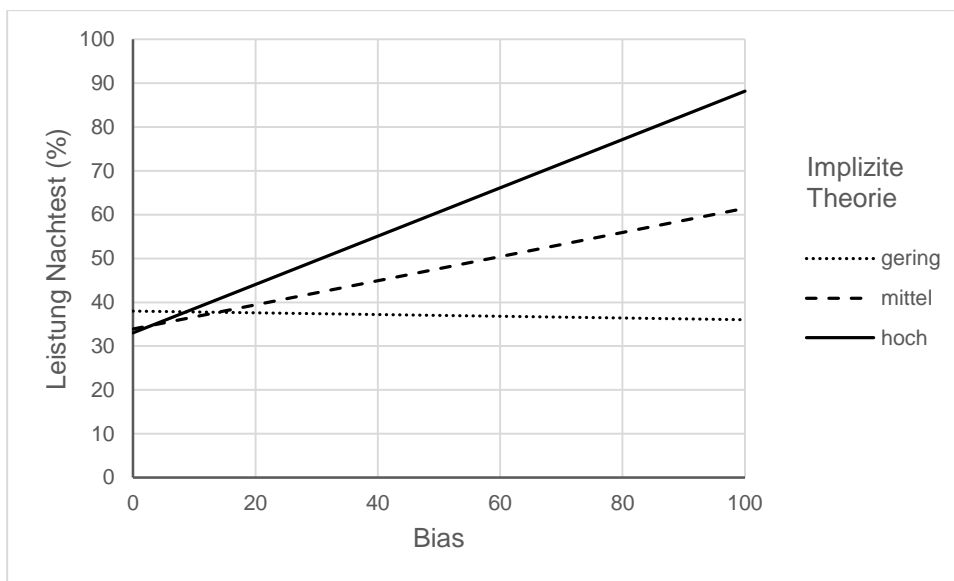


Abbildung C.2b. Moderation des Zusammenhangs zwischen dem Bias bei Überschätzern und der Leistung durch die Implizite Theorie. Eine hohe Ausprägung der Skala Implizite Theorie entspricht der Annahme der Veränderbarkeit statistischer Kompetenz.

#### C.5.2.2 Veränderung der Überschätzung

Zum Ende des Kurses überschätzt in der Gesamtstichprobe immer noch die Mehrheit der Studierenden (69.8%) ihre statistische Kompetenz. 20.9% aller Studierenden unterschätzen sich, 9.3% schätzen sich korrekt ein (Tabelle C.1b im Anhang). Der mittlere Bias in der Gesamtstichprobe steigt zum Semesterende auf  $M = 12.98$  ( $SD = 22.14$ ) (Tabelle C.1a). Da der Bias in der Gesamtstichprobe die (Fehl-)Kalibrierung auf einer Skala von -100 (starke Unterschätzung) bis +100 (starke Überschätzung) angibt, spricht dies jedoch immer noch für eine insgesamt leichte Überschätzung.

Nach H2e sollte sich die Überschätzung in Abhängigkeit von der impliziten Theorie der Lerner verändern. Um die Hypothese zu testen wurde erneut nur die Teilstichprobe der  $n = 54$  Studierenden betrachtet, welche zu Semesterbeginn zur Gruppe der Überschätzer gehörten. Für diese wurde eine Varianzanalyse mit Messwiederholung für die Veränderung des Bias gerechnet, in der die implizite Theorie als Kovariate eingeführt wurde.

Zunächst zeigt sich, dass die Fehlkalibrierung in der Gruppe der Überschätzer von  $M = 21.09$  ( $SD = 13.47$ ) zu Beginn auf  $M = 18.02$  ( $SD = 23.08$ ) zum Ende des Semesters leicht sinkt (Tabelle C.1c). Diese Reduktion der Überschätzung ist jedoch nicht signifikant ( $F(1/52) = 0.15$ ,  $p = .70$ ;  $\eta^2_{\text{part}} = .00$ ). Demnach bleibt die Überschätzung im Verlauf des Semesters relativ konstant. Dies gilt unabhängig von der impliziten Theorie der Lerner: diese hat keinen Einfluss auf die Veränderung der Überschätzung ( $F(1/52) = 0.04$ ,  $p = .83$ ;  $\eta^2_{\text{part}} = .00$ ). Entgegen der Annahme in H2e geht eine inkrementelle Theorie der Kompetenz somit nicht mit einer Reduktion der Überschätzung einher.

**Tabelle C.1c**

*Deskriptive Statistik und Skalenkennwerte in der Teilstichprobe der Überschätzer<sup>1</sup> (N = 54)*

	Semesterbeginn (T1)		Semesterende (T2)		$\alpha_1$	$\alpha_2$	$n$
	$M$	$SD$	$M$	$SD$			
Aktuelle Selbstwirksamkeit	65.57	17.58	70.28	17.25	.77	.84	5
Implizite Theorien	77.31	14.29	75.23	13.26	.78	.80	6
Bias <sup>2</sup>	21.09	13.47	18.02	23.09			
Anstrengung	73.61	13.85			.70		9
Testleistung (Pretest bzw. Nachtest)	32.92	18.27	38.83	19.08			

*Anmerkungen.*  $n$  = Anzahl der Items.  $\alpha_1$  = Semesterbeginn.  $\alpha_2$  = Semesterende.

<sup>1</sup>. Als Überschätzer wurden alle Studierenden mit einem Wert von 0 bis + 100 auf der ursprünglichen Biasvariablen von -100 bis + 100 definiert. <sup>2</sup>. Die Biasvariable hat einen Wertebereich von -100 bis +100. Negative Werte stehen für eine Unterschätzung und positive Werte für eine Überschätzung.

## **C.6 Diskussion**

In der vorliegenden Untersuchung ging es um die Frage, welche Rolle die statistische Selbstwirksamkeitserwartung zu Beginn eines Kurses für die Steuerung des Lernverhaltens und den Kompetenzerwerb in einem Statistikkurs spielt. Ist es wichtig, dass Studierende eine hohe statistische Selbstwirksamkeitserwartung haben? Oder ist es auch wichtig, dass eine hohe statistische Selbstwirksamkeitserwartung den Kompetenzstand nicht überschätzt? Oder ist die statistische Selbstwirksamkeitserwartung nur dann wichtig, wenn Studierende gleichzeitig von der Unveränderbarkeit von statistischer Kompetenz ausgehen?

Die Ergebnisse der Untersuchung lassen darauf schließen, dass sich eine hohe statistische Selbstwirksamkeitserwartung positiv auf die Kompetenzentwicklung auswirkt. Dabei ist es unerheblich, ob die hohe statistische Selbstwirksamkeitserwartung den Kompetenzstand korrekt einschätzt oder überschätzt. Eine geringe statistische Selbstwirksamkeitserwartung zu Beginn des Studiums kann eine erfolgreiche Statistikausbildung also behindern. Die Untersuchungsergebnisse zeigen aber auch, dass dies eher dann der Fall sein dürfte, wenn Studierende mit geringer statistischer Selbstwirksamkeitserwartung zu Studienbeginn gleichzeitig von der Unveränderbarkeit statistischer Kompetenz überzeugt sind. Diese Studierenden glauben nicht nur, dass sie „schlecht“ in Statistik sind, sondern auch, dass sich daran nichts ändern lässt. Wird die statistische Kompetenz von den Studierenden demgegenüber als veränderbar wahrgenommen, ist es für den statistischen Kompetenzerwerb weniger wichtig, wie hoch die statistische Selbstwirksamkeit zu Beginn des Studiums ist. In diesem Fall ist es nämlich möglich, durch Lernen besser in Statistik zu werden. Zudem lässt sich für den Fall der Überzeugung von der Veränderbarkeit der statistischen Kompetenz ein interessanter Effekt beobachten: wenn die statistische Selbstwirksamkeitserwartung den tatsächlichen Kompetenzstand überschätzt, fördert dies sogar die Kompetenzentwicklung.

### **C.6.1 Selbstwirksamkeitserwartung und Lernverhalten**

In mehreren Untersuchungen wurde bereits ein Zusammenhang zwischen einer hohen Selbstwirksamkeitserwartung und der Anstrengung beim Lernen (z.B. bei Schwierigkeiten nicht sofort aufgeben) gefunden (Liem et al., 2008; Pajares & Graham, 1999; Stegers-Jäger et al., 2012). Dieser Zusammenhang konnte in der vorliegenden Untersuchung nicht repliziert werden: die aktuelle statistische Selbstwirksamkeitserwartung korreliert nicht wie angenommen mit dem selbstberichteten Lernaufwand für den Kurs. Auch die Überschätzung hat in der Untersuchung keinen Einfluss auf den Lernaufwand. Obwohl diese Befunde nicht mit den vorab getroffenen Annahmen übereinstimmen, sind sie durchaus positiv zu interpretieren: Lerner, die nicht von ihrer statistischen Kompetenz überzeugt sind oder unrealistische Kompetenzüberzeugungen haben, können sich unabhängig von diesen Überzeugungen anstrengen.



Zusätzliche Analysen zeigen, dass für die Anstrengung beim Lernen allein die implizite Theorie von der Veränderbarkeit der statistischen Kompetenz wichtig zu sein scheint: diese zeigt sowohl in der Gesamtstichprobe ( $r = .15$ ) als auch bei der Teilstichprobe der Überschätzer ( $r = .23$ ) geringe positive Zusammenhänge zum selbstberichteten Lernaufwand (Tabelle C.2a und C.2b im Anhang). Je flexibler die statistische Kompetenz gesehen wird, desto mehr Aufwand investieren die Lerner also. Dies gilt unabhängig von der Selbstwirksamkeit – der angenommene Moderationseffekt zwischen Kompetenzüberzeugung oder Überschätzung der Kompetenz und impliziter Theorie kann nämlich nicht beobachtet werden. Demnach ist die *aktuelle Selbstwirksamkeit* für die Vorbereitungsphase beim selbstregulierten Lernen (*forethought phase*) im Modell von Zimmermann (Zimmerman et al., 2017), in der Zielsetzung und Planung der Lernhandlung erfolgen, irrelevant. Stattdessen deuten die Ergebnisse an, dass eher Aspekte der *lernbezogenen Selbstwirksamkeitserwartung* wichtig sein könnten. Die lernbezogene Selbstwirksamkeitserwartung basiert nicht auf dem aktuellen Kompetenzzustand, sondern auf der Überzeugung, einen bestimmten Zustand an statistischer Kompetenz in Zukunft durch Lernen erreichen zu können (Finney & Schraw, 2003). Diese umfasst demnach ähnliche Inhalte wie die implizite Theorie der Veränderbarkeit statistischer Kompetenz. Für zukünftige Untersuchungen zum Einfluss der Selbstwirksamkeitserwartung auf das Lernverhalten kann es demnach sinnvoll sein, zwischen der Messung der *aktuellen Selbstwirksamkeitserwartung* und der *lernbezogenen Selbstwirksamkeitserwartung* zu unterscheiden.

Bei der Interpretation der Ergebnisse zum Lernaufwand in der Untersuchung ist zu beachten, dass die Skala *Anstrengung* möglicherweise nicht ausreichend valide ist. Es ist beispielsweise denkbar, dass einige Studierende mit hoher Kompetenzüberzeugung die Statistikveranstaltung überhaupt nicht besuchen und sich den Stoff zu Hause aneignen, d.h. Lernstrategien wie Notizen machen sind für diese Studierenden nicht relevant. Beispielsweise wurde auch bei Stegers-Jäger et al. (2012) kein Zusammenhang zwischen der Selbstwirksamkeitserwartung und der Anwesenheit in Lehrveranstaltungen oder der Teilnahme an Studienaufgaben gefunden. Ein Ausschluss der entsprechenden Items zum Lernverhalten in der Statistikveranstaltung aus der Skala zeigt in der vorliegenden Untersuchung z.B. einen positiven Zusammenhang der Anstrengung von  $r = .23$  ( $t(87) = 2.23$ ,  $p = .03$ ) zur aktuellen Selbstwirksamkeit. Die Überschätzung zeigt auch für die reduzierte Skala einen Nullzusammenhang von  $r = -.05$  ( $t(53) = 0.37$ ,  $p = .71$ ). Demnach wirkt sich eine unrealistische Kompetenzüberzeugung offenbar nicht negativ auf das Lernverhalten aus. Zudem muss angemerkt werden, dass die Skala *Anstrengung* auch keinen Zusammenhang zur Leistung zum Ende des Semesters aufweist (Tabelle C.2a und C.2b im Anhang). Dies weist ebenfalls darauf hin, dass die hier gemessene Anstrengung zu Beginn des Semesters kein valides Kriterium für die Vorhersage der Kompetenzentwicklung ist. Dies könnte unter anderem daran liegen, dass der Lernaufwand im Verlauf des Semesters aufgrund der Erfahrungen beim Lernen (z.B. ob man den Stoff verstanden hat oder nicht) angepasst wird. Zuletzt muss berücksichtigt werden, dass die Anstrengung durch die Rahmenbedingungen des Studiums vielleicht auch zu sehr determiniert ist, um Varianz in der Kompetenzentwicklung aufzuklären. Die Psychologiestudierenden stehen unter

starkem Druck, einen guten Abschluss zu machen um zum Masterstudium zugelassen zu werden. Infolgedessen könnte der Lernaufwand bei diesen Studierenden generell sehr hoch sein. Tatsächlich zeigt auch die Skala Anstrengung in der Gesamtstichprobe einen eher hohen Mittelwert ( $M = 73.80$ ) und im Verhältnis dazu geringe Varianz ( $SD = 13.75$ ) (Tabelle C.1a).

### **C.6.2 Selbstwirksamkeitserwartung und Leistung**

Analog zu den meisten bisherigen Befunden zur Selbstwirksamkeitserwartung im akademischen Bereich (Richardson et al., 2012; Robbins et al., 2004) zeigt sich in der Untersuchung, dass eine höhere aktuelle Selbstwirksamkeitserwartung in Statistik insgesamt mit besseren Leistungen einhergeht. Die eher kleine Korrelation von  $r = .25$  steht dabei im Einklang mit bisherigen Ergebnissen (Nolan et al., 2012), in denen die Messung der Selbstwirksamkeit zu Beginn des Semesters die Leistungen zum Ende weniger gut vorhersagen konnte, als die Messung der Selbstwirksamkeit zum Ende des Semesters. Dies könnte darauf zurückzuführen sein, dass sich die Einschätzung der eigenen Kompetenz im Verlauf der Statistikausbildung verändert. In der Untersuchung zeigt sich beispielweise über das Semester hinweg eine leichte Zunahme der aktuellen Selbstwirksamkeitserwartung von  $M = 64.66$  ( $SD = 17.25$ ) auf  $M = 70.00$  ( $SD = 17.04$ ), obwohl die aktuelle Selbstwirksamkeitserwartung bereits zu Beginn relativ hoch ist. Insbesondere Studierende zu Beginn der Statistikausbildung haben nur wenig Erfahrung mit Statistik und schätzen ihre Kompetenz zunächst möglicherweise eher gering ein. Mit zunehmender Erfahrung mit Statistik im Verlauf des Studiums kommt es dann nicht nur zum Kompetenzzuwachs, sondern auch zu einer Korrektur einer übermäßig negativen Einschätzung der eigenen Kompetenz. Tatsächlich kann in der Untersuchung auch im Querschnitt eine leichte Zunahme der aktuellen Selbstwirksamkeitserwartung vom Beginn des ersten ( $M = 62.27$ ,  $SD = 17.25$ ) bis zum Beginn des sechsten ( $M = 72.69$ ,  $SD = 16.02$ ) Semesters beobachtet werden. Eine solch höhere Selbstwirksamkeitserwartung von Statistikkursen für Anfänger im Vergleich zu Statistikkursen für Fortgeschrittene konnte auch bei Walker und Brakke (2017) beobachtet werden. Jedoch ist zu berücksichtigen, dass die aktuelle Selbstwirksamkeit in der vorliegenden Untersuchung sowohl im Längs- als auch im Querschnitt nur gering ansteigt und bereits von Anfang an relativ hoch ist.

Interessanterweise zeigt sich keine Veränderung der Überschätzung: diese bleibt in der Teilstichprobe der Überschätzer weitgehend konstant, während sie in der Gesamtstichprobe im Semesterverlauf sogar leicht zunimmt. Obwohl die Selbsteinschätzung der statistischen Kompetenz mit der Zeit also etwas zunimmt, verbessert sich die Präzision der Selbsteinschätzung nicht. Eine mögliche Erklärung hierfür könnte sein, dass die Studierenden im Verlauf des Semester nicht ausreichend systematische Rückmeldungen über ihren Kompetenzzustand erhalten bzw. die Rückmeldungen nicht ausreichend zur Korrektur ihrer Kompetenzüberzeugungen genutzt haben. In allen untersuchten Statistikkursen wurden zwar während des Semesters verschiedene Übungen durchgeführt, Rückmeldungen über die Leistung in den Übungen erfolgten aber zum Großteil selbstgeneriert (z.B. müssen die Studierenden ihre Lösungen selbst mit einem vorgegebenen Lösungsblatt vergleichen). Es

ist denkbar, dass dieses selbstgenerierte Feedback nur dann zu einer verbesserten Übereinstimmung zwischen selbsteingeschätztem und tatsächlichem Kompetenzzustand führt, wenn die Lerner diese Rückmeldungen auch nutzen können. In einer Untersuchung von Hacker et al. (2000) zeigte sich beispielsweise, dass nur Lerner mit hoher Kompetenz ihre Kalibrierung durch Rückmeldungen verbessern konnten.

Andererseits ist es auch möglich, dass Studierende ihre überhöhten Kompetenzeinschätzungen deswegen nicht anpassen, weil sie keine negativen Auswirkungen auf die Kompetenzentwicklung haben. In der vorliegenden Untersuchung konnte beispielsweise kein negativer Zusammenhang zwischen der Überschätzung zu Beginn des Semesters und der Leistung zum Ende des Semesters gefunden werden. Es ist zudem durchaus möglich, dass eine leicht überhöhte Selbsteinschätzung in einem bestimmten Rahmen sogar positive Effekte haben könnte. Beispielsweise zeigen die Moderationsanalysen, dass eine Überschätzung bei einer flexiblen Theorie der statistischen Fähigkeiten sogar einen positiven Zusammenhang mit der Leistung hat. Dies könnte unter anderem darauf zurückzuführen sein, dass eine nur leichte Überschätzung zusätzlich eine motivationale Wirkung entfaltet. In der untersuchten Stichprobe von Psychologiestudierenden zeigt sich tatsächlich, dass das Ausmaß der Überschätzung sowohl in der Gesamtstichprobe ( $M = 6.36$  bei einem Range von -100 bis +100, wobei positive Werte für eine Überschätzung stehen) als auch in der ausgewählten Teilstichprobe der Überschätzer ( $M = 21.09$  bei einem Range von 0 bis +100, wobei positive Werte für eine Überschätzung stehen) eher gering ausgeprägt ist. Die geringe Überschätzung könnte darauf zurückzuführen sein, dass nicht nur Kompetenzüberzeugungen, sondern auch die tatsächliche statistische Kompetenz bei den Psychologiestudierenden überdurchschnittlich ausgeprägt ist. In diesem Fall sind einerseits die Selbstüberschätzungen der Kompetenz eventuell schwächer ausgeprägt (Buckelew et al., 2013). Andererseits werden Überschätzungen auch aufgrund der verwendeten Messmethode kleiner. In der vorliegenden Untersuchung wurde die Überschätzung als Differenz zwischen der Kompetenzüberzeugung und dem Kompetenzzustand gemessen (Pajares & Graham, 1999). Diese Differenz zwischen selbsteingeschätztem und tatsächlichem Kompetenzzustand wird umso geringer, je höher jemand auf der Skala der tatsächlichen Kompetenz liegt.

### **C.6.3 Implizite Theorien**

Nach dem Sozialkognitiven Ansatz der Motivation (Dweck & Leggett, 1988) sollte eine implizite Theorie über die Stabilität von Kompetenz für den statistischen Kompetenzerwerb hinderlich sein. Nehmen Lerner an, dass die statistische Kompetenz stabil ist, d.h. nicht verändert werden kann, sollte eine geringe Überzeugung von der statistischen Kompetenz sowie eine starke Überschätzung der eigenen Kompetenz am Semesteranfang langfristig schlechte Leistungen begünstigen. Bei Annahme der Flexibilität der statistischen Kompetenz sollten diese Selbsteinschätzungen hingegen keine Rolle spielen. Die Ergebnisse deuten tatsächlich darauf hin, dass die Annahme von der Veränderbarkeit von

statistischer Kompetenz im Gegensatz zur Annahme der Unveränderbarkeit den Kompetenzerwerb fördert.

Für den Zusammenhang zwischen der aktuellen Selbstwirksamkeit und der Leistung lässt sich dabei genau der angenommene Moderationseffekt beobachten: bei einer impliziten Theorie, die stark an der Stabilität der Kompetenz orientiert ist, geht eine hohe aktuelle Selbstwirksamkeit mit besseren Leistungen einher. Demgegenüber zeigt sich bei einer eher an der Flexibilität orientierten impliziten Theorie kein Zusammenhang zwischen aktueller Selbstwirksamkeit und Leistung. Demnach scheint die Überzeugung, dass die statistische Kompetenz veränderbar ist, die negativen Auswirkungen einer geringen Kompetenzüberzeugung zu Beginn des Semesters kompensieren zu können.

Im Fall einer Überschätzung zeigt sich ebenfalls die positive Wirkung der Überzeugung von der Veränderbarkeit von Kompetenz: während die Überschätzung bei einer Theorie der Stabilität keinen Zusammenhang zur Leistung hat, zeigt sich bei einer Theorie der Flexibilität sogar ein positiver Zusammenhang. Demnach wird unter letzterer Annahme die Kompetenzentwicklung sogar begünstigt, je stärker Lerner sich überschätzen. Obwohl diese Ergebnisse nicht dem angenommenen Moderationseffekt entsprechen, wonach sich bei einer impliziten Theorie der Stabilität sogar ein negativer Zusammenhang zur Leistung zeigen sollte, wird demnach auch hier deutlich, dass Lerner davon überzeugt werden müssen, dass Statistik erlernbar ist.

Insgesamt ist zu den Moderationsanalysen anzumerken, dass es sich wie bei allen Ergebnissen in der Untersuchung um sehr kleine Effekte handelt. Es ist denkbar, dass diese ebenfalls auf Besonderheiten der untersuchten Stichprobe zurückgehen: aufgrund der Zulassungsbeschränkungen zum Psychologiestudium werden vor allem Studierende ausgewählt, die bereits in der Schule erfolgreich waren, d.h. sehr gute kognitive und motivationale Lernvoraussetzungen haben. Bei solchen Studierenden ist die implizite Theorie möglicherweise insgesamt stärker an der Flexibilität von Kompetenz orientiert. Dafür spricht auch, dass der Wert auf der Skala *Implizite Theorie* in der Gesamtstichprobe im Mittel mit  $M = 76.28$  relativ hoch liegt und eine geringe Varianz aufweist ( $SD = 13.86$ ). Die Werte in der Stichprobe der Überschätzer sind relativ ähnlich ( $M = 77.31$ ,  $SD = 14.19$ ). Demnach wird in der Stichprobe das gesamte Kontinuum an Überzeugungen von der Veränderbarkeit der Kompetenz nur unzureichend abgebildet: während flexible Annahmen der Kompetenz übermäßig vertreten sind, gibt es kaum Studierende mit einer stabilen Annahme der Kompetenz. Demnach ist die Variabilität in den Ausprägungen des Moderators eher gering, was die kleinen Effekte erklären könnte.

Die bereits sehr günstigen Voraussetzungen der Studierenden könnten auch eine Erklärung dafür sein, dass die implizite Theorie im Verlauf des Semesters stabil bleibt. Nachfolgenden Analysen zeigen, dass die Studierenden in der Gesamtstichprobe sowohl zu Beginn ( $M = 76.29$ ,  $SD = 13.86$ ) als auch zum Ende des Semesters ( $M = 75.80$ ,  $SD = 14.14$ ) im gleichem Ausmaß von der Veränderbarkeit von Kompetenz überzeugt sind ( $t(87) = 0.40$ ,  $p = .691$ ). Dabei könnte es sich einerseits um einen methodischen „Deckeneffekt“ handeln: ein Großteil der Studierenden hat bereits die Überzeugung, dass statistische Kompetenz veränderbar ist, sodass eine weitere Verbesserung schwer möglich ist.

Andererseits ist es auch denkbar, dass es sich bei der impliziten Theorie – z.B. im Vergleich zur aktuellen Selbstwirksamkeit – um ein weniger änderungssensitives Konstrukt handelt. Möglicherweise müssen Studierende über einen längeren Zeitraum immer wieder die Erfahrung machen, dass sie statistische Inhalte verstehen, erlernen und anwenden können. Erst dann verändert sich ihre Überzeugung, dass statistische Kompetenz nicht von Anfang an „gegeben“ ist. Dazu müssen in Statistikkursen auch Lerngelegenheiten bereitgestellt werden, die solche Kompetenzerfahrungen ermöglichen (z.B. Lerngelegenheiten, die an das aktuelle Kompetenzniveau der Studierenden angepasst sind und Rückmeldung über den Kompetenzzuwachs ermöglichen). Es ist unklar, ob in den Kursen der vorliegenden Studie ausreichend solcher Lerngelegenheiten gegeben waren.

#### **C.6.4 Praktische Implikationen**

Zusammenfassend lässt sich ableiten, dass es für eine erfolgreiche Kompetenzentwicklung vor allem von Vorteil ist, wenn Studierende von der Veränderbarkeit statistischer Kompetenz überzeugt sind. Diese Überzeugung kann nicht nur eine anfänglich geringe Selbstwirksamkeitserwartung in Statistik in Bezug auf die spätere Leistung kompensieren, sondern selbst bei einer Überschätzung zu besseren Leistungen führen. Maßnahmen zur Unterstützung der statistischen Kompetenzentwicklung aus motivationaler Sicht sollten demnach darauf fokussieren, bei den Studierenden eine implizite Theorie von der Flexibilität bzw. Veränderbarkeit statistischer Kompetenz zu fördern. Solche positiven Kompetenzüberzeugungen werden vor allem dann gefördert, wenn sich die Studierenden bei der Kompetenzanwendung als erfolgreich und somit ihre Kompetenz als veränderbar erleben (z.B. wenn sie merken, dass sie die Inhalte verstehen oder dass sie die Übungsaufgaben lösen können).

Besonders zielführend könnte es dabei sein, diese Maßnahmen in der Statistikausbildung so früh wie möglich zu implementieren. Zu Beginn des Studiums starten die Studierenden möglicherweise vor allem mit vom mathematischen Schulunterricht geprägten Überzeugungen in die Statistikausbildung. Je stärker sich der universitäre Statistikkunterricht davon unterscheidet (z.B. durch weniger mathematische Formeln und mehr praktische Anwendung), desto eher werden die Studierenden diesen als „neuen“ Bereich wahrnehmen, in dem sie neue positive Kompetenzüberzeugungen ausbilden.

# D. UNTERSUCHUNG II – DIE FÖRDERUNG DER ENTWICKLUNG STATISTISCHER KOMPETENZ DURCH INDIVIDUALISIERTES LERNEN IM INVERTED CLASSROOM

## D.1 Einleitung

Traditionelle Lehrformate wie Vorlesungen eignen sich aufgrund der Beschränkung in Zeit und Ressourcen oft nicht dazu, Lernprozesse an die unterschiedlichen Voraussetzungen der Teilnehmer anzupassen. Eine mögliche Lösung dieses Problems bildet die Kombination von E-Learning und Präsenzlehre (Blended Learning) (Graham, 2006). Dadurch können Lernzeit und -geschwindigkeit vom Lerner selbst gesteuert werden, beispielsweise indem mit Onlinelernmaterialien zu Hause gelernt wird (Boelens et al., 2017; McKenzie et al., 2013; Spanjers et al., 2015). Diese Onlineinhalte erlauben nicht nur die Bereitstellung einer großen Auswahl an Materialien für unterschiedliche Lerner, sondern auch die Nutzung technologischer Möglichkeiten wie Simulationen, welche das Lernen erleichtern können (Berney & Bétrancourt, 2016; Liu et al., 2010).

Die am häufigsten verwendete und effektivste Form der Einbindung von E-Learning in die Präsenzlehre ist der Inverted bzw. Flipped Classroom (Margulieux, McCracken & Catrambone, 2016; Thai, DeWever & Valcke, 2017). Das Hauptkennzeichen von Inverted Classroom (IC) Modellen ist die Auslagerung der traditionell in Vorlesungen stattfindenden Wissensvermittlung aus der Lehrveranstaltung, sodass die Veranstaltungszeit für die Übung und Anwendung des Gelernten genutzt werden kann (Abeysekera & Dawson, 2015; DeLozier & Rhodes, 2017; Jensen et al., 2015; Kim et al., 2014; O’Flaherty & Phillips, 2016). Stattdessen bereiten sich die Studierenden mit Hilfe von Onlinematerialien, oft in Form von Vorlesungsaufzeichnungen oder kurzen Lehrvideos, zu Hause auf den nächsten Veranstaltungstermin vor (Chen et al., 2016; O’Flaherty & Phillips, 2016; Peterson, 2016; Roehling et al., 2017). So können sich die Lerner beim IC nicht nur je nach Leistungsfähigkeit und Vorwissen individuell mit Hilfe online bereitgestellter Lernmaterialien vorbereiten (Blau & Shamir-Inbal, 2017; Kim et al., 2014). Auch die Lernaktivitäten in der Veranstaltung können in Abhängigkeit vom aktuellen Wissensstand unterstützt werden (Brewer & Movahedazarhouli, 2018; Kim et al., 2014). Beispielsweise kann durch Gruppenarbeiten, Diskussionen und Erklärungen durch einen Tutor während der Übung und Anwendung das Verständnis komplexen Materials gefördert werden (Fung et al., 2021).

Bisherige Befunde deuten darauf hin, dass Studierende im IC tatsächlich bessere Leistungen erreichen (Chen et al., 2017; O’Flaherty & Phillips, 2016; Zheng et al., 2020) und mehr Motivation (z.B. Interesse am Fach, höhere Selbstwirksamkeit) entwickeln (Liebert et al., 2016; Thai et al., 2017; Zheng et al., 2020). Unklar ist allerdings, ob der IC tatsächlich eine individuelle Anpassung an unterschiedliche Voraussetzungen von Lernern und so bessere Lernergebnisse für alle Studierenden ermöglicht. Bisher finden sich zum einen Untersuchungen, in denen leistungsschwächere Studierende den IC als positiver bewerten (Nouri, 2016; Roehling et al., 2017). In anderen Studien waren es Studierende mit hoher Motivation, die den IC als

vorteilhaft wahrnahmen, und Studierende mit besserer Leistung, die besonders von diesem Lehrformat profitierten (Maciejewski, 2016; Wenliang et al., 2016).

Lehrformate wie der IC könnten besonders für Statistikgrundlagenkurse besser geeignet sein, als traditionelle Vorlesungen (Chili-Turner, 2015; Haughton & Kelly, 2015; Wilson, 2013). Studierende sind zu Beginn ihrer Statistikausbildung teilweise nur wenig von ihrer Kompetenz in Statistik überzeugt (Dykeman, 2011; Waples, 2016; Walker & Brakke, 2017): für sie ist Statistik schwer zu verstehen und sie fühlen sich in Statistikkursen schnell überfordert. Zudem ist das Interesse der Studierenden an Statistik meist eher gering, sodass sie auch wenig motiviert sind, sich mit dem schwierigen Stoff auseinanderzusetzen (Connors et al., 1998; Sciotto, 1995; Vittengl et al., 2004). Werden diese Studierenden im Kurs dann zusätzlich noch mit zu hohen Anforderungen konfrontiert (z.B. wenn in einer Vorlesung sehr viel komplexer Stoff in einer kurzen Zeit präsentiert wird), werden sie überlastet und verlieren vollständig die Motivation. Eine Anpassung des Kurses an die Voraussetzungen der Lerner, wie sie im IC stattfindet, könnte anstatt einer Überforderung ein positives Erleben der eigenen Kompetenz beim Lernen im Kurs fördern und damit die Kompetenzüberzeugungs-, Interessens- und Leistungsentwicklung unterstützen (Bandura, 1994; Deci & Ryan, 1985; Sweller et al., 1998). In Statistikkursen sind die Voraussetzungen der teilnehmenden Studierenden dabei oft sehr heterogen (Chili-Turner, 2015; Griffith et al., 2012; Mutz & Daniel, 2011), da solche Kurse in vielen unterschiedlichen Fächern wie Wirtschafts-, Sozial- oder Gesundheitswissenschaften Teil der Ausbildung sind (Haughton & Kelly, 2015; McLaughlin & Kang, 2017; Onwuegbuzie & Wilson, 2003). Für Statistikgrundlagenkurse ist deshalb besonders von Vorteil, dass das Lehrformat IC auch eine individuelle Anpassung an das Niveau unterschiedlicher Lerner ermöglicht (z.B. sowohl Studierende mit hohem Vorwissen, als auch Studierende mit geringem Vorwissen).

In der vorliegenden Untersuchung geht es daher um die Frage, ob ein individualisiertes Lehrformat wie der IC im Gegensatz zu einem traditionellen Vorlesungsformat besser für Statistikgrundlagenkurse für stark heterogene Studierendengruppen geeignet ist. Dazu wird einerseits untersucht, ob Lerner im IC durch die Individualisierung im Kurs tatsächlich weniger überlastet und motivierter sind und dies als Folge zum Ende des Kurses mit besseren Lernergebnissen einhergeht. Andererseits interessiert auch, ob die Anpassung an einzelne Lerner dazu führt, dass die Lernergebnisse des Kurses im IC für alle Studierenden gleich sind.

## **D.2 Theoretischer Hintergrund**

### **D.2.1 Wissenserwerb durch Individualisierung im IC**

Statistikkurse, die nach der Struktur des IC unterrichtet werden, haben sich bezüglich der Leistung zum Ende des Kurses bisher häufig als effektiver erwiesen, als traditionelle Kurse (Burgoyne & Eaton, 2018; Chili-Turner, 2015; Nielsen et al., 2018; Peterson, 2016; Wilson, 2013). Auch mehrere Monate nach Abschluss des Kurses erzielen die Teilnehmer eines IC bessere Leistungen (Winquist & Carlson, 2014). Bisher gibt es allerdings wenige Untersuchungen, warum Studierende im IC besser abschneiden (Jensen et al., 2015; Nishigawa et al.,

2017). Theoretisch werden durch die Individualisierung im IC sowohl aus kognitions- als auch aus motivationspsychologischer Sicht Lernprozesse unterstützt, was darauffolgend mit besseren Lernergebnissen einhergeht (Abeysekera & Dawson, 2015; Fung et al., 2021): Einerseits sollte durch die selbstbestimmte Lerngeschwindigkeit und die Anpassung des Lernmaterials an unterschiedliches Vorwissen der Lerner einer kognitiven Überlastung durch komplexe Lerninhalte vorgebeugt werden. Die geringere kognitive Auslastung kann vor allem den Erwerb von konzeptuellem, anwendbarem Statistikwissen erleichtern (Mayer, 2008). Andererseits sollte das Erleben von Autonomie, Kompetenz und der sozialen Einbindung bei der gemeinsamen aktiven Beschäftigung mit den Lerninhalten auch die Motivation fördern. Insbesondere für die Entwicklung der Motivation in Form positiver Kompetenzüberzeugungen ist es günstig, wenn Lerner sich selbst mit dem Stoff auseinandersetzen und dabei erleben, dass sie kompetent genug sind, um sich die Lerninhalte anzueignen – anstatt zu erleben, dass die Lerninhalte sie überfordern (Bandura, 1994; Bong & Skaalvik, 2003; Waples, 2016). Demnach wird auch ein positives (Erfolgs-)Erleben beim Lernen und infolgedessen die Motivation in Statistik gefördert, wenn die Lernaktivitäten an den Voraussetzungen der Lerner orientiert sind. Motivationen wie Zielorientierungen, Interesse oder Kompetenzüberzeugungen sind wiederum entscheidend dafür, dass die statistischen Inhalte angewendet werden können (Pugh & Bergin, 2006).

#### D.2.2 Förderung des Wissenserwerbs durch Individualisierung aus kognitiver Sicht

Das Ziel von Statistikkursen ist es im Regelfall, dass Lerner ihr Wissen auf statistische Problemstellungen in Studium, Beruf und Alltag anwenden können (Gal et al., 1997), also beispielsweise für die eigene Forschungsarbeit das korrekte statistische Verfahren auswählen, durchführen und interpretieren können.

Um statistisches Wissen erfolgreich auf neue Situationen transferieren zu können, ist es notwendig, dass das statistische Wissen in kognitive Schemata integriert und automatisiert wurde (Broers, 2001; Paas, 1992). Ein Schema kann dabei als eine übergeordnete kognitive Struktur verstanden werden, in der mehrere Wissenseinheiten miteinander verknüpft sind (Broers, 2001; 2002). Erst durch diese Verknüpfung von Wissenseinheiten kann es zum Verständnis eines statistischen Konzepts kommen (Broers, 2002). Wenn Lerner also zum Ende eines Statistikkurses isoliertes Wissen z.B. die Definition der Standardabweichung auswendig abrufen oder diese mit Hilfe einer Formel berechnen können, bedeutet dies nicht, dass sie das Konzept verstanden – d.h. ein Schema des Konzepts aufgebaut – haben (Clark et al., 2007). Infolgedessen hätten diese Lerner auch Schwierigkeiten ihr Wissen über die Standardabweichung auf eine praktische Problemstellung anzuwenden z.B. um die inhaltliche Bedeutung unterschiedlicher Standardabweichungen in zwei Gruppen zu beurteilen. Daher ist es wichtig, dass die Präsentation der Wissensinhalte und die Lernaktivitäten in einem Statistikkurs so gestaltet sind, dass sie *meaningful learning* fördern (Mayer, 2008).

*Meaningful learning* heißt, dass neue Wissensinhalte im Arbeitsgedächtnis gemeinsam mit Inhalten aus dem Langzeitgedächtnis in eine kohärente kognitive Struktur integriert werden. Dadurch



wird schematisches Wissen erworben, welches zur Lösung statistischer Problemstellungen eingesetzt werden kann. Ob es beim Lernen in Statistikkursen zum *meaningful learning* kommt oder ob die statistischen Inhalte später nur auswendig abgerufen werden können, ist von der kognitiven Belastung (*cognitive load*) des Arbeitsgedächtnisses beim Lernen abhängig. Die Cognitive Load Theorie (Paas et al., 2010; Sweller et al., 1998) geht davon aus, dass das Arbeitsgedächtnis zeitlich und in seiner Kapazität begrenzt ist – d.h. es kann immer nur eine bestimmte Anzahl an Informationselementen gleichzeitig verarbeitet werden. *Meaningful learning* tritt nur dann auf, wenn genug freie Kapazität für die Verarbeitung (d.h. Selektion, Organisation und Integration) lernrelevanter Information („germane cognitive load“) genutzt werden kann (Mayer, 2008). Je mehr Arbeitsgedächtniskapazität also von Faktoren außerhalb des „germane cognitive load“ beansprucht wird, desto schwieriger wird die Informationsverarbeitung. In diesem Fall können statistische Konzepte nur unvollständig oder fehlerhaft erlernt bzw. auf ein statistisches Problem angewendet werden. Die kognitive Belastung im Arbeitsgedächtnis erhöht sich beispielsweise dann, wenn Lernmaterial schlecht gestaltet ist („extraneous cognitive load“). Zudem ist umso mehr Arbeitsgedächtniskapazität notwendig, je komplexer das Lernmaterial ist („intrinsic cognitive load“). Komplexe Lernaufgaben enthalten viele Informationselemente, die alle gleichzeitig im Arbeitsgedächtnis strukturiert und integriert werden müssen. Einfache Lernaufgaben erfordern nur die Verarbeitung eines Informationselements bzw. die serielle Verarbeitung von Information (Sweller et al., 1998). Beim Lernen in Statistikkursen ist es häufig notwendig, eine große Menge an komplexer Information gleichzeitig zu verarbeiten. Insofern ist es nicht verwunderlich, dass die Kapazität des Arbeitsgedächtnisses schnell an die Grenzen gerät und das *meaningful learning* behindert wird. Dies gilt vor allem für Lerner mit geringem Vorwissen im Bereich Statistik, welche die statistischen Konzepte komplett neu erlernen müssen. Bei Lernern mit hohem Vorwissen wird bei der Verarbeitung weniger Arbeitsgedächtniskapazität durch komplexes Lernmaterial („intrinsic load“) beansprucht, da die Information in bereits vorhandene Strukturen (Schemata) im Langzeitgedächtnis eingeordnet werden kann (Sweller et al., 1998).

Lerner in Statistikkursen haben häufig eine unterschiedliche schulische oder universitäre Ausbildung erfahren, sodass sich der Stand des Vorwissens zu Beginn des Kurses stark unterscheidet. Manchen Studierenden fällt es zudem leicht, statistische Konzepte bereits in der Veranstaltung zu begreifen, während andere Lernzeit über die Veranstaltung hinaus benötigen (Haughton & Kelly, 2015). Um eine kognitive Überlastung zu verhindern, sodass es zum *meaningful learning* kommt, müssen diese Unterschiede in den kognitiven Voraussetzungen der Studierenden berücksichtigt werden. Andernfalls wird der Abstand zwischen dem Vorwissen der Lerner während des Kurses noch vergrößert, was insbesondere im weiteren Verlauf der Statistikausbildung (z.B. in weiterführenden Statistikkursen) zum Problem werden kann (Haughton & Kelly, 2015): während Lerner mit hohem Vorwissen die statistischen Konzepte erfolgreich erlernen und anwenden können, bleibt es bei Lernern mit geringem Vorwissen nur bei oberflächlichem Wissen. Um dies zu verhindern, können unterschiedliche Methoden eingesetzt werden, welche Lerner mit geringem Vorwissen unterstützen. Beispielsweise kann das Lernmaterial

sequentiell dargeboten werden oder bei komplexen Lernaufgaben das notwendige Vorwissen systematisch aufgebaut oder ergänzt werden (Mayer, 2008). Das Erlernen der Interpretation von statistischen Signifikanztests fällt beispielsweise leichter, wenn als Vorwissen zunächst die notwendigen Konzepte erlernt werden und diese dann erst für die schrittweise Ableitung der Ergebnisse verwendet werden (Hong & O'Neill, 1992).

Die Anpassung von Kursen an das Vorwissen der Lerner kann die kognitive Belastung nicht nur direkt, sondern auch indirekt über die Reduktion der Statistikangst verringern. Neuere Ansätze der Cognitive Load Theorie gehen davon aus, dass auch Emotionen Einfluss auf die kognitive Belastung haben (Moreno & Mayer, 2007; Plass & Kaplan, 2016). Beispielsweise fassen Plass und Kaluyga (2019) Angst als eine Form der extrinsischen kognitiven Belastung auf, welche ebenfalls Kapazität im Arbeitsgedächtnis beansprucht. Untersuchungen aus dem Bereich Mathematikangst deuten beispielsweise darauf hin, dass die negative Korrelation zwischen Mathematikangst und Mathematikleistung unter anderem darauf zurückzuführen ist, dass die sorgenvollen Gedanken kognitive Kapazität im Arbeitsgedächtnis beanspruchen (Shi & Liu, 2016; Ashcraft & Kirk, 2001; Skagerlund et al., 2019). Bei Lernern mit hoher Statistikangst wird das *meaningful learning* also zusätzlich erschwert. Angst vor Statistik entsteht dann, wenn die Lerner nicht davon überzeugt sind, die Anforderungen im Fach bewältigen zu können (Condrón et al., 2018; Dykeman, 2011; González et al., 2016; Waples, 2016). Werden die Anforderungen einer Lernaufgabe nicht an das Vorwissen der Lerner angepasst, erhöht dies also auch die Statistikangst und behindert dadurch den Wissenserwerb. Zudem wird weniger mentaler Aufwand in eine Aufgabe investiert, wenn ein Lerner nicht der Überzeugung ist, diese erfolgreich lösen zu können (Feldon et al., 2019).

### D.2.3 Förderung des Wissenserwerbs durch Individualisierung aus motivationaler Sicht

Mehrere empirische Befunde deuten darauf hin, dass eine Passung zwischen Lernanforderungen und Fähigkeiten des Lerners positive lernrelevante Emotionen auslöst, während eine Über- oder Unterforderung zu negativem Affekt wie Langeweile führt. Beispielsweise langweilen sich begabtere Schüler in normalen Klassen, weil sie sich unterfordert fühlen (Preckel et al., 2010). Diese Langeweile hat negative Effekte auf das Lernen (Tze et al., 2010). Langeweile tritt im akademischen Kontext umgekehrt aber auch dann auf, wenn Schüler das Gefühl geringer Kontrolle über eine Aufgabe haben, d.h. sich überfordert fühlen (Pekrun et al. 2010). Bei schlechteren Schülern ist die Langeweile beispielsweise besonders hoch, wenn die selbsteingeschätzten Fähigkeiten gering sind, die wahrgenommenen Aufgabenanforderungen aber hoch (Fulmer & Tulis, 2016). Zudem nimmt bei niedrigen Fähigkeiten auch der positive Affekt mit steigender Herausforderung durch die Aufgabe ab (Schweinle et al., 2008). Bei Lernern mit hohen Fähigkeiten hingegen steigt der positive Affekt bei höherer wahrgenommener Herausforderung durch eine Aufgabe leicht an.

Das Erleben positiven Affekts bei einer Lernaufgabe ist eines der Kennzeichen eines Zustands von situativem Interesse, d.h. Interesse in der konkreten Lernsituation (Linnenbrink-Garcia et al., 2010). Das Erleben von Interesse in einzelnen Lernsituationen begünstigt die Entwicklung von stabilisiertem,

persönlichem Interesse am Fach (Krapp et al., 1992; Hidi & Renniger, 2006). Wenn die Anforderungen im Statistikkurs zu den Fähigkeiten des Lerners passen, erlebt dieser sich als kompetent und hat mehr Spaß beim Lernen. Infolgedessen beginnt er möglicherweise sogar, sich für Statistik zu interessieren und sein Wissen in diesem Bereich zu erweitern. In einer Studie von Durik und Matarazzo (2008) zeigte sich beispielsweise, dass sich nach der Arbeit mit einem Lernprogramm das entwickelte Interesse in Abhängigkeit von wahrgenommener Komplexität des Stoffs und fachlichen Fähigkeiten unterscheidet. Bei hoher wahrgenommener Aufgabenkomplexität, entwickelten nur Schüler mit hohen fachlichen Fähigkeiten Interesse für das Thema und wollten sich weiter mit dem Thema beschäftigen. Für Schüler mit geringen Fähigkeiten zeigte sich ein gegensätzlicher Effekt.

Die Bedeutung der Passung zwischen Aufgabenanforderungen und Fähigkeiten für die Interessensentwicklung findet sich auch in den Annahmen der Selbstdeterminationstheorie (Deci & Ryan, 1985) wieder. Das Kompetenzerleben in Folge einer Übereinstimmung der Anforderungen und Fähigkeiten bei der Ausführung einer Lernaktivität ist neben der Wahrnehmung von Autonomie/Kontrolle über diese Lernaktivität die Voraussetzung für die Entwicklung intrinsischer Motivation (Abeysekera & Dawson, 2015; Ryan & Moller, 2017). Intrinsische Motivation bedeutet, dass Lernaktivitäten in Statistik ausgeführt werden, weil die Lerner Spaß an der Lernaktivität haben. Diese intrinsische Motivation entsteht jedoch nur dann, wenn die Aufgabenanforderungen nicht zu hoch oder zu niedrig sind. Werden Aufgaben als zu leicht wahrgenommen, bleibt das Kompetenzerleben nämlich ebenfalls aus (Ryan & Moller, 2017).

Die Anpassung der Anforderungen beim Lernen an die Voraussetzungen des Lerners – d.h. weder eine Über- noch eine Unterforderung des Lerners – ist somit auch für die Motivation beim Lernen von Bedeutung. Eine Kursgestaltung, die an diesem Prinzip ausgerichtet ist, sorgt für ein positives affektives Lernerleben, welches Ausdruck von situativem Interesse in einer Lernsituation ist und die Entwicklung fachlichen Interesses begünstigt. Studierende mit höherem Interesse wenden wiederum mehr Zeit für das Lernen in einem Fach auf und nutzen tiefenorientierte Lernstrategien, die das konzeptuelle Verständnis fördern (Schiefele, Wild & Winteler, 1995).

Sowohl aus kognitiver als auch aus motivationaler Sicht erscheint es daher sinnvoll, das Lernen in Statistikkursen individuell anzupassen. In einem Kurs mit heterogenen Voraussetzungen bedeutet dies einerseits, dass die Anforderungen des Kurses für Lerner mit geringem Vorwissen und hoher Statistikangst reduziert werden müssen, um eine kognitive Belastung und Demotivierung zu vermeiden. Andererseits muss die inhaltliche Komplexität des statistischen Lernmaterials beibehalten werden, um Lernern mit günstigeren Voraussetzungen weiterhin genügend Herausforderung zu bieten. Zudem können manche vereinfachte Darstellungen statistischer Konzepte auch Fehlkonzepte begünstigen (Hatfield et al., 2006; Watkins et al., 2014).

## **D.2.2 Methoden zur individuellen Förderung des Wissenserwerbs im IC**

Im IC gibt es unterschiedliche Methoden, um Lernprozesse zu individualisieren. Diese Methoden zielen vor allem darauf ab, Lerner mit ungünstigen Voraussetzungen beim Lernen zusätzlich zu unterstützen. Individualisierung bedeutet in diesem Fall also, dass anstatt das Anforderungsniveau des Kurses generell zu senken, dieses flexibel an die Lerner angepasst werden kann.

### **D.2.2.1 Individuelle Unterstützung in der Vorbereitungsphase**

Eines der zentralen Kennzeichen des Lernens im IC ist, dass die Wissensvermittlung nicht mehr in der Lehrveranstaltung, sondern vorab zu Hause stattfindet (Abeysekera & Dawson, 2015; DeLozier & Rhodes, 2016; Kim et al., 2014). Häufig werden dazu Videos (z.B. Vorlesungsaufzeichnungen oder Lernvideos) genutzt (DeLozier & Rhodes, 2016; O'Flaherty & Phillips, 2016; Peterson, 2016). Eine aktuelle Untersuchung von Jensen et al. (2018) zeigt, dass solche Vorlesungsvideos zur Wissensvermittlung im IC zum Ende des Kurses mit einer besseren Leistung im Abschlusstest einhergehen, als die Vorbereitung mit Lehrbuchtexten oder (inter-)aktivem Lernen mit Tutorials.

Videos eignen sich aus mehreren Gründen besonders gut als Medium zur Vermittlung von statistischen Wissensinhalten. Erstens erlauben es Videos im Gegensatz zu (Online-)Texten Bild- und Audiomaterial gleichzeitig zu präsentieren. Nach der kognitiven Theorie des Multimedialernens (Mayer & Moreno, 2003) verhindert dies eine kognitive Überlastung im visuellen Speicher und führt dadurch zu besseren Lernergebnissen. Empirische Befunde zeigen, dass dies vor allem für komplexes Lernmaterial gilt (Ginns, 2005; Mattis, 2015). Insofern sollte eine Instruktion per Video statt durch Texte vor allem bei Lernern mit wenig Vorwissen günstig sein, da diese bei komplexem Lernmaterial schneller kognitiv überlastet sind (Zumbach, 2010). Zweitens ist es in Statistik häufig wichtig, korrekte mentale Modelle aufzubauen und Fehlkonzepte zu vermeiden. Dazu werden häufig technikgestützte Simulationen verwendet (Liu et al., 2010; Tyroller, 2005), die auch einfach in Lernvideos integriert werden können. Eine aktuelle Metaanalyse kam zu dem Ergebnis, dass solch animiertes Lernmaterial vor allem Verstehen und Anwendung des Gelernten verbessert (Berney & Bétrancourt, 2016). Drittens erlauben es Videos im Vergleich zu (Live-)Vorlesungen zwischendurch zu stoppen (z.B. um Notizen zu machen) und Inhalte wiederholt anzusehen. Dadurch kann die Lerngeschwindigkeit vom Lerner selbst gesteuert werden. Von Videoaufzeichnungen profitieren deshalb vor allem Studierende mit geringem Vorwissen, welche in Vorlesungen schnell kognitiv überfordert sind. Beispielsweise zeigt eine Untersuchung von Williams et al., (2016), dass die Häufigkeit der Nutzung von Vorlesungsaufzeichnungen kaum Varianz in den Noten über bisherige Leistungen hinaus erklären kann. Demnach nutzen Studierende mit geringem Vorwissen die Vorlesungsaufzeichnungen offenbar häufiger und können dadurch die Vorwissensunterschiede kompensieren. Im Gegensatz dazu, sehen sich sehr gute Studierende Vorlesungsaufzeichnungen oft nur einmal an und überspringen bereits bekannte Passagen (Owston et al., 2011). Durch die Verwendung von Videos kann die Phase der Wissensvermittlung im IC demnach besser individualisiert werden als in einer Vorlesung: Lerner mit höherem Vorwissen brauchen die Videos nur einmal zu sehen, während

Lerner mit geringem Vorwissen diese beliebig oft wiederholen können. Neben der Unterstützung des statistischen Wissenserwerbs sind Videos möglicherweise auch ein besonders motivierendes Instruktionsmedium (Jensen et al., 2018). Beispielsweise wurde in zahlreichen Untersuchungen eine positive Wirkung von Videos auf die Einstellung von Lernern gefunden (Kay, 2012).

Es ist anzumerken, dass Videos nur eine Methode zur Wissensvermittlung in der Vorbereitungsphase darstellen. Häufig werden Video- oder Textmaterialien zur Vorbereitung noch durch Übungsaufgaben ergänzt, um die aktive Auseinandersetzung mit dem Stoff zu fördern und sicherzustellen, dass alle Lerner sich auf die Veranstaltung vorbereiten (Chili-Turner, 2015; Loux et al., 2016; McLaughlin & Kang, 2017). Unabhängig von der gewählten Methode gilt jedoch, dass Zeit und Geschwindigkeit beim Lernen im IC im Gegensatz zu traditionell aufgebauten Kursen durch den Lerner selbst gesteuert werden können – die Wissensvermittlung findet also individualisiert statt.

#### D.2.2.2 Individuelle Unterstützung in der Lehrveranstaltung

Dadurch, dass die Lerner im IC die Wissensinhalte für die Lehrveranstaltung zu Hause vorbereiten, steht in der Veranstaltung Zeit für die Anwendung der Inhalte zur Verfügung. In Statistikkursen wird ein Teil der Veranstaltungszeit meist dazu genutzt, gemeinsam anhand praktischer Problemstellungen zu üben (Burgoyne & Eaton, 2018; Chili-Turner, 2015; Loux et al., 2016; McLaughlin & Kang, 2017; Nielsen et al., 2018; Peterson, 2016; Wilson, 2013). Empirische Befunde zeigen, dass das Lernen im Fach Statistik anhand realistischer Probleme in anwendungsbezogenen Testaufgaben zu weit besseren Ergebnissen führt, als traditionelle Vorlesungen (Karpiak, 2011). Zudem können Lerner in Statistikkursen, die mit authentischen Alltagsmaterialien (wissenschaftliche Artikel) arbeiten, ihr Wissen nicht nur in akademischen, sondern vor allem in praktischen Situationen eher anwenden (Daniel & Braasch, 2013). Entscheidend für den Lernerfolg ist jedoch auch, dass dieses Üben mit Problemstellungen im IC innerhalb der Veranstaltung stattfindet. In einer Studie von Shinaberger (2017) wurde schrittweise jedes Semester eine Veränderung in der Struktur eines Statistikkurses vorgenommen. Die Umstellung vom Vorlesungsformat auf den IC hatte dabei eine stärkere Verbesserung der Kursleistungen zur Folge, als die zuvor erfolgte Anreicherung der Übungsaufgaben für das Lernen zu Hause durch praktische Problemstellungen.

Eines der Kennzeichen des IC, welches Lernaktivitäten in der Veranstaltungsphase erfolgreich macht, ist die Anwesenheit einer professionellen Lehrperson (Chen et al., 2014; Kim et al., 2014). Diese Lehrperson hat im IC vor allem die Aufgabe, die Aktivität der Lerner individuell anzuleiten und zu unterstützen, beispielsweise indem sie Fragen der Lerner beantwortet, den Lerner bei der Auswahl von Material unterstützt, Rückmeldungen gibt, individuell Hilfestellung bei der Aufgabenbearbeitung leistet, den Fortschritt der Lerner überwacht, die Kommunikation zwischen Lernern anregt usw. (Blau & Shamir-Inbal, 2017). Dabei ist es im Vergleich zu Online-Tutoring auch möglich, dass die anwesende Lehrperson auf die sozialen und emotional-motivationalen Bedürfnisse der Lerner eingeht (Lee et al., 2017). Es ist wichtig, dass die Unterstützung durch die Lehrperson in Abhängigkeit von den Voraussetzungen des Lerners erfolgt, beispielsweise nach dem Ansatz der *Cognitive Apprenticeship*

(Collins et al., 1989). In diesem fördert die Lehrperson als Experte die kognitiven und metakognitiven Prozesse beim Lernen, wobei die Unterstützung immer weiter reduziert wird, je mehr Expertise der Lerner gewinnt (sog. *Scaffolding*). Für die kognitive Belastung der Lerner kann es entscheidend sein, dass die Art der instruktionalen Unterstützung an das Vorwissen der Lerner angepasst wird. Studien zum Lernen mit praktischen Problemstellungen zeigen beispielsweise, dass vor allem schwächere Lerner von einer zusätzlichen Anleitung beim Problemlösen (z.B. detailliert ausgearbeitete Lösungsschritte) profitieren. Lerner mit mehr Expertise schneiden hingegen besser ab, wenn sie weniger zusätzliche Anleitung erhalten (z.B. weniger ausgearbeitete Lösungsschritte) (Kalyuga & Sweller, 2004; Reisslein, 2005; Reisslein et al., 2006). Dieser Expertisen-Umkehr-Effekt tritt auf, weil die zusätzliche Instruktion bei höherer Expertise überflüssige Information darstellt, welche die kognitive Belastung (in diesem Fall den *extraenous load*) im Arbeitsgedächtnis erhöht (Kalyuaga, 2017). Diese Ergebnisse sind vermutlich auch auf andere Unterstützungsmethoden übertragbar (Kalyuaga, 2017). Es ist also nicht sinnvoll, wenn die statistischen Konzepte noch einmal für alle erklärt werden. Diese Erklärungen sollten nur bei den Lernern erfolgen, die dafür Bedarf haben. Wichtig ist deshalb, dass die Lehrperson neben fachlichem Wissen über pädagogische Expertise verfügt. Eine Studie von DeGrave et al. (1999) zeigt beispielsweise, dass Studierende eher das Gefühl haben, von einem Tutor zu profitieren, der ihren Lernprozess anregt, als von einem Tutor, der vor allem seine Fachexpertise einsetzt.

Ein zweiter Faktor zur Unterstützung des Lernens mit statistischen Problemen in der Veranstaltung ist das gemeinsame Lernen (Loux et al., 2016). In einer Studie von Kuo et al. (2011) konnte beispielsweise das problembasierte Lernen mit einer Lernumgebung noch über *Scaffolding* durch die Lehrperson hinaus verbessert werden, wenn die Lerner gemeinsam in heterogenen Kleingruppen arbeiten. Durch die Gruppenarbeit erzielten die Lerner nicht nur bessere Leistungen beim Problemlösen, sondern waren auch weniger mental belastet und hatten zum Ende eine positivere Einstellung gegenüber dem Fach.

Der Vorteil beim kollaborativen Lernen mit praktischen Problemen ist, dass der einzelne Lerner nicht über das gesamte für die Problemlösung notwendige Wissen verfügen und dieses gleichzeitig im Arbeitsgedächtnis verarbeiten muss (Janssen et al., 2010; Paas et al., 2010). Stattdessen kann die kognitive Belastung durch die Problemlöseaufgabe auf ein gemeinsames Arbeitsgedächtnis der Gruppenmitglieder aufgeteilt werden (Kirschner et al., 2009b). Dies reduziert beim einzelnen Lerner zwar die intrinsische kognitive Belastung durch die Komplexität der Aufgabeninhalte kann jedoch zu zusätzlicher extrinsischer Belastung aufgrund der Zusammenarbeit in der Gruppe führen (Janssen et al., 2010; Kirschner et al., 2009a). Das bedeutet, dass einzelne Lerner nur dann von der Gruppenarbeit profitieren, wenn die Komplexität der Aufgabeninhalte ihre kognitive Kapazität übersteigt (Jansen et al., 2010). Beispielsweise schneiden Lerner bei Aufgaben, die einen weiten Transfer von Lerninhalten erfordern, besser ab und müssen weniger mentalen Aufwand investieren, wenn sie die Lerninhalte in der Gruppe erworben haben (Kirschner et al., 2009b). Zudem zeigt sich, dass gemeinsames Lernen bei komplexen Aufgaben zum nahen Transfer zwar nicht zu besseren Leistungen führt, die Lerner aus der Gruppe aber

für die gleiche Leistung weniger mentalen Aufwand investieren müssen, als individuelle Lerner (Kirschner et al., 2011). Demnach sollte das Lernen in der Gruppe besonders Lerner mit geringerem Vorwissen unterstützen, für welche die eigenständige Lösung komplexer Aufgaben eine (zu) hohe Belastung des Arbeitsgedächtnisses bedeutet. In der Studie von Kuo et al. (2011) wurden die positiven Effekte der Gruppenarbeit auf Problemlösen und mentale Belastung beispielsweise nur für Lerner im niedrigen und mittleren Leistungsbereich beobachtet. Auch Metaanalysen zeigen, dass schlechtere Lerner am meisten von der Gruppenarbeit profitieren (Lou et al., 1996; 2001). Jedoch scheint gemeinsames Üben nicht immer zu nachhaltigen Effekten zu führen. Beispielsweise schneiden Lerner, die statistische Problemlöseaufgaben bearbeiten zwar besser ab, als wenn sie alleine arbeiten, dieser Effekt verschwindet jedoch wenn die Gruppenmitglieder einzeln getestet werden (Pociask & Rajaram, 2014). Eine mögliche Erklärung für diesen Befund ist, dass die Gruppen in der Studie durch die Lehrperson zusammengestellt wurden, sodass die Gruppenmitglieder in der gemeinsamen Lernphase zu wenig motiviert waren. Lerner sind vor allem dann intrinsisch motiviert, wenn sie sich ihre Gruppenzusammenstellung selbst aussuchen können (Ciani et al., 2008). Deswegen ist es aus motivationaler Sicht möglicherweise besser, wenn Lerner selbst entscheiden können, ob und in welchen Gruppen sie arbeiten möchten.

### **D.3 Hypothesen**

Der IC bietet einen weitaus größeren Spielraum zur Individualisierung des Lernens als traditionelle Lehrveranstaltungen wie beispielsweise Vorlesungen. Besonders Lerner mit geringem Vorwissen und/oder hoher Angst im Fach Statistik können im IC durch unterschiedliche Methoden (z.B. selbststeuerbare Lernvideos, individuelles Tutoring, kollaboratives Lernen) individuell so unterstützt werden, dass die kognitive Belastung beim Lernen reduziert wird (Janssen et al., 2010; Kalyaga, 2017; Mayer & Moreno, 2003; Paas et al., 2010). Dadurch sollte der Erwerb konzeptuellen und praktisch anwendbaren Statistikwissens gefördert werden (Mayer, 2008; Moreno & Mayer, 2007; Sweller et al., 1998).

H1a: Der statistische Wissenserwerb ist im IC höher, als bei einer Statistikvorlesung.

H1b: Diese Unterschiede werden über die geringere kognitive Belastung im IC mediiert.

Neben dem Verhindern einer kognitiven Überlastung kann die Individualisierung auch ein positives affektives Erleben beim Lernen begünstigen (Durik & Matarazzo, 2008; Ryan & Moller, 2017). Dieses positive affektive Erleben wiederum sorgt dafür, dass sich Lerner freiwillig mit statistischen Inhalten beschäftigen (Hidi & Renniger, 2006; Linnebrink-Garcia et al., 2010; Ryan & Moller, 2017) und fördert die vertiefte Verarbeitung der Inhalte und damit den Erwerb konzeptuellen Wissens (Schiefele et al., 1995).

H1c: Der höhere Wissenserwerb im IC im Vergleich zu einer Statistikvorlesung wird über das affektive Erleben beim Lernen gefördert.

Im IC können die Kursanforderungen flexibel an die Voraussetzungen unterschiedlicher Gruppen von Lernern (z.B. Lerner mit geringem vs. Lerner mit hohem Vorwissen) angepasst werden. In Statistikkursen wurden oft solche Untergruppen von Lernern mit eher günstigen und eher ungünstigen Voraussetzungen (z.B. Vorwissen, Statistikangst, Statistikeinstellung, Interesse) identifiziert (Lapka et al., 2010; Mutz & Daniel, 2011; Stark & Mandl, 2002). Im Gegensatz zu einer traditionellen Statistikvorlesung sollte der IC geeignet sein, solch heterogene Voraussetzungen zu kompensieren. Das bedeutet, im IC sollten alle Lerner unabhängig von ihren Voraussetzungen zum Ende des Kurses über einen ähnlichen Level an konzeptuellem, anwendbarem Wissen verfügen.

H2: Der statistische Wissenserwerb ist im IC homogener, als bei einer Statistikvorlesung.

Der IC sollte als Kursformat nicht nur den Wissenserwerb, sondern auch die positive Entwicklung der Statistikmotivation (z.B. Selbstwirksamkeit und Interesse) unterstützen (Liebert et al., 2016; Thai et al., 2017; Zheng et al., 2020). Im IC kann das Lernen so angepasst werden, dass die aktuellen Anforderungen genau dem Vorwissen eines Lerners entsprechen. Dadurch kommt es häufiger zu Erfolgserlebnissen beim Lernen. Eigene Erfolgserlebnisse sind der entscheidende Faktor, welcher die Überzeugung von der Erlernbarkeit von Statistik steuert (Bandura, 1994). Diese Überzeugung beeinflusst wiederum, ob Lerner auch motiviert an schwierigere Lernaufgaben herantreten – und reduziert nebenbei die Angst vor Statistik (Pajares, 1996). Zudem ist die Arbeit mit praktischen Problemstellungen im IC eine gute Methode, um den Lernern die praktische Relevanz von Statistik aufzuzeigen und damit Interesse am Fach Statistik zu wecken (Harackiewicz & Knogler, 2017). Dies wird zusätzlich gefördert, wenn das Lernen mit Problemstellungen beispielsweise in Kleingruppen oder mit individueller Betreuung erfolgt (Hidi & Renniger, 2006). Langfristig wird sich dieses Interesse jedoch nur stabilisieren, wenn Lerner die Erfahrung machen, dass sie das notwendige statistische Wissen zur Lösung solcher Probleme auch erlernen können, d.h. wenn die Lerner auch die notwendige Selbstwirksamkeit haben. Viele Untersuchungen aus dem Bereich der fachlichen und beruflichen Interessensentwicklung deuten darauf hin, dass diese Selbstwirksamkeit mit dem Interesse im Zusammenhang steht (Wright & Holtum, 2012; Bierer et al., 2014). Die Individualisierung im IC begünstigt diese Selbstwirksamkeitsentwicklung. Im Gegensatz zu traditionellen Kursen sollte der IC also auch die Motivation im Fach Statistik fördern.

H3: Die Statistikmotivation im IC ist zum Ende des Kurses höher, als bei einer Statistikvorlesung.



## D.4 Methode

### D.4.1 Untersuchungsdesign

Für die Evaluation wurde ein quasi-experimentelles Design mit Messwiederholung verwendet, in dem ein Statistikkurs im Inverted-Classroom-Format (IC) mit einem Statistikkurs im traditionellen Vorlesungsformat (TL) verglichen wurde. Bei beiden Kursen handelte es sich um ein zusätzliches Angebot außerhalb des regulären Curriculums.

Zudem wurde eine Kontrollgruppe evaluiert, welche nicht an einem zusätzlichen Statistikkurs teilnahm. Dadurch sollte die Wirksamkeit der Statistikkurse über im Befragungszeitraum aufgetretene Veränderungen durch andere Einflussfaktoren (z.B. curriculare Statistikkurse) hinaus beurteilt werden.

Bei allen Untersuchungsteilnehmern wurde zu Beginn und zum Ende des Semesters die Statistikmotivation gemessen. Zum Semesterende wurde zudem ein Test zur Messung der Statistikleistung durchgeführt. Zudem wurden die statistische Vorbildung der Studierenden und die allgemeine kognitive Leistungsfähigkeit zu Semesterbeginn als Kontrollvariablen erhoben.

Bei den Teilnehmern der Statistikkurse wurde zudem im Verlauf des Semesters das Lernerleben im Kurs gemessen. Außerdem wurde nach Abschluss des dritten Moduls ein modulspezifischer Leistungstest durchgeführt. Der Untersuchungsaufbau ist schematisch in Abbildung D.1 dargestellt.

Modul 1: Deskriptive Statistik			Modul 2: Population & Stichprobe			Modul 3: Effekt & Signifikanztest			Modul 4: Beta-Fehler & Teststärke			Modul 5: Korrelation & Regression		
	Wissens- vermittlung	Wissens- anwendung		Wissens- vermittlung	Wissens- anwendung		Wissens- vermittlung	Wissens- anwendung		Wissens- vermittlung	Wissens- anwendung		Wissens- vermittlung	Wissens- anwendung
IC	außerhalb der Lehr- veranstaltung	innerhalb der Lehr- veranstaltung	IC	außerhalb der Lehr- veranstaltung	innerhalb der Lehr- veranstaltung	IC	außerhalb der Lehr- veranstaltung	innerhalb der Lehr- veranstaltung	IC	außerhalb der Lehr- veranstaltung	innerhalb der Lehr- veranstaltung	IC	außerhalb der Lehr- veranstaltung	innerhalb der Lehr- veranstaltung
TL	innerhalb der Lehr- veranstaltung	außerhalb der Lehr- veranstaltung	TL	innerhalb der Lehr- veranstaltung	außerhalb der Lehr- veranstaltung	TL	innerhalb der Lehr- veranstaltung	außerhalb der Lehr- veranstaltung	TL	innerhalb der Lehr- veranstaltung	außerhalb der Lehr- veranstaltung	TL	innerhalb der Lehr- veranstaltung	außerhalb der Lehr- veranstaltung

#### Statistikmotivation

- SATS
- Statistische Selbstwirksamkeitserwartung
- Vermeidung

#### Kontrollvariablen

- Vorwissen
- Studienfach
- Fachsemester
- Statistikausbildung

#### Kognitive Belastung

- Affektives Lernereleben

#### Statistikleistung

- Zwischentest

#### Statistikleistung

- Nachtest
- Statistikmotivation
- SATS
- Statistische Selbstwirksamkeitserwartung
- Vermeidung

Abbildung D.1. Schematische Darstellung des Untersuchungsaufbaus für die beiden Statistikkurse im Lehrformat Inverted Classroom (IC) und im Lehrformat traditionelle Vorlesung (TL).

## D.4.2 Struktur der Statistikkurse

In allen Statistikkursen (IC und TL) wurden statistische Grundlagen wie deskriptive Statistik, Stichproben und Populationen, Signifikanztest, Teststärke und Korrelation behandelt (Giesbrecht et al., 1997; Landrum, 2005). Jeder Kurs war thematisch in fünf Module unterteilt. Jedes Modul umfasste einen gemeinsamen Veranstaltungstermin und eine Phase des Lernens außerhalb der Veranstaltung. Die Kurstermine fanden im Abstand von zwei Wochen statt und dauerten jeweils 90 min. Der Aufbau der Veranstaltungsphase und der Phase des selbstständigen Lernens unterschied sich in Abhängigkeit vom Kursformat (IC oder TL). Der IC wurde so entwickelt, dass das Lernen so flexibel wie möglich an die Voraussetzungen der Lerner angepasst werden konnte, während der Individualisierungsgrad der TL geringer sein sollte. Alle anderen Variablen wie Lernmaterial, zeitlicher Aufwand und Kursleitung wurden in beiden Kursen so ähnlich wie möglich gestaltet.

### D.4.2.1 Inverted Classroom

Im IC bereiteten die Teilnehmer die Inhalte des Moduls zu Hause mit Hilfe von zwei oder drei Lernvideos vor, die auf der Lernplattform Moodle zugänglich gemacht wurden. Jedes Video dauerte zwischen 10 – 15 min. Die Lernvideos wurden von mehreren Experten entwickelt, die langjährige pädagogische Erfahrung im Unterrichten von Statistik haben.

Da Studierende aber auch Wissenschaftler in der Forschung und Praxis häufig mangel- oder fehlerhafte statistische Konzepte aufweisen (Clark et al., 2007; Haller & Kraus, 2002; Lavigne et al., 2008; Sotos et al., 2007; Huck, 2009), wurde besonderer Wert darauf gelegt, durch die Videos den Aufbau korrekten schematischen Wissens zu fördern. Vor der Einführung in die mathematischen Grundlagen wurde dazu zunächst versucht, ein mentales Modell des Konzepts aufzubauen. Dazu wurden u.a. grafische Darstellungen der Konzepte verwendet, die sich in anderen Untersuchungen bewährt haben (Hong & O'Neill, 1992; Aguinis & Branstetter, 2007). Soweit es möglich war wurde zudem versucht, multiple grafische Repräsentationen zu verwenden, da dies die Bildung mentaler Modelle unterstützen kann (Schnotz & Bannert, 2003). Zudem wurden Simulationen eingesetzt, welche das Verständnis statistischer Konzepte fördern und die Entstehung von Fehlkonzepten verhindern können (Mills, 2004; Novak, 2013; Liu et al., 2010; Tyroller, 2005). Falls Formeln in den Videos präsentiert wurden, wurden diese in einer Schreibweise verwendet, welche das konzeptuelle Verständnis unterstützt (Atkinson et al., 2003). Für die Videos wurden auch allgemeine Gestaltungsempfehlungen für Lernvideos (z.B. kurze Beschreibung der Inhalte jeden Videos zu Beginn, Einbettung des erklärten statistischen Konzepts in einen empirischen Kontext, Zusammenfassungen) verwendet (Brar & van der Meij, 2017).

Zu jedem Video wurden auf der Lernplattform Moodle mehrere Leitfragen zu den zentralen Inhalten bereitgestellt. Diese Leitfragen können das Lernen aus Videos unterstützen (Lawson et al., 2006). Jeder Veranstaltungstermin im IC begann zunächst mit einer kurzen Wiederholung der Inhalte aus den Lernvideos (ca. 10 – 15 min), bei der die Teilnehmer Fragen zu den Videos beantworten konnten. In der restlichen Zeit bearbeiteten die Teilnehmer des IC auf der Lernplattform Moodle bereitgestellte

Übungsaufgaben. Um das Lernen in der Veranstaltung maximal an die Voraussetzungen der Lerner anzupassen, standen drei Aufgabentypen zur Verfügung. Diese Aufgaben lassen sich theoretisch anhand der von Dunham et al. (2015) für das Fach Statistik angepassten Lerntaxonomie von Bloom (Anderson et al., 2001) hierarchisch klassifizieren. Auf dem ersten Level wird mit *Wissensaufgaben* das Faktenwissen (z.B. Definition der Standardabweichung) geprüft. Auf dem zweiten Level erfordern *Verständnisaufgaben* ein statistisches Konzept in anderen Worten zu beschreiben, die Bestandteile zu identifizieren und das Konzept mit anderen zu vergleichen (z.B. Standardabweichung von Varianz und Standardfehler abgrenzen). Auf dem dritten Level soll das Wissen in Beispielfragestellungen angewendet werden (*Anwendungsfragen*) (z.B. Vergleich der Standardabweichung in unterschiedlichen Stichproben). Ziel dieser hierarchischen Struktur ist es, dass Lerner mit ungünstigen Voraussetzungen durch die Wissens- und Verständnisaufgaben ihr statistisches Wissen zunächst automatisieren und in Form konzeptueller Schemata strukturieren können, bevor sie dieses beim Problemlösen anwenden (Paas, 1992). Direktes Feedback über die richtige Aufgabenlösung erfolgte über eine online verfügbare Musterlösung.

Die Bearbeitung der Übungsaufgaben wurde aktiv durch die verantwortliche Lehrperson unterstützt. Diese gab beispielsweise Hinweise zur Auswahl des richtigen Aufgabentyps, Feedback während der Aufgabenbearbeitung, beantwortete aber auch offene Fragen zu den Inhalten des Moduls.

In der Veranstaltung wurde den Teilnehmern freigestellt, ob sie die Aufgaben alleine oder in Gruppen bearbeiten möchten.

Alle Übungsaufgaben standen den Teilnehmern auch nach dem Kurstermin weiter zur Verfügung.

#### D.4.2.2 Traditionelle Vorlesung

Jedes Modul der TL begann mit einer Vorlesung zu den statistischen Inhalten. Als Vorlesungsmaterial wurden die PPT-Präsentationen verwendet, die auch in den Lernvideos zum Einsatz kamen. Nach jedem Termin wurde das Material auf der Lernplattform Moodle als pdf-Datei zur Verfügung gestellt. Allerdings hatten die Lerner keinen Zugriff auf Erklärungen zum Material wie es in den Lernvideos der Fall war.

Im Anschluss an den Veranstaltungstermin sollten die Teilnehmer an der TL die Inhalte selbstständig mit Hilfe von Übungsaufgaben nachbereiten. Die Übungsaufgaben mit Musterlösungen wurden wie im IC auf der Lernplattform Moodle bereitgestellt. Im Gegensatz zum IC hatten die Teilnehmer jedoch nur Zugriff auf die Anwendungsaufgaben zu praktischen Beispielfragestellungen und nicht auf Aufgabentypen für Lerner mit geringerem Vorwissen (Wissens- und Verständnisaufgaben).

### D.4.3 Durchführung

Zwischen dem SoSe 2016 und dem SoSe 2018 wurden an der Universität Greifswald zwei Statistikkurse als IC und ein Kurs als TL durchgeführt. Alle Kurse wurden als außercurriculare Lehrveranstaltung für alle Fachdisziplinen (außer Mathematik und Psychologie) angeboten. Die Rekrutierung erfolgte über Werbeplakate und Flyer an zentralen Begegnungspunkten der Universität (Mensa, Bibliothek, Institute etc.) sowie per Rundmail an die Fachschaften. Die Kurse wurden auch gezielt in Lehrveranstaltungen verschiedener empirisch arbeitender Fächer beworben. Insgesamt wurden im SoSe 2016  $n = 31$  (IC), im WiSe 2016/17  $n = 40$  (IC) und im SoSe 2018  $n = 79$  (TL) Studierende als Kursteilnehmer angeworben.

Als Voraussetzung für die Kursteilnahme war neben regelmäßiger Anwesenheit die Teilnahme an den Befragungen zur Evaluation des Kurses verpflichtend. Diese Befragungen wurden online über [www.socisurvey.de](http://www.socisurvey.de) (Leiner, 2019) durchgeführt. Bei Erfüllung der Voraussetzungen wurde den Teilnehmern zum Ende des Semesters ein Kurszertifikat ausgestellt. Zu Beginn jedes Kurses erfolgte eine Aufklärung über die freiwillige Teilnahme an der Untersuchung sowie die Anonymisierung und den Datenschutz.

Für die Kontrollgruppe wurde im WiSe 2016/17 und im WiSe 2017/18 parallel zu den Statistikkursen für eine Umfrage zu Statistikeinstellungen geworben. Im Rahmen dieser Umfrage wurde zunächst eine Vorbefragung durchgeführt, um Studienfach und bisherige Statistikausbildung der Umfrageteilnehmer zu ermitteln. Aus den  $n = 50$  (WiSe 2016/17) bzw.  $n = 23$  (WiSe 2017/18) Teilnehmern der Vorbefragung wurden dann  $n = 20$  bzw.  $n = 12$  Teilnehmer für die Kontrollgruppe ausgewählt. Bei dieser Auswahl wurden Fachbereich und Fachsemester der Umfrageteilnehmer berücksichtigt. So sollte eine möglichst hohe Vergleichbarkeit zwischen den Kursteilnehmern des IC und den Kontrollgruppenteilnehmern hinsichtlich ihres fachlichen Hintergrunds und damit ihrer curricularen Statistikausbildung erreicht werden. An die Kontrollgruppenteilnehmer wurde dann zu Beginn und zum Ende des Semesters ein Onlinefragebogen über das Programm SoSciSurvey versendet. Als Aufwandsentschädigung erhielt jeder Kontrollgruppenteilnehmer einen Büchergutschein im Wert von 20 Euro.

### D.4.4 Stichprobe

Nach Ausschluss der Teilnehmer mit unvollständigen Datensätzen (z.B. durch Dropout aus Kursen und Befragung, fehlende Angaben) umfasst die für die Datenanalysen verwendete Gesamtstichprobe  $n = 94$  Studierende. Die Teilstichprobe der Teilnehmer am Statistikkurs mit traditioneller Vorlesung (TL) ist mit  $n = 43$  fast doppelt so groß wie der Anteil der Teilnehmer an den Kursen im Inverted Classroom (IC) ( $n = 27$ ) und den Studierenden aus der Kontrollgruppe ohne Statistikkurs ( $n = 24$ ).

In der Gesamtstichprobe befinden sich 30.8% männliche und 69.2% weibliche Studierende. Die meisten Studierenden sind zwischen 20 und 26 Jahre alt ( $M = 23.54$ ,  $SD = 3.76$ ) und befinden sich zwischen dem zweiten und fünften Semester ( $M = 4.01$ ,  $SD = 2.78$ ). Den größten Teil der Stichprobe bilden Studierende aus dem Fachbereich Landschaftsökologie (29.8%) und Politik-

/Kommunikationswissenschaften (31.95%). Die drittgrößte Gruppe stellen die Wirtschaftswissenschaftler (14.9%). Der Rest der Untersuchungsteilnehmer stammt überwiegend aus anderen naturwissenschaftlichen Fachbereichen (siehe Tabelle D.1).

Zwei Drittel der Untersuchungsteilnehmer (64.9%) hatten bereits einen Statistikkurs besucht. Die Beschreibung dieser Kurse lässt darauf schließen, dass es sich vorwiegend um Statistik auf Grundlagenniveau handelte (siehe Tabelle D.6 im Anhang). Etwa ein Drittel der Gesamtstichprobe (28.7%) nahm während des Semesters in dem Kurse und Befragungen stattfanden parallel an einem weiteren (curricularen) Statistikkurs teil.

**Tabelle D.1**

*Verteilung der Studienfächer*

	IC	TL	KG	Gesamt	
Biochemie	0	1	0	1	1.1 %
Biologie	1	3	0	4	4.3 %
Geisteswissenschaften	0	3	0	3	3.2 %
Geografie	0	3	1	4	4.3 %
Humanbiologie	0	2	1	3	3.2 %
Landschaftsökologie	8	14	7	30	31.9 %
Medizin	1	1	1	3	3.2 %
Pharmazie	0	0	1	1	1.1 %
Politik-/Kommunikationswissenschaften	12	6	10	28	29.8 %
Psychologie	0	0	1	1	1.1 %
Wirtschaftswissenschaften	2	10	1	14	14.9 %
Promotionsstudium	2	0	0	2	2.1 %
<b>Gesamt</b>	<b>27</b>	<b>43</b>	<b>24</b>	<b>94</b>	

## D.4.5 Messinstrumente

### D.4.5.1 Statistikleistung

Die Statistikleistung zum Ende des Semesters (Nachttest) wurde durch mehrere Testaufgaben im Multiple Choice Format gemessen, welche die Anwendung (Transfer) des statistischen Wissens auf ein praktisches Beispiel erfordern (z.B. Auswahl der korrekten deskriptiven Statistik zur Beschreibung von Daten, Interpretation der Ergebnisse eines Signifikanztests). Das Design der Aufgaben orientierte sich am *Statistical Reasoning Assessment* (SRA; Garfield, 1998). Dieses Instrument soll nicht nur die Ableitung korrekter Schlussfolgerungen messen, sondern enthält in den vorgegebenen Antwortalternativen auch typische Fehlschlüsse (Garfield, 1998; Garfield & Chance, 2000). Als Maß der Leistung wurde die prozentuale Anzahl korrekt gelöster Testaufgaben verwendet.

Die Testaufgaben des Nachttests bezogen sich auf Inhalte aus allen Lernmodulen (außer „Beta-Fehler & Teststärke“). Für die Messung der Leistung im Verlauf des Semesters (Zwischentest) wurden

nur jene Testaufgaben verwendet, welche sich auf die Anwendung der Inhalte des vorangegangenen Lernmoduls („Effekt & Signifikanztest“) bezogen.

#### D.4.5.2 Statistikmotivation

Für die Messung der Statistikmotivation wurde die deutsche Übersetzung der Survey of Attitudes towards Statistics (SATS-28; Schau, Stevens, Dauphinee & Del Vecchio, 1995, deutsch: Zimprich, 2012) verwendet. Dieses Instrument wird häufig in Studien zur Messung motivationaler und emotionaler Lernergebnisse in Statistikkursen eingesetzt und wurde bereits umfassend validiert (Emmioğlu & Capa-Aydin, 2011; Nolan, Beran & Hecker, 2012). Die SATS-28 enthält insgesamt 28 Aussagen zur Einstellung zum Fach Statistik die jeweils auf einer Skala von 1 (*trifft nicht zu*) bis 7 (*trifft voll zu*) bewertet werden sollen. Davon fragen sechs Items nach der Einschätzung der eigenen statistischen Kompetenz (*Cognitive Competence*), sieben nach der wahrgenommenen Schwierigkeit von Statistik als Fach (*Difficulty*), neun Items nach der Relevanz von Statistik für Beruf und Alltag (*Value*) und weitere sechs Items nach den mit dem Fach Statistik verbundenen positiven und negativen Affekten (*Affect*). Letztere Komponente misst somit auch die Statistikangst, die in vielen Studien problematisiert wird (Schau et al., 1995; Stark & Mandl, 2000). Die stichprobenspezifische interne Konsistenz der meisten Skalen liegt in der Untersuchung im zufriedenstellenden Bereich ( $\alpha = .76$  bis  $\alpha = .90$ ) (Tabelle D.2a). Lediglich die Skala *Difficulty* erweist sowohl bei der Messung zu Beginn ( $\alpha = .58$ ) als auch zum Ende des Semesters ( $\alpha = .71$ ) als weniger intern konsistent.

**Tabelle D.2a**

*Interne Konsistenz Statistikmotivation*

	Itemanzahl	Semesterbeginn	Semesterende
		Cronbach's $\alpha$	Cronbach's $\alpha$
Affect	6	.86	.90
Difficulty	7	.58	.71
Cognitive Competence	6	.77	.83
Value	9	.85	.90
Statistische Selbstwirksamkeit	8	.85	.94
Vermeidende Lernstrategien	7	.86	.89

*Anmerkungen. N = 94*

Neben der SATS-28 wurden zwei weitere Skalen zur Messung der Statistikmotivation eingesetzt. Die erste Skala fragt nach der Bewertung von acht Aussagen zur *Statistischen Selbstwirksamkeit*. Für die Bewertung wurde ein fünfstufiges Antwortformat von 0 (*stimme gar nicht zu*) bis 100 (*stimme voll zu*) verwendet. Die Skala zur statistischen Selbstwirksamkeit enthält sowohl Items, welche die Überzeugung von der aktuellen statistischen Kompetenz erfragen („Ich traue mir zu, mein statistisches Wissen auch auf praktische Fragestellungen anzuwenden.“), als auch Items zur Überzeugung von der Veränderbarkeit der statistischen Kompetenz („Egal wie sehr ich mich anstrengte, Statistik werde ich

wahrscheinlich nie verstehen.“) wie sie auch bei anderen Autoren zu finden sind (Duncan & McKeachie, 2005; Finney & Shraw, 2003; Schunk, 1996). Diese Items zur statistischen Selbstwirksamkeit wurden bereits in einer früheren Untersuchung validiert (Bünger, 2017). In der Stichprobe liegt die interne Konsistenz bei  $\alpha = .85 - .91$ . In der zweiten Skala soll die Zustimmung zu sieben Aussagen bezüglich *vermeidender Lernstrategien* bewertet werden („Ich versuche meist gar nicht erst komplizierte statistische Zusammenhänge zu verstehen.“). Auch hier ist die stichprobenspezifische interne Konsistenz hoch ( $\alpha = .86 - .89$ ).

#### D.4.5.3 Kognitive Belastung und affektives Lernerleben

Die *kognitive Belastung* im Kurs wurde mit fünf Items gemessen, in denen nach der subjektiven Überforderung durch die Kursinhalte gefragt wird („Ich habe Schwierigkeiten, im Seminar mitzukommen.“). Eine hohe kognitive Belastung sollte dann resultieren, wenn die Anforderungen des Kurses die Voraussetzungen des Lernalters übersteigen. Hohe Werte auf der Skala kognitive Belastung bedeuten also eine Überforderung während geringe sowohl für eine angemessene Anforderung als auch für eine Unterforderung stehen können. Die interne Konsistenz der Skala ist zufriedenstellend ( $\alpha = .87$ ) (Tabelle D.2b).

Das Erleben von positivem Affekt beim Lernen ist ein Aspekt, der das situative Interesse kennzeichnet (Linnenbrink-Garcia et al., 2010). Dieses *affektive Interesse* wurde als zweiter Indikator der Passung von Lernvoraussetzungen und Anforderungen des Kurses gemessen. Für die Studie wurden dazu vier Aussagen formuliert, welche die Freude am Lernen im Kurs beinhalten („Ich habe Freude daran, mich mit den Inhalten des Seminars auseinanderzusetzen.“). Auch hier ist die interne Konsistenz der Skala zufriedenstellend ( $\alpha = .88$ ).

Sowohl für die kognitive Belastung als auch für den positiven Affekt wurde das fünfstufige Antwortformat von 0 bis 100 verwendet.

**Tabelle D.2b**

*Interne Konsistenz Lernerleben*

	Itemanzahl	Cronbach's $\alpha$
Kognitive Belastung	4	.87
Affektives Interesse	5	.88

*Anmerkungen. N = 70.*

#### D.4.5.4 Kontrollvariablen

Das Vorwissen der Teilnehmer wurde über einen Multiple-Choice-Test gemessen, in dem ebenso wie im Nachtest praktische Beispiele gelöst werden sollten. Unterschiede in der statistischen Vorbildung wurden über die bisher besuchten *Statistikveranstaltungen* und die *Statistiknote* in diesen Veranstaltungen erhoben. Zudem wurden *Studienfach* und aktuelles *Fachsemester* erfragt. Je nach Fach kann es nicht nur Unterschiede in den statistischen Lerninhalten, sondern auch in der didaktischen

Aufbereitung der Veranstaltungen geben. Infolgedessen unterscheidet sich in Abhängigkeit vom Fachbereich nicht nur das statistische Vorwissen der Studierenden, sondern auch die Statistikmotivation, welche von der Art der Vermittlung von Lerninhalten abhängt. Außerdem kann angenommen werden, dass Studierende in höheren Semestern in ihrer Statistikausbildung weiter fortgeschritten sind. Neben der bisherigen Statistikausbildung wurde auch erhoben, ob die Studierenden während des Untersuchungszeitraums an einem *curricularen Statistikkurs* teilnahmen.

Die allgemeine kognitive Leistungsfähigkeit wurde über die *Abiturnote* und die *Mathematiknote* der Teilnehmer operationalisiert.



## D.5 Ergebnisse

### D.5.1 Voranalysen: Unterschiede zwischen den Gruppen

Vor Beginn der Hauptanalysen wurde getestet, ob sich relevante Gruppenunterschiede in den Kontrollvariablen ergeben. Dazu wurden die beiden Gruppen IC (Inverted Classroom) und TL (Traditionelle Vorlesung) hinsichtlich der Kontrollvariablen miteinander verglichen. Zudem wurde untersucht, ob sich die Gruppen mit Statistikkurs als IC und TL von einer Kontrollgruppe ohne Kurs (KG) unterscheiden. Diese Kontrollgruppe sollte den beiden Gruppen mit Kurs (IC und TL) hinsichtlich ihrer Statistikausbildung möglichst ähnlich sein.

Ein Vergleich der Fächerverteilung zwischen den Gruppen zeigt, dass sich das Studienfach in der Kontrollgruppe weder von der Gruppe mit IC ( $X^2(9) = 7.28, p = .61$ ) noch von der Gruppe mit TL ( $X^2(10) = 14.80, p = .14$ ) signifikant unterscheidet. Vergleicht man den IC mit dem TL zeigt sich jedoch, dass sich in der Gruppe mit TL deutlich mehr Studierende aus der BWL befinden ( $X^2(9) = 17.69, p = .04$ ) (Tabelle D.1). Zudem lässt sich deskriptiv erkennen, dass die Kontrollgruppenteilnehmer durchschnittlich in einem höheren Fachsemester studieren, bereits mehr Statistikkurse belegt haben und diese Kurse mit einer besseren Note abgeschlossen haben, als die Teilnehmer am IC und der TL (Tabelle D.3a, Tabelle D.3b und Tabelle D.3c). Der statistische Vergleich der Gruppen zeigt allerdings, dass das aktuelle Fachsemester in der KG nicht signifikant höher ist, als in den anderen Gruppen ( $F(2/90) = 11.78, p = 0.49$ ). Dasselbe gilt für die Angaben zur bisherigen Statistikausbildung: es ergeben sich weder signifikante Unterschiede in der Anzahl an Studierenden, die bereits einen Statistikkurs belegt haben und Statistikanfängern ( $X^2(2) = 2.89, p = .24$ ), noch Unterschiede in der Statistiknote ( $F(2/40) = 3.27, p = .45$ ). Auch die Leistung im Test vor Beginn des Statistikkurses unterscheidet sich nicht signifikant zwischen den beiden Kursen im IC und TL oder zur Kontrollgruppe ( $F(2/91) = 0.21, p = .98$ ). Obwohl die Kontrollgruppenteilnehmer also möglicherweise eine umfangreichere statistische Vorbildung erfahren haben, ist ihre statistische Kompetenz nicht höher als bei den anderen Gruppen.

**Tabelle D.3a**

*Statistikkurse*

	IC	TL	KG	Gesamt
bisherige Statistikkurse	59.3%	60.5%	79.2%	64.9 %
aktuelle Statistikkurse	33.3%	25.6%	30.4%	29.0 %

**Tabelle D.3b***Fachsemester*

	IC			TL			KG			Gesamt		
	<i>M (SD)</i>	<i>Min</i>	<i>Max</i>	<i>M (SD)</i>	<i>Min</i>	<i>Max</i>	<i>M (SD)</i>	<i>Min</i>	<i>Max</i>	<i>M (SD)</i>	<i>Min</i>	<i>Max</i>
Fachsemester	4.37 (3.05)	1	12	3.63 (2.26)	1	10	4.30 (3.32)	1	11	4.01 (2.79)	1	12

**Tabelle D.3c***Statistische, mathematische und allgemeine kognitive Fähigkeiten*

	IC	TL	KG	Gesamt
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Vorwissenstest <sup>a</sup>	42.96 (22.67)	43.72 (25.55)	42.50 (24.54)	43.19 (24.24)
Statistiknote <sup>b</sup>	3.01 (1.51)	3.20 (1.57)	2.53 (1.12)	2.93 (1.41)
Mathematiknote	2.70 (1.07)	2.42 (0.91)	2.67 (0.96)	2.56 (0.97)
Abiturnote	2.15 (0.60)	1.95 (0.69)	2.04 (0.75)	2.03 (0.68)

*Anmerkungen.* <sup>a</sup>. Leistung in Prozentpunkten. <sup>b</sup>. Nur bei bisherigem Statistikkurs.

In allen drei Gruppen (IC, TL und KG) besuchte ein etwa gleich großer Anteil an Studierenden einen weiteren Statistikkurs im Befragungszeitraum ( $X^2(2) = 0.51, p = .77$ ). Infolgedessen ist die Statistikausbildung der Kontrollgruppe mit den Gruppen mit Statistikkurs vergleichbar.

Bezüglich allgemeiner Leistungsfähigkeit zeigen sich ebenfalls keine Gruppenunterschiede: sowohl die Mathematiknoten ( $F(2/91) = 1.69, p = .41$ ) als auch die Abiturnoten ( $F(2/91) = 0.63, p = .51$ ) sind ähnlich.

### D.5.1 Statistikleistung in IC und TL

Theoretisch sollte der individualisierte Statistikkurs im IC im Vergleich zum Kurs als TL mit einem höheren Erwerb an praktisch anwendbarem Statistikwissen einhergehen (H1a). Um diese Hypothese zu testen, wurde untersucht, ob sich die Leistung im Zwischentest und die Leistung im Nachtest im IC von der Leistung im TL unterscheidet. Dazu wurde eine Regressionsanalyse mit der Zwischen- bzw. Nachtestleistung als Kriterium und dem Kursformat als dummykodierten Prädiktor (TL vs. IC) durchgeführt. Dieses Vorgehen wurde gewählt, um den totalen Effekt des Kursformats besser mit dem Mediationseffekt vergleichen zu können. Bei beiden Analysen wurde für die Leistung im Vorwissenstest kontrolliert.

Für die Leistung im Zwischentest ergibt sich nur ein sehr kleiner positiver Zusammenhang mit der Art des Kursformats ( $\beta = .13$ ). Demnach schnitten die Teilnehmer des IC beim Leistungstest nach dem Lernblock „Effekt & Signifikanztest“ etwas besser ab als Teilnehmer des TL. Während letztere im Test durchschnittlich knapp 19% der Aufgaben lösen konnten ( $M = 18.60, SD = 26.78$ ), sind es in der Gruppe mit IC 26% der Teilnehmer ( $M = 25.93, SD = 28.99$ ) (Tabelle D.4a). Dieser Unterschied ist nicht signifikant ( $t(67) = 1.07, p = .29$ ).

Im Leistungstest nach Abschluss des Kurses ergeben sich ähnliche Unterschiede. Die Teilnehmer des IC erreichen mit  $M = 38.54$  ( $SD = 21.04$ ) durchschnittlich neun Prozent mehr Punkte, als die Teilnehmer am TL ( $M = 47.62$ ,  $SD = 14.29$ ). Es zeigt sich ein signifikanter positiver Zusammenhang zwischen Kursformat und Testleistung ( $t(67) = 2.21$ ,  $p = .03$ ;  $\beta = .24$ ).

Damit sprechen die Ergebnisse teilweise für H1a. Der statistische Kompetenzerwerb ist im IC höher als beim TL. Allerdings sind die Unterschiede relativ gering.

**Tabelle D.4a**  
Statistikleistung

	IC	TL	$d^a$
	$M$ ( $SD$ )	$M$ ( $SD$ )	
Zwischentest	25.93 (28.99)	18.60 (26.78)	0.27
Nachtest	47.62 (14.29)	38.54 (21.04)	0.48

Anmerkungen. <sup>a</sup>. Standardisierte Mittelwertsdifferenz (Hedges  $g$ ) für den Unterschied in der Testleistung zwischen IC und TL.

Die beobachteten Unterschiede im statistischen Wissen zwischen IC und TL sollten darauf zurückzuführen sein, dass die kognitive Belastung im IC geringer (H1b) und das affektive Erleben beim Lernen höher (H1c) ist, als beim TL. Die in H1b und H1c angenommenen Mediationseffekte von kognitiver Belastung und affektivem Lernerleben wurden mit dem SPSS Makro PROCESS (Hayes, 2018) geprüft. Diese Vorgehen zum Test des Mediationseffekts erlaubt neben dem Signifikanztest des direkten Effekts des Kursformats auf die Leistung auch eine Prüfung des indirekten Effekts des Kursformats auf Signifikanz per Bootstrapping. Letztere prüft das Vorliegen eines Mediationseffekts.

Beim Test aller Mediationseffekte wurde für den Einfluss der Leistung im Vorwissenstest kontrolliert.

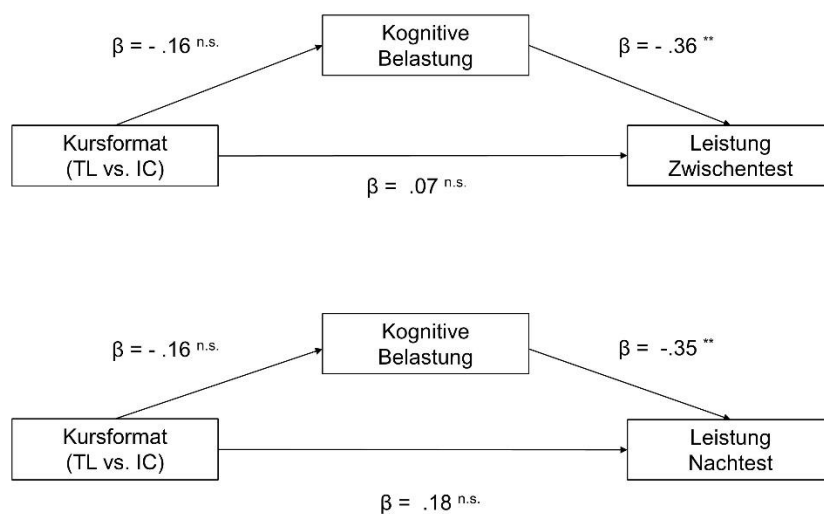


Abbildung D.2a Mediationsmodell der kognitiven Belastung. \*\*,  $p < .01$ . n.s.,  $p > .05$ .

Zunächst kann beobachtet werden, dass eine hohe kognitive Belastung im Kurs von Nachteil ist: bei höherer Überlastung ist sowohl die Leistung im Zwischentest ( $\beta = -.38$ ;  $t(67) = 3.23$ ,  $p = .00$ ) als auch im Nachtest ( $\beta = -.35$ ;  $t(67) = 3.37$ ,  $p = .00$ ) signifikant geringer.

Es zeigt sich zudem, dass sich die Unterschiede in der Testleistung zwischen den Kursformaten reduzieren, wenn die kognitive Belastung im Kurs berücksichtigt wird. Der Zusammenhang zwischen Kursformat und die Leistung im Zwischentest reduziert sich von einem totalen Effekt von  $\beta = .13$  (s.o.) auf einen direkten Effekt von  $\beta = .07$  ( $t(67) = 0.61$ ,  $p = .54$ ). Auch im Nachtest ist der direkte Effekt des Kursformats mit  $\beta = .18$  kleiner als der totale Effekt von  $\beta = .24$  (s.o.). Zudem ist der direkte Zusammenhang zwischen Kursformat und Leistung im Nachtest sogar nicht mehr signifikant ( $t(67) = 1.80$ ,  $p = .08$ ), wenn die kognitive Leistung als zusätzliche Variable betrachtet wird. Der indirekte Effekt des Kursformats auf die Leistung im Zwischentest beträgt allerdings nur  $\beta = .06$  und ist nicht signifikant. Auch für den Nachtest liegt der indirekte Effekt des Kursformats nahe Null ( $\beta = .06$ ) und ist nicht signifikant. Damit kann die beobachtete Reduktion des Zusammenhangs zwischen Kursformat und Leistung nicht im Sinne einer Mediation interpretiert werden. Infolgedessen muss H1b abgelehnt werden. Der Unterschied in der statistischen Kompetenz zwischen den Kursen wird nicht durch die geringere kognitive Belastung im IC mediiert.

Zudem muss berücksichtigt werden, dass es nur einen kleinen und insignifikanten negativen Zusammenhang zwischen Kursformat und kognitiver Belastung gibt ( $t(68) = 1.32$ ,  $p = .19$ ;  $\beta = -.16$ ). Die kognitive Belastung im IC ist mit  $M = 21.11$  ( $SD = 15.95$ ) insgesamt also nur wenig geringer, als beim TL mit  $M = 26.98$  ( $SD = 19.89$ ) (Tabelle D.4b).

Das Mediationsmodell der kognitiven Belastung ist in Abbildung D.2a dargestellt.

**Tabelle D.4b**  
*Lernerleben in IC und TL*

	IC	TL
	<i>M (SD)</i>	<i>M (SD)</i>
Kognitive Belastung	21.11 (15.95)	26.98 (19.88)
Affektives Interesse	67.13 (21.48)	50.73 (21.92)

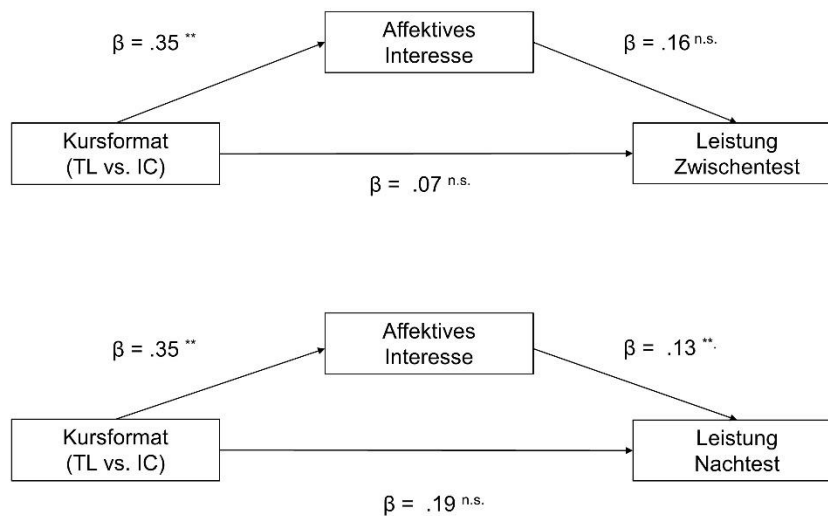


Abbildung D.2b. Mediationsmodell des affektiven Interesses.  $^{**}$ .  $p < .01$ .  $^{n.s.}$ .  $p > .05$ .

Auch beim Test des Mediationseffekts für das affektive Interesse verringert sich der totale Zusammenhang zwischen Kursformat und Testleistung nach Berücksichtigung des Einflusses des potenziellen Mediators. Beim Zwischentest sinkt der totale Effekt von  $\beta = .13$  (s.o.) auf einen direkten Effekt von  $\beta = .07$  ( $t(67) = 0.56$ ,  $p = .58$ ), für den Nachttest auf einen direkten Effekt von  $\beta = .24$  (s.o.) auf  $\beta = .19$ , wobei dieser Zusammenhang nicht mehr signifikant ist ( $t(67) = 1.66$ ,  $p = .10$ ) (Abbildung D.2b). Allerdings ist auch hier der über den Mediator vermittelte indirekte Effekt des Kursformats auf die Testleistung extrem gering: der indirekte Effekt auf den Zwischentest liegt bei  $\beta = .06$ , der indirekte Effekt auf den Nachttest bei  $\beta = .05$ . Keiner der beiden Effekte ist signifikant. Damit gilt auch das affektive Lernerleben, dass dieses nicht als Mediator des Zusammenhangs zwischen Kursformat und Leistung interpretiert werden kann. H1c muss infolgedessen ebenfalls abgelehnt werden. Interessanterweise kann für das affektive Interesse generell kein signifikanter Zusammenhang zur Leistung im Zwischentest ( $t(67) = 1.26$ ,  $p = .21$ ;  $\beta = .16$ ) oder im Nachttest ( $t(67) = 1.15$ ,  $p = .26$ ;  $\beta = .13$ ) beobachtet werden. Demnach scheint mehr „Spaß“ am Lernen nur wenig mit besseren Leistungen einherzugehen. Allerdings ist hervorzuheben, dass das Lernen im IC signifikant positiver erlebt wird, als im TL ( $t(68) = 3.11$ ,  $p = .00$ ;  $\beta = .35$ ). Im IC liegt das affektive Interesse bei  $M = 67.13$  ( $SD = 21.48$ ), beim TL nur bei  $M = 50.23$  ( $SD = 21.92$ ) (Tabelle D.4b).

## D.5.2 Homogenisierung im IC und TL

Im Gegensatz zu einer TL, welche nur für eine Gruppe von Lernern mit bestimmten Voraussetzungen geeignet ist, sollte der IC für alle Lerner zu einem ähnlichen Level an statistischem Wissen führen. Die Leistung bei der Bearbeitung der anwendungsbezogenen Testaufgaben sollten im IC also homogener sein, als bei der TL (H3).

Um diese Hypothese zu überprüfen wurde für die Messung des statistischen Wissens zu Beginn des Kurses (Vorwissenstest), die Messung des erworbenen Wissens im dritten Lernmodul (Zwischentest) und das Statistikwissen zum Ende des Kurses (Nachtest) jeweils die Varianz in der Gruppe mit IC mit der Varianz in der Gruppe mit TL verglichen. Dazu wurde der Levene-Test auf Varianzhomogenität angewendet (Bortz, 2005).

Die Varianz der Testleistung zu Beginn des Semesters ist in beiden Gruppen relativ ähnlich. Im Vorwissenstest liegt die Standardabweichung im IC bei  $SD = 22.67$  und bei der TL bei  $SD = 25.55$  Prozentpunkten (Tabelle d.3c). Dieser Unterschied ist nicht signifikant ( $F(1/68) = 0.68, p = .408$ ). Das Gleiche gilt für die Leistung im Zwischentest ( $F(1/68) = 1.11, p = .297$ ). Hier ist die Varianz mit  $SD = 28.99$  (TL) und  $SD = 26.78$  (IC) sehr ähnlich. Demgegenüber zeigt der Levene-Test für die Leistung im Nachtest, dass die Annahme der Varianzhomogenität verworfen werden muss ( $F(1/68) = 6.65, p = .012$ ). Im IC ist die Standardabweichung der Leistung im Nachtest mit  $SD = 14.29$  geringer, als in der Gruppe mit TL ( $SD = 21.04$ ). Die vor dem Kurs bestehenden Unterschiede im Statistikwissen der Studierenden konnten im IC also verringert werden. In Übereinstimmung mit H3 ist die Leistung im IC homogener als beim TL.

### D.5.3 Statistikmotivation im IC und TL

Durch die Individualisierung sollte der IC im Vergleich zum TL auch die Statistikmotivation fördern (H2). Es wurde deshalb geprüft, ob der Zuwachs an positiver Einstellung und statistischer Selbstwirksamkeitserwartung und die Verringerung der Vermeidung von Statistik im IC höher sind, als beim TL. Dazu wurde jeweils eine Varianzanalyse mit Messwiederholung durchgeführt. Die vier Skalen zur Motivation aus der SATS (*Affect, Difficulty, Cognitive Competence, Value*) wurden von der ursprünglichen Skalierung von 1 bis 7 in eine Skalierung von 0 bis 100 transformiert, um die Ergebnisse besser mit denen der anderen Motivationsskalen vergleichen zu können.

Es zeigen sich signifikante Unterschiede in der Veränderung der Selbstwirksamkeitserwartung zwischen den beiden Kursen ( $F(1/62) = 4.73, p = .03, \eta_{\text{part}} = .07$ ). Obwohl die Selbstwirksamkeitserwartung in beiden Kursen im Semesterverlauf ansteigt, ist der Zuwachs an Selbstwirksamkeit im IC signifikant und deutlich größer ( $t(26) = 3.29, p = .00; d = 0.52$ ) (Tabelle D.5a). Im Kurs mit TL zeigt sich kein signifikanter Zuwachs ( $t(42) = 1.56, p = .06; d = 0.32$ ). Geringe Unterschiede in der Veränderung der Statistikmotivation zwischen den Kursen zeigen sich auch für die *Cognitive Competence*. Hier werden 3 % Varianz durch das Kursformat erklärt ( $\eta_{\text{part}} = .03$ ). Der Unterschied ist jedoch nicht signifikant ( $F(1/62) = 1.61, p = .21$ ). Im Hinblick auf die Art der Veränderungen im Kursverlauf zeigt sich bei der *Cognitive Competence* ein ähnliches Bild wie bei der statistischen Selbstwirksamkeitserwartung: die *Cognitive Competence* steigt in beiden Statistikkursen zwar signifikant an ( $t(26) = 3.18, p = .000$  bzw.  $t(42) = 2.04, p = .02$ ) jedoch entspricht der Anstieg im TL nur einem kleinen Effekt ( $d = 0.37$ ), während der Anstieg im IC bereits mittelgroß ist ( $d = 0.68$ ) (Tabelle D.5a).

**Tabelle D.5a**

Veränderung der Statistikmotivation in IC und TL

		Semesterbeginn	Semesterende	$t^a$	$df$	$p^b$	$d^c$
		$M (SD)$	$M (SD)$				
Affect	IC	40.84 (20.06)	57.41 (25.04)	3.83	26	.000	0.76
	TL	35.47 (19.51)	48.26 (21.43)	4.83	42	.000	0.86
Difficulty	IC	38.18 (11.20)	43.83 (15.91)	1.95	26	.031	0.35
	TL	38.81 (12.84)	42.03 (13.80)	1.76	42	.043	0.30
Cognitive Competence	IC	49.69 (16.94)	59.88 (19.53)	3.18	26	.002	0.68
	TL	47.54 (17.21)	52.32 (19.22)	2.04	42	.024	0.37
Value	IC	74.97 (16.01)	80.18 (14.98)	2.28	26	.016	0.56
	TL	69.77 (16.78)	70.76 (20.72)	0.59	42	.278	0.17
Statistische Selbstwirksamkeit	IC	49.65 (17.98)	65.74 (23.55)	3.29	26	.001	0.52
	TL	47.97 (17.25)	51.67 (22.27)	1.56	42	.063	0.32
Vermeidung	IC	45.24 (21.50)	44.05 (24.62)	0.29	26	.387	0.06
	TL	49.50 (23.26)	52.66 (21.81)	1.27	42	.106	0.27

Anmerkungen. <sup>a</sup>. T-Test für abhängige Stichproben für die Veränderung der Statistikmotivation von Semesterbeginn bis Semesterende in IC bzw. TL. <sup>b</sup>. Einseitiger p-value. <sup>c</sup>.  $d$  für Messwiederholung nach Cohen (1988) (vgl. Lakens, 2013)

Auch bei der Veränderung des *Value* im Kursverlauf zeigt sich eine kleiner, wenn auch nicht signifikanter Unterschied zwischen TL und IC ( $F(1/62) = 2.55, p = .12, \eta_{\text{part}} = .04$ ). Der wahrgenommene Wert von Statistik ändert sich nicht durch die TL ( $t(42) = 0.59, p = .28; d = 0.17$ ), während er im IC signifikant zunimmt ( $t(26) = 2.28, p = .02; d = 0.56$ ) (Tabelle D.5a).

Beim positiven Affekt zeigen sich keinerlei Unterschiede in der Veränderung in Abhängigkeit vom Kurformat ( $F(1/62) = 1.06, p = .31, \eta_{\text{part}} = .02$ ). In beiden Kursen nimmt der positive Affekt bis zum Ende des Semesters deutlich zu (Tabelle D.5a). Sowohl im IC ( $t(26) = 3.83, p = .00; d = 0.76$ ) als auch im TL ( $t(42) = 4.83, p = .00; d = 0.86$ ) ist diese Zunahme signifikant.

Es lassen sich ebenfalls keine Gruppenunterschiede für die Veränderung auf der Einstellungsskala *Difficulty* ( $F(1/62) = 0.41, p = .52, \eta_{\text{part}} = .01$ ) und die *Vermeidung* statistischer Inhalte ( $F(1/62) = 0.87, p = .36, \eta_{\text{part}} = .01$ ) beobachten. Der Wert der Skala *Difficulty* steigt sowohl im IC als auch im TL über den Kursverlauf hinweg leicht an ( $t(26) = 1.95, p = 0.31, d = 0.35$  bzw.  $t(42) = 1.76, p = .04, d = 0.30$ ). Dabei ist zu berücksichtigen, dass ein Anstieg in diesem Fall nicht als Anstieg, sondern als Abnahme in der wahrgenommenen Schwierigkeit interpretiert werden muss, da alle Skalen der SATS im Sinne einer positiven Einstellung gepolt sind. Die Vermeidungshaltung bleibt sowohl bei

den Teilnehmern am IC ( $t(26) = 0.29, p = 0.39, d = 0.06$ ) als auch bei den Teilnehmern am TL ( $t(42) = 1.27, p = .16, d = 0.27$ ) über den Kurs hinweg konstant (Tabelle D.5a).

Insgesamt zeigt sich damit zwar, dass der IC eine stärkere Wirkung auf die Statistikmotivation hat, als die TL. Die Unterschiede sind jedoch sehr gering. Bei vielen Variablen wie der statistischen Selbstwirksamkeit, der *Cognitive Competence*, der *Difficulty* und dem *Affect* lassen sich in beiden Kursen positive Veränderungen beobachten. Die Ergebnisse sprechen damit nur bedingt für H2.

#### **D.5.4 Explorative Datenanalyse: Unterschiede zur Kontrollgruppe**

Neben den hypothetisch angenommenen Unterschieden zwischen den beiden Kursformaten (IC und TL) wurde auch untersucht, wie sich Statistikwissen und Statistikmotivation während des Untersuchungszeitraums, in dem die Kurse angeboten wurden, in einer Kontrollgruppe ohne Kurs verändern.

Zunächst wurde untersucht, ob sich auch im Vergleich zur Kontrollgruppe ohne Kurs Unterschiede in der statistischen Kompetenz zum Ende des Semesters ergeben. Dazu wurde eine Regressionsanalyse zur Vorhersage der Leistung im Nachtest durchgeführt. Die beiden Gruppen mit Statistikkurs im IC und als TL wurden so dummykodiert, dass sie mit der Kontrollgruppe verglichen werden konnten. Zudem wurde in der Regression für die Vortestleistung kontrolliert.

Dabei ergibt sich ein signifikant positiver Zusammenhang für den Unterschied zwischen der Gruppe mit IC und der Kontrollgruppe ( $t(90) = 2.49, p = .02; \beta = .28$ ). Deskriptiv zeigt sich, dass die Kontrollgruppe nur 35% ( $M = 35.71, SD = 18.36$ ) der Aufgaben im Nachtest richtig löst, während es beim Statistikkurs im IC über 47% sind ( $M = 47.62, SD = 14.29$ ). Die Teilnehmer am Statistikkurs mit TL erreichen mit einem Testwert von gut 38% ( $M = 38.54, SD = 21.04$ ) keine besseren Leistungen als die Kontrollgruppe. Für diesen Unterschied ergibt sich kein Zusammenhang zum Nachtest ( $t(90) = 0.56, p = .58; \beta = .06$ ).

Im Anschluss wurden die Veränderungen in der Statistikmotivation in der Kontrollgruppe ohne Kurs mit den beiden Kursgruppen verglichen. Dazu wurden zunächst jeweils zwei ANOVAs mit Messwiederholung durchgeführt, in der die Veränderung in der Gruppe mit IC mit der Kontrollgruppe bzw. die Veränderung in der Gruppe mit TL mit der Kontrollgruppe verglichen wurden. Die Ergebnisse sind in Tabelle D.5b dargestellt.



**Tabelle D.5b***Veränderung der Statistikmotivation in der KG*

	Semesterbeginn	Semesterende	$t^a$	$df$	$p^b$	$d^c$
	$M(SD)$	$M(SD)$				
Affect	38.84 (23.15)	49.19 (25.11)	0.20	23	.422	0.12
Difficulty	42.86 (13.70)	44.14 (16.21)	0.54	23	.298	0.14
Cognitive Competence	56.83 (18.91)	58.21 (20.70)	0.58	23	.283	0.20
Value	65.74 (19.52)	64.27 (19.53)	0.88	23	.194	0.43
Statistische Selbstwirksamkeit	62.89 (23.39)	60.54 (24.59)	0.98	23	.168	0.42
Vermeidung	44.94 (21.60)	47.47 (22.55)	1.21	23	.119	0.49

*Anmerkungen.* <sup>a</sup>. T-Test für abhängige Stichproben für die Veränderung der Statistikmotivation von Semesterbeginn bis Semesterende in der KG. <sup>b</sup>. Einseitiger p-value. <sup>c</sup>.  $d$  für Messwiederholung nach Cohen (1988) (vgl. Lakens, 2013)

Die Ergebnisse zeigen einen signifikanten Unterschied in der Veränderung auf der Skala Affect sowohl im Vergleich zum IC ( $F(1/49) = 11.07, p = .00, \eta_{\text{part}} = .18$ ) als auch im Vergleich zum TL ( $F(1/65) = 10.83, p = .00, \eta_{\text{part}} = .14$ ). Im Gegensatz zu den beiden Statistikkursen, welche den positiven Affekt fördern, bleibt dieser in der Kontrollgruppe von Beginn bis zum Ende des Semesters ( $t(23) = 0.20, p = 0.42; d = 0.12$ ) konstant (Tabelle D.5b).

Wird die Veränderung in den Überzeugungen von der statistischen Kompetenz betrachtet, unterscheidet sich hingegen nur der IC signifikant von der Kontrollgruppe. Sowohl bei der Cognitive Competence ( $F(1/49) = 4.67, p = .04, \eta_{\text{part}} = .09$ ), als auch bei der statistischen Selbstwirksamkeit ( $F(1/49) = 10.61, p = .00, \eta_{\text{part}} = .18$ ) ergibt sich ein signifikanter Interaktionsterm. Dies gilt nicht, wenn die Veränderungen auf den beiden Skalen der Überzeugungen von statistischer Kompetenz zwischen TL und Kontrollgruppe verglichen werden ( $F(1/65) = 0.88, p = .35, \eta_{\text{part}} = .01; F(1/65) = 2.76, p = 0.10, \eta_{\text{part}} = .04$ ). Betrachtet man die Veränderungen der Überzeugungen von der statistischen Fähigkeit in der Kontrollgruppe, zeigt sich, dass Cognitive Competence und statistische Selbstwirksamkeit in der Kontrollgruppe kaum zunehmen ( $t(23) = 0.58, p = 0.28; d = 0.20$  bzw.  $t(23) = 0.98, p = 0.17; d = 0.42$ ) (Tabelle D.5b).

Auch auf der Skala *Value* ist in der Kontrollgruppe ebenso wie in der Gruppe mit Statistikkurs als TL keine signifikante Veränderung im Semesterverlauf zu beobachten ( $t(23) = 0.88, p = 0.19; d = 0.43$ ) (Tabelle D.5b). Infolgedessen ergeben sich auch keine Unterschiede zwischen TL und KG ( $F(1/65) = 0.921, p = .341, \eta_{\text{part}} = .01$ ). Dahingegen unterscheidet sich die Kontrollgruppe auch hier signifikant vom IC ( $F(1/49) = 5.32, p = .03; \eta_{\text{part}} = .10$ ), in dem der Value im Semesterverlauf zunimmt.

Die Difficulty in der Kontrollgruppe verändert sich nicht im Vergleich zum TL ( $F(1/65) = 0.40, p = .53; \eta_{\text{part}} = .01$ ) oder dem IC ( $F(1/49) = 1.30, p = .26; \eta_{\text{part}} = .03$ ). Diese bleibt von Beginn bis zum

Ende des Semesters konstant ( $t(23) = 0.54, p = 0.30; d = 0.14$ ). Ebenso zeigt sich keine Interaktion für den Unterschied zum TL ( $F(1/65) = 0.03, p = .87; \eta_{\text{part}} = .00$ ) oder dem IC ( $F(1/49) = 0.604, p = .44; \eta_{\text{part}} = .01$ ), wenn die vermeidenden Lernstrategien betrachtet werden. Die Vermeidung steigt in der Kontrollgruppe im Semesterverlauf leicht an, der Anstieg ist jedoch nicht signifikant ( $t(23) = 1.21, p = 0.12; d = 0.49$ ).

Insgesamt betrachtet bleibt die Statistikmotivation in der Kontrollgruppe im Semesterverlauf also weitgehend gleich

## D.6 Diskussion

In dieser Studie wurde untersucht, ob der Inverted Classroom (IC) als Kursformat für einen fachübergreifenden Statistikkurs besser geeignet ist als ein traditionelles Vorlesungsformat (TL). Es wurde angenommen, dass die höhere Individualisierung im IC unterschiedliche Voraussetzungen der Kursteilnehmer besonders gut kompensieren kann und sowohl mit einer besseren Anwendung des Statistikwissens als auch mit einer höheren Statistikmotivation einhergeht.

Insgesamt ergeben sich nur kleine Unterschiede in der Testleistung und in der Zunahme der Statistikmotivation zwischen den Kursen. Am Ende des Kurses sind vor allem das anwendungsbezogene Statistikwissen und die Kompetenzüberzeugungen als Teil der Statistikmotivation im IC höher als in der TL. Festzuhalten ist auch, dass sich die Unterschiede im Statistikwissen zwischen den Kursteilnehmern im IC zum Semesterende reduzieren. Zudem zeigt ein Vergleich der beiden Statistikkurse mit einer Kontrollgruppe ohne zusätzlichen Statistikkurs, dass der IC beim Tests des anwendungsbezogenen Wissens und der Motivation insgesamt besser abschneidet als die Kontrollgruppe, während sich für den Statistikkurs im TL und die Gruppe ohne Kurs kaum Unterschiede ergeben. Unklar bleibt, warum der der IC das Statistikwissen und die Statistikmotivation besser fördert als die TL: Hinweise auf eine Mediation durch die theoretisch angenommenen kognitions- und motivationspsychologischen Prozesse (z.B. geringere kognitive Belastung und positiveres affektives Lernerleben), welche den Wissenserwerb im IC fördern, wurden nicht gefunden.

### D.6.1 Förderung des statistischen Wissenserwerbs im IC

Im Einklang mit den Ergebnissen vergleichbarer Studien (Chili-Turner, 2015; Peterson, 2016) erreichen die Teilnehmer des IC nach Abschluss des Lernmoduls zum Signifikanztest eine um 7% und zum Ende des Kurses eine um 9% bessere Testleistung, als Teilnehmer des traditionellen Vorlesungsformats. Dabei ist festzuhalten, dass der standardisierte Unterschied im Test der Inhalte des Lernmoduls zum Signifikanztest mit  $d = 0.26$  deutlich kleiner ist als im Gesamttest der Kursinhalte zum Ende des Semesters mit  $d = 0.50$ . Dies deutet darauf hin, dass individualisierte Kursformate wie der IC nicht nur kurzfristig, sondern vor allem längerfristig zu besseren Lernergebnissen führen. Um schematisches Wissen zu erwerben, das erfolgreich in praxisbezogenen Aufgaben angewendet werden kann, ist viel Zeit erforderlich. Ein Test direkt nach Abschluss eines Moduls ist möglicherweise wenig geeignet, dieses Wissen zu messen. Dafür spricht auch, dass auch die Teilnehmer des IC nur 25 Prozentpunkte im Test nach dem Modul zum Signifikanztest erreichten. Allerdings waren diese Teilnehmer im Vergleich zur TL durch Spaß am Lernen im Kurs (s.u.) möglicherweise motivierter, sich auch nach Abschluss des Kurses bzw. des Moduls mit den statistischen Inhalten auseinanderzusetzen. Beispielsweise könnten Teilnehmer mit eher wenig Vorwissen Übungsaufgaben, die sie im Kurs nicht geschafft haben, im Kursverlauf zu Hause nachbereitet haben. Genauso wie es in der hierarchischen Konzeption der bereitgestellten Übungsaufgaben vorgesehen war, haben diese Teilnehmer im Kurs also zunächst mit

den einfachen Wissensaufgaben begonnen und die Anwendungsaufgaben erst gelöst, nachdem sie genug Vorwissen aufgebaut hatten.

Es ist außerdem anzunehmen, dass sich vor dem Kurs bestehende Unterschiede im statistischen Wissen über die Zeit hinweg immer mehr vergrößern (Haughton & Kelly, 2015). Es fällt schwer, immer komplexere statistische Konzepte zu verstehen, wenn das notwendige Grundlagenwissen fehlt. Zu Beginn des Kurses können Wissenslücken also leichter kompensiert werden, als gegen Ende. Die Ergebnisse heben infolgedessen noch einmal die Notwendigkeit einer Homogenisierung des Statistikwissens vor allem in Kursen zu statistischen Grundlagen hervor. Solche Kurse dienen als Vorbereitung für die weitere statistisch-methodische Ausbildung (z.B. Arbeit an Forschungsprojekten, Kurse zu fortgeschrittenen statistischen Verfahren). Es ist daher wichtig, dass sie die Schere zwischen den Kursteilnehmern durch den Statistikkurs nicht zu vergrößern, sondern zu verringern.

Die Ergebnisse der Studie weisen darauf hin, dass der Statistikkurs im IC zum Ende tatsächlich eine solche Homogenisierung erreichen konnte. Dafür sprechen die Unterschiede in den Standardabweichungen der Leistung zwischen den Gruppen im Kursverlauf: während die Leistung zu Beginn des Kurses und zur Kursmitte im IC und in der TL ähnlich heterogen ist, zeigen sich zum Ende des Kurses im IC deutlich weniger Leistungsunterschiede zwischen den Studierenden innerhalb des Kurses. Die individuellen Unterstützungsmöglichkeiten von Lernern mit ungünstigen Voraussetzungen im IC scheinen also tatsächlich zu fördern, dass diese Lerner zum Ende des Kurses einen ebenso hohen Level an anwendungsbezogenem Statistikwissen erreichen, wie ihre Mitstudierenden. Hingegen weist die Heterogenität im traditionellen Vorlesungsformat darauf hin, dass dieser nur für einen Teil der Lerner geeignet war. Wie die Unterschiede in der Testleistung scheint sich die Homogenisierung in der Leistung erst nach einem längeren Zeitraum (nämlich zum Kursende) bemerkbar zu machen.

Die geringere Heterogenität in der Statistikleistung zum Ende des Kurses deutet darauf hin, dass der IC mittelbar besser als traditionelle Veranstaltungsformate dazu geeignet ist, individuelle Voraussetzungen wie Unterschiede im Vorwissen zu berücksichtigen. Diese Individualisierung sollte sich auch unmittelbar während des Kurses in einer geringeren kognitiven Belastung und positiverem Lernerleben bemerkbar machen.

Die Ergebnisse zeigen jedoch, dass die kognitive Belastung im IC nur geringfügig höher ist, als bei der TL. Die bessere Leistung in den Anwendungstests kann somit nicht dadurch erklärt werden, dass die Teilnehmer im IC durch das individualisierte Lernen weniger überfordert sind, als in der Vorlesung. Stattdessen zeigt sich, dass die subjektive kognitive Überlastung mit einem Wert von 21 (IC) bzw. 27 (TL) auf einer Skala von 0 - 100 in beiden Kursen eher niedrig ist. Es ist denkbar, dass das Anforderungsniveau der Statistikkurse unabhängig vom Kursformat (IC oder TL) allgemein geringer war, als dies in Grundkursen im Fach Statistik die Regel ist. Diese decken beispielsweise meist sehr viel mehr unterschiedliche Themen ab, als dies in der Studie der Fall war (Giesbrecht et al., 1997; Landrum, 2005). Infolgedessen stand auch während des Kurses mit traditioneller Vorlesung genug Zeit zur Verfügung, die statistischen Konzepte ausführlich vorzustellen und Fragen zu den Konzepten zu beantworten.

Zudem wurde das Material für Vorlesung und Lernvideos so gestaltet, dass eine kognitive Überlastung vermieden und das *meaningful learning* der statistischen Konzepte gefördert wird. Damit konnten auch Teilnehmer, die eher weniger Vorwissen hatten, der Vorlesung möglicherweise besser folgen. Es ist anzunehmen, dass die Unterschiede in der kognitiven Belastung von IC zur Vorlesung größer sind, wenn das generelle Anforderungsniveau des Kurses höher ist. Gleichzeitig unterstreichen die Ergebnisse aber auch, dass es bei einer sorgfältigen Auswahl und Gestaltung des Kursmaterials nach kognitionspsychologischen Leitlinien für die kognitive Belastung weniger entscheidend ist, nach welchem Kursformat Statistik unterrichtet wird.

Für den Spaß beim Statistikhernen scheint das Kursformat allerdings eine Rolle zu spielen: In der Studie zeigt sich, dass das Erleben von Interesse in der Lernsituation im IC höher ist, als im Kurs nach dem traditionellen Format. Dabei ist der positive Zusammenhang zwischen dem IC und dem erlebten Interesse sogar dann noch bedeutsam, wenn neben der statistischen Kompetenz auch für die Statistikmotivation zu Beginn des Kurses kontrolliert wird ( $t(61) = 3.26, p = .002; \beta = .31$ ). Die Teilnehmer im IC haben also mehr Spaß beim Lernen, ganz egal wie ängstlich oder interessiert an Statistik sie sind.

Das hier erhobene affektive Interessenserleben scheint allerdings keinen unmittelbaren Einfluss auf die Leistung zum Ende des Kurses zu haben, sodass nicht davon ausgegangen werden kann, dass es als Mediator zwischen der Individualisierung im IC und dem Erwerb von anwendbarem Statistikwissen wirkt. Auch in anderen Untersuchungen hatte das situationale Interesse im IC nur einen geringen Einfluss auf die Leistung (Chen et al., 2016). Es ist denkbar, dass das erlebte Interesse in einer Lernsituation kurzfristig zunächst nur die Entwicklung von stabilem Interesse am Fach Statistik fördert (Krapp et al., 1992; Hidi & Renninger, 2006). Trifft diese Annahme zu, sollte das Erleben von Interesse im Kurs nicht die Leistung im Zwischen- und Nachtest, sondern der Zusammenhang zum Interesse am Fach Statistik beeinflussen. Dieses Interesse zeichnet sich im Allgemeinen durch eine positive affektive und eine wertbezogene Komponente aus (Krapp, 1999). Post hoc durchgeführte Analysen zeigen, dass das Interesse beim Lernen im Kurs (nach Kontrolle der Statistikmotivation zu Beginn des Kurses) tatsächlich mit den Werten auf der Skala *Affect* ( $t(61) = 4.15, p = .000, \beta = .44$ ) und *Value* ( $t(61) = 2.71, p = .008, \beta = .25$ ) zusammenhängt. Das im IC entwickelte Interesse an Statistik beeinflusst dann möglicherweise erst über einen längeren Zeitraum nach Abschluss des Kurses, ob sich die Teilnehmer ihr statistisches Wissen weiterentwickeln und in Beruf und Alltag anwenden. In zukünftigen Untersuchungen wäre es dementsprechend sinnvoll zu untersuchen, ob sich auch nach längerer Zeit Unterschiede zwischen den Kursteilnehmern an IC und TL zeigen (z.B. bei der Bachelor- oder Masterarbeit).

## D.6.2 Förderung der Motivation im IC

Einer der Hauptunterschiede zwischen dem Lernen in den beiden Kursen in dieser Studie ist, dass der IC die Überzeugung, Statistik erlernen und anwenden zu können (*Statistische Selbstwirksamkeit* und *Cognitive Competence*) in einem stärkeren Ausmaß fördert als die traditionelle Vorlesung. Zudem

zeigen sich signifikante Unterschiede des IC zu einer Kontrollgruppe ohne zusätzlichen Statistikkurs, in der sich die Kompetenzüberzeugungen in Statistik im Verlauf des Semesters nicht veränderten. Dies unterstreicht die praktische Relevanz des IC als Kursformat – besonders in Statistikkursen zu Beginn der Ausbildung. Eines der Hauptprobleme in diesen Kursen ist, dass die Teilnehmer mit der Überzeugung in den Kurs kommen, dass sie Statistik niemals erlernen können, weil das Fach viel zu komplex und mathematisch ist. Die Studierenden sollten während des Kurses die Erfahrung machen, dass sie ihre statistischen Fähigkeiten effektiv verändern können. Dies gelingt nur dann, wenn die Anforderungen des Kurses graduell an den aktuellen Wissens- und Fähigkeitslevel angepasst werden. Dazu sind Kursformate wie der IC notwendig, in denen eine solche Individualisierung des Lernens möglich ist.

Überraschenderweise geht der hohe Zuwachs am Vertrauen in die statistischen Fähigkeiten im IC nicht mit einer höheren Abnahme an Statistikangst einher. Der *Affect*, welcher in dieser Untersuchung als Indikator der Statistikangst diente, verändert sich im IC und der traditionellen Vorlesung gleich stark, indem der positive Affect zu bzw. der negative Affect abnimmt. Im Vergleich dazu bleibt er in der Kontrollgruppe, die nicht am Kurs teilgenommen hat, konstant. Es ist denkbar, dass dieses Ergebnis ebenfalls auf das reduzierte Anforderungsniveau der Statistikkurse zurückzuführen ist. Bei einer ausführlichen und verständlichen Vermittlung der Inhalte werden die Lerner auch in einer Vorlesung nicht überlastet und haben das Gefühl, die Anforderungen bewältigen zu können. Im Einklang damit reduziert sich die wahrgenommene Schwierigkeit (*Difficulty*) der statistischen Inhalte sowohl im IC als auch in der TL etwas im Semesterverlauf, während sie in der Kontrollgruppe ohne Kurs konstant bleibt. Interessant ist dennoch, dass eine geringe kognitive Belastung im TL möglicherweise die Angst reduziert, aber nicht in gleichem Ausmaß die Überzeugungen von der statistischen Kompetenz erhöht. Dies könnte dadurch bedingt sein, dass im IC mehr aktives Lernen stattfand als in der TL. Obwohl die Teilnehmer an der TL in der Studie auch die Möglichkeit hatten, ihr statistisches Wissen in Übungsaufgaben mit direktem Feedback zu Hause anzuwenden, gab es keinerlei Kontrolle. Viele Teilnehmer haben sich deshalb möglicherweise gar nicht mit den Anwendungsaufgaben auseinandergesetzt. Die Erfahrung der eigenen Kompetenz (z.B. bei der Bearbeitung von Aufgaben mit Feedback über den eigenen Lernerfolg) ist jedoch die beste Methode, um das Vertrauen in die eigenen Fähigkeiten zu fördern.

Obwohl das Vertrauen in die eigenen Fähigkeiten gefördert werden konnte, vermeiden die Teilnehmer der Statistikkurse im IC und Vorlesung auch nach dem Kurs weiterhin, sich freiwillig vertieft mit statistischen Inhalten auseinanderzusetzen. Zum Ende des Kurses verwenden die Teilnehmer immer noch im selben Ausmaß vermeidende Lernstrategien (*Vermeidung*). Ähnliche Ergebnisse finden sich auch in anderen Studien, in denen eine Zunahme der Kompetenzüberzeugung nicht mit einer positiveren Einstellung einherging (Walker & Brakke, 2017). Hier muss sich jedoch die Frage gestellt werden, welche Ziele für die Motivation der Teilnehmer in einem Statistikkurs sinnvoll sind. Sicher wäre es von Vorteil, bei einem Großteil der Lerner Interesse zu wecken, sodass diese sich auch freiwillig

im Fach Statistik weiterzubilden und nicht nur deswegen, weil sie die Statistikprüfung bestehen müssen. Allerdings muss dabei auch berücksichtigt werden, dass Statistikkurse hauptsächlich dazu dienen, die Studierenden auf das empirische Arbeiten im Rahmen ihres Studiums im eigenen Fach vorzubereiten. Dazu ist es viel wichtiger, statistisch kompetente Studierende auszubilden, die bei einem Problem beim statistischen Arbeiten nicht vorschnell aufgeben, sondern dieses aufgrund des Vertrauens in die eigenen Fähigkeiten selbstständig nach angemessenen Lösungsstrategien suchen. Zudem zeigt sich, dass im Vergleich zur TL zumindest der Statistikkurs im IC die wahrgenommene Relevanz von Statistik für Studium und Alltag (*Value*) erhöhen konnte. Die Vermittlung der Relevanz eines Themas in einer Lehrveranstaltung kann wiederum die Grundlage für den Start des Prozesses der Interessensentwicklung schaffen (Harackiewicz & Knogler, 2017).

### D.6.3 Limitationen und Ausblick

Eines der Hauptprobleme der Studie ist, dass es sich beim untersuchten Statistikkurs um eine freiwillige Lehrveranstaltung außerhalb des normalen Curriculums handelte. Es ist daher anzunehmen, dass sich die Teilnehmer an den Kursen von der generellen Studierendenpopulation unterscheiden. Beispielsweise könnten vor allem an einer Statistikweiterbildung interessierte Studierende an den Kursen teilgenommen haben. Fraglich ist, ob dieses „Interesse“ am Kurs dadurch motiviert wird, dass die Studierenden eine Erweiterung ihrer statistischen Kompetenz anstreben oder aus Verzweiflung und Angst vor dem Versagen in Statistik alle Lernangebote nutzen, die von der Universität angeboten werden. Die Vorabanalyse zeigt, dass die bisherigen Statistiknoten der Teilnehmer in IC und TL zwar nicht signifikant, aber deskriptiv eine halbe Notenstufe schlechter sind, als in der Kontrollgruppe (Tabelle D.3c). Zudem scheint bei den Kursteilnehmern die Statistikangst (*Affect*) deutlich geringer ( $t(91) = 2.17, p = .032$ ) und die Kompetenzüberzeugungen (*Cognitive Competence* und *Statistische Selbstwirksamkeit*) zu Beginn des Semesters deutlich schlechter ( $t(91) = 1.96, p = .053$  bzw.  $t(91) = 3.08, p = .003$ ) zu sein, als in der Kontrollgruppe (Tabelle D.5a und D.5b). Demnach könnten an den Kursen vor allem solche Studierende teilgenommen haben, die in Statistikkursen bisher schlechte Erfahrungen gemacht haben (z.B. starke Verständnisschwierigkeiten hatten oder durch die Prüfung gefallen sind) – also Studierende mit ungünstigen Voraussetzungen. Obwohl der IC besonders solche Studierende ansprechen sollte ist unklar, ob sich die Ergebnisse auch auf Studierende mit günstigeren Voraussetzungen übertragen lassen.

Auch der extrem hohe Dropout, der im Verlauf des Kurses beobachtet werden konnte lässt darauf schließen, dass es Studierende gibt, die nicht mit der Art des Kursformats zufrieden sind. Beispielsweise haben 58.4% der Teilnehmer am TL den Kurs beendet, während es beim IC nur 44.2% waren. Obwohl dieser Unterschied nur gering und statistisch nicht signifikant ist ( $X^2(1) = 2.94, p = .086$ ) könnte dies ein Hinweis darauf sein, dass der IC nicht für alle Studierenden in der Population geeignet ist. Zusätzliche Analysen zeigen beispielsweise, dass im IC der *Value* zu Beginn des Semesters signifikant negativ mit dem Dropout korreliert ist ( $r = -.25, t(68) = 2.13, p = .037$ ): Studierende, für die

Statistik keine Relevanz für Studium und Alltag hat, machen sich offenbar auch nicht den Aufwand an einem zusätzlichen Statistikkurs teilzunehmen, in dem sie selbst auch noch aktiv mitarbeiten müssen (wie das im IC der Fall ist). In zukünftigen Studien sollte daher unbedingt untersucht werden, ob die gefundenen Ergebnissen auch auf curriculare Kurse übertragbar sind, in denen die Teilnehmerzusammensetzung (z.B. die Motivation für die Kursteilnahme) anders ist.

Problematisch für die Untersuchung der Unterschiede zwischen IC und traditionellem Kurs ist auch, dass sich die angebotene Veranstaltung vermutlich von den Pflichtveranstaltungen, die die Studierenden bisher belegt hatten, unterschied. Beispielsweise erfolgte keine studienrelevante Leistungsbewertung im Kurs und auch die statistischen Inhalte waren didaktisch vermutlich anders aufbereitet (z.B. weniger Lernstoff, starker Praxisbezug). Die Verbesserung der Motivation und Leistung kann also möglicherweise dadurch erklärt werden, dass überhaupt ein Kurs stattfand. Darauf weisen auch die Unterschiede zur Kontrollgruppe hin, die im Verlauf des Semesters keinen zusätzlichen Kurs erhielt.

Insgesamt zeigt die Untersuchung, dass Kursformate mit individuellen Lernangeboten wie der IC das Potenzial haben, die Kompetenzentwicklung in Statistikkursen zu fördern. Nicht in allen Curricula wird es möglich sein, Statistikkurse in Form eines solchen IC anzubieten. Dies bedeutet nicht, dass eine Verbesserung des Lernens in anderen Kursformaten nicht möglich ist. In der Untersuchung zeigte sich beispielsweise, dass sich auch Studierende im Statistikkurs im „traditionellen Vorlesungsformat“ während des Lernens nicht überfordert fühlten und sich ihre Angst bzw. ihr negativer Affekt im Kursverlauf positiv entwickelte. Dies lässt sich dadurch erklären, dass die statistischen Inhalte auch in der „traditionellen Vorlesung“ der Untersuchung einfach und verständlich vermittelt wurden. Demnach ist eine veränderte Aufbereitung der Inhalte in Statistikvorlesungen bereits der erste Schritt bei der Unterstützung der Kompetenzentwicklung. Für den Lernerfolg in Statistik scheint es außerdem wichtig zu sein, Studierende während eines Kurses zum „aktiven“ Lernen anzuregen (Jensen et al., 2015; Shinaberger, 2017). Solche aktivierenden Elemente können mit wenig Aufwand auch in Vorlesungen eingebaut werden (z.B. kurze Quizfragen oder Übungsaufgaben während der Vorlesung) (Gier & Kreiner, 2009; Goode et al., 2018; Mallenby & Mallenby, 2003; Winstone & Millward, 2012). Zudem können auch nur einzelne Teile einer Vorlesung als IC angeboten werden – z.B. bei besonders komplexen Inhalten (Burgoyne & Eaton, 2018).



## E. GESAMTDISKUSSION

Ziel der vorliegenden Arbeit ist es Empfehlungen abzuleiten, wie die Kompetenzentwicklung in Statistikkursen zu Beginn des Studiums unterstützt werden kann. Dabei interessiert besonders, wie und welche Aspekte der Statistikmotivation bei der Unterstützung der Kompetenzentwicklung berücksichtigt werden sollten. Was können Lehrende also tun, damit ihre Studierenden ihren ersten Statistikkurs überzeugt, interessiert und fähig das Gelernte in Situationen in Studium, Beruf und Alltag anzuwenden, verlassen?

Betrachtet man die Voraussetzungen für statistische Kompetenz aus einer theoretischen Perspektive, wird Lernen und Leistung besonders durch Kompetenzüberzeugungen motiviert (Marsh et al., 2017; Nolan et al., 2012; Pugh & Bergin, 2006; Richardson et al., 2012; Robbins et al., 2004; Zimmerman et al., 2017). Unglücklicherweise sind Studierende zu Beginn ihrer Statistikausbildung oft eher wenig davon überzeugt, statistische Inhalte verstehen, erlernen und anwenden zu können (Dykeman, 2011; Walker & Brakke, 2017; Waples, 2016). Ein Ansatz die statistische Kompetenz zu verbessern könnte demnach darin bestehen, diese Kompetenzüberzeugungen in Statistikkursen zu verändern.

Die Ergebnisse der ersten Untersuchung deuten darauf hin, dass für gute Leistungen vor allem eine hohe Kompetenzüberzeugung zu Beginn des Kurses wichtig ist – auch wenn diese den tatsächlichen Stand der Kompetenz überschätzt. Zudem kann laut Untersuchungsergebnissen besonders eine implizite Theorie der Veränderbarkeit statistischer Kompetenz („Growth Mindset“) hilfreich sein: diese kann kompensieren, wenn Studierende zu Beginn eines Kurses noch an ihrer aktuellen Kompetenz zweifeln und führt besonders dann, wenn Studierende sich überschätzen, zu guten Leistungen. In der Praxis scheint es daher zielführend zu sein, Studierende in Statistikkursen zu überzeugen, dass sie Statistik bewältigen können – d.h. dass ihre statistische Kompetenz veränderbar ist.

Dazu können beispielsweise Interventionen zur Förderung eines „Growth Mindset“ eingesetzt werden. In diesen werden die Lerner beispielsweise zu Beginn einer Kurses über „neueste Ergebnisse der Gehirnforschung“ aufgeklärt, nach denen sich biologische Strukturen im Gehirn und damit auch Kompetenz ändern lassen (z.B. Blackwell et al., 2007; Bostwick & Becker-Blease, 2018). Ziel der Interventionen ist, dass die Lerner diese Informationen verinnerlichen und dadurch ihr eigenes „Mindset“ ändern. Die Interventionen können sogar inhaltlich auch an das „Mindset“ in einem bestimmten Bereich (z.B. Statistik) angepasst werden (z.B. Smith & Capuzzi, 2019; Burnette et al., 2020). Ein zusätzlicher positiver Nebeneffekt neben der Förderung eines „Growth Mindset“ ist, dass aufgrund eines möglichen Zusammenhangs zwischen impliziter Theorie und Kompetenzüberzeugungen letztere ebenfalls durch die Intervention erhöht werden können (Diseth et al., 2018; Huang et al., 2017; Burnette et al., 2020). Zudem können solche Interventionen, wenn sie zu Beginn des Studiums stattfinden, möglicherweise nicht nur ein „Growth Mindset“ in Statistik, sondern auch allgemein eine positive Überzeugung von der Erlernbarkeit von Kompetenz unterstützen.

Neben den genannten Interventionen, welche direkt an einer Veränderung der Kompetenzüberzeugungen ansetzen gibt es auch solche, welche diese indirekt oder kombiniert mit einer Förderung des Kompetenzzuwachses erhöhen (O'Mara, Marsh & Craven, 2006). Kompetenz und Kompetenzüberzeugungen bilden sich besonders dadurch aus, dass Studierende beim Erlernen und Anwenden der statistischen Inhalte erfolgreich sind (Lent et al., 1991; Lent et al., 1996). Statistikkurse sollten demnach so gestaltet sein, dass Studierende beim Lernen dieses Erfolgserleben haben.

Die Ergebnisse der zweiten Untersuchung lassen darauf schließen, dass eine Förderung von Kompetenzüberzeugungen in Statistik beispielsweise mit Lehrformaten wie dem „Inverted Classroom“ (IC) gelingen kann, in denen Lernen individualisiert werden kann. Individualisiert bedeutet dabei, dass die Anforderungen beim Lernen flexibel an Voraussetzungen unterschiedlicher Lerner angepasst werden können. Dadurch werden letztere weder über- noch unterfordert, was zu mehr Erfolgserleben beim Lernen und besseren Lernergebnissen führen sollte. Die zweite Untersuchung zeigt entsprechend, dass neben höheren Kompetenzüberzeugungen auch ein homogenerer Kompetenzzuwachs durch das Lernen im IC möglich ist.

Das Lehrformat IC stellt nicht die einzige Option zur Individualisierung des Statistikkernens dar. Insbesondere durch die Ergänzung von Lehrveranstaltungen durch E-Learning kann flexibler auf die Voraussetzungen und Bedürfnisse von Lernern eingegangen werden (Boelens et al., 2017). In Statistik werden beispielsweise schon längere Zeit digitale Lernumgebungen eingesetzt, welche das Verständnis des komplexen Stoffs unterstützen sollen (z.B. Austerschmidt & Bebermeier, 2019; Schober et al., 2006; Tyroller, 2005). Solche digitalen Lernumgebungen können im Rahmen einer Lehrveranstaltung beispielsweise genutzt werden, um Lerner auch in „traditionellen Vorlesungen“ auf die Veranstaltungstermine vorzubereiten, sodass alle mit ähnlichem Vorwissen in die Vorlesung starten (Cerbin, 2018). Wichtig bei der Nutzung von Lernumgebungen ist dabei einerseits, dass diese nicht einfach alleine von den Studierenden bearbeitet, sondern in das Konzept der Lehrveranstaltung mit eingebunden werden. Dadurch können Studierende bei Schwierigkeiten gezielt unterstützt werden (z.B. wenn diese Probleme haben, das Lernmaterial zu verstehen oder zu strukturieren). Der durch die Lehrveranstaltung mögliche direkte Kontakt mit Lehrenden ist zudem wichtig für die Motivation und Zufriedenheit der Studierenden mit der Veranstaltung (Boelens et al., 2017; Fryer & Bovee, 2016; Poelmans & Wessa, 2015). Andererseits ist von Bedeutung, dass die eingesetzten Lernumgebungen nicht nur den aktuellen digitalen Standards entsprechen, sondern auch nach didaktischen Leitlinien (z.B. Brar & van der Meij, 2017; Noetel et al., 2021) sinnvoll gestaltet sind (Rock et al., 2016). Beispielsweise kann das Material in Lernumgebungen ähnlich wie im IC der zweiten Untersuchung nach Lernzieltaxonomien (z.B. Dunham et al., 2015) strukturiert werden: Lerner mit wenig Vorwissen können so mit Übungsaufgaben beginnen, welche zunächst eine „Festigung“ der statistischen Grundlagen zum Ziel haben, während Lerner mit mehr Wissen Übungsaufgaben wählen können, deren Ziel die Auswahl des richtigen statistischen Verfahrens für eine Fragestellung ist. Solche strukturierten Lernumgebungen können nicht nur für die Individualisierung des Lernens bei Statistikanfängern, sondern auch bei Fortgeschrittenen eingesetzt werden.

Beispielsweise wurde in einer Untersuchung von Stock und Hiemisch (2016) ein Kurs mit einer solchen individualisierten Lernumgebung mit einem wenig individualisierten textbasierten Kurs für ein Seminar zu weiterführender Statistik bei Psychologiestudierenden verglichen. Im textbasierten Kurs lasen alle Studierenden einen Auszug aus einem Statistikkbuch und diskutierten anschließend gemeinsam die Konzepte. Im Kurs mit der Lernumgebung konnten die Studierenden frei zwischen Übungsaufgaben mit unterschiedlichen, aufeinander aufbauenden Lernzielen wählen (z.B. Wiederholungsfragen zu den zentralen statistischen Konzepten, Übungsaufgaben zur Interpretation von statistischen Ergebnissen und Anwendungsaufgaben zur Arbeit an realen Forschungsfragen mit SPSS), diese gemeinsam bearbeiten und sich bei Fragen an die Lehrperson wenden. Es zeigte sich, dass die Studierenden im individualisierten Kurs während des Kurses eine geringere kognitive Belastung, sowie ein höheres Erleben von affektivem Interesse und ein höheres Kompetenzerleben aufwiesen. Zudem war das Wissen und die statistische Kompetenz in anwendungsbezogenen Aufgaben zum Ende des Kurses bei diesen Studierenden höher. Ähnlich wie im IC scheint also auch hier die individuelle Anpassung der Anforderungen an die Lerner Kompetenzüberzeugungen und Kompetenzzuwachs zu unterstützen. Es ist zudem anzumerken, dass die Studierenden in der Untersuchung von Stock und Hiemisch (2016) alle eine ähnliche statistische Grundlagenausbildung hatten. Dennoch scheint es Unterschiede in den kognitiven und motivationalen Voraussetzungen der Lerner gegeben zu haben, für die der individualisierte Statistikkurs besser geeignet war.

Studierende müssen ihre Statistikausbildung also nicht voller Angst und statistischer Fehlkonzepte abschließen. Vor allem durch den sinnvollen Einsatz von E-Learning können Statistikkurse so gestaltet werden, dass das Verständnis und die Anwendbarkeit von Statistik gefördert wird und Studierende Vertrauen in ihre statistische Kompetenz entwickeln. Dies gilt nicht nur für Statistikgrundlagenkurse zu Beginn des Studiums, sondern auch für spätere Kurse. Die Anforderungen an die Studierenden werden dort zunehmend komplexer. Zudem ist davon auszugehen, dass in Kursen im späteren Studienverlauf (z.B. Kursen für Promotionsstudierende), die Unterschiede in statistischem Vorwissen und Motivation noch stärker ausgeprägt sind, als bei Statistikanfängern. Zusätzliche individuelle E-Learning Angebote (z.B. Lernumgebungen mit strukturiertem Material für Anfänger und Fortgeschrittene) können helfen, mit der Heterogenität in solchen Kursen besser umzugehen und allen Studierenden einen Kompetenzzuwachs zu ermöglichen.

## LITERATUR

- Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and call for research. *Higher Education Research & Development, 34*(1), 1–14. <https://doi.org/10.1080/07294360.2014.934336>
- Adedokun, O.A., Bessenbacher, A.B., Parker, L.C., Kirkham, L.L., & Burgess, W.D. (2013). Research skills and STEM undergraduate research students' aspirations for research careers: Mediating effects of research self-efficacy. *Journal of Research and Science Teaching, 50*(8), 940–951. <https://doi.org/10.1002/tea.21102>
- Aguinis, H., & Branstetter, S.A. (2007). Teaching the concept of the sampling distribution of the mean. *Journal of Management Education, 31*(4), 467–483. <https://doi.org/10.1177/1052562906290211>
- Alao, S., & Guthrie, J.T. (1999). Predicting conceptual understanding with cognitive and motivational variables. *Journal of Educational Research, 92*(4), 243–253. <https://doi.org/10.1080/00220679909597602>
- Ali, A.Z., & Iqbal, F. (2012). Statistics anxiety among psychology graduates. An analysis. *International Proceedings of Economics, Development and Research, 53*(25), 113–117.
- Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., and Wittrock, M.C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman.
- Arens, A.K., Marsh, H.W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effect across five waves and three achievement tracks. *Journal of Educational Psychology, 109*(5), 621–634. <https://doi.org/10.1037/edu0000163>
- Ashcraft, M.H., & Kirk, E.P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology, 130*(2), 224–237. <https://doi.org/10.1037/0096-3445.130.2.224>
- Atkinson, R. K., Catrambone, R., & Merrill, M. M. (2003). Aiding Transfer in Statistics: Examining the Use of Conceptually Oriented Equations and Elaborations During Subgoal Learning. *Journal of Educational Psychology, 95*(4), 762–773. <https://doi.org/10.1037/0022-0663.95.4.762>
- Austerschmidt, K. L., & Bebermeier, S. (2019). Flexible Unterstützungsangebote in Statistik: Implementation und Effekte auf Studienerfolg. *Zeitschrift für Hochschulentwicklung, 14*(3), 137–155.
- Baartman, L.K.J., Bastiaens, T.J., Kirschner, P.A., & van der Vleuten, C.P.M. (2007). Evaluating quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review, 2*(2), 114–129. <https://doi.org/10.1016/j.edurev.2007.06.001>
- Bachiochi, P., Everton, W., Evans, M., Fugere, M., Escoto, C., Letterman, M. & Leszcynski, J. (2011). Using empirical article analysis to assess research methods courses. *Teaching of Psychology, 38*(1), 5–9. <https://doi.org/10.1177/0098628310387787>
- Baloğlu, M. (2004). Statistics anxiety and mathematics anxiety: Some interesting differences I. *Educational Research Quarterly, 27*(3), 38–48. <https://search.ebscohost.com/login.aspx?direct=true&db=pbh&AN=37376471&site=ehost-live>
- Baltes-Götz, B. (2018). *Mediator- und Moderatoranalyse mit SPSS und PROCESS* [Unveröffentlichtes Manuskript]. Zentrum für Informations-, Medien- und Kommunikationstechnologie (ZIMK). Universität Trier. <https://www.uni-trier.de/universitaet/wichtige-anlaufstellen/zimk/downloads-broschueren/statistik/mediator-und-moderatoranalyse-mit-spss-und-process>
- Bandalos, D.L., Finney, S.J., & Geske, J.A. (2003). A model of statistics performance based on achievement goal theory. *Journal of Educational Psychology, 95*(3), 604–616. <https://doi.org/10.1037/0022-0663.95.3.604>
- Bandalos, D.L., Yates, K., & Thorndike-Christ, T. (1995). Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety. *Journal of Educational Psychology, 87*(4), 611–623. <https://doi.org/10.1037/0022-0663.87.4.611>
- Bandura, A. (1977). Self-Efficacy. Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1986). *Social foundations of thought and action. A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

- Bandura, A. (1994). Self-efficacy. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71-81). New York: Academic Press.
- Bandura, A., & Schunk, D.H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, *41*(3), 586–598. <https://doi.org/10.1037/0022-3514.41.3.586>
- Barron, K.E., & Apple, K.J. (2014). Debating curricular strategies for teaching statistics and research methods: What does current evidence suggest? *Teaching of Psychology*, *41*(3), 187–194. <https://doi.org/10.1177/0098628314537967>
- Benson, A., & Blackman, D. (2003). Can research methods ever be interesting? *Active Learning in Higher Education*, *4*(1), 39–55. <https://doi.org/10.1177/1469787403004001004>
- Berney, S., & Bétrancourt, M., (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, *101*, 150–167. <https://doi.org/10.1016/j.compedu.2016.06.005>
- Bi, Y., Li, P., Liu, S., & Zhang, B. (2017). The enhancing effect of overconfidence on subsequent learning performance: Potential mediation effects of dissatisfaction and effort. *Journal of Behavioral Decision Making*, *30*(2), 555–568. <https://doi.org/10.1002/bdm.1974>
- Bierer, S. B., Prayson, R. A., & Dannefer, E. F. (2015). Association of research self-efficacy with medical student career interests, specialization, and scholarship: a case study. *Advances in Health Sciences Education*, *20*(2), 339–354. <https://doi.org/10.1007/s10459-014-9531-7>
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246 – 263. <https://doi.org/10.1111/j.1467-8624.2007.00995.x>
- Blanch-Hartigan, D. (2011). Medical students’ self-assessment of performance. Results from three meta-analyses. *Patient Education and Counseling*, *84*(1), 3–9. <https://doi.org/10.1016/j.pec.2010.06.037>
- Blanco, A. (2011). Applying social cognitive career theory to predict interests and choice goals in statistics among Spanish psychology students. *Journal of Vocational Behavior*, *78*(1), 49–58. <https://doi.org/10.1016/j.jvb.2010.07.003>
- Blau, I., & Shamir-Inbal, T. (2017). Re-designed flipped learning model in an academic course: The role of co-creation and co-regulation. *Computers & Education*, *69*–81. <https://doi.org/10.1016/j.compedu.2017.07.014>
- Blömeke, S. (2013). *Validierung als Aufgabe im Forschungsprogramm „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“* (KoKoHs Working Papers, 2). Berlin & Mainz: HumboldtUniversität & Johannes Gutenberg-Universität. [https://www.kompetenzen-im-hochschulsektor.de/files/2018/05/KoKoHs\\_WP\\_02\\_Bloemeke\\_2013.pdf](https://www.kompetenzen-im-hochschulsektor.de/files/2018/05/KoKoHs_WP_02_Bloemeke_2013.pdf)
- Boekharts, M., & Rozendaal, J.S. (2011). Using multiple calibration indices in order to capture the complex picture of what affects students’ accuracy of feeling of confidence. *Learning and Instruction*, *20*(5), 372–382. <https://doi.org/10.1016/j.learninstruc.2009.03.002>
- Boelens, R., De Wever, B., & Voet, M. (2017). Four key challenges on the design of blended learning: A systematic literature review. *Educational Research Review*, *22*, 1–18. <https://doi.org/10.1016/j.edurev.2017.06.001>
- Böttcher, F., & Thiel, F. (2017). Evaluating research-oriented teaching: a new instrument to assess university students’ research competences. *Higher Education*, *75*, 91–110. <https://doi.org/10.1007/s10734-017-0128-y>
- Bol, L., & Hacker, J.D. (2012). Calibration research. Where do we go from here. *Frontiers in Psychology*, *3*(Art.229). <https://doi.org/10.3389/fpsyg.2012.00229>
- Bong, M. (2002). Predictive utility of subject-, task-, and problem-specific self-efficacy judgements for immediate and delayed academic performances. *The Journal of Experimental Education*, *70*(2), 133–162. <https://doi.org/10.1080/00220970209599503>
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy. How different are they really. *Educational Psychology Review*, *15*(1), 1–40. <https://doi.org/10.1023/A:1021302408382>
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer.
- Bortz, J., & Döring, N. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin, Heidelberg: Springer.

- Bostwick, K. C., & Becker-Blease, K. A. (2018). Quick, easy mindset intervention can boost academic achievement in large introductory psychology classes. *Psychology Learning & Teaching, 17*(2), 177-193. <https://doi.org/10.1177/1475725718766426>
- Brar, J., & van der Meij, H. (2017). Complex software training: Harnessing and optimizing video instruction. *Computers in Human Behavior, 70*, 475–485. <https://doi.org/10.1016/j.chb.2017.01.014>
- Braten, I., & Olaussen, B. S. (1998). The relationship between motivational beliefs and learning strategy use among Norwegian college students. *Contemporary Educational Psychology, 23*(2), 182–194. <https://doi.org/10.1006/ceps.1997.0963>
- Brewer, R., & Movahedazarhouli, S. (2018). Successful stories and conflicts: A literature review on the effectiveness of flipped learning in higher education. *Journal of Computer Assisted Learning, 34*(4), 409–416. <https://doi.org/10.1111/jcal.12250>
- Broers, N.J. (2001). Analyzing propositions underlying the theory of statistics. *Journal of Statistics Education, 9*(3). <https://doi.org/10.1080/10691898.2001.11910536>
- Broers, N.J. (2002). Selection and use of propositional knowledge in statistical problem solving. *Learning and Instruction, 12*(3), 323–344. [https://doi.org/10.1016/S0959-4752\(01\)00025-1](https://doi.org/10.1016/S0959-4752(01)00025-1)
- Broers, N. (2009). Using propositions for the assessment of structural knowledge. *Journal of Statistics Education, 17*(2). <https://doi.org/10.1080/10691898.2009.11889513>
- Buckelew, S. P., Byrd, N., Key, C. W., Thornton, J., & Merwin, M. M. (2013). Illusions of a good grade: Effort or luck? *Teaching of Psychology, 40*(2), 134–138. <https://doi.org/10.1177/0098628312475034>
- Bünger, U. (2017). *Validierung eines Fragebogens zur Erfassung der statistischen Selbstwirksamkeit*. [Unveröffentlichte Diplomarbeit]. Universität Greifswald.
- Burgoynes, S., & Eaton, J. (2018). The partially flipped classroom: The effects of flipping a module on “junk science” in a large methods course. *Teaching of Psychology, 45*(2), 154–157. <https://doi.org/10.1177/0098628318762894>
- Burnette, J. L., Pollack, J. M., Forsyth, R. B., Hoyt, C. L., Babij, A. D., Thomas, F. N., & Coy, A. E. (2020). A growth mindset intervention: Enhancing students’ entrepreneurial self-efficacy and career development. *Entrepreneurship Theory and Practice, 44*(5), 878 – 908. <https://doi.org/10.1177/1042258719864293>
- Cerbin, W. (2018). Improving student learning from lectures. *Scholarship of Teaching and Learning in Psychology, 4*(3), 151–163. <https://doi.org/10.1037/stl0000113>
- Chen, F., Lu, A.M., & Martinell, S.M. (2017). A systematic review of the effectiveness of flipped classrooms in medical education. *Medical Education, 51*, 585 – 597. <https://doi.org/10.1111/medu.13272>
- Chen, J.A., & Tutwiler, M.S. (2017). Implicit theories of ability and self-efficacy. Testing alternative social cognitive models to science motivation. *Zeitschrift für Psychologie, 225*(2), 127–136. <https://doi.org/10.1027/2151-2604/a000289>
- Chen, P.P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*(1), 79–92. <https://doi.org/10.1016/j.lindif.2003.08.003>
- Chen, S.-C., Yang, J.H., & Hsiao, C.-C. (2016). Exploring student perceptions, learning outcome and gender differences in a flipped mathematics course. *British Journal of Educational Technology, 47*(6), 1096–1112. <https://doi.org/10.1111/bjet.12278>
- Chen, Y., Wang, Y., & Chen, N.-S. (2014). Is FLIP enough? Or should we use the FLIPPED model instead? *Computers & Education, 79*, 16 – 27. <https://doi.org/10.1016/j.compedu.2014.07.004>
- Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students’ statistics achievement. *Statistics Education Research Journal, 9*(1), 6–26. <https://doi.org/10.52041/serj.v9i1.385>
- Chili-Turner, E. (2015). Measuring learning outcomes and attitudes in a flipped introductory statistics course. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies, 25*(9-10), 833–846. [10.1080/10511970.2015.1046004](https://doi.org/10.1080/10511970.2015.1046004)
- Choi, N. (2005). Self-efficacy and self-concept as predictors of college students’ academic performance. *Psychology in the Schools, 42*(2), 197–205. <https://doi.org/10.1002/pits.20048>

- Ciani, K.D., Summers, J.J., Easter, M.A., & Sheldon, K.M. (2008). Collaborative learning and positive experiences: does letting students choose their own groups matter? *An International Journal of Experimental Educational Psychology*, 28(6), 627–641. <https://doi.org/10.1080/01443410802084792>
- Ciarocco, N. J., Lewandowski Jr, G. W., & Van Volkom, M. (2013). The impact of a multifaceted approach to teaching research methods on students' attitudes. *Teaching of Psychology*, 40(1), 20–25. <https://doi.org/10.1177/0098628312465859>
- Clark, J.M., Kraut, G., Mathews, D., & Wimbish, J. (2007). The „fundamental“ theorem of statistics: Classifying student understanding of basic statistical concepts [Unpublished Manuscript]. [https://moam.info/fundamental-theorem-of-statistics-classifying-student-citeseerx\\_5a2b5eac1723ddd7574195e2.html](https://moam.info/fundamental-theorem-of-statistics-classifying-student-citeseerx_5a2b5eac1723ddd7574195e2.html)
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York: Routledge.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Condron, D.J., Becker, J.H., & Bzhetaj, L. (2018). Sources of students' anxiety in a multidisciplinary social statistics course. *Teaching Sociology*, 46(4), 346 – 355. <https://doi.org/10.1177/0092055X18780501>
- Connors, F.A., McCown, S.M., & Roskos-Ewoldson, B. (1998). Unique challenges in undergraduate statistics. *Teaching of Psychology*, 25(1), 40–42. [https://doi.org/10.1207/s15328023top2501\\_12](https://doi.org/10.1207/s15328023top2501_12)
- Cooper, L.L., & Shore, F.S. (2008). Students' misconceptions in interpreting center and variability of data representations via histograms and stem-and-leaf plots. *Journal of Statistics Education*, 16(2). <https://doi.org/10.1080/10691898.2008.11889559>
- Cruise, R., Cash, R., & Bolton, D. (1985). Development and Validation of an Instrument to Measure Statistical Anxiety. *Paper presented at the annual meeting of the American Statistical Association Statistics Education Section*. Las Vegas, NV.
- Cury, F., Elliott, A.J., Fonseca, D.D., & Moller, A.C. (2006). The social-cognitive model of achievement motivation and the 2x2 achievement goal framework. *Personality Processes and Individual Differences*, 90(4), 666–679. <https://doi.org/10.1037/0022-3514.90.4.666>
- Dahling, J.J., & Ruppel, C.L. (2016). Learning goal orientation buffers the effects of negative normative feedback on test self-efficacy and reattempt interest. *Learning and Instruction*, 50, 296–301. <https://doi.org/10.1016/j.lindif.2016.08.022>
- Daniel, F., & Braasch, J. L. (2013). Application exercises improve transfer of statistical knowledge in real-world situations. *Teaching of Psychology*, 40(3), 200–207. <https://doi.org/10.1177/0098628313487462>
- Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
- DeGrave, W.S., Dolmans, D.H., & van der Vleuten (1999). Profiles of effective tutors in problem-based learning: scaffolding student learning. *Medical Education*, 33(12), 901–906. <https://doi.org/10.1046/j.1365-2923.1999.00492.x>
- DelMas, R., Garfield, J., Ooms, A., & Chance, B. (2006). Assessing students' conceptual understanding after a first course in statistics. [Conference Paper]. Annual Meetings of The American Educational Research Association. San Francisco.
- DelMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58. <https://doi.org/10.52041/serj.v6i2.483>
- DeLozier, S.J., & Rhodes, M.G. (2017). Flipped classroom: a review of key ideas and recommendations for practice. *Educational Psychology Review*, 29(1), 141–151. <https://doi.org/10.1007/s10648-015-9356-9>
- Dimotakis, N., Mitchell, D., & Maurer, T. (2017). Positive and negative feedback in relation to development self-efficacy, feedback seeking and promotion. *Journal of Applied Psychology*, 102, 1514–1527. <https://doi.org/10.1037/apl0000228>

- Diseth, A. (2011). Self-efficacy, goal orientations and learning strategies as mediators between preceding and subsequent academic achievement. *Learning and Individual Differences, 21*(2), 191–195. <https://doi.org/10.1016/j.lindif.2011.01.003>
- Diseth, A., Breidablik, H. J., & Meland, E. (2018). Longitudinal relations between perceived autonomy support and basic need satisfaction in two student cohorts. *Educational Psychology, 38*(1), 99–115. <https://doi.org/10.1080/01443410.2017.1356448>
- Du, C., Qin, K., Wang, Y. & Xin, T. (2021). Mathematics interest, anxiety, self-efficacy and achievement. Examining reciprocal relations. *Learning and Individual Differences, 91*. <https://doi.org/10.1016/j.lindif.2021.102060>
- Duncan, T.G., & McKeachie, W.J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist, 40*(2), 117–128. [https://doi.org/10.1207/s15326985ep4002\\_6](https://doi.org/10.1207/s15326985ep4002_6)
- Dunham, B., Yapa, G., & Yu, E. (2015). Calibrating the difficulty of an assessment tool: The blooming of a statistics exam. *Journal of Statistics Education, 23*(3). <https://doi.org/10.1080/10691898.2015.11889745>
- Dunning, D., Heath, C. & Suls, J. (2004). Flawed self-assessment. Implications for health, education and the workplace. *Psychological Science in the Public Interest, 5*(3), 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Durik, A.M., & Matarazzo, K.L. (2008). Revved up or turned off? How domain knowledge changes the relations of perceived task complexity and task interest. *Learning and Individual Differences, 19*(1), 155–159. <https://doi.org/10.1016/j.lindif.2008.08.005>
- Dweck, C. (1999). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia: Psychology Press.
- Dweck, C.S. (2006). *Mindset: The new psychology of success*. New York: Random House.
- Dweck, C.S., Chiu, C.Y., & Hong, Y.Y. (1995). Implicit theories and their role in judgements and reactions. A world from two perspectives. *Psychological Inquiry, 6*(4), 267–285. [https://doi.org/10.1207/s15327965pli0604\\_1](https://doi.org/10.1207/s15327965pli0604_1)
- Dweck, C.S., & Leggett, E.L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273. <https://doi.org/10.1037/0033-295X.95.2.256>
- Dweck, C.S., & Molden, D.C. (2017). Mindsets. Their impact on competence motivation and acquisition. In A.J. Elliot, C.S. Dweck & D.S. Yeager (Eds.), *Handbook of competence and motivation. Theory and application* (2nd ed., pp. 135–154). New York: The Guilford Press.
- Dykeman, B.F. (2011). Statistics anxiety: Antecedents and instructional interventions. *Education, 132*(2), 441–446.
- Ehrlinger, J., Mitchum, A.L., & Dweck, C.S. (2016). Understanding overconfidence. Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology, 63*, 94–100. <https://doi.org/10.1016/j.jesp.2015.11.001>
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden*. Weinheim, Basel: Beltz.
- Elliott, E.S., & Dweck, C.S. (1998). Goals. An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(5), 5–12. <https://doi.org/10.1037//0022-3514.54.1.5>
- Elliott, A.J., Dweck, C.S. & Yeager, D.S. (2017). *Handbook of competence and motivation. Theory and application* (2nd ed.). New York: The Guilford Press.
- Emmioğlu, E. & Capa-Aydin, Y. (2011). A Meta-Analysis on Students' Attitudes toward Statistics. In *Proceedings of the 58th World Statistical Congress* (pp. 2795 – 2801). Dublin. <https://2011.isiproceedings.org/papers/650325.pdf>
- Estepa, A., & Batanero, C. (1996). Judgments of correlation in scatterplots: Students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education, 4*(25), 25–41.
- Fan, J., Meng, H., Billings, R.S., Litchfield, R.C., & Kaplan, I. (2008). On the role of goal orientation traits and self-efficacy in the goal-setting process. Distinctions that make a difference. *Human Performance, 21*(4), 354–382. <https://doi.org/10.1080/08959280802347122>
- Feldon, D.F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review, 31*, 319–337. <https://doi.org/10.1007/s10648-019-09464-6>



- Ferla, J. Valcke, M. & Cai, Y. (2009). Academic self-efficacy and academic self-concept: reconsidering structural relationships. *Learning and Individual Differences, 19*(4), 499–505. <https://doi.org/10.1016/j.lindif.2009.05.004>
- Finney, S.J. & Shraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*(2), 161–186. [https://doi.org/10.1016/S0361-476X\(02\)00015-2](https://doi.org/10.1016/S0361-476X(02)00015-2)
- Fong G. T., Krantz D. H., Nisbett R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*(3), 253–292. [https://doi.org/10.1016/0010-0285\(86\)90001-0](https://doi.org/10.1016/0010-0285(86)90001-0)
- Forester, M., Kahn, J., H. & Hesson-McInnis, M.S. (2004). Factor structure of three measures of research self-efficacy. *Journal of Career Assessment, 12*(1), 3–16. <https://doi.org/10.1177/1069072703257719>
- Foster, N.L., Was, C.A., Dunlowksy, J., & Isaacson, R.M. (2017). Even after thirteen class exams students are still overconfident: the role of memory for past exam performance in student predictions. *Metacognition Learning, 12*, 1–19. <https://doi.org/10.1007/s11409-016-9158-6>
- Frenzel, A. C., Götz, T. & Pekrun, R. (2020). Emotionen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 211-234). Berlin, Heidelberg: Springer.
- Fryer, L. K., & Bovee, H. N. (2016). Supporting students' motivation for e-learning: Teachers matter on and offline. *The Internet and Higher Education, 30*, 21–29. <https://doi.org/10.1016/j.iheduc.2016.03.003>
- Fryer, L.K., Bovee, B.H. & Nakao, K. (2022). Self-efficacy latent growth trajectories' longitudinal links with achievement and interest. Both baseline and growth rate are important for interest outcomes. *British Journal of Educational Psychology, 92*(2). <https://doi.org/10.1111/bjep.12473>
- Fulmer, S.M., & Tulis, M. (2016). Adding nuance to the challenge-skill relationship: The interaction of perceived and actual skill. *International Journal of Educational Research, 77*, 143–154. <https://doi.org/10.1016/j.ijer.2016.04.002>
- Fung, C.-H., Besser, M., & Poon, K.-K. (2021). Systematic literature review of flipped classroom in mathematics. *EURASIA Journal of Mathematics, Science and Technology Education, 17*. <https://doi.org/10.29333/ejmste/10900>
- Gal, I. (2002). Adults' statistical literacy. Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–51. <https://doi.org/10.1111/j.1751-5823.2002.tb00336.x>
- Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics. Towards an assessment framework. *Journal of Statistics Education, 2*(2), 1–15. <https://doi.org/10.1080/10691898.1994.11910471>
- Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In Gal, I. & Garfield, J. B. (Eds.). *The Assessment Challenge in Statistics Education* (pp. 37 – 51). Washington, D.C.: IOS Press.
- Galligan, L. (2013). A systematic approach to embedding academic numeracy at university. *Higher Education Research & Development, 32*(5), 734–747. <https://doi.org/10.1080/07294360.2013.777037>
- Garfield, J. B. (1998). The Statistical Reasoning Assessment: Development and Validation of a Research Tool. In *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 781-786). Singapore. <https://iase-web.org/documents/papers/icots5/Topic6u.pdf?1402524957>
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3). <https://doi.org/10.1080/10691898.2002.11910676>
- Garfield, J. B., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning, 2*(1-2), 99–125. [https://doi.org/10.1207/S15327833MTL0202\\_5](https://doi.org/10.1207/S15327833MTL0202_5)
- Gess, C., Geiger, C., & Ziegler, M. (2018). Social-scientific research competency: Validation of test score interpretations for evaluative purposes in higher education. *European Journal of Psychological Assessment, 35*(5), 737–750. <https://doi.org/10.1027/1015-5759/a000451>
- Gier, V. S., & Kreiner, D. S. (2009). Incorporating active learning with PowerPoint-based lectures using content-based questions. *Teaching of Psychology, 36*(2), 134–139. <https://doi.org/10.1080/00986280902739792>
- Giesbrecht, N., Sell, Y., Scialfa, C., Sandals, L. & Ehlers, P. (1997). Essential Topics in Introductory Statistics and Methodology Courses. *Teaching of Psychology, 24*(4), 242–246. [https://doi.org/10.1207/s15328023top2404\\_2](https://doi.org/10.1207/s15328023top2404_2)

- Giins, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313–331. <https://doi.org/10.1016/j.learninstruc.2005.07.001>
- González, A., Rodríguez, Y., Faílde, J.M. & Carrera, M.V. (2016). Anxiety in the statistics class: Structural relations with self-concept, intrinsic value, and engagement in two samples of undergraduates. *Learning and Individual Differences, 45*, 214–221. <https://doi.org/10.1016/j.lindif.2015.12.019>
- Goode, C. T., Lamoreaux, M., Atchison, K. J., Jeffress, E. C., Lynch, H. L., & Sheehan, E. (2018). Quantitative skills, critical thinking, and writing mechanics in blended versus face-to-face versions of a research methods and statistics course. *Teaching of Psychology, 45*(2), 124–131. <https://doi.org/10.1177/0098628318762873>
- Graham, C.R. (2006). Blended Learning Systems. Definition, Current Trends, and Future Directions. In C.J. Bonk & C.R. Graham (Eds.), *Handbook of Blended Learning. Global Perspectives, Local Designs* (pp. 3–21). San Francisco: Pfeiffer Publishing.
- Griffith, J.D., Adams, L.T., Gu, L.L., Hart, C.L., & Nichols-Whitehead, P. (2012). Students' attitudes toward statistics across the disciplines: A mixed-methods approach. *Statistics Education Research Journal, 11*(2), 45–56. <https://doi.org/10.52041/serj.v11i2.328>
- Groß Ophoff, J., Wolf, R., Schladitz, S., & Wirtz, M. (2017). Assessment of Educational Research Literacy in Higher Education: Construct validation of the factorial structure of an assessment instrument comparing different treatments of omitted responses. *Journal of Educational Research Online, 9*(2), 37–68. <https://doi.org/10.25656/01:14896>
- Gulikers, J. T. M., Bastiaens, Th. J. & Kirschner, P. A. (2007). *Defining authentic assessment: five dimensions of authenticity*. In A. Havnes & L. McDowell (Eds.). *Balancing dilemmas in assessment and learning in contemporary education* (pp. 73–86). New York: Routledge.
- Hacker, D.J., & Bol, L. (2019). Calibration and self-regulated learning. Making the connection. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 647–677). Cambridge: Cambridge University Press.
- Hacker, D.J., Bol, L., Horgan, D., & Rakow, E.A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160–170. <https://doi.org/10.1037/0022-0663.92.1.160>
- Hadwin, A.F., & Webster, E.A. (2013). Calibration and goal setting: Examining the nature of judgements of confidence. *Learning and Instruction, 24*, 37 – 47. <https://doi.org/10.1016/j.learninstruc.2012.10.001>
- Hall, S.R., Stephens, J.R., Seaby, E.G., Gesteira Andrade, M., Lowry, A.F., Parton, W.J.C., Smith, C.F. & Border, S. (2010). Can medical students accurately predict their learning? A study comparing perceived and actual performance in neuroanatomy. *Anatomical Sciences Education, 9*(5), 488–495. <https://doi.org/10.1002/ase.1601>
- Haller, H. & Kraus, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research, 7*(1), 1–20.
- Harackiewicz, J.M., & Knogler, M. (2017). Interest: Theory and Application. In A.J. Elliot, C.S. Dweck & D.S. Yeager (Eds.), *Handbook of competence and motivation. Theory and application (2nd ed., pp. 335–352.)*. New York: The Guilford Press.
- Harris, A.H.S., Reeder, R., & Hyun, J.K. (2009). Common statistical research design problems in manuscripts submitted to high-impact psychiatry journals: What editors and reviewers want authors to know. *Journal of Psychiatric Research, 43*(15), 1231–1234. <https://doi.org/10.1016/j.jpsychires.2009.04.007>
- Hatfield, J., Faunce, G. J., & Job, R. F. S. (2006). Avoiding confusion surrounding the phrase “correlation does not imply causation.”. *Teaching of Psychology, 33*(1), 49–51.
- Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction, 24*, 62–66. <https://doi.org/10.1016/j.learninstruc.2012.05.009>
- Haughton, J., & Kelly, A. (2015). Student performance in an introductory business statistics course: Does delivery mode matter? *Journal of Education for Business, 9*(1), 31–43. <https://doi.org/10.1080/08832323.2014.968518>
- Hawker, M.J., Dysleski, L., & Rickey, D. (2016). Investigation general chemistry students metacognitive monitoring of their exam performance by measuring postdiction accuracies over time. *Journal of Chemical Education, 93*, 832 – 846. <https://doi.org/10.1021/acs.jchemed.5b00705>

- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach* (2nd ed.). New York: The Guilford Press.
- Hidi, S., & Renninger, K.A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. [https://doi.org/10.1207/s15326985ep4102\\_4](https://doi.org/10.1207/s15326985ep4102_4)
- Hiemisch, A. (2012). The validity of self-assessments of competence in academic course evaluation. *Zeitschrift für Hochschulentwicklung*, 7, 60 – 70.
- Hong, E., & O'Neill, H.F.Jr. (1992). Instructional strategies to help learners build relevant mental models in inferential statistics. *Journal of Educational Psychology*, 84(2), 150–159. <https://doi.org/10.1037/0022-0663.84.2.150>
- Hong, Y.Y., Chiu, C.Y., Dweck, C.S., Lin, M.S., & Wang, W. (1999). Implicit theories, attributions, and coping. A meaning system approach. *Journal of Personality and Social Psychology*, 77(3), 588–599. <https://doi.org/10.1037/0022-3514.77.3.588>
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49(5), 505–528. <https://doi.org/10.1016/j.jsp.2011.07.001>
- Huang, X., Zhang, J., & Hudson, L. (2019). Impact of math self-efficacy, math anxiety, and growth mindset on math and science career interest for middle school students: The gender moderating effect. *European Journal of Psychology of Education*, 34(3), 621–640. <https://doi.org/10.1007/s10212-018-0403-z>
- Huberty, C.J., Dresden, J., & Bak, B-G. (1993). Relations among dimensions of statistical knowledge. *Educational and Psychological Measurement*, 53(2), 523–532. <https://doi.org/10.1177/0013164493053002022>
- Huck, S.W. (2009). *Statistical misconceptions*. London: Routledge.
- Hughes, A., Galbraith, D., & White, D. (2011). Perceived competence: A common core for self-efficacy and self-concept? *Journal of Personality Assessment*, 93(3), 278–289. <https://doi.org/10.1080/00223891.2011.559390>
- Jansen, M., Lüdkte, O., & Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology*, 46, 116–127. <https://doi.org/10.1016/j.cedpsych.2016.05.004>
- Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences. Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, 41, 13–24. <https://doi.org/10.1016/j.cedpsych.2014.11.002>
- Janssen, J., Kirschner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load approaches. *Educational Psychology Review*, 22(2), 139–154. <https://doi.org/10.1007/s10648-010-9131-x>
- Jensen, J.L., Holt, E.A., Sowards, J.B., & West, R.E. (2018). Investigating strategies for pre-class content learning in a flipped classroom. *Journal of Science Education and Technology*, 27, 523–535. <https://doi.org/10.1007/s10956-018-9740-6>
- Jensen, J.L., Kummer, T.A., & Godoy, P.D. (2015). Improvements from a flipped classroom may simply be the fruits of active Learning. *Life Sciences Education*, 14(1), 1–12. <https://doi.org/10.1187/cbe.14-08-0129>
- Kalyuaga, S. (2017). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19, 509–539. <https://doi.org/10.1007/s10648-007-9054-3>
- Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, 96(3), 558 – 568. <https://doi.org/10.1037/0022-0663.96.3.558>
- Karpiak, C.P. (2011). Assessment of problem-based learning in the undergraduate statistics course. *Teaching of Psychology*, 38(4), 251–254. <https://doi.org/10.1177/0098628311421322>
- Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, 28(3), 820–831. <https://doi.org/10.1016/j.chb.2012.01.011>
- Keren, G. (1991). Calibration and probability judgments. Conceptual and methodological issues. *Acta Psychologica*, 77(3), 217–273. [https://doi.org/10.1016/0001-6918\(91\)90036-Y](https://doi.org/10.1016/0001-6918(91)90036-Y)

- Kim, M.K., Kim, S.M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: an exploration of design principles. *Internet and Higher Education*, 22, 37–50. <https://doi.org/10.1016/j.iheduc.2014.04.003>
- Kintsch, W. (1998). The representation of knowledge in minds and machines. *International Journal of Psychology*, 33(6), 411–420. <https://doi.org/10.1080/002075998400169>
- Kirschner, F., Paas, F., & Kirschner, P.A. (2009a). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21, 31–24. <https://doi.org/10.1007/s10648-008-9095-2>
- Kirschner, F., Paas, F., & Kirschner, P.A. (2009b). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, 25(2), 306–314. <https://doi.org/10.1016/j.chb.2008.12.008>
- Kirschner, F., Paas, F., & Kirschner, P.A. (2011). Task complexity as a driver for collaborative learning efficiency: the collective working-memory effect. *Applied Cognitive Psychology*, 25(4), 615–624. <https://doi.org/10.1002/acp.1730>
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903. <https://doi.org/10.25656/01:4493>
- Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, 32(5), 448–470. <https://doi.org/10.2307/749801>
- Koepfen, K., Harting, J., Klieme, E. & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie*, 216(2), 61–73. <https://doi.org/10.1027/0044-3409.216.2.61>
- Kopp, B. (2020). Lernen, selbstgesteuertes Lernen. In M.A. Wirtz (Hrsg.), *Dorsch- Lexikon der Psychologie*. <https://m.portal.hogrefe.com/dorsch/lernen-selbstgesteuertes/>
- Krapp, A. (1999). Interest, motivation and learning. An educational-psychological perspective. *European Journal of Psychology and Education*, 14(1), 23–40. <https://doi.org/10.1007/BF03173109>
- Krapp, A., Hidi, S., & Renniger, K.A. (1992). Interest, learning and development. In K.A. Renniger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3 – 25), Hillsdale, NJ: Erlbaum.
- Kuo, F.-R., Hwang, G.-J., & Lee, C.-C. (2011). A hybrid approach to promoting students' web-based problem-solving competence and learning attitude. *Computers & Education*, 58(1), 351–364. <https://doi.org/10.1016/j.compedu.2011.09.020>
- Labuhn, A.S., Zimmerman, B.J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance. The influence of feedback and self-evaluative standards. *Metacognition and Learning*, 5, 173–194. <https://doi.org/10.1007/s11409-010-9056-2>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4(Art. 863). <https://doi.org/10.3389/fpsyg.2013.00863>
- Landmann, M., Perels, F., Otto, B., Schnick-Vollmer, K., & Schmitz, B. (2014). Selbstregulation und selbstreguliertes Lernen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 46–65). Berlin, Heidelberg: Springer.
- Landrum, R. E. (2005). Core terms in undergraduate statistics. *Teaching of Psychology*, 32(4), 249–251. [http://dx.doi.org/10.1207/s15328023top3204\\_8](http://dx.doi.org/10.1207/s15328023top3204_8)
- Lapka, D. (2010). „Methodische Entwicklungen zur Qualitätssicherung von Programm-Evaluationen am Beispiel eines Lernprogramms im Hochschulbereich: Vienna E-Lecturing“. [Dissertation, Universität Wien]. Hochschulschriften-Repository der Universität Wien. <http://dx.doi.org/10.25365/thesis.13858>
- Lapka, D., Wagner, P., Schober, B., Grading, P., Reimann, R. & Spiel, C. (2010). Methodenlehre. Alptraum oder Herausforderung für Psychologiestudierende. Eine Typologie auf Basis des sozialkognitiven Motivationsmodells von Dweck. *Psychologie in Erziehung und Unterricht*, 57(3), 209–222. <https://doi.org/10.2378/peu2010.art15d>
- Lavigne, N.C., Salkind, S.J., & Yan, J. (2008). Exploring college students' mental representations of inferential statistics. *Journal of Mathematical Behavior*, 27(1), 11–32. <https://doi.org/10.1016/j.jmathb.2007.10.003>

- Lawson, T.J., Bodle, J.H., Houlette, M.A., & Haubner, R.R. (2006). Guiding questions enhance student learning from educational videos. *Teaching of Psychology, 33*(1), 31–33. [https://doi.org/10.1207/s15328023top3301\\_7](https://doi.org/10.1207/s15328023top3301_7)
- Lee, J., Lim, C., & Kim, H. (2017). Development of an instructional design model for flipped learning in higher education. *Education Technology Research and Development, 65*, 427–453. <https://doi.org/10.1007/s11423-016-9502-1>
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06). [Computer software]. <https://www.socisurvey.de>
- Lent, R.W., Brown, S.D. & Gore, P.A., Jr. (1997). Discriminant and predictive validity of academic self-concept, academic self-efficacy, and mathematics-specific self-efficacy. *Journal of Counseling Psychology, 44*(3), 307–315. <https://doi.org/10.1037/0022-0167.44.3.307>
- Lent, R. W., Brown, S. D., Gover, M. R., & Nijjer, S. K. (1996). Cognitive assessment of the sources of mathematics self-efficacy: A thought-listing analysis. *Journal of Career Assessment, 4*(1), 33–46. <https://doi.org/10.1177/106907279600400102>
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*(1), 79–122. <https://doi.org/10.1006/jvbe.1994.1027>
- Lent, R. W., Lopez, F. G., & Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of Counseling Psychology, 38*(4), 424–430. <https://doi.org/10.1037/0022-0167.38.4.424>
- Lent, R.W., Lopez, F.G., & Bieschke, K.J. (1993). Predicting mathematics-related choice and success behaviors. Test of an expanded social cognitive model. *Journal of Vocational Behavior, 42*(2), 223–236. <https://doi.org/10.1006/jvbe.1993.1016>
- Lichtenberg, J.W., Bebeau, M.J., Nelson, P.D., Smith, I.L., Portnoy, S.M., Leigh, I.W., Rubin, N.J., & Kaslow, N.J. (2007). Challenges to the assessment of competence and competencies. *Professional Psychology Research and Practice, 38*(5), 474–478. <https://doi.org/10.1037/0735-7028.38.5.474>
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: the state of the art to 1980. In D. Kahneman, P. Slavic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306 – 334). New York: Cambridge University Press.
- Liebert, C.A., Lin, D.T., Mazer, L.M., Bereckney, P.H., & Lau, J.N. (2016). Effectiveness of the surgery core clerkship flipped classroom: a prospective cohort trial. *The American Journal of Surgery, 211*(2), 451–457. <https://doi.org/10.1016/j.amjsurg.2015.10.004>
- Liem, A.D., Lau, S., & Nie, Y. (2008). The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome. *Contemporary Educational Psychology, 33*(4), 486–512. <https://doi.org/10.1016/j.cedpsych.2007.08.001>
- Linnenbrink-Garcia, L., Durik, A.M., Conley, A.M., Barron, K.E., Tauer, J.M., Karabenick, S.A. & Harackiewicz, J.M. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement, 70*, 647 – 671. <https://doi.org/10.1177/0013164409355699>
- Liu, T.C., Lin, Y.C., & Kinshuk (2010). The application of simulation-assisted learning statistics (SALS) for correcting misconceptions and improving understanding of correlation. *Journal of Computer Assisted Learning, 26*(2), 143–158. <https://doi.org/10.1111/j.1365-2729.2009.00330.x>
- Lou, Y., Abrami, P.C., & d’Appollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*(3), 449–521. <https://doi.org/10.3102/00346543071003449>
- Lou, Y., Abrami, P.C., Spence, J.C, Poulsen, C., Chambers, B., & d’Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423–458. <https://doi.org/10.3102/00346543066004423>
- Loux, T.M., Varner, S.E., & VanNatta, M. (2016). Flipping an introductory biostatistics course: A case study of student attitudes and confidence. *Journal of Statistics Education, 24*(1), 1–7. <https://doi.org/10.1080/10691898.2016.1158017>
- Lovett, M. (2001). A Collaborative Convergence on Studying Reasoning Processes: A Case Study in Statistics. In S.M. Carver & D. Klahr (Eds.), *Cognition and Instruction* (pp. 347 – 384). London: Routledge.

- Lüftenecker, M., & Chen, J.A. (2017). Conceptual issues and assessment of implicit theories. *Zeitschrift für Psychologie*, 225(2), 99–106. <https://doi.org/10.1027/2151-2604/a000286>
- Macher, D., Paechter, M., Papousek, I., & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European Journal of Psychology of Education*, 27(4), 483–498. <https://doi.org/10.1007/s10212-011-0090-5>
- Macher, D., Paechter, M., Papousek, I., Ruggeri, K., Freudenthaler, H. H., & Arendasy, M. (2013). Statistics anxiety, state anxiety during an examination, and academic achievement. *British Journal of Educational Psychology*, 83(4), 535 – 549. <https://doi.org/10.1111/j.2044-8279.2012.02081.x>
- Maciejewski, W. (2016). Flipping the calculus classroom: an evaluative study. *Teaching Mathematics and Its Applications*, 35(4), 187–201. <https://doi.org/10.1093/teamat/hrv019>
- Mallenby, D.W., & Mallenby, M.L. (2003). Use of brief collaborative quizzes on new quantitative lecture material. *Decision Sciences Journal of Innovative Education*, 1(1), 141–144. <https://doi.org/10.1111/1540-5915.00011>
- Margulieux, L.E., McCracken, W.M., & Catrambone, R. (2016). A taxonomy to define courses that mix face-to-face and online learning. *Educational Research Review*, 19, 104–118. <https://doi.org/10.1016/j.edurev.2016.07.001>
- Marsh, H.W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76(2), 397–416. <https://doi.org/10.1111/j.1467-8624.2005.00853.x>
- Marsh, H.W., Martin, A.J., Yeung, A.S., & Craven, R.G. (2017). Competence self-perceptions. In A.J. Elliot, C.S. Dweck & D.S. Yeager (Eds.), *Handbook of competence and motivation. Theory and application* (2nd ed., pp. 85–115). New York: The Guilford Press.
- Mattis, K.V. (2015). Flipped classroom versus traditional textbook instruction: Assessing accuracy and mental effort at different levels of mathematical complexity. *Technology, Knowledge and Learning*, 22, 231–248. <https://doi.org/10.1007/s10758-014-9238-0>
- Mayer, R.E. (2008). *Learning and instruction* (2nd ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Mayer, R.E., & Moreno, R. (2003). Nine way to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52. [https://doi.org/10.1207/S15326985EP3801\\_6](https://doi.org/10.1207/S15326985EP3801_6)
- McKenzie, W.A., Perini, E., Rohlf, V., Toukhsati, S., Conduit, R., & Sanson, G. (2013). A blended learning lecture delivery model for large and diverse undergraduate cohorts. *Computers & Education*, 64, 136–126. <https://doi.org/10.1016/j.compedu.2013.01.009>
- McLaughlin, J.E., & Kang, I. (2017). A flipped classroom model for a biostatistics short course. *Statistics Education Research Journal*, 16(2), 441–453. <https://doi.org/10.52041/serj.v16i2.200>
- Mills, J.D. (2004). Learning abstract statistics concepts using simulation. *Educational Research Quarterly*, 28(4), 18–33.
- Möller, J., & Trautwein, U. (2015). Selbstkonzept. In E. Wild (Hrsg.), *Pädagogische Psychologie* (S. 177–199). Berlin, Heidelberg: Springer.
- Möller, J., & Trautwein, U. (2020). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 187-209). Berlin, Heidelberg: Springer.
- Moores, T.T., & Chang, J. C.-J. (2009). Self-efficacy, overconfidence, and the negative effect on subsequent performance. A field study. *Information & Management*, 46(2), 69–76. <https://doi.org/10.1016/j.im.2008.11.006>
- Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309 –326. <https://doi.org/10.1007/s10648-007-9047-2>
- Mutz, R., & Daniel, H.-D. (2011). University and student segmentation: Multilevel latent-class analysis of students' attitudes towards research methods and statistics. *British Journal of Educational Psychology*, 83(2), 280–304. <https://doi.org/10.1111/j.2044-8279.2011.02062.x>
- Nielsen, P.L., Bean, N.W., & Larson, R.A.A. (2018). The impact of a flipped classroom model of learning on a large undergraduate statistics class. *Statistics Education Research Journal*, 17(1), 121–140. <https://doi.org/10.52041/serj.v17i1.179>

- Niemvitra, M., & Tapola, A. (2007). Self-efficacy, interest, and task performance. Within-task changes, mutual relationships and predictive effects. *Zeitschrift für Pädagogische Psychologie*, *21*(3-4), 241–250. <https://doi.org/10.1024/1010-0652.21.3.241>
- Nietfeld, J.L., Cao, L., & Osborne, J.W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning*, *1*, 159–179. <https://doi.org/10.1007/s10409-006-9595-6>
- Nishigawa, K., Omoto, K., Hayama, R., Okura, K., Tajima, T., Suzuki, Y., Hosoki, M., Shigemoto, S., Ueda, M., Rodis, O.M.M., & Matsuka, Y. (2017). Comparison between flipped classroom and team-based learning in fixed prosthodontic education. *Journal of Prosthodontic Research*, *61*, 217 – 222. <https://doi.org/10.1016/j.jpjpor.2016.04.003>
- Noetel, M., Griffith, S., Delaney, O., Harris, N. R., Sanders, T., Parker, P., Cruz, B.D.P, & Lonsdale, C. (2022). Multimedia design for learning: An overview of reviews with meta-meta-analysis. *Review of Educational Research*, *92*(3), 413–454. <https://doi.org/10.3102/00346543211052329>
- Nolan, M., Beran, T. & Hecker, K.G. (2012). Surveys assessing students' attitudes toward statistics. A systematic review of validity and reliability. *Statistics Education Research Journal*, *11*(2), 103–123. <https://doi.org/10.52041/serj.v11i2.333>
- Nouri, J. (2016). The flipped classroom: for active, effective and increased learning – especially for low achievers. *International Journal of Technology in Higher Education*, *13*(33). <https://doi.org/10.1186/s41239-016-0032-z>
- Novak, E. (2013). Effects of simulation-based learning on student's statistical factual, conceptual and application knowledge. *Journal of Computer Assisted Learning*, *30*(2), 148–158. <https://doi.org/10.1111/jcal.12027>
- O'Flaherty, J., & Phillips, C. (2016). The use of flipped classrooms in higher education: A scoping review. *Internet and Higher Education*, *25*, 85–95. <https://doi.org/10.1016/j.iheduc.2015.02.002>
- O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, *41*(3), 181–206. [https://doi.org/10.1207/s15326985ep4103\\_4](https://doi.org/10.1207/s15326985ep4103_4)
- Onwuegbuzie, A. J. (2003). Modeling statistics achievement among graduate students. *Educational and Psychological measurement*, *63*(6), 1020–1038. <https://doi.org/10.1177/0013164402250989>
- Onwuegbuzie, A.J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education*, *29*(1), 3–19. <https://doi.org/10.1080/0260293042000160384>
- Onwuegbuzie, A.J., DaRos, D. & Ryan, J. (1997). The components of statistics anxiety. A phenomenological study. *Focus on Learning Problems in Mathematics*, *19*(4), 11–35.
- Onwuegbuzie, A. J., & Seaman, M. A. (1995). The Effect of Time Constraints and Statistics Test Anxiety on Test Performance in a Statistics Course. *The Journal of Experimental Education*, *63*(2), 115–124. <https://doi.org/10.1080/00220973.1995.9943816>
- Onwuegbuzie, A.J., & Wilson, V.A. (2003). Statistics anxiety, nature, etiology, antecedents, effects, and treatments - A comprehensive review of the literature. *Teaching in Higher Education*, *8*(2), 195–209. <https://doi.org/10.1080/1356251032000052447>
- Owston, R. Lupshenyuk, D., & Wideman, H. (2011). Lecture capture in large undergraduate classes: Student perceptions and academic performance. *Internet and Higher Education*, *14*(4), 262–268. <https://doi.org/10.1016/j.iheduc.2011.05.006>
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, *84*(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>
- Paas, F., van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspective. *Educational Psychology Review*, *22*, 115–121. <https://doi.org/10.1007/s10648-010-9133-8>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578. <https://doi.org/10.3102/00346543066004543>
- Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory Into Practice*, *41*(2), 116–125. [https://doi.org/10.1207/s15430421tip4102\\_8](https://doi.org/10.1207/s15430421tip4102_8)

- Pajares, F. & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, 24(2), 124–139. <https://doi.org/10.1006/ceps.1998.0991>
- Pekrun, R., Goetz, T., Daniels, L.M., Stupnisky, R.H., & Perry, R.P. (2010). Boredom in achievement settings: Exploring control-value antecedents and performance outcomes of neglected emotion. *Journal of Educational Psychology*, 102(3), 531–549. <https://doi.org/10.1037/a0019243>
- Peterson, D.J. (2016). The flipped classroom improves student achievement and course satisfaction in a statistics course. *Teaching of Psychology*, 43(1), 10–15. <https://doi.org/10.1177/0098628315620063>
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 451–502). San Diego, CA: Academic Press.
- Pintrich, P.R. & DeGroot, E.V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. <https://doi.org/10.1037/0022-0663.82.1.33>
- Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
- Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review*, 31(2), 339–359. <https://doi.org/10.1007/s10648-019-09473-5>
- Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131–162). New York: Elsevier.
- Pociask, S., & Rajaram, S. (2014). The effects of collaborative practice on statistical problem solving: Benefits and boundaries. *Journal of Applied Research in Memory and Cognition*, 3(4), 252–260. <https://doi.org/10.1016/j.jarmac.2014.06.005>
- Poelmans, S., & Wessa, P. (2015). A constructivist approach in a blended e-learning environment for statistics. *Interactive Learning Environments*, 23(3), 385–401. <https://doi.org/10.1080/10494820.2013.766890>
- Preckel, F., Götz, T., & Frenzel, A. (2010). Ability grouping of gifted students: Effects on academic self-concept and boredom. *British Journal of Educational Psychology*, 80(3), 451–472. <https://doi.org/10.1348/000709909X480716>
- Pugh, K. J., & Bergin, D. A. (2006). Motivational influences on transfer. *Educational psychologist*, 41(3), 147 - 160. [https://doi.org/10.1207/s15326985ep4103\\_2](https://doi.org/10.1207/s15326985ep4103_2)
- Quilici, J.L., & Mayer, R.E (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, 16(3), 325–342. <https://doi.org/10.1002/acp.796>
- Rabinowitz, M., & Hogan, T.M. (2008). Experience and problem representation in statistics. *American Journal of Psychology*, 121(3), 395–407. <https://doi.org/10.2307/20445474>
- Raufelder, D. & Ringeisen, T. (2016). Self-perceived competence and test anxiety. The role of academic self-concept and self-efficacy. *Journal of Individual Differences*, 37(3), 159–167. <https://doi.org/10.1027/1614-0001/a000202>
- Reisslein, J. (2005). *Learner achievement and attitudes under varying paces of transitioning to independent problem solving*. [Doctoral dissertation, Arizona State University]. ProQuest Dissertations Publishing.
- Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, 16(2), 92–103. <https://doi.org/10.1016/j.learninstruc.2006.02.008>
- Richard, E.M., Diefendorff, J.M., & Martin, J.H. (2006). Revisiting the within-person self-efficacy and performance relation. *Human Performance*, 19(1), 67–87. [https://doi.org/10.1207/s15327043hup1901\\_4](https://doi.org/10.1207/s15327043hup1901_4)
- Richardson, M., Abraham, C. & Bond, R. (2012). Psychological correlates of university students' academic performance. A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387. <https://doi.org/10.1037/a0026838>
- Rinne, L. F., & Mazzocco, M.M.M., (2014). Knowing right from wrong in mental arithmetic judgements: Calibration of confidence predicts the development of accuracy. *PLoS ONE*, 9(e98663). <https://doi.org/10.1371/journal.pone.0098663>



- Robbins, R.W., & Pals, J.L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity*, 1(4), 313–336. <https://doi.org/10.1080/15298860290106805>
- Robbins, S.B., Lauver, K., Le, H., Davis, D., Langley, R. & Carlston, A. (2004). Do psychological and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261–288. <https://doi.org/10.1037/0033-2909.130.2.261>
- Roberts, D.M., & Saxe, J.E. (1982). Validity of a statistics attitude survey: A follow-up study. *Educational and Psychological Measurement*, 42(3), 907–912. <https://doi.org/10.1177/001316448204200326>
- Roberts, L.D. (2016). Editorial: Research methods pedagogy: Engaging psychology students in research methods and statistics. *Frontiers in Psychology*, 7(Art. 1430). <https://doi.org/10.3389/fpsyg.2016.01430>
- Rock, A.J., Coventry, W.L., Morgan, M.I. & Loi, N.M. (2016). Teaching research methods and statistics in elearning environments: Pedagogy, practical examples, and possible futures. *Frontiers in Psychology*, 7(Art.399 ). <https://doi.org/10.3389/fpsyg.2016.00339>
- Roehling, P.V., Root Luna, L.M., Richie, F.J., & Shaughnessy, J.J. (2017). The benefits, drawbacks, and challenges of using the flipped classroom in an introduction to psychology course. *Teaching of Psychology*, 44(3), 183–192. <https://doi.org/10.1177/0098628317711282>
- Rumsey, D.J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910678>
- Rutherford, T. (2017). Within and between person associations of calibration and achievement. *Contemporary Educational Psychology*, 49(3), 226–237. <https://doi.org/10.1016/j.cedpsych.2017.03.001>
- Ryan, R.M., & Moller, A.C. (2017). Competence as central, but not sufficient, for high-quality motivation. In A.J. Elliot, C.S. Dweck & D.S. Yeager (Eds.), *Handbook of competence and motivation. Theory and application* (2nd. ed., pp. 214–231). New York: The Guilford Press.
- Sabbag, A., Garfield, J., & Zieffler, A. (2018). Assessing statistical literacy and statistical reasoning: The REALI instrument. *Statistics Education Research Journal*, 17(2), 141–160. <https://doi.org/10.52041/serj.v17i2.163>
- Scepansky, J., & Carkenord, D.M. (2004). Senior year retention of methods and statistics concepts. *Teaching of Psychology*, 31(1), 9–12. [https://doi.org/10.1207/s15328023top3101\\_3](https://doi.org/10.1207/s15328023top3101_3)
- Schau, C. (2003). *Students attitudes: The „other” important outcome in statistics education* [Conference Paper]. Joint Statistics Meetings. San Francisco. <http://statlit.org/pdf/2003SchauASA.pdf>
- Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, 55(5), 868–875. <https://doi.org/10.1177/0013164495055005022>
- Schladitz, S., Groß Ophoff, J., & Wirtz, M (2015). Konstruktvalidierung eines Tests zur Messung bildungswissenschaftlicher Forschungskompetenz. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden* (S. 167–184). Weinheim u.a.: Beltz Juventa. <https://doi.org/10.25656/01:15509>
- Schlösser, T., Dunning, D., Johnson, K.L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the “signal extraction” counterexplanation for the Dunning-Kruger-effect in self-evaluation of performance. *Journal of Economic Psychology*, 39, 85–100. <https://doi.org/10.1016/j.joep.2013.07.004>
- Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183–212). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schiefele, U. & Pekrun, R. (1996). Psychologische Modelle des fremdgesteuerten und selbstgesteuerten Lernens. In F. E. Weinert (Hrsg.), *Psychologie des Lernens und der Instruktion*. (S. 249–278). Göttingen: Hogrefe.
- Schiefele, U., Wild, K.P., & Winteler, A. (1995). Lernaufwand und Elaborationsstrategien als Mediatoren der Beziehung von Studieninteresse und Studienleistung. *Zeitschrift für Pädagogische Psychologie*, 9(3-4), 181–188.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)

- Schober, B., Wagner, P., Reimann, R., Atria, M., & Spiel, C. (2006). Teaching research methods in an internet-based blended-learning setting: Vienna e-lecturing (VEL). *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(2), 73–82. <https://doi.org/10.1027/1614-2241.2.2.73>
- Schunk, D. H. (1985). Self-efficacy and classroom learning. *Psychology in the Schools*, 22(2), 208–223. [https://doi.org/10.1002/1520-6807\(198504\)22:2<208::AID-PITS2310220215>3.0.CO;2-7](https://doi.org/10.1002/1520-6807(198504)22:2<208::AID-PITS2310220215>3.0.CO;2-7)
- Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review*, 1, 173–208. <https://doi.org/10.1007/BF01320134>
- Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, 26(3-4), 207–231. <https://doi.org/10.1080/00461520.1991.9653133>
- Schunk, D. H. (1996). *Self-Efficacy for Learning and Performance* [Conference Paper]. Annual Conference of American Educational Research Association. New York.
- Schweinle, A., Turner, J.C., & Mexer, D.K. (2008). Understanding young adolescents' optimal experiences in academic settings. *The Journal of Experimental Education*, 77(2), 125–143. <https://doi.org/10.3200/JEXE.77.2.125-146>
- Schweizer, K., Steinwascher, M., Moosbrugger, H., & Reiss, S. (2011). The structure of research methodology competency in higher education and the role of teaching teams and course temporal distance. *Learning and Instruction*, 21(1), 68–76. <https://doi.org/10.1016/j.learninstruc.2009.11.002>
- Sciutto, M.J. (1995). Student-centered methods for decreasing anxiety and increasing interest level in undergraduate statistics courses. *Journal of Instructional Psychology*, 22(3), 277–280.
- Sciutto, M.J. (2002). The methods and statistics portfolio. A resource for the introductory course and beyond. *Teaching of Psychology*, 29(3), 213–215. [https://doi.org/10.1207/S15328023TOP2903\\_07](https://doi.org/10.1207/S15328023TOP2903_07)
- Shi, Z., & Liu, P. (2016). Worrying thoughts limit working memory capacity in math anxiety. *PLOS ONE*, 11(10). <https://doi.org/10.1371/journal.pone.0165644>
- Shinaberger, L. (2017). Components of a flipped classroom influencing student success in an undergraduate business statistics course. *Journal of Statistics Education*, 25(3), 122–130. <https://doi.org/10.1080/10691898.2017.1381056>
- Sitzman, T., & Johnson, S.K. (2012). Where is ignorance bliss? The effects of inaccurate self-assessments of knowledge on learning and attrition. *Organizational Behavior and Human Processes*, 117(1), 192–207. <https://doi.org/10.1016/j.obhdp.2011.11.004>
- Sitzman, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology*, 66(3), 531–568. <https://doi.org/10.1111/peps.12035>
- Sizemore, O.J., & Lewandowski, G.W. (2009). Learning might not equal liking: research methods course changes knowledge but not attitudes. *Teaching of Psychology*, 36(2), 90–95. <https://doi.org/10.1080/00986280902739727>
- Skagerlund, K., Östergren, R., Västfjäll, D., & Träff, U. (2019). How does mathematics anxiety impair mathematical abilities? Investigating the link between math anxiety, working memory, and number processing. *PLOS ONE*, 14(1). <https://doi.org/10.1371/journal.pone.0211283>
- Smith, T. F., & Capuzzi, G. (2019). Using a mindset intervention to reduce anxiety in the statistics classroom. *Psychology Learning & Teaching*, 18(3), 326–336. <https://doi.org/10.1177/1475725719836641>
- Sotos, A.E.C., Vanhoof, S., Van den Noortgate, W. & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2(2), 98–113. <https://doi.org/10.1016/j.edurev.2007.04.001>
- Spanjers, I.A., Könings, K.D., Leppink, J., Verstegen, D.M., Jong, N. de, Czabanowska, K., & Merriënboer, J.J.G. van (2015). The promised land of blended learning: Quizzes as a moderator. *Educational Research Review*, 15, 59–74. <https://doi.org/10.1016/j.edurev.2015.05.001>
- Sproesser, U., Engel, J. & Kuntze, S. (2016). Fostering self-concept and interest for statistics through specific learning environments. *Statistics Education Research Journal*, 15(1), 28–54. <https://doi.org/10.52041/serj.v15i1.256>

- Stark, R. & Mandl, H. (2000). 8 - Training in empirical research method: Analysis of problems and intervention from a motivational perspective. In J. Heckhausen (ed.), *Advances in Psychology* (pp. 165 – 183). Amsterdam: Elsevier. [https://doi.org/10.1016/S0166-4115\(00\)80011-4](https://doi.org/10.1016/S0166-4115(00)80011-4)
- Stark, R. & Mandl, H. (2002). "Unauffällige", "Vorwissensschwache", "Unmotivierte" und "Musterschüler": homogene Untergruppen beim Lernen mit einem komplexen Lösungsbeispiel im Bereich empirischer Forschungsmethoden (Forschungsbericht Nr. 147) [Unveröffentlichter Forschungsbericht]. München: LMU, Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie. [https://pub.uni-muenchen.de/255/1/FB\\_147.pdf](https://pub.uni-muenchen.de/255/1/FB_147.pdf)
- Stegers-Jager, K.M., Cohen-Schotanus, J., & Themmen, A.P.N. (2012). Motivation, learning strategies, participation and medical school performance. *Medical Education*, 46(7), 678 – 688. <https://doi.org/10.1111/j.1365-2923.2012.04284.x>
- Sticca, F., Goetz, T., Nett, U.E., Hubbard, K. & Haag, L. (2017). Short- and Logterm effects of over-reporting of grades on academic self-concept and achievement. *Journal of Educational Psychology*, 109(2), 842–854. <https://doi.org/10.1037/edu0000174>
- Stock, R., & Hiemisch, A. (2016). Moving e-learning back to class. In *Proceedings of the INTED2016 Conference* (pp. 3899–3904). Valencia, Spain.
- Stone, D.N. (1994). Overconfidence in initial self-efficacy judgements: Effects on decision processes and performance. *Organizational Behavior and Human Performance*, 59(3), 452–474. <https://doi.org/10.1006/obhd.1994.1069>
- Stone, N.J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12, 437–475. <https://doi.org/10.1023/A:1009084430926>
- Stoof, A., Martens, R.L., van Merriënboer & Bastiaens, T.J. (2002). The boundary approach of competence. A constructivist aid for understanding and using the concept of competence. *Human Resource Development Review*, 1(3), 345–365. <https://doi.org/10.1177/1534484302013005>
- Stratmann, J., Preussler, A., & Kerrens, M. (2009). Lernerfolg und Kompetenz bewerten. Didaktische Potenziale von Portfolios in Lehr-/Lernkontext. MedienPädagogik. *Zeitschrift für Theorie und Praxis der Medienbildung*, 18, 1–19 <https://doi.org/10.21240/mpaed/18/2009.12.18.X>
- Sweller, J., van Merriënboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Taberero, C., & Wood, R.E. (1999). Implicit theories versus the social construal of ability in self-regulation and performance on a complex task. *Organizational Behavior and Human Decision Processes*, 78(2), 104–127. <https://doi.org/10.1006/obhd.1999.2829>
- Thai, N.T.T., DeWever, B., & Valcke, M. (2017). The impact of flipped classroom design on learning performance in higher education: Looking for the best “blend” of lectures and guiding questions with feedback. *Computers & Education*, 107, 113–126. <https://doi.org/10.1016/j.compedu.2017.01.003>
- Thiel, F., & Böttcher, F. (2014). Modellierung fächerübergreifender Forschungskompetenzen. Das RMKR-W-Modell als Grundlage der Planung und Evaluation von Formaten forschungsorientierter Lehre. In B. Behrendt, A. Fleischmann, G. Salmhofer, N. Schaper & B. Szczyrba (Hrsg.), *Neues Handbuch Hochschullehre* (I 2.10, S. 109–124). <https://www.nhhl-bibliothek.de/>
- Tremblay, P.F., Gardner, R.C., & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude and performance in introductory statistics. *Canadian Journal of Behavioral Science*, 32(1), 40–48. <https://doi.org/10.1037/h0087099>
- Tyroller, M. (2005). *Effekte metakognitiver Prompts beim computerbasierten Statistikkernen*. [Dissertation, LMU München]. Elektronische Hochschulschriften der LMU München. <https://doi.org/10.5282/edoc.5514>
- Tze, V.M., Daniels, L.M., & Klassen, R.M. (2010). Evaluating the relationship between boredom and academic outcomes: A meta-analysis. *Educational Psychology Review*, 28, 119–144. <https://doi.org/10.1007/s10648-015-9301-y>
- Usher, E.L., & Pajares, F. (2008). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, 34(1), 89 – 101. <https://doi.org/10.1016/j.cedpsych.2008.09.002>

- Valentine, J.C., DuBois, D.L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement. A meta-analytic review. *Educational Psychologist, 39*(2), 11–133. [https://doi.org/10.1207/s15326985ep3902\\_3](https://doi.org/10.1207/s15326985ep3902_3)
- Vancouver, J.B., & Kendall, L.N. (2006). When self-efficacy negatively relates to motivation and performance in learning context. *Journal of Applied Psychology, 91*(5), 1146–1153. <https://doi.org/10.1037/0021-9010.91.5.1146>
- Vancouver, J.B., Thompson, C.M., & Williams, A.A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86*(4), 605–620. <https://doi.org/10.1037/0021-9010.86.4.605>
- Vancouver, J.B., Thompson, C.M., Tischner, E.C., & Putka, D.J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*(3), 506–516. <https://doi.org/10.1037/0021-9010.87.3.506>
- Viljaranta, J., Tolvanen, A., Aunola, K., & Nurmi, J. E. (2014). The developmental dynamics between interest, self-concept of ability, and academic performance. *Scandinavian Journal of Educational Research, 58*(6), 734–756. <https://doi.org/10.1080/00313831.2014.904419>
- Vittengl, J.R., Bosley, C.Y., Brescia, S.A., Eckardt, E.A., Neidig, J.M., Shelver, K.S., & Sappenoff, L.A. (2004). Why are some undergraduates more (and others less) interested in psychological research? *Teaching of Psychology, 31*(2), 91–97. [https://doi.org/10.1207/s15328023top3102\\_3](https://doi.org/10.1207/s15328023top3102_3)
- Williams, A.E., Aguilar-Roca, N.M., & O’Dowd, D.K. (2016). Lecture capture podcasts: differential student use and performance in a large introductory course. *Educational Technology, Research and Development, 64*, 1–12. <https://doi.org/10.1007/s11423-015-9406-5>
- Walker, C.O., Green, B.A., & Mansell, R.A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences, 16*(1), 1–12. <https://doi.org/10.1016/j.lindif.2005.06.004>
- Walker, E. R., & Brakke, K. E. (2017). Undergraduate psychology students’ efficacy and attitudes across introductory and advanced statistics courses. *Scholarship of Teaching and Learning in Psychology, 3*(2), 132–140. <https://doi.org/10.1037/stl0000088>
- Waples, J.A. (2016). Building emotional rapport with students in statistics courses. *Scholarship of Teaching and Learning in Psychology, 2*(4), 285–293. <https://doi.org/10.1037/stl0000071>
- Watkins, A.E., Bargagliotti, A., & Franklin, C. (2014). Simulation of sampling distribution of the mean can mislead. *Journal of Statistics Education, 22*(3). <https://doi.org/10.1080/10691898.2014.11889716>
- Weinert, F. E. (2001). Concept of competence: a conceptual clarification. In D. S. Rychen, & L. H. Salanik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Ashland, OH: Hogrefe & Huber.
- Wenliang, H., Holton, A., Farkas, G. & Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction, 45*, 61–71. <https://doi.org/10.1016/j.learninstruc.2016.07.001>
- Wilson, S.G. (2013). The flipped class: A method to address the challenges of an undergraduate statistics course. *Teaching of Psychology, 40*(3), 193–199. <https://doi.org/10.1177/0098628313487461>
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: Lawrence Erlbaum.
- Winquist, J.R., & Carlson, K.A. (2014). Flipped statistics class results: Better performance than lecture over one year later. *Journal of Statistics Education, 22*(3). <https://doi.org/10.1080/10691898.2014.11889717>
- Winstone, N., & Millward, L. (2012). Reframing perceptions of the lecture from challenges to opportunities: Embedding active learning and formative assessment into the teaching of large classes. *Psychology Teaching Review, 18*(2), 31–41.
- Wise, S.L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement, 45*(2), 401–405. <https://doi.org/10.1177/001316448504500226>

- Wright, A.B., & Holttum, S. (2012). Gender identity, research self-efficacy and research intention in trainee clinical psychologists in the UK. *Clinical Psychology and Psychotherapy*, 19(1), 46–56. <https://doi.org/10.1002/cpp.732>
- Zheng, L., Bhagat, K. K., Zhen, Y., & Zhang, X. (2020). The effectiveness of the flipped classroom on students' learning achievement and learning motivation: A meta-analysis. *Educational Technology & Society*, 23(1), 1–15. <https://www.jstor.org/stable/26915403>
- Ziegler, L.A. (2014). *Reconceptualizing statistical literacy: Developing an assessment for the modern introductory statistics course*. [Doctoral dissertation, University of Minnesota]. University of Minnesota Digital Conservancy. <https://hdl.handle.net/11299/165153>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA US: Academic Press.
- Zimmerman, B.J., Bandura, A., & Martinez-Pons, M. (1992). Self-Motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663–676. <https://doi.org/10.3102/00028312029003663>
- Zimmermann, W.A., & Goins, D.D. (2012). Calibration of self-efficacy for conducting a chi-squared test of independence. *Statistics Education Research Journal*, 14(2), 76–92. <https://doi.org/10.52041/serj.v14i2.262>
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 299–315). New York: Routledge.
- Zimmerman, B. J., Schunk, D. H., & DiBenedetto, M. K. (2017). The role of self-efficacy and related beliefs in self-regulation of learning and performance. In A.J. Elliot, C.S. Dweck & D.S. Yeager (Eds.), *Handbook of competence and motivation. Theory and application* (2nd ed., pp. 41–50). New York: The Guilford Press.
- Zimprich, D. (2012). Attitudes towards statistics among Swiss psychology students. *Swiss Journal of Psychology*, 71(3), 149–155. <https://doi.org/10.1024/1421-0185/a000082>
- Zlatkin-Troitschanskaia, O. & Pant, H.A. (2016). Measurement Advances and Challenges in Competency Assessment in Higher Education. *Journal of Educational Measurement*, 53(3), 253–264. <https://doi.org/10.1111/jedm.12118>
- Zlatkin-Troitschanskaia, O., Blömeke, S., Kuhn, C., & Buchholtz, C. (2012). Wirksamkeitsprüfungen im Hochschulbereich – Aufgaben und Herausforderungen des BMBF-Forschungsprogramms „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“. *Zeitschrift für Evaluation*, 11(1), 95–103.
- Zumbach, J. (2010). *Lernen mit neuen Medien. Instruktionspsychologische Grundlagen*. Stuttgart: Kohlhammer.

# ANHANG UNTERSUCHUNG I

## Messinstrumente

### Skala Aktuelle Selbstwirksamkeit

Die Art und Weise wie ich denke, macht es schwer für mich, Statistik zu verstehen.

Ich verstehe Statistik kaum.

In Statistik mache ich viele Mathematik-Fehler.

Ich kann Statistik-Gleichungen recht gut verstehen.

Ich habe Schwierigkeiten Grundbegriffe der Statistik zu verstehen.

### Skala Implizite Theorie

Jeder Mensch hat ein bestimmtes Ausmaß an statistischen Fähigkeiten, daran kann man selbst nicht viel ändern.

Jeder Mensch kann sein statistisches Können substanziell verbessern.

Wie gut oder schlecht man in Statistik ist, kann man nicht wirklich selbst beeinflussen.

Die eigenen Fähigkeiten in Statistik sind grundsätzlich veränderbar.

Man kann in Statistik zwar etwas Neues dazulernen, aber nur wenig an seinen grundlegenden statistischen Fähigkeiten ändern.

Egal wie schlecht man in Statistik ist, man kann immer selbst etwas an seinen Fähigkeiten ändern.

### Skala Anstrengung

Ich mache mir während der Statistikveranstaltung Notizen.

Ich denke während der Statistikveranstaltung aktiv mit.

Ich bin während der Statistikveranstaltung aufmerksam.

Ich bin während der Statistikveranstaltung regelmäßig anwesend.

Ich bearbeite regelmäßig die in der Statistikveranstaltung gestellten Aufgaben (Hausaufgaben, Übungsaufgaben etc.).

Ich strenge mich beim Statistiklernen sehr an.

Ich höre in Statistik oft mit dem Lernen auf, bevor ich geschafft habe, was ich mir vorgenommen habe.

Wenn ich in Statistik Verständnisprobleme habe, gebe ich meist vorschnell auf oder lerne nur oberflächlich.

Ich bemühe mich, den Statistikstoff genau zu verstehen.

**Tabelle C.2a***Korrelationen in der Gesamtstichprobe (N = 88)*

	(11)	(10)	(9)	(8)	(7)	(6)	(5)	(4)	(3)	(2)	(1)
Aktuelle Selbstwirksamkeit T1 (1)	.18	.01	-.04	.15	.21	.70	.11	.18	.28	.68	
Aktuelle Selbstwirksamkeit T2 (2)	.38	.19	.15	-.02	.21	-.04	.01	.31	.27		
Implizite Theorie T1 (3)	.16	-.05	.15	-.12	-.13	.00	.07	.68			
Implizite Theorie T2 (4)	.12	.01	.19	-.13	-.11	.00	-.01				
Bias <sup>1</sup> T1 (5)	-.14	-.63	-.00	.25	.15	.46					
Bias <sup>1</sup> T2 (6)	-.59	-.22	.06	.60	.04						
Bias Kalibrierung <sup>2</sup> T1 (7)	.12	-.10	-.10	.22							
Bias Kalibrierung <sup>2</sup> T2 (8)	-.33	-.21	.07								
Anstrengung (9)	-.04	-.04									
Pretest (10)	.41										
Nachtest (11)											

*Anmerkungen.* Alle Korrelationen  $|r| > .21$  sind signifikant ( $p < .05$ ). <sup>1</sup>. Positive Werte stehen für eine Über- und negative Werte für eine Unterschätzung. <sup>2</sup>. Höhere Werte stehen für eine schlechtere Kalibrierung.

**Tabelle C.2b***Korrelationen in der Teilstichprobe der Überschätzer (n = 54)*

	(9)	(8)	(7)	(6)	(5)	(4)	(3)	(2)	(1)
Aktuelle Selbstwirksamkeit T1 (1)	.18	-.02	-.10	.10	.21	.28	.31	.67	
Aktuelle Selbstwirksamkeit T2 (2)	.41	.23	.05	.03	.18	.34	.32		
Implizite Theorie T1 (3)	.17	-.07	.23	-.05	-.14	.75			
Implizite Theorie T2 (4)	.03	-.10	.19	.07	-.05				
Bias <sup>1</sup> T1 (5)	.05	-.36	-.08	.32					
Bias <sup>1</sup> T2 (6)	-.60	-.07	-.02						
Anstrengung (7)	.00	.06							
Pretest (8)	.32								
Nachtest (9)									

*Anmerkungen.* Alle Korrelationen  $|r| > .25$  sind signifikant ( $p < .05$ ). <sup>1</sup>. Höhere Werte im Bias stehen für eine stärkere Überschätzung.

# **ANHANG UNTERSUCHUNG II**

## **Messinstrumente**

### **Skala Statistische Selbstwirksamkeit**

Auch wenn ich in Statistik nicht sofort alles verstehe, bin ich überzeugt, dass ich mir den Stoff erarbeiten kann.

Ich traue mir zu, meinen KommilitonInnen statistische Inhalte zu erklären.

Auch wenn ich mich mal mit statistischen Inhalten beschäftige, bringt das meistens doch nichts.

Ich bin überzeugt, dass ich selbstständig eine statistische Aufgabenstellung bearbeiten kann.

Egal wie sehr ich mich anstrengende, Statistik werde ich wahrscheinlich nie verstehen.

Ich traue mir zu, mein statistisches Wissen auch auf praktische Fragestellungen anzuwenden.

Ich denke, dass ich genug Vorwissen habe, um neue statistische Inhalte verstehen zu können.

Ich bin mir sicher, dass ich Fragen zu grundlegenden statistischen Inhalten beantworten kann.

### **Skala Vermeidende Lernstrategien**

Wenn ich etwas in Statistik nicht verstehe, gebe ich auf.

Ich gebe mich häufig damit zufrieden, ein Konzept in Statistik „ungefähr“ begriffen zu haben.

Ich beschäftige mich nur dann mit Statistik, wenn es unbedingt sein muss.

In Statistik mache ich nicht mehr als unbedingt nötig.

Ich vermeide es soweit wie möglich, mich mit statistischen Inhalten zu beschäftigen.

Ich muss mich jedes Mal zwingen, mich mit statistischen Inhalten auseinanderzusetzen.

Ich versuche meist gar nicht erst komplizierte statistische Zusammenhänge zu verstehen.

### **Skala Kognitive Belastung**

Im Seminar wird zu viel Vorwissen vorausgesetzt.

Ich habe Schwierigkeiten, im Seminar mitzukommen.

Während des Seminars habe ich häufig das Gefühl, überfordert zu sein.

Die Bearbeitung der im Seminar gestellten Aufgaben/Arbeitsaufträge fällt mir leicht.

Die Schwierigkeit der Seminarinhalte ist zu hoch.

### **Skala Affektives Interesse**

Es macht mir Spaß, die an mich gestellten Aufgaben/Arbeitsaufträge im Seminar zu lösen.

Ich finde die Inhalte des Seminars interessant.

Ich habe Freude daran, mich mit den Inhalten des Seminars auseinanderzusetzen.

Es fällt mir leicht, mich zu einer Beschäftigung mit den Inhalten des Seminars zu motivieren.



## Deskriptive Statistik

Tabelle D.6

*Bisher belegte Statistikkurse*

IC	TL	KG
"Einführung in die Statistik" (für Politikwissenschaftler)	Advanced Statistical Methods	1
Biomathematik für Mediziner	ataStS	An der HNEE im BSC
Statistik für Politikwissenschaftler	B.Sc.: Statistik für Biologen (Anwendungsorientiert)	Einführungsveranstaltung (Vorlesung)
Statistik <sup>a</sup>	einführende	Grundkurs Statistik (Dr. Uhlemann)
Statistik 1/2	Geostatistik 1	Mathematik/ Statistik
Statistik für Geographen	Grundlagen wissenschaftlichen Arbeitens (im Bachelor)	Klausur/Mündliche Prüfung (Vorlesung, Übung)
Statistik für Geographen (2004)	Mathematik für Naturwissenschaftler	Methoden der Politikwissenschaft
Statistik für Landschaftsökologen	Mathematik und Statistik, Biostatistik	Quantitative Methods
Statistik für Naturwissenschaftler	Mathematik/Statistik	QB1 Epidemiologie, medizinische Biometrie und medizinische Informatik
Statistik für Politikwissenschaftler <sup>a</sup>	Methoden der Politikwissenschaft	Statistik 1 (Deskriptive Statistik)
statistik in der powi	Methoden der Politikwissenschaft im Bachelor an der Uni Rostock und Methoden der Organisationskommunikation hier an der Uni Greifswald	Statistik 1 für Betriebswirte
Statistische Übungen I	PC-Kurs	Statistik als Nebenfach
Vegetationsökologie	R- Kurs	Statistik der Politikwissenschaft
	Sta 1	Statistik für B.A. Powi
	Statistik (für BSc LaÖk)	Statistik für Politikwissenschaftler
	Statistik 1 <sup>a</sup>	Statistik für Politikwissenschaftler
	Statistik für Landschaftsökologen	Statistik I + II für Psychologen
	Statistik I <sup>a</sup>	Statistische Methoden
	Statistik Vorlesung und Übung im 1. Semester	Vorlesung Statistik
	Statistische Methoden I <sup>a</sup>	

*Anmerkungen.* Dargestellt sind die Antworten auf die Frage „Bitte geben Sie hier an, welche Statistikveranstaltung(en) Sie [bisher] belegt haben.“ <sup>a</sup>. Mehrfach genannt.

## **DANKSAGUNG**

Ich danke meiner Betreuerin Anette Hiemisch, die mich stets mit Geduld und guten Ratschlägen unterstützt hat.

Meinen Kollegen Stefan Schelske und Vincent Exner danke ich für die Mithilfe bei der Gestaltung der Statistikkurse.

Besonderer Dank gilt Ina Lams, Marie Bischoff und Sandra Lange für die inhaltlichen, formalen und sprachlichen Hinweise sowie Devu Mahesan für die Hilfe bei der Übersetzung.

Zuletzt danke ich allen Studierenden der Universität Greifswald, die mich immer wieder dazu motiviert haben, in der Lehre mein Bestes zu geben.