Entwicklung und Evaluierung eines spezifischen Anfängerpraktikums für Lehramtsstudierende im Fach Physik

Dissertation zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät der
Christian-Albrechts-Universität zu Kiel

vorgelegt von
Jasmin Andersen

Kiel, 2020

Erster Gutachter: Prof. Dr. Knut Neumann

Zweite Gutachterin: Prof. Dr. Dr. h.c. Ilka Parchmann

Tag der mündlichen Prüfung: 17. November 2020

Inhaltsverzeichnis

Zusammenfassung 1

I Einführung 3

1 Einleitung 4

2 Physikalische Laborpraktika 7

3 Wirkung adressatenspezifischer Laborpraktika 14

4 Ziele und Forschungsfragen 27

II Lehramtspraktikum 31

5 Rahmenbedingungen 32

6 Neukonzeption Lehramtspraktikum 44

7 Synopse zur Praktikumskonzeption 67

III Wahrgenommene Relevanz 69

8 Methodisches Vorgehen 70

9 Analyse der wahrgenommenen Relevanz 106

10 Synopse zur wahrgenommenen Relevanz 139

IV Wirkung 143

11 Methodisches Vorgehen 144

12 Prüfung des Wirkmodells 183

13 Synopse zum Wirkmodell 248

V Abschließende Gesamtbetrachtung 251

14 Diskussion der Befunde im Forschungskontext 252

15 Implikationen für die Praxis 257

16 Desiderat 260

Literaturverzeichnis **263**

A Anhang **287**

Publikationen und Tagungsbeiträge **407**

Danksagung **409**

Eidesstattliche Erklärung **411**

B Digitaler Anhang **413**

Zusammenfassung

In der universitären Lehramtsausbildung in Deutschland werden die Trennung von Fach und Fachdidaktik und der hohe fachwissenschaftliche Anteil in der Studieneingangsphase kritisch diskutiert. An der Christian-Albrechts-Universität zu Kiel wurde daher beschlossen, die Lehramtsausbildung im Fach Physik neu auszurichten und die Bereiche Fach und Fachdidaktik früher im Studium zu vernetzen, um die kumulative Entwicklung der fachlichen und fachdidaktischen Komponenten des Professionswissens zu verbessern und die wahrgenommene Relevanz der Lerninhalte zu fördern. Dazu wurde im Rahmen dieser Arbeit ein physikalisches Laborpraktikum für Lehramtsstudierende mit 25 neuen, lehramtsspezifischen Praktikumsversuche konzipiert, das sich an aktuellen fachdidaktischen Gestaltungsprinzipien orientiert. Der Entwicklung liegt ein theoretisch abgeleitetes Wirkmodell zugrunde, wonach eine Förderung der wahrgenommenen Relevanz sich positiv auf die Lernmotivation und damit wiederum förderlich auf den Lernerfolg auswirkt.

Das Forschungsziel dieser Arbeit ist, Erkenntnisse über die optimale Gestaltung von Lehramtspraktika zu gewinnen und die Wirkannahmen quantitativ zu prüfen. Hierzu wurde im neuen Lehramtspraktikum eine Vollerhebung vorgenommen, sodass eine ausreichende Stichprobe und eine Vielzahl der Bedingungsfaktoren erfasst werden konnten. Mithilfe eines dafür neu entwickelten Fragebogens wurden dabei auch die Eigenschaften der Praktikumsversuche erfasst, sodass in der begleitenden Beforschung mittels Mehrebenenanalyse Erkenntnisse über die Eigenschaften relevant wahrgenommener Versuche gewonnen werden konnten. Zentrale Ergebnisse der Arbeit sind folgende: Eine Schlüsselrolle für die wahrgenommene Relevanz spielt dabei neben dem schulrelevanten Versuchsthema und der Anschaulichkeit vor allem die Kontextualisierung der Versuchsbeschreibungen. Eine Mehrebenen-Strukturgleichungsmodellierung des Wirkmodells hat gezeigt, dass die wahrgenommene Relevanz das situationale Interesse an den Praktikumsinhalten hypothesenkonform positiv beeinflusst. Zudem konnte anhand von empirischen Indizien aus weiteren Modellierungen der Wirkzusammenhänge geschlossen werden, dass sich die wahrgenommene Relevanz ebenfalls förderlich auf den Lernerfolg in Bezug auf experimentelle Kompetenz, Selbstwirksamkeitserwartungen bzgl. des Experimentierens und Fachwissen der Studierenden auswirkt. Basierend auf diesen Erkenntnissen konnte damit in dieser Arbeit gezeigt werden, dass ein lehramtsspezifisches Laborpraktikum eine wirksame und umsetzbare Möglichkeit ist, um den Studierenden die Relevanz der Lerninhalte für ihre zukünftige Lehrtätigkeit zu verdeutlichen, die Lernmotivation und den Lernerfolg zu fördern und dadurch die Lehramtsausbildung in den naturwissenschaftlichen Fächern zu verbessern.

Abstract

In teacher education programs at universities in Germany the separation of academic subject and subject education as well as the high subject-specific proportion in the introductory phase are critical discussed. Therefore, at Kiel University the teacher education program in physics was redesigned with an emphasis on linking academic subject and subject education early, in order to improve the cumulative development of content knowledge and pedagogical content knowledge and to enhance the perceived professional relevance of learning contents. Within the scope of this work a physics laboratory course for pre-service teachers with 25 new teaching-specific practical experiments was developed on the basis of recent didactical design principles. A base for this new development is a theoretically derived causal model, according to which an enhancement of the perceived relevance has positive effects on the learning motivation and in turn on the learning success.

The goal of this work is to better understand the optimal design of teacher's laboratory courses and to quantitatively verify the hypothesized causal model. On the basis of an adequate sample an exhaustive survey including relevant covariates was administered in the new teacher's laboratory course. With a newly developed questionnaire the characteristics of the experiments were identified, so that as part of this research, knowledge about the characteristics of experiments perceived as relevant could be gained by multilevel analysis. Main results of this work are: Besides a topic with relevance for school and a demonstrative character of the experiment, the contextualisation of the experiment instruction plays a key role for the perceived relevance of the experiment. Supporting the causal hypothesis, multilevel structural equation modelling showed, that the perceived relevance has a positive effect on the situational interest in the learning contents of the experiments. Additionally, from further modelling of the presumed causal relations indicates, that the perceived relevance also beneficially affects the learning success with regard to experimental competences, self-efficacy expectations concerning experimentation, and content knowledge of the students. These findings of this work show, that a teaching-specific laboratory course is an effective and feasible opportunity to clarify the relevance of the learning contents to the students for their future teaching, to enhance the learning motivation and the learning success, and thus to improve the teacher education programs in natural science subjects.

Teil I
Einführung

1 Einleitung **4**

2 Physikalische Laborpraktika **7**

2.1 Lernziele in physikalischen Praktika 7

2.2 Organisationsform physikalischer Praktika 9

2.3 Konzeption adressatenspezifischer Praktika 10

3 Wirkung adressatenspezifischer Laborpraktika **14**

3.1 Lernerfolg 14

3.2 Motivation und Interesse 16

3.3 Wahrgenommene Relevanz 19

3.4 Kontextualisierung 21

3.5 Das Wirkmodell 22

3.6 Beforschung adressatenspezifischer Praktika 24

4 Ziele und Forschungsfragen **27**

4.1 Ziele der Entwicklungsarbeit 27

4.2 Forschungsfragen 28

4.3 Struktur der Arbeit 30

1 Einleitung

Ein bedeutsamer Einflussfaktor für guten Unterricht ist die professionelle Handlungskompetenz von Lehrkräften (Baumert & Kunter, 2006). Zu den zentralen Facetten der professionellen Handlungskompetenz zählt das Professionswissen, das unter anderem das Wissen und Können in den Bereichen Fach, Fachdidaktik und Pädagogik umfasst (Shulman, 1986; Baumert & Kunter, 2006). Diese drei Inhaltsbereiche werden im Lehramtsstudium klassischerweise getrennt voneinander vermittelt (z. B. Neumann et al., 2017). Auch Lehramtsstudierende des Faches Physik besuchen voneinander getrennte fachliche, fachdidaktische und bildungswissenschaftliche Lehrveranstaltungen. In den ersten Semestern des Lehramtsstudiums Physik liegt der Fokus dabei meist auf den fachlichen Veranstaltungen mit Grundlagenvorlesungen, zugehörigen Übungen und Laborpraktika (z. B. J. Bauer et al., 2012). Der fachwissenschaftliche Teil der Lehrkräftebildung Physik findet dabei in vielen Universitäten und so auch an der Christian-Albrechts-Universität zu Kiel (CAU) hauptsächlich in polyvalenten Lehrveranstaltungen ohne Professionsorientierung für die Lehramtsstudierenden statt. Das bedeutet, dass die Lehrveranstaltungen der Lehramtsstudierenden des Faches Physik (2-Fach-Studiengang) in den ersten Semestern identisch sind mit denen des Studiums der Physik (1-Fach-Studiengangs). Im fortschreitenden Studium nehmen dann die Veranstaltungen der Physikdidaktik zu. Damit umfassen die Lehrveranstaltungen im Lehramtsstudium jeweils etwa 40 % je Fach für fachliche und fachdidaktische Inhalte und etwa 20 % für Pädagogik.

In der Lehramtsausbildung des Faches Physik sind die Trennung von Fach und Fachdidaktik, die polyvalenten fachlichen Lehrveranstaltungen ohne Professionsorientierung und der hohe fachwissenschaftliche Anteil in der Studieneingangsphase Teil kritischer Diskussionen (z. B. Schecker & Ralle, 2009). Bedingungsfaktor erfolgreichen Unterrichtens ist nicht nur eine umfangreiche, sondern auch eine gut vernetzte Wissensbasis (z. B. Baumert & Kunter, 2011). Fundiertes Fachwissen ist sogar eine wichtige Voraussetzung für die Entwicklung fachdidaktischen Wissens (z. B. Baumert & Kunter, 2006; Sorge, Kröger et al., 2019). Mit einer strikten Trennung von Fach und Fachdidaktik in den Lehrveranstaltungen sowie einem hohen Anteil fachphysikalischer Veranstaltungen im Bachelorstudium und einem zunehmenden Anteil fachdidaktischer Veranstaltungen erst im Masterstudium ist das Lehramtsstudium aber nicht auf eine erforderliche kumulative Entwicklung von Fachwissen und fachdidaktischem Wissen ausgerichtet. Aus diesem Grund wird eine stärkere Vernetzung fachlicher und fachdidaktischer Ausbildungsanteile gefordert (z. B. Schecker & Ralle, 2009). Die Deutsche Physikalische Gesellschaft (DPG) und das nationale MINT-Forum empfehlen sogar eine eigenständige Lehramtsausbildung (Großmann & Urban, 2006; MINT-Forum, 2013; Großmann & Hertel, 2014), die durchgehend ohne polyvalente Lehrveranstaltungen auf eine optimale Entwicklung des Professionswissens von Lehrkräften ausgerichtet ist.

Der hohe fachphysikalische Anteil und die geringe Vernetzung von Fach und Fachdidaktik wird insbesondere in der Studieneingangsphase auch in Bezug auf motivationale Faktoren kritisch betrachtet. Lehramtsstudierenden kann es in dieser Phase besonders schwer fallen, die Relevanz der hauptsächlich fachlichen Lehrangebote für die spätere Tätigkeit als Lehrkraft zu erkennen, da diese auf die Anforderungen von Studierenden des 1-Fach-Studienganges Physik ausgerichtet sind und damit keinen klar ersichtlichen Bezug zu ihrer künftigen Lehrtätigkeit in der Schule aufweisen (z. B. Lersch, 2006; AG-Studienqualität, 2011; Makrinus, 2012; Woitkowski, 2017). Aus einer geringen wahrgenommenen Relevanz kann eine mangelnde Lernmotivation folgen (z. B. Schiefele, 2004). Diese kann sich wiederum negativ auf die Lernleistung auswirken (z. B. Schiefele, Krapp & Schreyer, 1993). In Hinblick auf die hohen Leistungsanforderungen des Physikstudiums kann die Konsequenz sogar ein Studienabbruch sein (Heublein et al., 2017).

Um diesen Problemen zu begegnen, ist an der CAU das Vorhaben entstanden, durch stärkere Vernetzung von Fach und Fachdidaktik im Lehramtsstudium Physik die kumulative Entwicklung der fachlichen und fachdidaktischen Komponenten des Professionswissens zu unterstützen (*Projekt Lehramtsausbildung Physik 2.0 oder kurz LaP 2.0*, Andersen, Block, Neumann & Wendlandt, 2018). Diese stärkere Vernetzung soll anhand einer Neukonzeption der physikalischen Laborpraktika der Lehramtsausbildung umgesetzt werden. Die Laborpraktika wurden zum einen ausgewählt, um die Entwicklung des fachlichen Wissens und Könnens besonders im Bereich des Experimentierens zu unterstützen und dadurch dem als mangelhaft bewerteten Einsatz von Experimenten im Physikunterricht (z. B. Harlen, 1999; Tesch & Duit, 2004) entgegenzuwirken. Zum anderen wurde die mangelnde Vernetzung von Fach und Fachdidaktik an der CAU besonders im fachdidaktisch ausgerichteten Laborpraktikum im Masterstudium deutlich, da das Fachwissen für das schulorientierte Experimentieren nicht vorausgesetzt werden konnte. Dadurch hatte sich der eigentliche Fokus von fachdidaktischen Fragestellungen zu fachlichen Fragestellungen verschoben (z. B. zur Nutzung schultypischer Messtechnik). Warum entsprechendes Fachwissen nicht bereits im Anfängerpraktikum im Bachelorstudium vermittelt wurde, könnte daran liegen, dass die Lehramtsstudierenden üblicherweise am fachlichen Anfängerpraktikum der 1-Fach-Physiker teilnehmen, das auf das Tätigkeitsprofil von Physiker*innen ausgerichtet ist und damit nicht, wie von der DPG gefordert (Großmann & Hertel, 2014), auf die spätere berufliche Tätigkeit als Physiklehrkraft. Dies wiederum dürfte die wahrgenommene Relevanz und in der Folge auch die Wirksamkeit nachhaltig beeinträchtigen (z. B. Woitkowski, 2017).

Ziel dieser Arbeit ist daher die Neuentwicklung eines spezifischen Anfängerpraktikums für Lehramtsstudierende des Faches Physik. Diese Arbeit ist Teil der Kieler Programminitiative *CAU - Lehramt in Bewegung* (CAU-LiB) und verfolgt übergeordnet das Ziel, das Lehramtsstudium an der CAU zu verbessern. Bei der Neukonzeption des Anfängerpraktikums soll die Trennung von Fach und Fachdidaktik in der Studienanfangsphase durch nachhaltige Vernetzungselemente aufgebrochen werden.

Es sollen dabei vor allem gezielt die spezifischen Anforderungen der Lehramtsstudierenden berücksichtigt werden, um zum einen das fachliche und fachdidaktische Wissen zu vermitteln, das für die spätere berufliche Tätigkeit als Physiklehrkraft notwendig ist. Zum anderen sollen dadurch wahrgenommene Relevanz, Lernmotivation und Lernerfolg der Lehramtsstudierenden gefördert werden. Die Neuentwicklung des lehramtsspezifischen Anfängerpraktikums orientiert sich damit an der von der DPG geforderten Lehramtsausbildung „sui generis“ (Großmann & Hertel, 2014).

Aus diesem formulierten Entwicklungsziel ergeben sich zwei grundlegende Aufgabenstellungen: Zum einen muss herausgefunden werden, wie ein solches physikalisches Anfängerpraktikum spezifisch für Lehramtsstudierende des Faches Physik idealerweise gestaltet sein sollte. Zum anderen ist zu klären, wie sich so ein spezifisches Laborpraktikum tatsächlich auf die motivationalen Faktoren und den Lernerfolg der Studierenden auswirkt.

Auf Basis der vorangegangenen grundlegenden Einführung in die Lehramtsausbildung im Fach Physik im Teil I dieser Arbeit sollen nun zwei erforderliche Schritte vorgenommen werden, um das Ziel dieser Arbeit erreichen zu können. Der erste wichtige Schritt ist die Erfassung des aktuellen Stands zur Umsetzung und Beforschung von (adressatenspezifischen) Laborpraktika (Abschnitt 2), um Randbedingungen bei der Gestaltung von Praktika zu identifizieren und einen Kriterienkatalog für die Neukonzeption eines Laborpraktikums spezifisch für Lehramtsstudierende aufzustellen. Der zweite notwendige Schritt ist eine Auseinandersetzung mit dem Stand der Forschung zur Relevanzwahrnehmung, zur Lernmotivation und zum Lernerfolg, um Hypothesen über die zugrundeliegenden Wirkmechanismen von adressatenspezifischen Laborpraktika abzuleiten (Abschnitt 3). Vor diesem Hintergrund werden dann die zentralen Forschungsfragen dieser Arbeit formuliert (Abschnitt 4). Hieraus ergibt sich dann die weitere Strukturierung der Teile II, III, IV und V dieser Arbeit.

2 Physikalische Laborpraktika

Ein physikalisches Laborpraktikum ist eine universitäre Lernumgebung, in der Studierende praktisch Experimente durchführen und auswerten. Da das Experimentieren eine zentrale Methode der Naturwissenschaft zur Erkenntnisgewinnung ist (z. B. Feynman et al., 1963), nimmt das physikalische Praktikum eine wichtige Rolle in der physikalischen Ausbildung ein. Physikalische Laborpraktika, oder im Folgenden auch kurz nur Physikpraktika oder Praktika, sind von Schulpraktika, den praktischen Phasen des Lehramtsstudiums in der Schule, abzugrenzen. Sofern nicht anders gekennzeichnet, wird in dieser Arbeit Bezug genommen auf das physikalische Anfängerpraktikum, auch Grundpraktikum, das in den ersten Studiensemestern durchgeführt wird.
Für die Neukonzeption eines physikalischen Laborpraktikums mit spezifischer Ausrichtung für Lehramtsstudierende ist sowohl die Beachtung gewisser Rahmenbedingungen als auch die Berücksichtigung von Erfahrungen und Erfolgen vorhandener Physikpraktika erforderlich. Dafür werden zum einen die grundlegenden Lernziele (Abschnitt 2.1) und die klassische Organisationsform (Abschnitt 2.2) physikalischer Laborpraktika betrachtet. Zum anderen werden die Besonderheiten adressatenspezifischer Laborpraktika herausgearbeitet, um schließlich einen Kriterienkatalog für die Neukonzeption des lehramtsspezifischen Anfängerpraktikums abzuleiten (Abschnitt 2.3).

2.1 Lernziele in physikalischen Praktika

Die Gestaltung physikalischer Laborpraktika folgt in langer Tradition dem Praktikumskonzept von Westphal (1937) (vgl. z. B. den Überblick von Neumann, 2004, für die historische Entwicklung). Diese Gestaltung wird fast ebenso lange kritisch in Bezug auf das Erreichen der Lernziele diskutiert (vgl. z. B. Hucke, 2000; Holmes & Wieman, 2018). In der Folge entstanden mehrere Studien zur Bestimmung der Lernziele (z. B. Boud et al., 1980; AAPT, 1998; Nagel et al., 2018) und zur nicht-trivialen Untersuchung der Effektivität von Lernumgebungen im Kontext des Experimentierens (vgl. Hofstein & Lunetta, 1982; Toothacker, 1983; Hofstein & Lunetta, 2004). Dazu zählen auch Studien des EU-Projektes „Labwork in Science Education“ (Welzel et al., 1998; K. Haller, 1999; Psillos & Niedderer, 2002). In diesen Betrachtungen zeigt sich, dass in Praktika viele unterschiedliche Ziele verfolgt werden, die in den Studien unterschiedlich spezifisch formuliert und unterschiedlich gewichtet werden. Ein guter Konsens ist in der Untersuchung von Welzel et al. (1998) zu finden. Welzel et al. (1998) haben Lehrende aus den Fachbereichen Biologie, Chemie und Physik in sechs europäischen Ländern nach den Zielen von naturwissenschaftlichen Praktika befragt und dabei fünf fächerübergreifende Hauptlernziele mit jeweils sechs bis zwölf Unterzielen ermittelt. Auf dieser Basis hat Neumann (2004) die Hauptlernziele anhand einer offenen Expertenbefragung weiter für Anfängerpraktika für Physiker

spezifiziert und die Hauptziele von Welzel et al. (1998) bestätigt. Die Lernziele im Physikpraktikum, mit abnehmender Wichtigkeit, sind nach Neumann (2004):

1. Experimentelle Fähigkeiten erwerben
2. Methoden wissenschaftlichen Denkens kennenlernen
3. Theorie und Praxis verbinden

Das Praktikum zielt damit in erster Linie auf das Erreichen fachlicher Lernziele ab. Im Detail wird dabei nicht nur der Erwerb experimentbezogener Fähigkeiten allgemein verstanden, sondern insbesondere auch Fähigkeiten und Kenntnisse speziell im Umgang mit Messgeräten. Bei den Methoden wissenschaftlichen Denkens wird nach (Neumann, 2004) von den Befragten besonderer Fokus auf die Planung, Auswertung und Interpretation experimenteller Daten gelegt. Diese Lernziele werden oftmals unter dem Begriff „experimentelle Kompetenzen“ zusammengefasst. Nach Weinert (2001) können Kompetenzen als die kognitiven Fähigkeiten und Fertigkeiten sowie die motivationalen, volitionalen und sozialen Bereitschaften verstanden werden, die notwendig sind, um bestimmte Probleme zu lösen. Der Begriff der experimentellen Kompetenz wird daher auch in dieser Arbeit genutzt, um insbesondere deutlich zu machen, dass es nicht nur theoretischen Wissens bedarf, sondern darüber hinaus auch der Befähigung zur Handlung, um im Experimentierprozess Herausforderungen zu begegnen und Erkenntnisse zu gewinnen (siehe weiterführend Abschnitt 3.1). So lassen sich die experimentellen Lernziele auch begrifflich vom dritten Lernziel bzgl. des theoretischen Fachwissens abgrenzen. Als Drittes sollen die Studierenden als Teil der Verbindung von Theorie und Praxis vor allem theoretische Kenntnisse aus den Grundlagenvorlesungen vertiefen. Aus diesem Grund werden im Anfängerpraktikum in der Regel die Themenbereiche Mechanik, Elektrizitätslehre, Wärmelehre, Atomphysik und Optik als zentrale Themen der physikalischen Grundlagenvorlesungen in den ersten Semestern der Physikausbildung behandelt (vgl. Neumann, 2004).

Neben den im Praktikum angestrebten fachlichen Aspekten diskutieren Neumann (2004) und Welzel et al. (1998) auch den affektiven Charakter, wonach das Praktikum im Hinblick auf die Ausbildung motivieren und Interesse wecken soll. Dies ist insbesondere für die Neukonzeption eines adressatenspezifischen Praktikums ein wichtiger Punkt, da motivationale Aspekte per se eine besondere Rolle spielen (Abschnitt 2.3). Folglich sollte neben der Entwicklung verschiedener experimenteller Kompetenzen und der Vertiefung von Fachwissen ein modernes Praktikum auch eine motivierende Aufgabe erfüllen.

2.2 Organisationsform physikalischer Praktika

Die Organisationsform ist ein wesentliches Merkmal eines Laborpraktikums und oftmals eine über die Modulbeschreibung fest im Studienplan verankerte und damit weitgehend unveränderliche Randbedingung. Die Kenntnis der Praktikumsstruktur ist somit erforderlich, um eine bessere Vorstellung von den Möglichkeiten für die später beschriebene lehramtsspezifische Konzeption des Praktikums zu erlangen.

Die klassische Organisationsform eines traditionellen physikalischen Anfängerpraktikums (vgl. Rehfeldt, 2018a) trifft auf das bisherige Anfängerpraktikum an der CAU (vgl. CAU, 2017c), an dem bisher 1-Fach-Bachelor Physik und Lehramtsstudierende gleichermaßen teilgenommen haben, zu. Davon abweichende Praktikumsstrukturen wie zum Beispiel beim Paderborner Physik Praktikum (3P) (A. Bauer & Sacher, 2018) sind im bisherigen Anfängerpraktikum der CAU nicht zu finden.

Anfängerpraktika finden meist in den ersten Studiensemestern statt und gehören damit zu den ersten praktischen experimentellen Lehrveranstaltungen im Bachelorstudium. An den Praktika nehmen nicht nur Studierende des 1-Fach- und 2-Fach-Studienganges Physik teil, sondern auch Studierende anderer Fachrichtungen wie zum Beispiel Materialwissenschaften oder Meteorologie. An die Teilnahme sind bestimmte Zulassungsvoraussetzungen geknüpft, wie zum Beispiel an der CAU das Bestehen der ersten physikalischen Grundlagenvorlesung. An der CAU erstreckt sich das Praktikum über zwei Semester, in deren Verlauf die Studierenden mehrere Praktikumsversuche durchführen. Die Praktikumstermine umfassen dabei eine Präsenzzeit von drei bis vier Stunden pro Woche.

Die Versuchstage beginnen in der Regel mit einem Vorgespräch der Studierenden mit einem der Betreuenden. Die Betreuenden prüfen darin, ob die Studierenden sich mithilfe der gegebenen Versuchsbeschreibung und den Literaturhinweisen angemessen auf den Versuch vorbereitet haben und den Versuch dadurch sinnvoll und sicher durchführen können. Eine unzureichende Vorbereitung der Studierenden kann zu einem Ausschluss vom Versuch führen. Anschließend führen die Studierenden meist in einem Zweierteam einen der Praktikumsversuche durch. Dafür bauen die Studierenden zu Beginn des Versuchstages den Versuch anhand von bereitgestellten Versuchsmaterialien auf. Für den Aufbau und die Durchführung steht ihnen wieder die Versuchsbeschreibung mit entsprechenden Hinweisen und Aufgabenstellungen zur Verfügung. Die Versuchsdurchführung und die aufgenommenen Messdaten werden von den Studierenden in einem Laborbuch dokumentiert. Die Studierenden werden bei der Versuchsdurchführung von Betreuenden und wissenschaftlichen Hilfskräften unterstützt. Nach der Durchführung des Versuches fertigen die Praktikumsgruppen anhand der Aufzeichnungen und einer Auswertung der Messdaten ein Praktikumsprotokoll an. Für den erfolgreichen Abschluss des Moduls müssen alle zugeteilten Versuche durchgeführt und die zugehörigen Protokolle gegebenenfalls nach einmaliger Korrektur vollständig und korrekt sein.

Durch die Angabe von Themengebieten, Lernzielen und -inhalten, Semesterwochenstunden, Zulassungsvoraussetzungen, Prüfungsformen und -umfang ist dieser Praktikumsablauf in der Modulbeschreibung abgebildet und im Weiteren nur sehr eingeschränkt veränderbar.

2.3 Konzeption adressatenspezifischer Praktika

Um ein physikalisches Anfängerpraktikum spezifisch für Lehramtsstudierende des Faches Physik zu entwickeln, ist es hilfreich zunächst zu betrachten, wie ein solches Praktikum idealerweise gestaltet sein sollte. Nachdem in den vorangegangenen Abschnitten Rahmenbedingungen für die Gestaltung des Praktikums identifiziert wurden, wird in diesem Abschnitt eine Synthese bestehender adressatenspezifischer Praktika vorgenommen, um einen Kriterienkatalog für die Neukonzeption des lehramtsspezifischen Anfängerpraktikums an der CAU abzuleiten.

Physikalische Praktika werden für eine Reihe von Studierenden unterschiedlicher Studienrichtungen angeboten. Neben den typischen 1-Fach-Studierenden des Faches Physik können dies zum Beispiel auch Studierende in den Bereichen Meteorologie, Medizin oder Biologie sein. Trotz unterschiedlicher Zielgruppe unterscheiden sich die Praktika oftmals nur geringfügig vom Physikpraktikum für Physiker. Eine Anpassung an das Fach wird dabei oft nur durch Streichen einzelner Versuchsteile oder ganzer Versuche vorgenommen (vgl. Theyßen, 2000; Kissmann et al., 2010). Die ursprünglich für Studierende der Physik entwickelten Praktika können damit nicht die physikalischen Kompetenzen vermitteln, die für andere Studienrichtungen relevant sind (z. B. Theyßen, 2000, 2005). Aufgrund dieser Kritik wurden verstärkt adressatenspezifische Praktika entwickelt (z. B. Plomer et al., 2010; Nagel & Wolny, 2013; Huber & Jessen, 2013). Durch die stärkere Ausrichtung dieser Praktika auf die zukünftigen Tätigkeitsprofile der entsprechenden Studiengänge wurde das Ziel verfolgt, die wahrgenommene Relevanz der Studierenden in Bezug auf die Praktika zu erhöhen und dadurch einen positiven Einfluss auf deren Motivation und deren Lernerfolg zu nehmen (z. B. Theyßen, 2000; Nagel & Wolny, 2013).

Den Auftakt der Entwicklung adressatenspezifischer Praktika hat Theyßen (2000) mit der systematischen Neuentwicklung und Evaluation eines Praktikums für Studierende der Medizin an der Heinrich-Heine-Universität Düsseldorf gemacht. Theyßen (2000) hat dazu die Methode der didaktischen Rekonstruktion nach Kattmann et al. (1997) basierend auf der konstruktivistischen Lerntheorie (z. B. Aufschnaiter & Welzel, 1996) auf die Entwicklung eines physikalischen Laborpraktikums in der Hochschullehre übertragen. Basis des Ansatzes ist die Klärung der fachlichen Inhalte und Zielsetzung einer Lernumgebung (fachliche Klärung) sowie der kognitiven und motivationalen Voraussetzungen (z. B. Vorwissen) der Lernenden und gegebenenfalls auch der Vorstellung der Lernenden bezogen auf den Lerngegenstand (Lernendenperspektive). Diese zwei gleichberechtigten Komponenten fachliche Klärung

und Lernendenperspektive bilden die Grundlage für die didaktische Strukturierung, also die Entwicklung des Unterrichtsgegenstandes. Deshalb hat Theyßen (2000) die fachliche Klärung und die Lernendenperspektive mit Fokus auf die kognitiven und motivationalen Faktoren als Basis für die didaktische Strukturierung des Praktikums für Medizinstudierende eingesetzt.

Als eines der Ziele ihres Projektes nennt Theyßen (2000) in der abschließenden Diskussion ihrer Dissertation, dass ihre Arbeit weitere „Neuentwicklungen initiieren“ soll. Obwohl sich Neuentwicklungen von Praktika finanziellen und strukturellen Herausforderungen stellen müssen, wurden in Folge dieser Pionierarbeit eine Reihe adressatenspezifischer Praktika entwickelt. Der Ansatz von Theyßen (2000) hatte sogar Auswirkungen auf die Praktika für Physiker. So hat Neumann (2004) Theyßens (2000) Vorgehensweise der Methode der didaktischen Rekonstruktion aufgegriffen, um an der Heinrich-Heine-Universität Düsseldorf mit einem physikalischen Praktikum spezifisch entwickelt für Physiker der Diskrepanz zwischen formulierten Lernzielen und tatsächlich im Praktikum Gelernten entgegenzuwirken (siehe auch Neumann & Welzel, 2007). Neumanns (2004) Praktikum teilt sich dazu in drei aufeinander aufbauende Teile mit den Schwerpunkten Messgeräte, Methoden und Projektarbeit auf. In den ersten Praktikumsteilen werden die Studierenden dabei durch den Erwerb experimenteller Kompetenzen anhand Aufgaben mit zunehmender Komplexität auf das eigenständige Arbeiten in der Projektphase vorbereitet. Auch in der Entwicklung des physikalischen Praktikums für Lehramtsstudierende der RWTH Aachen (Frühwein & Heinke, 2005, 2007) und in einer aktuellen Weiterentwicklung dieses Praktikums (Joußen et al., 2019) erhöht sich im Laufe des Praktikums der Schwierigkeitsgrad, indem die Unterstützung der Lernenden mit zunehmendem Lernfortschritt schrittweise zurück genommen wird. Weitere Beispiele für adressatenspezifische Praktika, die auf Theyßens (2000) Ansatz basieren, sind in den Studienrichtungen Meteorologie (T. Haller, 2011), Ernährungswissenschaften (Nagel & Wolny, 2013) und Biologie (Kissmann et al., 2010; Huber & Jessen, 2013; Schwarz et al., 2013a, 2013b) zu finden.

Die genannten Neuentwicklungen folgten durchweg der von Theyßen (2000) etablierten Methode der didaktischen Rekonstruktion. Ausgangspunkt dieser Methode ist die Klärung der fachlichen Zielsetzung. Die Klärung zur Ausrichtung des Praktikums für die Studierendengruppe kann dabei über Befragungen fachlicher Experten und Lehrender sowie über Analyse von Lehrbüchern und Modulbeschreibungen weiterführender Lehrveranstaltungen erfolgen (z. B. Theyßen, 2000; Neumann, 2004; Nagel & Wolny, 2013). Zur fachlichen Klärung zählt nicht nur die Feststellung spezifischer Lernziele, sondern auch die Auswahl von fachlichen Themen (z. B. Neumann, 2004) und Kontexten (z. B. T. Haller, 2011), die aus Sicht der Experten der Studierendengruppe gerecht werden. Neumann (2004) nimmt hier auch die Auswahl von Messgeräten und Methoden hinzu, die authentisch für die späteren Tätigkeiten der Studierendengruppe sind. Aus der fachlichen Klärung kann auch die Forderung

nach einem transparenten Praxisbezug und relevanten Versuchen in Hinsicht auf das Berufsfeld der Studienrichtung resultieren, die sich unter anderem in der Versuchsauswahl widerspiegelt (z. B. Theyßen, 2000). Dadurch versprechen sich die Befragten und Entwickler*innen das Erreichen der spezifischen Lernziele der Lernenden und eine höhere Lernmotivation (z. B. Theyßen, 2000; Nagel & Wolny, 2013). Theyßen (2000) richtet die Versuche des Praktikums für Mediziner*innen beispielsweise stärker auf medizinische Diagnostik (z. B. Versuch zur Röntgenstrahlung mit Bezug auf Röntgendiagnostik) und anschauliche Funktionsmodelle (z. B. Aufbau zur Strömungsmechanik als Modell für den Blutkreislauf) aus. Auch die Interpretation der physikalischen Ergebnisse im studienfachbezogenen Kontext kann gefordert werden, was wiederum die Praktikumsaufgaben prägt. So macht Plomer (2011) den Bedarf einer Vernetzung von physikalischen Begriffen und Zusammenhängen mit studienfachbezogenen Anwendungen deutlich und verknüpft in den Versuchsanleitungen physikalische und studienfachbezogene Fragestellungen.

Aus der fachlichen Klärung der vorherigen Arbeiten können damit drei elementare Gestaltungselemente abgeleitet werden: Die Festlegung der Lernziele für die Studierendengruppe, die Auswahl relevanter Versuchsinhalte sowie relevanter Messtechnik in Hinsicht auf das Berufsfeld der Studienrichtung und die Vernetzung physikalischer und studienfachbezogener Fragestellungen.

Die zweite Komponente der Vorgehensweise von Theyßen (2000) ist die Berücksichtigung der Lernendenperspektive. Ein wesentliche Aspekt ist dabei das Vorwissen der Lernenden, das in einer adressatengerechten Versuchsplanung berücksichtigt werden muss (Theyßen, 2000). Um die kognitiven und auch motivationalen Voraussetzungen der Studierendengruppe zu erfassen und in der Entwicklung berücksichtigen zu können, wurden Befragungen der Lernenden (z. B. Nagel & Wolny, 2013), Lernprozessuntersuchungen (z. B. Theyßen, 2000) und als Anhaltspunkt auch Inhaltsanalysen der Grundvorlesungen vor dem Praktikum (z. B. T. Haller, 2011) vorgenommen. Die Berücksichtigung des Vorwissens und damit eine Anpassung des Schwierigkeitsgrades kann über ein geeignetes Maß an Unterstützungsmaßnahmen beispielsweise in Form entsprechender Fragestellungen in den Versuchsbeschreibungen vorgenommen werden, um einen Lernfortschritt zu ermöglichen (z. B. Theyßen, 2000; Nagel & Wolny, 2013). Je nach Studierendengruppe kann diese Unterstützungsmaßnahme auch schrittweise reduziert werden, um die spezifischen Lernziele zu erreichen (z. B. Neumann, 2004; Frühwein & Heinke, 2005). Im Physikpraktikum für Physiker von Neumann (2004) beispielsweise bauen die Versuche stark aufeinander auf, um einen geführten Übergang von einer stark unterstützten Einarbeitung in die physikalischen Messgeräte zu einer selbstständigen Projektarbeit zu schaffen.

Die Lernendenperspektive umfasst zudem auch motivationale Aspekte. Nagel und Wolny (2013) zeigen, dass Studierende der Ernährungswissenschaften vor und nach Teilnahme an einem klassischen physikalischen Praktikum die Relevanz der Lehrveranstaltung kaum erkennen. Nicht nur von Lehrenden, sondern auch von Ler-

nenden wird daher der Bezug der Praktikumsversuche zum Studienfach und zum Tätigkeitsfeld der Studienrichtung gefordert (z. B. T. Haller, 2011). Um den Studierenden die Relevanz der physikalischen Inhalte und Methoden des Praktikums und den Bezug zum Studienfach transparent zu verdeutlichen, werden in allen adressatenspezifischen Praktikumsversuchen sinnstiftende Kontexte aus dem Tätigkeitsumfeld des Studienfaches genutzt. Dazu wird bereits in einem einleitenden Kapitel der Versuchsbeschreibung der Bezug zum Studienfach und die Relevanz der physikalischen Inhalte für Herausforderungen des späteren Berufs expliziert (z. B. Theyßen, 2000). Dieser Kontextualisierung wird oft ein Schwerpunkt bei der Entwicklung adressatenspezifischer Praktikumsversuche eingeräumt, um durch die verdeutlichte Relevanz die Lernmotivation der Studierenden zu steigern (z. B. T. Haller, 2011).

Damit können aus den Erfahrungen der bisherigen Arbeiten bezogen auf die Lernendenperspektive weitere elementare Gestaltungselemente abgeleitet werden: Die Anknüpfung an das Vorwissen der Lernenden, die Anpassung des Schwierigkeitsgrades an die Studierendengruppe und Bezugnahme auf Kontexte aus dem Interessenfeld der Studierendengruppe.

Zusammengenommen können daher anhand der langen Reihe an Neuentwicklungen sechs allgemeine Gestaltungsprinzipien für adressatenspezifische Praktika und Praktikumsversuche als Grundlage für die Neuentwicklung eines spezifischen Anfängerpraktikums für Lehramtsstudierende des Faches Physik an der CAU identifiziert werden:

- Festlegung der spezifischen **Lernziele** für die Studierendengruppe
- Anknüpfung an das **Vorwissen** der Lernenden und entsprechende Anpassung des Schwierigkeitsgrades an die Studierendengruppe
- Auswahl relevanter **Versuchsinhalte** in Hinblick auf das Berufsfeld der Studienrichtung
- Nutzung von **Messtechnik**, die typisch ist für das Berufsfeld der Studienrichtung
- Bezugnahme auf **Kontexte** aus dem Interessenfeld der Studierendengruppe
- **Vernetzung** physikalischer und studienfachbezogener Fragestellungen

3 Wirkung adressatenspezifischer Laborpraktika

Eingangs wurde als zweite grundlegende Aufgabenstellung dieser Arbeit die Frage formuliert, wie sich ein adressatenspezifisches Laborpraktikum tatsächlich auf die motivationalen Faktoren und den Lernerfolg der Studierenden auswirkt. Wie im vorangegangen Abschnitt berichtet wurde, zielen die adressatenspezifischen Praktika auf eine Steigerung des Lernerfolgs und der Motivation durch eine Erhöhung der wahrgenommenen Relevanz des Praktikums für die Studierenden ab (z. B. Theyßen, 2000; Nagel & Wolny, 2013). Um dieser Annahme nachzugehen ist zunächst eine Auseinandersetzung mit der Forschung zu Lernerfolg (Abschnitt 3.1), Motivation und Interesse (Abschnitt 3.2), Relevanzwahrnehmung (Abschnitt 3.3) und Kontextualisierung (Abschnitt 3.4) notwendig. Nur vor diesem Hintergrund können theoriebasierte Hypothesen über die zugrundeliegenden Wirkmechanismen von adressatenspezifischen Laborpraktika und damit ein Wirkmodell abgeleitet werden (Abschnitt 3.5). Basierend auf dem aktuellen Stand der Forschung zu adressatenspezifischen Praktika werden dann die offene Fragen im Forschungsfeld identifiziert und damit die Ziele dieser Arbeit abgeleitet (Abschnitt 3.6).

3.1 Lernerfolg

Die Weiter- oder Neuentwicklung einer Lernumgebung, wie einem Physikpraktikum, geht mit dem Ziel einher, die Effektivität zu verbessern. Die Effektivität eines physikalischen Praktikums ist nicht allgemein und eindeutig definiert und kann aus verschiedenen Perspektiven betrachtet werden (z. B. Nagel, 2009). Damit ist die Untersuchung der Effektivität eines Praktikums keine triviale Aufgabe. Für diese Studie wird die Effektivität des Praktikums am Lernerfolg, also der Entwicklung der in den Lernzielen angestrebten Kompetenzen und Kenntnisse, gemessen. Die grundlegenden fachlichen Lernziele eines Physikpraktikums sind die Wiederholung und Vertiefung von Fachwissen aus den Grundlagenvorlesungen durch Verbindung mit der Praxis und der Erwerb verschiedener experimenteller Kompetenzen (Neumann, 2004, siehe Abschnitt 2.1). Im Folgenden sollen das Fachwissen sowie die experimentelle Kompetenz näher betrachtet werden, um den Lernerfolg im Praktikum weiter zu definieren.

Das Fachwissen umfasst Kenntnisse über die akademischen Inhalte der jeweiligen Fachrichtung, über übergreifende Konzepte und auch über Methoden zur Gewinnung neuer Kenntnisse im Fachbereich (Gess-Newsome, 2015; Shulman, 1986). Die Vertiefung dieses Fachwissens wird im Praktikum als eines der grundlegenden Lernziele und damit als Lernerfolg gewertet.

Das zweite grundlegende Lernziel im Praktikum ist der Erwerb experimenteller Kompetenzen (Abschnitt 2.1; Neumann, 2004). Da es im Gegensatz zum Fachwissen keine allgemeingültige Definition experimenteller Kompetenz gibt, ist es an dieser Stelle für weitere Betrachtungen der Effektivität von Praktika erforderlich, das ak-

tuelle Verständnis von experimenteller Kompetenz in dieser Arbeit kurz auszuführen. Nach Weinert (2001) werden Kompetenzen als die kognitiven Fähigkeiten und Fertigkeiten sowie die motivationalen, volitionalen und sozialen Bereitschaften aufgefasst, die erforderlich sind, um Problemstellungen zu lösen. Wenn diese Probleme durch Experimente gelöst werden, die entsprechende Evidenz zur Beantwortung der Aufgabenstellung liefern, dann sind experimentelle Kompetenzen erforderlich (vgl z. B. Heidrich, 2016). Experimentieren ist allerdings ein komplexer Prozess mit vielen einzelnen zu lösenden Problemstellungen und durchzuführenden Teilhandlungen, die wiederum auch vom Experiment selbst abhängen (vgl. z. B. Hofstein & Lunetta, 2004; Heidrich, 2016). Damit ist auch die experimentelle Kompetenz nicht als einzelne Fähigkeit beschreibbar. Dies ist vermutlich auch der Grund, warum eine allgemeingültige und einheitliche Definition nicht zu finden ist (vgl. z. B. Hodson, 1996; Hofstein & Lunetta, 1982, 2004; Emden, 2011). Die Komplexität des Experimentierprozesses und damit auch der Erfassung experimenteller Kompetenzen zeigt sich auch in der Vielzahl der Modelle der experimentellen Kompetenz (vgl. z. B. Emden, 2011; Schreiber, 2012; Maiseyenka et al., 2013; Heidrich, 2016). In den Modellen wird experimentelle Kompetenz typischerweise in Teilkompetenzen zerlegt, die die verschiedenen Teilaufgaben des Experimentierprozesses betreffen. Ein Experimentierprozess ist allerdings nicht als linear mit fester Reihenfolge der Handlungsschritte zu verstehen, sondern kann auch unvollständig oder iterativ sein, wenn zum Beispiel Probleme auftreten und der Versuchsplan variiert werden muss (z. B. Schreiber et al., 2009; Heidrich, 2016; Dounas-Frazer et al., 2016). In dieser Arbeit orientiert sich der Begriff experimentelle Kompetenz am Modell des Projektes eXkomp (Schreiber et al., 2009; Schreiber, 2012), das auch Rehfeldt (2018a) zur Untersuchung der Qualität von Laborpraktika in Hinblick auf experimentelle Kompetenzen angewendet hat. Das eXkomp-Modell strukturiert die experimentelle Kompetenz (nach Hammann, 2004; Klahr et al., 2000) anhand der drei Phasen Planung, Durchführung und Auswertung des Experimentierprozesses, die für viele Beschreibungen des Experimentierens zu finden sind (vgl. Emden, 2011). Insgesamt wurden diese drei Phasen wiederum in 13 Teilaspekte unterteilt (Schreiber et al., 2009; Schreiber, 2012), zu denen beispielsweise der Aufbau der Versuchsanordnung und die Dokumentation der Messungen in der Durchführungsphase gehören. Die experimentelle Kompetenz lässt sich nach dem Vorschlag von Heidrich (2016) anhand von drei Kriterien bewerten: Richtigkeit der Anwendung einzelner Fähigkeiten und Fertigkeiten (z. B. Maiseyenka et al., 2010), der Strukturiertheit bei der Abfolge von Fähigkeiten und Fertigkeiten (z. B. Emden, 2011), und der Zielorientiertheit beim Experimentierprozess mit Blick auf die Lösungsstrategie (z. B. Klahr et al., 2000). Damit geht die Bewertung der experimentellen Kompetenz über ein bloßes „richtig oder falsch?“ hinaus und verdeutlicht die Schwierigkeit der Erfassung dieser Kompetenz. Die Zunahme experimenteller Kompetenz als Teil des Lernerfolg im Praktikum ist damit eine vielschichtige Größe.

Zusammengefasst ist das Ziel der Neuentwicklung des adressatenspezifischen Praktikums, einen möglichst hohen Lernerfolg zu erreichen. Der angestrebte Lernerfolg des Lehramtspraktikums an der CAU wird in erster Linie als das Erreichen der Hauptlernziele verstanden und daher als Vertiefung von Fachwissen und Erwerb verschiedener experimenteller Kompetenzen definiert.

3.2 Motivation und Interesse

In einem nächsten Schritt ist nun zu klären, wie die Verbesserung dieses Lernerfolgs bzgl. Fachwissen und experimenteller Kompetenz durch die Konzeption des (Lehramts-)Praktikums erreicht werden kann. Zum einen sind die Entwickler*innen natürlich daran interessiert, die Qualität der Lernumgebung, in diesem Fall also der Praktikumsversuche, zu optimieren. Zum anderen können zusätzlich leistungsbeeinflussende Merkmale der Studierenden Betrachtung finden. Kognitive Faktoren wie Vorwissen und Intelligenz haben einen hohen positiven Einfluss auf den Lernerfolg (vgl. Schiefele, Krapp & Schreyer, 1993), können aber von außen wenig beeinflusst werden. Aber auch motivationale Faktoren der Lernenden verbessern den Lernerfolg (z. B. Schiefele, Krapp & Schreyer, 1993). In adressatenspezifischen Praktika kommt daher der Förderung motivationaler Aspekte eine besondere Rolle zu (z. B. Theyßen, 2000), um den Lernerfolg zu verbessern.

Motivation wird definiert als Ausrichtung einer Handlung auf einen angestrebten Zustand als Folge dieser Handlung (Rheinberg, 2002). Das zentrale Definitionskriterium für Motivation ist dabei der Bezug zur Handlung oder Handlungsbereitschaft einer Person (z. B. Krapp, 1992; Schiefele, 2009). Die im Kontext dieser Studie zur Erhöhung des Lernerfolgs angestrebte Lernmotivation wird entsprechend als Zustand verstanden, sich Wissen über einen Lerninhalt aneignen zu wollen. Gründe der Lernenden für das Erstreben des Wissens können beispielsweise sein, dadurch soziale Anerkennung oder Kompetenzerleben zu erreichen (z. B. Schiefele, 2008).

Die Begriffe Motivation und Interesse werden im alltäglichen Sprachgebrauch oft synonym verwendet, müssen aber im Rahmen dieser Arbeit sowie in der Fachdidaktik generell unterschieden werden. Interesse bezieht sich anders als die Motivation auf eine spezifische Beziehung zwischen einer Person und einem Gegenstandsbereich (z. B. Krapp, 1992; Hidi & Renninger, 2006; Schiefele, 2009; Schiefele & Schaffner, 2015). Interesse am Lerninhalt ist ein Bedingungsfaktor oder sogar Voraussetzung für das (intrinsisch) motivierte Handeln bzw. Lernen (z. B. Rheinberg et al., 2001; Schiefele, 2008).

Beim Interesse wird unterschieden zwischen individuellem Interesse und situationalem Interesse. Das individuelle Interesse wird als zeitlich relativ stabiles Merkmal einer Person interpretiert. Es entwickelt sich durch eine Auseinandersetzung mit einem Gegenstand oder Lerninhalt über eine längere Zeitdauer. Zum Ausdruck kommt Interesse unter anderem durch positive Emotionen und persönliche Wertschätzung des Gegenstandsbereiches.

Das situationale Interesse wird dagegen kurzfristig durch Interaktion einer Person mit einem Gegenstand oder einem Lerninhalt in einer spezifischen Situation hervorgerufenen. Nach dem Vier-Phasen-Modell zur Entwicklung von Interesse von Hidi und Renninger (2006) ist das situationale Interesse Basis für die Entwicklung von individuellem Interesse. Das situationale Interesse kann auch als State-Komponente und das individuelle als Trait-Komponente aufgefasst werden (z. B. Knogler et al., 2015). Für das situationale Interesse werden zwei ineinander übergehende Phasen unterschieden, die als Catch und Hold (Mitchell, 1993; Knogler et al., 2015) oder als Triggered und Maintained (Hidi & Renninger, 2006) bezeichnet werden. Das situationale Interesse an einem Gegenstandsbereich wird zunächst angeregt (Triggered/Catch) und kann anschließend weiter aufrecht erhalten sein (Maintained/Hold), um sich dann gegebenenfalls zu weniger oder stärker ausgeprägtem individuellen Interesse zu entwickeln. Die Catch-Phase des situationalen Interesses ist dabei geprägt von kurzfristig erhöhter Aufmerksamkeit und positiven Emotionen in Bezug auf den Gegenstandsbereich, während in der langfristigeren Hold-Phase der persönliche Wert und der Wunsch, mehr über den Gegenstandsbereich erfahren zu wollen, zum Ausdruck kommt (z. B. Hidi & Renninger, 2006; Knogler et al., 2015).

Da Interesse am Lerninhalt Bedingungsfaktor oder sogar Voraussetzung für (intrinsische) Lernmotivation ist (z. B. Rheinberg et al., 2001; Schiefele, 2008), kann das situationale Interesse folglich als ein Prädiktor für kurzfristige Lernmotivation aufgefasst werden. Das ist in diesen Betrachtungen insbesondere deshalb relevant, weil die Lernmotivation bei der Bearbeitung einzelner Praktikumsversuche schwer zu erfassen ist. Da die Motivation vor einer Handlung, also in diesem Fall vor der Durchführung des Praktikumsversuches erfasst wird (z. B. Rheinberg et al., 2001), kann auf diese Weise der direkte Einfluss des Praktikumsversuches auf die Lernmotivation nicht erfasst werden. Aus diesem Grund wird in diesen Betrachtungen das situationale Interesse (Catch), das direkt in Bezug auf den Praktikumsversuch erfasst werden kann, als Proxy für die während der Bearbeitung des Praktikumsversuches angeregte Lernmotivation genutzt.

Der positive Effekt von Interesse am Lerninhalt auf den Lernerfolg konnte insbesondere im schulischen Kontext in einigen Studien gezeigt werden (z. B. Schiefele, Krapp & Schreyer, 1993; Köller et al., 2006; Rotgans & Schmidt, 2011). Schiefele, Krapp und Schreyer (1993) fanden in ihrer Metaanalyse einen mittleren Zusammenhang zwischen dem Interesse am Schulfach und der Leistung darin. Im akademischen Kontext konnte ein ähnlicher positiver Zusammenhang gezeigt werden (Rotgans & Schmidt, 2011). Köller et al. (2006) fanden für die Sekundarstufe II ebenfalls Effekte von Interesse auf Lernleistung unter Berücksichtigung diverser Kontrollvariablen. Damit ist das (situationale) Interesse am Lerninhalt unabhängig von Lernvoraussetzungen bedeutsam für den Lernerfolg.

Nachdem nun die Rolle der Lernmotivation bzw. des situationalen Interesses als Bedingungsfaktor für den Lernerfolg im Wirkmodell ausdifferenziert wurde, kommt den motivationalen Aspekten nun zusätzlich eine zweite Rolle im Wirkmodell adressatenspezifischer Praktika zu. Physikalische Praktika haben auch die pädagogische Aufgabe, im Hinblick auf die Ausbildung langfristiger zu motivieren und Interesse zu wecken (z. B. Welzel et al., 1998, siehe auch Abschnitt 2.1). Da motivationale Aspekte in adressatenspezifischen Praktika zentral sind (z. B. Theyßen, 2000), wird die langfristige Steigerung des Interesses und der Lernmotivation ebenfalls als wichtiger (Lern-)Erfolg eines adressatenspezifischen Praktikums betrachtet.

Nach dem Vier-Phasen-Modell zur Entwicklung von Interesse (Hidi & Renninger, 2006) entwickelt sich das langfristigere individuelle Interesse aus dem kurzfristigeren situationalen Interesse am Lerninhalt des Praktikumsversuches. Damit besteht das individuelle Interesse an (experimenteller) Physik über das Praktikum hinaus und wird in dieser Arbeit als Lernerfolg des Praktikums verstanden. Das individuelle Interesse kann wiederum auch als Proxy für entsprechend langfristigere Lernmotivation aufgefasst werden.

Aufgrund der Studienwahl wäre es plausibel, dass Studierende physik-naher Studienrichtungen (z. B. Meteorologie, Physik Lehramt) grundsätzlich im Studium ein sehr großes individuelles Interesse an (experimenteller) Physik haben (z. B. T. Haller, 2011, siehe auch Abschnitt 12.1.1). Wenn das individuelle Interesse der Studierendengruppe sehr hoch ist, dann kann dies wiederum eine messtechnische Herausforderung des per se relativ stabilen Konstrukts bedeuten, wenn dadurch Deckeneffekte kombiniert mit geringen Änderungen über den verhältnismäßig kurzen Zeitraum des Praktikums auftreten. In diesem Fall besteht auch die Möglichkeit, das situationale Interesse Hold als Proxy für das individuelle Interesse heranzuziehen, da es das aufrecht erhaltene Interesse aufgrund des Praktikumsversuches umfasst und nach dem Vier-Phasen-Modell die Vorstufe für die Entwicklung des individuellen Interesses darstellt (Hidi & Renninger, 2006). Das situationale Interesse Hold kann zudem in gewissem Maß als Proxy für aufrecht erhaltene Lernmotivation aufgefasst werden. Im Folgenden wird aber zur besseren Unterscheidung auf das individuelle Interesse verwiesen, wenn Bezug auf den motivationalen (Lern-)Erfolg des Praktikums genommen wird, der durch eine Steigerung des situationalen Interesses (Catch) im adressatenspezifischen Praktikum gefördert werden soll.

Aus den vorangegangenen Betrachtungen folgt somit die erste Wirkannahme für die Entwicklung adressatenspezifischer Praktika:

Wirkannahme 1:
Der Lernerfolg von Studierenden im adressatenspezifischen Praktikum in Form von experimenteller Kompetenz, Fachwissen und individuellem Interesse wird durch Erhöhung des situationalen Interesses am Praktikumsinhalt und damit mittelbar durch Erhöhung der Lernmotivation verbessert.

3.3 Wahrgenommene Relevanz

An dieser Stelle stellt sich nun die Frage, wie man das situationale Interesse im Praktikum und in der Folge die Lernmotivation systematisch steigern kann. Maßnahmen zur Förderung von Interesse bzw. Lernmotivation umfassen die Förderung der Kompetenzwahrnehmung, der Selbstbestimmung, der sozialen Bezogenheit und die Verdeutlichung der Bedeutsamkeit des Lerninhaltes (z. B. Deci & Ryan, 1993; Schiefele, 2004). Während die ersten drei dieser Maßnahmen das Interesse indirekt durch Befriedigung grundlegender psychologischer Bedürfnisse fördern, kann die Steigerung der Relevanz und des subjektiven Wertes des Lerninhaltes für die Lernenden das Interesse direkt und damit gegebenenfalls auch langfristiger über das Praktikum hinaus erhöhen (Schiefele, 2004). Ein weiterer Anhaltspunkt für die wahrgenommene Relevanz als sinnvolle Stellschraube im Praktikum ist, dass Studierendengruppen das herkömmliche Physikpraktikum als wenig relevant erachten (z. B. Nagel & Wolny, 2013). Daher ist eine Steigerung des situationalen Interesses durch eine Erhöhung der von den Studierenden wahrgenommenen Relevanz des Praktikums naheliegend.

Der Begriff Relevanz wird in den Bildungswissenschaften häufig verwendet, eine einheitliche Definition und Verwendung des Begriffes ist jedoch nicht zu finden (Stuckey et al., 2013). Daher ist es erforderlich, im Folgenden den Begriff Relevanz für diese Arbeit zu klären. Grundsätzlich wird die wahrgenommene Relevanz eines Themas oder eines Gegenstandes so verstanden, dass diesem Bedeutsamkeit, Wichtigkeit und Wert beigemessen wird, um angestrebte Ziele zu erreichen (z. B. Bergin, 1999). Lernende nehmen folglich dann einen Lerninhalt als persönlich bedeutsam und relevant wahr, wenn der Lerninhalt mit ihren Zielen und Interessen verbunden ist (z. B. Schiefele, 2004). Stuckey et al. (2013) haben, basierend auf einer Analyse der Literatur seit der 1950er Jahre, eine systematische Differenzierung des Begriffs Relevanz für den naturwissenschaftlichen Unterricht vorgeschlagen. Dabei werden von Stuckey et al. (2013) drei grundlegende Bedeutungsebenen unterschieden: Persönliche, soziale und berufliche Relevanz. Persönlich relevant sind demnach Lerninhalte, wenn diese mit dem Interesse des Lernenden verbunden sind und Wissen und Fähigkeiten für das alltägliche Leben bieten. Sozial bedeutsam werden Lerninhalte gesehen, die darauf vorbereiten aktiv und reflektiert an gesellschaftlichen Diskussionen teilnehmen zu können. In der dritten Bedeutungsebene wird die Relevanz für die Orientierung und Vorbereitung der Schüler*innen für das berufliche Leben betrachtet.

In Bezug auf adressatenspezifische Physikpraktika in der universitären Ausbildung überschneiden sich diese schulisch ausgerichteten Bedeutungsebenen. Aufgrund der Studienfachwahl wird angenommen, dass das Studienfach und die damit verbundene zukünftige berufliche Tätigkeit der gemeinsame Nenner für das Interesse und das angestrebte Ziel der Studierendengruppe im adressatenspezifischen Praktikum sind. Hat sich jemand beispielsweise für die Studienrichtung Physik Lehramt ent-

schieden, dann wird davon ausgegangen, dass die Studienwahl im Interesse an Physik und dem persönlichen Berufsziel Physiklehrkraft begründet ist. Dann folgt daraus, dass Lerninhalte, die in diesem Interessenbereich liegen und entsprechende Kompetenzen zum Erreichen dieses Zieles offensichtlich fördern, als relevant wahrgenommen werden. Sofern nicht anders verwiesen wird, bezeichnet daher wahrgenommene Relevanz oder auch kurz Relevanz im Folgenden die durch die Studierenden wahrgenommene Relevanz in Bezug auf ihre spätere berufliche Tätigkeit basierend auf ihrer Studienfachwahl.

Eine Erhöhung der wahrgenommenen Relevanz des Lerninhaltes im adressatenspezifischen Praktikum wird angestrebt, um das situationale Interesse der Studierenden und damit den Lernerfolg zu erhöhen, so wie es im Abschnitt 3.2 als Ziel formuliert wurde, denn die Steigerung der wahrgenommenen Relevanz wird neben anderen als eine wichtige Maßnahme zur Förderung von (situationalem) Interesse bzw. Lernmotivation betrachtet (z. B. Deci & Ryan, 1993; Prenzel, 1995; Bergin, 1999; Schiefele, 2004). Beispielsweise haben Lewalter und Geyer (2009) einen hohen Effekt der wahrgenommenen inhaltlichen Relevanz auf das situationale Interesse für Schulklassen in naturwissenschaftlich-technischen Museen festgestellt. Prenzel und Drechsel (1996) haben das Interesse in der kaufmännischen Ausbildung untersucht und konnten für diese Lerngruppe zeigen, dass interessiertes Lernen und inhaltliche Relevanz hoch korrelieren. Auch für Studierende unterschiedlicher Studienrichtungen konnte ein positiver Zusammenhang zwischen Relevanz und Studieninteresse gezeigt werden (Schiefele et al., 2003). Schneider und Preckel (2017) haben in einer Übersichtsarbeit über die Einflussfaktoren auf den Wissenserwerb im Studium festgestellt, dass das bedeutungsvolle Lernen mittlere bis hohe Effekte auf den Wissenserwerb hat. Dabei wurde nicht explizit für das Interesse kontrolliert, aber auch ein direkter Einfluss der Relevanz auf den Wissenserwerb ist nicht auszuschließen, da eine Relevanz des Lerninhaltes, auch ohne persönliches Interesse anzuregen, eine Notwendigkeit zum Lernen aufzeigen kann. Daraus folgt, dass die Steigerung der wahrgenommenen Relevanz von Lerninhalten eine Förderung von situationalem Interesse und damit des Lernerfolgs begünstigt. Damit ist die zweite Wirkannahme für die Entwicklung adressatenspezifischer Praktika:

Wirkannahme 2:
Das situationale Interesse der Studierenden und damit mittelbar die Lernmotivation wird durch Steigerung der wahrgenommenen Relevanz des Praktikumsinhaltes gefördert.

3.4 Kontextualisierung

Um als Lernende beurteilen zu können, ob ein Lerninhalt den Zielvorstellungen entspricht und damit relevant ist, muss die Bedeutung des Lerninhaltes für die Lernenden nachvollziehbar sein (z. B. Schiefele, 2004). Damit hat nicht nur der Lerninhalt selbst Einfluss auf die wahrgenommene Relevanz, sondern auch die Gestaltung der Lerngelegenheit. Ein etabliertes Vorgehen zur Verdeutlichung der Relevanz ist daher die Kontextualisierung von Lerninhalten (vgl. z. B. van Vorst et al., 2015; Parchmann & Kuhn, 2018).

Auch der Begriff Kontext wird in den Bildungswissenschaften nicht einheitlich verwendet (z. B. Van Oers, 1998; Parchmann & Kuhn, 2018), sodass an dieser Stelle dargestellt werden muss, wie der Begriff in dieser Arbeit genutzt wird. Finkelstein (2005) hat eine systematische Differenzierung des Begriffs Kontext mit drei Ebenen ausgearbeitet. Auf den zwei äußeren Ebenen wird demnach Kontext als soziokulturelle Rahmung des Lernens und der Lernsituation (z. B. mit beteiligten Personen, Arbeitsformen und materiellen Gegebenheiten) verstanden. In anderen Betrachtungen werden diese zwei Ebenen auch zusammengefasst als Kontextualisierung durch die Lernumgebung (z. B. Parchmann & Kuhn, 2018). Die innere, dritte Ebene umfasst nach Finkelstein (2005) die Gestaltung der Aufgabenstellung. Der Kontext ist dabei der Zugang zu den fachlichen Fragestellungen und bettet den Lerninhalt abgestimmt auf die Lernenden zum Beispiel in nicht-fachliche Situationen und Erlebnisse, realitätsnahe Probleme oder alltägliche Anwendungsmöglichkeiten ein (z. B. van Vorst et al., 2015). Um diese inhaltlichen Kontexte weiter zu spezifizieren, haben van Vorst et al. (2015) basierend auf Literaturanalysen ein Modell zu den Merkmalen von Kontexten (z. B. Bekanntheit, Authentizität) erstellt. Daraus wird besonders die Interaktion zwischen den Eigenschaften von Kontext und Lernenden und die erforderliche Abstimmung dieser deutlich. Für adressatenspezifische Physikpraktika ist die zuletzt beschriebene inhaltliche Kontextebene relevant, da diese den größten Gestaltungsspielraum bietet, um die Praktikumsinhalte mit abgestimmten Kontexten an die Studienfachwahl und die daraus resultierenden beruflichen Ziele der Lernenden anzuknüpfen.

Der Erfolg kontextorientierter Lerngelegenheiten als relevanzfördernd und auch motivations- und zum Teil lernfördernd konnte für Schüler*innen in verschiedenen Projekten wie zum Beispiel „Chemie im Kontext“ (Demuth et al., 2008) und anderen Studien (z. B. Berger, 2002; Häussler & Hoffmann, 2002; Lubben et al., 2005; Bennett et al., 2007) gezeigt werden. Theyßen (2000) hat festgestellt, dass sich eine Kontextualisierung auch im Studium für das adressatenspezifische Physikpraktikum für Mediziner positiv auf die wahrgenommene Relevanz auswirkt. Bei der Kontextualisierung von adressatenspezifischen Praktikumsversuchen wird Bezug genommen auf das Studienfach und die damit verbundenen beruflichen Ziele der Studierendengruppe, da diese den gemeinsamen Nenner bzgl. der Interessen der Studierenden

bilden (Abschnitt 2.3). Eine adressatenspezifische Kontextualisierung von Praktikumsversuchen bietet insbesondere den Vorteil, dass sie mit verhältnismäßig geringem Aufwand und bei eng gesteckten Rahmenbedingungen umgesetzt werden können. Die Kontextualisierung ist allerdings nicht das einzige Gestaltungselement von adressatenspezifischen Praktika (Abschnitt 2.3). Relevanz kann nicht nur durch einen geeigneten Kontext, sondern auch durch geeignete Lerninhalte (d. h. die Versuchsthemen) und -medien (z. B. Messgeräte) verdeutlicht werden. Bei der Formulierung der dritten Wirkannahme für adressatenspezifische Physikpraktika werden daher die Versuchseigenschaften insgesamt und die Kontextualisierung im Besonderen einbezogen:

Wirkannahme 3:
Die wahrgenommene Relevanz des Praktikumsinhaltes wird durch geeignete Gestaltung der Praktikumsversuche insbesondere durch eine Kontextualisierung mit Bezug auf das Studienfach und die zukünftige berufliche Tätigkeit sowie durch eine passende Auswahl an Lerninhalten und -medien gefördert.

3.5 Das Wirkmodell

In den vorangegangen Abschnitten wurden auf Grundlage theoretischer Überlegungen drei Wirkannahmen für die Steigerung des Lernerfolgs (Abschnitt 3.1) durch adressatenspezifische Praktika abgeleitet. Diese Wirkannahmen können nun in ein Modell für die Wirkung adressatenspezifischer Praktika zusammengeführt werden (Abb. 1).

Durch eine adressatenspezifische Gestaltung der Praktikumsversuche, insbesondere durch die Kontextualisierung der Lerninhalte bezüglich der Studienwahl und der zukünftigen beruflichen Tätigkeit, soll die von den Studierenden wahrgenommene Relevanz der Praktikumsinhalte gesteigert werden (Wirkannahme 3, Abschnitt 3.4). Eine höhere wahrgenommene Relevanz der Praktikumsversuche soll wiederum das situationale Interesse der Studierenden an den Praktikumsinhalten und damit mittelbar die Lernmotivation fördern (Wirkannahme 2, Abschnitt 3.3). Eine Erhöhung des situationalen Interesses an den Praktikumsinhalten soll wiederum zu einer Steigerung des Lernerfolgs in Bezug auf experimentelle Kompetenz, Fachwissen und individuelles Interesse führen (Wirkannahme 1, Abschnitt 3.2). Auch ein direkter positiver Einfluss der wahrgenommenen Relevanz der Praktikumsversuche auf den Lernerfolg ist möglich.

Das spannende an diesem Wirkmodell ist, dass diese Wirkkette als Grundlage dienen kann, ein adressatenspezifisches Physikpraktikum zu entwickeln, um den Lernerfolg der Studierenden zu verbessern. Hierbei muss jedoch berücksichtigt werden, dass, auch wenn für die einzelnen Wirkannahmen empirische Belege vorliegen (Abschnitte 3.2 - 3.4), das Wirkmodell als Ganzes und in Bezug auf adressatenspezifische Praktika empirisch zu prüfen ist.

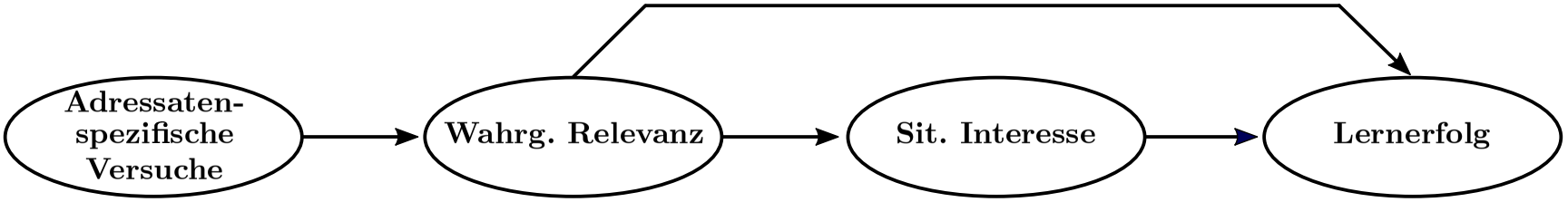


Abbildung 1: *Das theoriebasierte Wirkmodell für adressatenspezifische Praktika. Die Eigenschaften der adressatenspezifischen Versuche (z. B. Kontextualisierung) wirken demnach positiv auf die wahrgenommene Relevanz. Diese fördert wiederum das situationale Interesse und damit den Lernerfolg.*

3.6 Beforschung adressatenspezifischer Praktika

Der letzte Schritt auf dem Weg zur Formulierung der Forschungsfragen dieser Arbeit ist, den aktuellen Stand der Beforschung von adressatenspezifischen Praktika vor dem Hintergrund des zuvor theoretisch hergeleiteten Wirkmodells (Abschnitt 3.5) zu sichten und dadurch offene Fragen in Hinsicht auf die Wirkannahmen zu identifizieren. Da die Neuentwicklung eines adressatenspezifischen Praktikums ein umfangreiches und zeitintensives Projekt ist, ist eine umfassende Beforschung des Praktikums in Bezug auf die angenommene Wirkung nicht erfolgt. Die bisherigen Studien haben jedoch einzelne Teile des Wirkmodells in Bezug auf die Steigerung der wahrgenommenen Relevanz und/oder des Lernerfolges untersucht.

Theyßen (2000) untersuchte im Physikpraktikum für Studierende der Medizin mithilfe einer Befragung zur Selbsteinschätzung zum einen, ob die Studierenden ($n = 184$) die Relevanz fünf neuer Praktikumsversuche im Vergleich zu fünf klassischen Versuchen erkennen. Dabei zeigte sich, dass die Versuche mit medizinischem Kontext in Bezug auf die Relevanz deutlich besser bewertet wurden als die klassischen Versuche. Für diese Versuche stellte Theyßen zudem gemäß dem Wirkmodell eine höhere Motivation der Studierenden fest. Eine Selbsteinschätzung der Studierenden zeigte, dass den Studierenden die physikalischen Zusammenhänge für den kontextualisierten Teil der Praktikumsversuche klarer war. Dies weist zwar auf eine Steigerung des Lernerfolgs durch Kontextualisierung hin. Für eine vollständige Untersuchung des Wirkmodells fehlt allerdings ein Nachweis des Lernerfolgs als Vertiefung von Fachwissen und Erwerb von experimentellen Kompetenzen. Diese Lücke versucht Plomer et al. (2010) durch Untersuchungen im Physikpraktikum für Mediziner*innen in München zu schließen. Für zwei Versuche konnten dort anhand von Wissenstests und Concept Maps bessere Lernerfolge ($d = 1.8$) sowohl in Bezug auf das erwartete fächerübergreifende Wissen, als auch für das physikalische Wissen aufgrund der physiologischen Ausrichtung der Versuche festgestellt werden (Plomer et al., 2010). Diese Ergebnisse unterstützen Theyßens Hinweise auf den Lernerfolg und liefern weitere Indizien für die Wirkannahmen. Für eine Prüfung des gesamten Wirkmodells fehlt bei Plomer et al. (2010) allerdings die Untersuchung der wahrgenommenen Relevanz der Versuche.

Die Versuche von Plomer et al. (2010) wurden in München auch im Physikpraktikum für Biologen eingesetzt und in Bezug auf wahrgenommene Relevanz untersucht (Huber & Jessen, 2013). Die Untersuchung von Huber und Jessen (2013) ist die einzige, in der die Erfassung der wahrgenommenen Relevanz für die einzelnen klassischen und adressatenspezifischen Praktikumsversuche vorgenommen und verglichen wurde, während in den anderen Studien die Relevanz für Versuchsgruppen erhoben wurden. Die Analyse der Relevanz der einzelnen Versuche von Huber und Jessen (2013) zeigt, dass die Relevanz eines biologiebezogenen Versuches (Thema Nervenzellen) deutlich höher ist als für die klassischen Versuche. Der zweite biologiebezogene Versuch (Thema akustische und elektrische Signale) unterscheidet sich

hingegen in Bezug auf die Relevanz nicht von den klassischen Versuchen. Diese unerwartet unterschiedliche Bewertung der Relevanz der kontextualisierten Praktikumsversuche zeigt den Bedarf, tiefer in die Untersuchung der Wirkung auf Ebene der einzelnen adressatenspezifischen Praktikumsversuche einzutauchen. Es stellt sich die Frage, inwieweit die Kontextualisierung und auch andere Eigenschaften der einzelnen spezifisch ausgerichteten Praktikumsversuche ausschlaggebend sind, um die Bewertung der Relevanz positiv zu beeinflussen.

Eine umfassendere Begleitforschung wurde für das Physikpraktikum für Studierende der Ernährungswissenschaft in Wien von Nagel und Wolny (2013) vorgenommen. Die Erfassung der Lernendenperspektive hat ergeben, dass die Relevanz des klassischen Physikpraktikums von den Studierenden kaum erkannt wurde. Die darauf aufbauende Beforschung des neuentwickelten Physikpraktikums zeigt, dass die Lerninhalte des Praktikums von Studierenden, die hauptsächlich Versuche mit Bezug auf Ernährungswissenschaften durchgeführt haben, im Mittel mit mittlerer Effektstärke als signifikant relevanter für ihr Studium eingeschätzt wurden als von Studierenden, die hauptsächlich klassische Praktikumsversuche durchgeführt haben. Auch der Lernerfolg als Zuwachs an Fachwissen zeigt sich bei Durchführung der adressatenspezifischen Praktikumsversuche im Mittel mit kleiner Effektstärke höher. Für die Motivation konnten, aufgrund des Erhebungsinstrumentes wie die Autoren Nagel und Wolny (2013) vermuten, keine Unterschiede zwischen den Studierendengruppen aufgelöst werden. Ohne belastbare Betrachtung der Lernmotivation bzw. des Interesses ist das in dieser Arbeit hergeleitete Wirkmodell ebenfalls nicht vollständig untersucht. Zudem wäre ein Einbezug der experimentellen Kompetenzen in die Untersuchung des Lernerfolgs wünschenswert.

T. Haller (2011), die eine sehr ähnliche Begleitforschung wie Nagel und Wolny (2013) für das Physikpraktikum für Studierende der Meteorologie in Wien eingesetzt hat, konnte hingegen keine Unterschiede in Bezug auf Relevanz, Motivation und Lernerfolg für die kleine Stichprobe ($n_{klassich} = 11$ & $n_{spezifisch} = 24$) nachweisen. An dieser Studie zeigt sich die Herausforderung der kleinen Stichproben bei der Praktikumsbeforschung. Je nach Standort und Studienrichtung nehmen nur wenige Studierende am Praktikum teil, sodass eine Studie entsprechend über mehrere Semester durchgeführt werden muss, um eine ausreichend große Stichprobe zu beforschen. In der Studie von Nagel und Wolny (2013) konnten im Vergleich insgesamt rund 200 Studierende der Ernährungswissenschaften befragt werden und damit Unterschiede in der Relevanz gezeigt werden. Theyßen (2000) hat in ihrer Untersuchung ebenfalls eine Stichprobe von knapp 200 Studierenden der Medizin. Gemäß Cohen (1992) sind rund 100 Probanden notwendig, um für Korrelationen auch mittlere Effekte signifikant nachweisen zu können, sodass sich daran der Studienumfang definiert.

Somit lässt sich feststellen, dass bei ausreichend großer Stichprobe Hinweise und auch Nachweise gewonnen wurden, dass die adressatenspezifischen Physikpraktika eine Steigerung von Lernerfolg und/oder wahrgenommener Relevanz im Mittel erreichen und zudem praxistauglich sind. Dabei bleibt aber die Frage offen, wie es sich mit der wahrgenommenen Relevanz für einzelne Versuche verhält. Huber und Jessen (2013) haben aufgezeigt, dass zur Gewinnung weiterer Erkenntnisse zur Gestaltung adressatenspezifischer Praktika Untersuchungen auf der Ebene der einzelnen Praktikumsversuche angestrebt werden müssen. Anhand des Ergebnisses stellt sich die Frage, welche adressatenspezifischen Gestaltungselemente und Merkmale die einzelnen Versuche aufweisen, die von den Studierenden als besonders relevant wahrgenommen werden. Besonders interessant ist dabei die Wirkung der Kontextualisierung, die alle beschriebenen adressatenspezifischen Versuche als gemeinsames Merkmal haben.

Weiterer Forschungsbedarf ergibt sich für die Wirkannahmen der adressatenspezifischen Praktika. In den Untersuchungen adressatenspezifischer Praktika wird die Steigerung von wahrgenommener Relevanz und Lernerfolg als zwei disjunkte Ergebnisse der Praktikumsneuentwicklungen betrachtet (z. B. Nagel & Wolny, 2013). Ein Zusammenhang der Größen Relevanz und Lernerfolg, wie er als Wirkhypothese in Abschnitt 3.5 formuliert wurde, wurde bisher in den Studien zu adressatenspezifischen Praktika nicht direkt geprüft. Der Lernerfolg wurde in den Studien als Zuwachs an Fachwissen oder einer Selbsteinschätzung erfasst. Eine Untersuchung eines Unterschieds in den experimentellen Fähigkeiten, die ein zentrales Lernziel physikalischer Praktika sind, wurde nicht vergleichend für klassische und adressatenspezifische Praktika vorgenommen. Selten wurde in den Studien ein Augenmerk auf die Verbesserung der Motivation oder des Interesses der Studierenden gelegt. Eine Untersuchung, inwieweit eine höhere wahrgenommene Relevanz der Versuche durch die Studierenden zu einem größeren situationalen Interesse bzw. größerer Lernmotivation führt und in welchem Maße dies tatsächlich zu einem besseren Lernerfolg führt, steht damit noch aus.

Als ein letzter Punkt muss noch angeführt werden, dass keine Erkenntnisse zu einer Beforschung lehramtsspezifischer Physikpraktika mit Bezug auf Relevanz und Lernerfolg veröffentlicht wurden. Damit bleiben die Fragen offen, ob ein Praktikum für Lehramtsstudierende des Faches Physik gemäß der Wirkhypothesen ebenfalls von einer Kontextualisierung der Praktikumsversuche profitieren würde und ob Lernziele durch eine lehramtsspezifische Ausrichtung besser erreicht werden können.

4 Ziele und Forschungsfragen

Das übergeordnete Ziel dieser Arbeit ist die Entwicklung eines spezifischen Anfängerpraktikums für Lehramtsstudierende des Faches Physik. Einführend wurden damit zwei grundlegende Aufgabenstellungen verbunden. Zum einen muss für die Entwicklung geklärt werden, wie ein solches physikalisches Anfängerpraktikum spezifisch für Lehramtsstudierende des Faches Physik idealerweise gestaltet sein sollte. Zum anderen muss herausgefunden werden, inwieweit ein derart gestaltetes Laborpraktikum tatsächlich auf die motivationalen Faktoren und den Lernerfolg der Studierenden wirkt. Nach der vorangegangenen Erarbeitung von Randbedingungen und Gestaltungselementen von adressatenspezifischen Praktika (Abschnitt 2) sowie der theoriebasierten Herleitung eines Wirkmodells und der Analyse des Standes der Forschung (Abschnitt 3) sollen diese grundlegenden Aufgabenstellungen an dieser Stelle nun durch ein Entwicklungsziel und drei konkrete Forschungsfragen ausgeschärft werden.

4.1 Ziele der Entwicklungsarbeit

Ziel dieser Arbeit ist die Entwicklung eines lehramtsspezifischen Anfängerpraktikums. Die Neuentwicklung soll sich dabei innerhalb der gegebenen Rahmenbedingungen (z. B. Organisationsform) an den herausgearbeiteten Kriterienkatalog der Gestaltungselemente orientieren (siehe Abschnitte 2.2 und 2.3).

Das Ziel des Lehramtspraktikums ist es, die kumulative Entwicklung des Professionswissens der Lehramtsstudierenden in den Bereichen Fach und Fachdidaktik bereits zu einem frühen Zeitpunkt des Studiums durch Vernetzung von fachlichen und fachdidaktischen Fragestellungen anzubahnen. Im Gegensatz zum herkömmlichen, rein fachlichen Anfängerpraktikum soll das lehramtsspezifische Anfängerpraktikum stärker auf die Vermittlung und Vertiefung von experimentellen Fähigkeiten und fachlichen Kenntnissen abzielen, die für Lehramtsstudierende im weiteren Studium und in der späteren Schulpraxis notwendig sind. Eine geeignete Auswahl relevanter Versuchsinhalte sowie relevanter Messgeräte und Methoden in Hinsicht auf das Berufsfeld der Studienrichtung soll dies unterstützen.

Durch die adressatenspezifische Konzeption sollen in der Anfangsphase des Bachelorstudiums Physik Lehramt neben den vielen rein fachlichen Lehrveranstaltungen erste Bezüge zum zukünftigen Tätigkeitsfeld als Physiklehrkraft hergestellt werden und dadurch die Wirksamkeit des Praktikums verbessert werden. Der hohe Anteil fachlicher Lehrveranstaltungen in der Studieneingangsphase ohne Bezug zur späteren Lehrtätigkeit dürfte die Lernmotivation nachhaltig beeinträchtigen (z. B. Woitkowski, 2017). Durch die Bezugnahme auf Kontexte aus dem Schulunterricht soll die wahrgenommene Relevanz der Studieninhalte für die Studierenden gefördert werden, um deren situationales Interesse und damit mittelbar deren Lernmotivation und somit den Lernerfolg zu verbessern (siehe Abschnitt 3).

4.2 Forschungsfragen

Bereits in den Abschnitten 1 bis 3 ist deutlich geworden, dass die Entwicklung des Lehramtspraktikums zwingend mit einer Beforschung verknüpft sein muss, um neue Erkenntnisse bezüglich der optimalen Gestaltung und der Wirkung von lehramtsspezifischen Praktika im Besonderen und adressatenspezifischen Praktika im Allgemeinen gewinnen zu können.

Bei dieser Forschung müssen einige Herausforderungen berücksichtigt werden. Ein Praktikum ist eine komplexe, multidimensionale Lernumgebung, sodass eine möglichst breite Erfassung der vielen Qualitätsmaße und Bedingungsfaktoren, ebenfalls sehr umfassend und nicht immer trivial (z. B. experimentelle Kompetenz, vgl. Heidrich, 2016) ist. Bereits die Entwicklungsphase ist oft zeitlich sehr umfangreich, sodass eine zeitlich parallel verlaufende Beforschung eine notwendige, aber zusätzlich herausfordernde Aufgabe ist. Für Praktika in eher kleinen Studiengängen, wie im Lehramtsstudium Physik, müssen zudem die geringen Studierendenzahlen berücksichtigt werden, um durch eine Beforschung über mehrere Praktikumsdurchläufe (Semester) eine ausreichend große Stichprobe ($n \approx 100$) zu erhalten.

4.2.1 Forschungsfragen zur wahrgenommenen Relevanz

In einem ersten Schritt soll die wahrgenommene Relevanz der Versuche untersucht werden. Dies baut auf dem ersten angenommenen Wirkzusammenhang für adressatenspezifische Praktika auf (siehe Abb. 2). Basierend auf den Erkenntnissen aus den Abschnitten 2.3 und 3.6 soll auf Ebene der einzelnen Praktikumsversuche die Wirkung der Kontextualisierung sowie weiterer Versuchseigenschaften auf die wahrgenommene Relevanz geklärt werden. Konkret lauten die Forschungsfragen zur wahrgenommenen Relevanz:

Forschungsfrage 1
In welchem Maße lässt sich durch das Aufzeigen eines Bezuges zum Lehrberuf die von den Studierenden wahrgenommene Relevanz von physikalischen Versuchen im Anfängerpraktikum Lehramt erhöhen?

Forschungsfrage 2
Welche Merkmale charakterisieren die Praktikumsversuche, die von den Lehramtsstudierenden als besonders relevant wahrgenommen werden?

4.2.2 Forschungsfrage zur Wirkung

Im zweiten Schritt, das heißt nach der erfolgreichen Untersuchung der Praktikumsversuche in Bezug auf die wahrgenommene Relevanz und den Versuchseigenschaften, soll die komplexere Untersuchung der übrigen Zusammenhänge im Wirkmodell (Abb. 2) angegangen werden. Es soll untersucht werden, inwieweit eine höhere wahr-

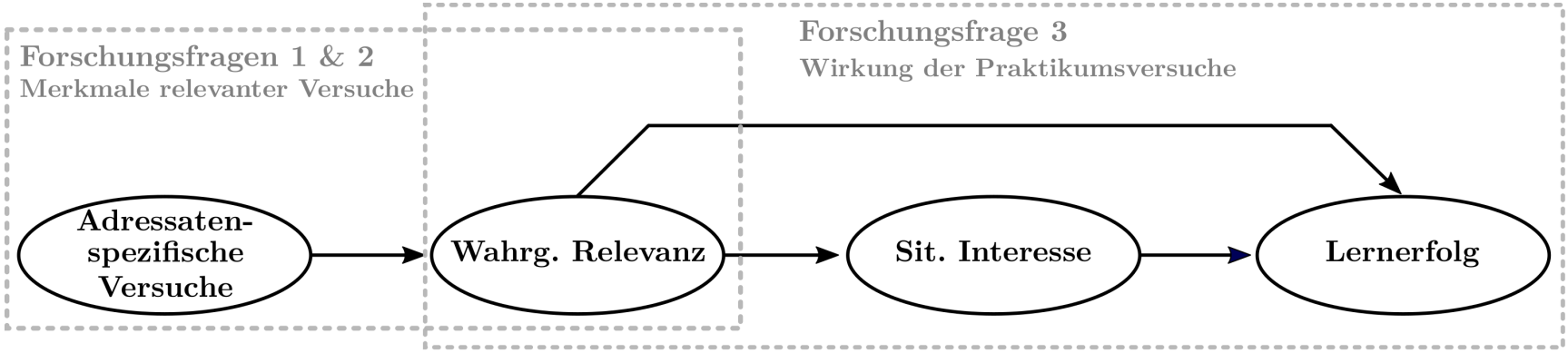


Abbildung 2: *Diese Abbildung ergänzt das das theoriebasierte Wirkmodell für adressatenspezifische Praktika aus Abbildung 1 um die Forschungsfragen. Die Forschungsfragen 1 und 2 umfassen die Untersuchung der Merkmale der als relevant wahrgenommenen Versuche. Die Forschungsfrage 3 befasst sich mit der Prüfung des Wirkmodells in Bezug auf den Zusammenhang von Relevanz, situationalem Interesse und Lernerfolg.*

genommene Relevanz der Versuche durch die Studierenden zu einem gesteigerten situationalem Interesse führt und in welchem Maße dies tatsächlich zu einem größeren Lernerfolg in Form von Kompetenz- und Wissenserwerb sowie einer Steigerung des individuellen Interesses führt. Damit ergibt sich für die Prüfung des Wirkmodells folgende Forschungsfrage:

Forschungsfrage 3
In welchem Umfang bildet sich eine höhere wahrgenommene Relevanz in einem gesteigerten situationalen Interesse und in einem größeren Lernerfolg ab?

4.3 Struktur der Arbeit

Diese Arbeit gliedert sich anhand der kombinierten Entwicklungs- und Forschungsarbeit in fünf große Teile (siehe Abb. 3). Dieser Abschnitt schließt den einführenden Teil I dieser Arbeit mit der Ausformulierung der konkreten Ziele und Forschungsfragen dieser Arbeit ab. Darauf aufbauend folgen nun drei Hauptteile, die die Entwicklungs- und Forschungsarbeit umfassen. Die Neukonzeption des lehramtsspezifischen Laborpraktikums, das gleichzeitig der Forschungsgegenstand zur anschließenden Untersuchung der aufgeworfenen Fragestellungen ist, findet sich in Teil II. Die Forschungsarbeit teilt sich anhand der oben formulierten Forschungsfragen in zwei Teile auf. Zunächst werden die Forschungsfragen 1 und 2 zur optimalen Gestaltung der lehramtsspezifischen Praktikumsversuche untersucht (Teil III). Darauf aufbauend wird anschließend Forschungsfrage 3 zur Wirkung des Lehramtspraktikums analysiert (Teil IV). Auf Basis dieser drei Abschnitte werden die Ergebnisse der Entwicklung und Beforschung diskutiert und Schlussfolgerungen für die Gestaltung und Wirkung adressatenspezifischer Laborpraktika daraus gezogen (Teil V).

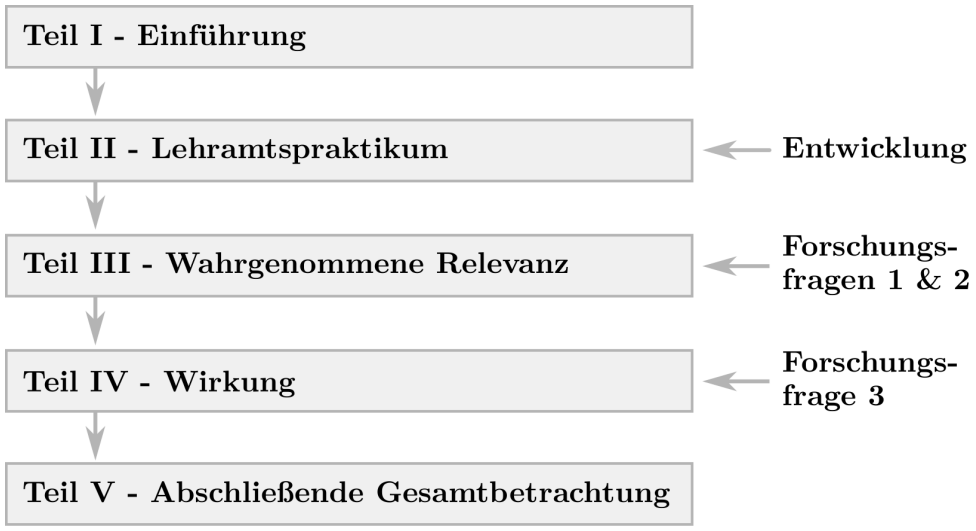


Abbildung 3: *Überblick über die Struktur dieser Arbeit. Nach der Einführung folgen drei Hauptteile zur Entwicklungs- und Forschungsarbeit. Abschließend wird eine Gesamtbetrachtung angestellt.*

Teil II
Lehramtspraktikum

Nachdem im Teil I allgemeine Rahmenbedingungen und Kriterien für die Gestaltung (adressatenspezifischer) Praktika identifiziert wurden (Abschnitt 2) und anhand dessen das Entwicklungsziel dieser Arbeit präzisiert wurde, soll nun in diesem Teil II das Lehramtspraktikum selbst im Fokus stehen. Die Neukonzeption des gesamten physikalischen Anfängerpraktikums für Lehramtsstudierende im Fach Physik an der Christian-Albrechts-Universität zu Kiel (CAU) stellt neben der Forschungsarbeit das Hauptziel dieser Arbeit dar (Abschnitt 4.1). Auf Basis der identifizierten allgemeinen Rahmenbedingungen für physikalische Anfängerpraktika, soll nun in einem ersten Schritt der Neukonzeption eine Spezifizierung der Rahmenbedingungen und damit auch der Gestaltungsfreiräume für diese konkrete Entwicklungsarbeit am Standort Kiel vorgenommen werden (Abschnitt 5). Erst innerhalb dieses aufgestellten Rahmens können dann im zweiten Schritt die identifizierten Gestaltungselemente adressatenspezifischer Praktika nacheinander spezifisch für die Lehramtsstudierenden des Faches Physik ausgeformt werden (Abschnitt 6). Abschließend ergibt sich daraus dann das Gesamtbild der Konzeption der 25 lehramtsspezifischen Praktikumsversuche (Abschnitt 7). Dieses im ersten Jahr des Projektes entstandene lehramtsspezifische Anfängerpraktikum stellt zudem den Untersuchungsgegenstand für die anschließende Forschungsarbeit in den Teilen III und IV dieser Arbeit dar.

5 Rahmenbedingungen **32**

5.1 Lernziele 32

5.2 Einordnung in das Curriculum der Lehramtsausbildung 33

5.3 Ablauf des Praktikums 38

5.4 Lehramtsstudierende 40

6 Neukonzeption Lehramtspraktikum **44**

6.1 Lernziele 44

6.2 Vorwissen 46

6.3 Versuchsinhalte 48

6.4 Messtechnik 54

6.5 Kontext und Vernetzung 57

6.6 Beispiel: Rutherford'scher Streuversuch 61

7 Synopse zur Praktikumskonzeption **67**

5 Rahmenbedingungen

In vielen Fällen kann sich eine Neukonzeption eines Praktikums nur in einem vorgegebenen Rahmen bewegen. Daher ist es notwendig, diese Rahmenbedingungen vor der Neukonzeption des Praktikums zu klären, damit diese dann angemessen berücksichtigt werden können. In diesem Abschnitt werden diese inhaltlichen und organisatorischen Rahmenbedingungen für die Neukonzeption des lehramtsspezifischen Anfängerpraktikums an der Christian-Albrechts-Universität zu Kiel spezifiziert. Dabei ist der Standort Kiel exemplarisch zu sehen, da diese Randbedingungen bei einer Neuentwicklung auch für viele andere Standorte eine Rolle spielen und anhand dieses Beispiels entsprechend übertragen und/oder angepasst werden können. Für die Spezifikation der Rahmenbedingungen für diese Entwicklungsarbeit werden vier wesentliche Eckpunkte betrachtet. Der erste Eckpunkt baut dabei auf den im ersten Teil identifizierten allgemeinen Hauptlernzielen und den damit verbundenen Themenbereichen für physikalische Anfängerpraktika auf (Abschnitt 2.1), die eine grundlegende Anforderung an das zu entwickelte Lehramtspraktikum stellen (Abschnitt 5.1). Der zweite Eckpunkt ist die Einordnung des Lehramtspraktikums in das Curriculum der Lehramtsausbildung (Abschnitt 5.2). Eine Analyse der Anknüpfungspunkte des Lehramtspraktikums an Lehrveranstaltungen vor und nach dem Lehramtspraktikum zeigt Voraussetzungen für die Neukonzeption in Bezug auf eine weitere Präzisierung der Lernziele und die Anknüpfung an das lehramtsspezifische Vorwissen. Der dritte Eckpunkt betrifft den Ablauf des Praktikums (Abschnitt 5.3). Dieser ist durch die Modulbeschreibung des zertifizierten Studienganges in weiten Teilen vorgegeben und hat damit auch mittelbar Auswirkungen auf die Beforschung des Praktikums. Als vierter Eckpfeiler dürfen die Lehramtsstudierenden nicht fehlen, die als Zielgruppe beispielsweise maßgebend für die Auswahl und die lehramtsspezifische Ausrichtung der Praktikumsversuche sind (Abschnitt 5.4). Diese vier Eckpunkte spannen zusammen nicht nur den festen Rahmen, sondern auch die Gestaltungsspielräume für die Neukonzeption im anschließenden Abschnitt auf, die im Folgenden ebenfalls betrachtet werden.

5.1 Lernziele

Hauptlernziele eines physikalischen Anfängerpraktikums sind die Entwicklung verschiedener experimenteller Kompetenzen und die Vertiefung von Fachwissen (Neumann, 2004), wie in Abschnitt 2.1 dargestellt wurde. Aus den physikalischen Grundlagenvorlesungen ergeben sich damit die inhaltlichen Themenbereiche Mechanik, Elektrizitätslehre, Wärmelehre, Atomphysik und Optik, die in einem Anfängerpraktikum in der Physikausbildung in der Regel behandelt werden (Neumann, 2004). Auch in den Anfängerpraktika der CAU für 1-Fach-Studierende und Lehramtsstudierende im Fach Physik herrscht ein Konsens unter den Modulverantwortlichen über diese Hauptlernziele und die klassischen Themenbereiche (CAU, 2007b, 2017c,

2007c, 2017d). Das Erreichen dieser Hauptlernziele stellt damit eine grundlegende Anforderung an die Neugestaltung eines Physikpraktikums.

Verschiedene Studien zeigen, dass die Hauptlernziele in Praktika unterschiedlich gewichtet werden können und im Rahmen dieser Hauptlernziele auch unterschiedliche Detailziele in Praktika verfolgt werden (Abschnitt 2.1, z. B. Boud et al., 1980; Welzel et al., 1998; Nagel et al., 2018). Im Detailbereich bieten die kompetenzorientierten Hauptlernziele und die groben inhaltlichen Themenbereiche damit einen gewissen Spielraum wie zum Beispiel die Auswahl von inhaltlichen Unterthemen und Messtechnik. Diese Spielräume zu nutzen, um neben den Hauptlernzielen auch lehramtsspezifische Lernziele zu adressieren, muss daher das Ziel des neu entwickelten Praktikums sein.

5.2 Einordnung in das Curriculum der Lehramtsausbildung

Da die Entwicklung eines (adressatenspezifischen) Praktikums selten mit einer Neuentwicklung oder einer Rezertifizierung eines Studienganges zusammenfällt, ersetzt das neue Praktikum meist eine andere Lehrveranstaltung. Dadurch gibt die Fachprüfungsordnung eine feste Verortung des neuen Praktikums im Studienplan vor. An die Semesterlage des Praktikums sind vor allem inhaltliche Rahmenbedingungen durch die vorangegangenen und nachfolgenden Lehrveranstaltungen geknüpft. Da diese für die Neukonzeption zum Beispiel eine Rolle für die Spezifizierung von Lernzielen und das Anknüpfen an Vorwissen haben, ist eine Einordnung des Anfängerpraktikums in das Curriculum der Lehramtsausbildung Physik der CAU erforderlich. Das Lehramtsstudium teilt sich an der CAU in drei Säulen auf, die Lehrveranstaltungen der Pädagogik, sowie des ersten und zweiten Faches enthalten. Hier wird der Fokus auf das Fach Physik gelegt, das im Bachelor wiederum in drei Säulen unterteilt werden kann: Vorlesungen und Übungen zu physikalischem Fachwissen (Abb. 4 links), zu mathematischen und theoretischen Grundlagen (Abb. 4 Mitte) und Laborpraktika (Abb. 4 rechts).

An der CAU sieht der Regelstudienplan vor, dass die Lehramtsstudierenden das Anfängerpraktikum im dritten und vierten Semester belegen (CAU, 2017b)[1]. Damit baut es hauptsächlich auf den physikalischen Grundlagenvorlesungen und den zugehörigen Übungen der ersten drei Bachelorsemester zu den Themenbereichen Mechanik und Wärmelehre (Physik I), Elektrizitätslehre und Optik (Physik II) sowie Atom- und Kernphysik (Physik der Materie I und II) auf (Abb. 4; CAU (2017b)). Parallel zu den Grundlagenvorlesungen erlernen die Lehramtsstudierenden im Fach

[1] Durch die Rezertifizierung der Studiengänge an der CAU konnte das Anfängerpraktikum erst mit Einführung der neuen Fachprüfungsordnung (CAU, 2017b), als erste praktische experimentelle Lehrveranstaltung des Bachelorstudiums, näher an den Studienbeginn verschoben werden. Nach der alten Fachprüfungsordnung (CAU, 2007a) wurde das Anfängerpraktikum in der ersten Phase nach der Neuentwicklung nach Regelstudienplan im fünften und sechsten Semester des 2-Fach-Bachelors Physik durchgeführt. Zum Vergleich, die 1-Fach-Studierenden des Faches Physik führen das Praktikum im vierten und fünften Semester durch (CAU, 2017a).

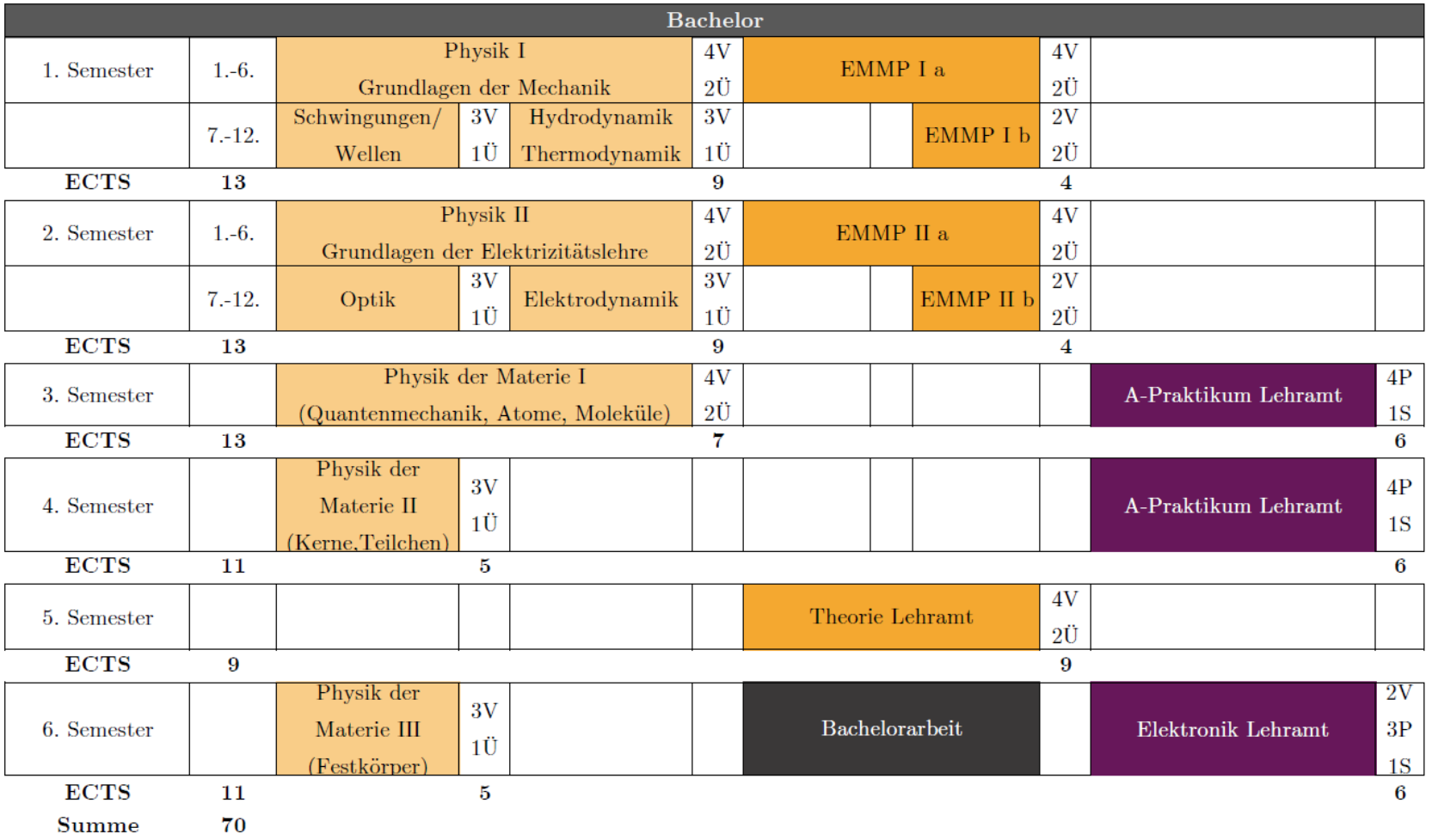


Abbildung 4: *In dieser Abbildung ist der Studienplan des Bachelorstudiums Physik Lehramt an der CAU basierend auf der aktuellen Fachprüfungsordnung CAU (2017b) dargestellt. Die fachlichen Vorlesungen (V) und Übungen (Ü) zu physikalischen Grundlagen sind in hellem orange markiert (links) und zu mathematischen Grundlagen in dunklerem orange (Mitte). Laborpraktika (P) mit ggf. zugehörigem Seminar (S) sind magenta eingefärbt (rechts). Das Anfängerpraktikum Lehramt (A-Praktikum Lehramt) findet sich im 3. und 4. Semester. Die Abschlussarbeit ist dunkelgrau markiert. Angegeben sind außerdem die Anzahl der Semesterwochenstunden (z. B. 4V: 4 Stunden Vorlesung) und Leistungspunkte (ECTS). Die Abkürzung EMMP steht für das Modul Elementare Mathematische Methoden der Physik.*

mathematisches Handwerkszeug im Modul Elementare Mathematische Methoden der Physik (EMMP in Abb. 4). Parallel zum Anfängerpraktikum findet im vierten Semester eine Lehrveranstaltung zur Vermittlung von erstem physikdidaktischem Grundlagenwissen statt (nicht in Abb. 4 dargestellt, da formal Teil der Bildungswissenschaften), die die Studierenden befähigen soll, Physikunterricht zu planen, durchzuführen und zu analysieren (CAU, 2018).

Der fachwissenschaftliche Teil der Lehramtsausbildung an der CAU nimmt damit vor und parallel zum Anfängerpraktikum einen hohen Anteil ein. Im Anfängerpraktikum kann man damit solide physikalische und mathematische Grundlagen voraussetzen. Die fachwissenschaftlichen Lehrveranstaltungen sind dabei identisch mit den Lehrveranstaltungen des 1-Fach-Studiums Physik. Die Lehramtsstudierenden des Faches Physik nehmen aber nicht an allen Modulen des konsekutiv aufgebauten Physikstudiums teil. Daraus folgt, dass die Lehramtsstudierenden im Vergleich zu 1-Fach-Studierenden im Anfängerpraktikum ein anderes Vorwissen mitbringen, das im Physikpraktikum Lehramt berücksichtigt werden muss. Anders als die 1-Fach-Bachelor des Faches Physik an der CAU durchlaufen die Lehramtsstudierenden vor Teilnahme am Anfängerpraktikum kein Elektronik-Grundpraktikum. Das bedeutet, dass das Anfängerpraktikum für die Lehramtsstudierenden die erste universitäre Lernumgebung zum praktischen Experimentieren im Fach Physik ist. Im Gegensatz zum 1-Fach-Studium wird im Lehramtsstudium außerdem keine Lerngelegenheit angeboten, Erfahrungen im Programmieren oder im Umgang mit Software zum Auswerten und Darstellen von Daten zu gewinnen. Da nur eine fachdidaktische Lehrveranstaltung im vierten Semester parallel zum Praktikum für die Lehramtsstudierenden vorgesehen ist, sind die fachdidaktischen Vorkenntnisse im Anfängerpraktikum sehr gering. Zudem lässt sich aus dem Studienplan in Kombination mit den Hauptlernzielen (Abschnitt 5.1) auch schlussfolgern, dass das Anfängerpraktikum eine geeignete Lehrveranstaltung für eine lehramtsspezifische Gestaltung ist, da hier gezielt das bestehende Fachwissen der Grundlagenvorlesungen (Abb. 4 links) vertieft werden soll.

Auf das Anfängerpraktikum folgen im Bachelorstudium die Veranstaltung Theorie Lehramt, eine weitere physikalische Grundlagenvorlesung zur Festkörperphysik (Physik der Materie III), das Elektronikpraktikum für Lehramtsstudierende sowie gegebenenfalls die Bachelorarbeit im Fach (Abb. 4). Das anschließende Masterstudium an der CAU besteht aus Lehrveranstaltungen zur Vertiefung der Fachdidaktik (Abb. 5 links), einem Wahlbereich zur modernen Physik (Abb. 5 Mitte), dem Fortgeschrittenenpraktika (Abb. 5 rechts) sowie einem Praxissemester und gegebenenfalls einer physikdidaktischen Masterarbeit. Damit muss das Anfängerpraktikum nicht explizit anschlussfähig an fachwissenschaftliche Module ab dem dritten Studienjahr sein (Abb. 4 links und 5 Mitte). Durch die lehramtsspezifische Konzeption des neuen Anfängerpraktikums besteht jedoch ein Anknüpfungspunkt an die Vertiefung der Fachdidaktik im Masterstudium (Abb. 5 links). Der direkte Anknüpfungspunkt ist das Fortgeschrittenenpraktikum für Lehramtsstudierende im ersten und zweiten

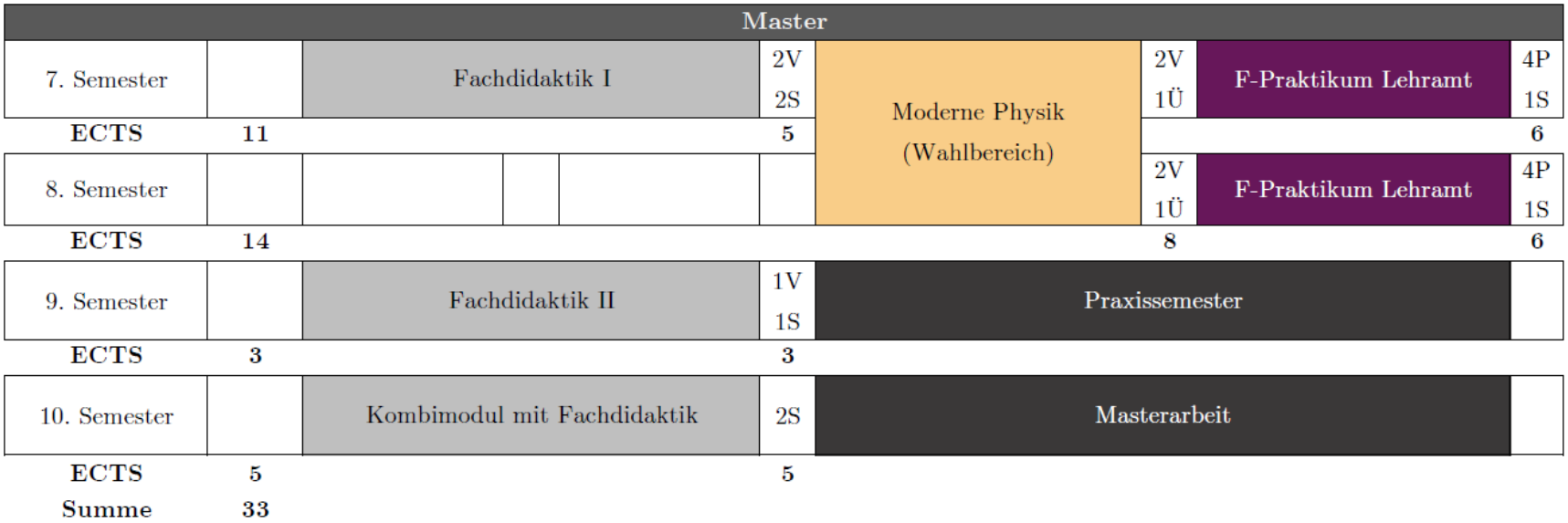


Abbildung 5: *In dieser Abbildung ist der Studienplan des Masterstudiums Physik Lehramt an der CAU zu Kiel basierend auf der aktuellen Fachprüfungsordnung CAU (2017b) dargestellt. Hellgrau eingefärbt sind fachdidaktische Vorlesungen (V) und Seminare (S). Die fachlichen Vorlesungen (V) und Übungen (Ü) im Fach Physik sind orange markiert. Fachliche Praktika (P) mit ggf. zugehörigem Seminar (S) sind magenta eingefärbt. Das Fortgeschrittenenpraktikum Lehramt (F-Praktikum Lehramt) findet sich im 1. und 2. Semester. Das Praxissemester und die Abschlussarbeit, die optional auch im zweiten Fach geschrieben werden kann, sind dunkelgrau markiert. Angegeben sind außerdem die Anzahl der Semesterwochenstunden (z. B. 4V: 4 Stunden Vorlesung) und Leistungspunkte (ECTS).*

Mastersemester an der CAU, das deutlich von einem klassischen Fortgeschrittenenpraktikum abweicht und als Praktikum zum schulorientierten Experimentieren konzipiert wurde (siehe auch *Projekt Lehramtsausbildung Physik 2.0 (LaP 2.0)*, Andersen, Block, Neumann & Wendlandt, 2018). Es baut nicht nur auf fachlichen, fachdidaktischen und bildungswissenschaftlichen Lehrveranstaltungen auf, sondern vor allem auch auf dem (lehramtsspezifischen) Anfängerpraktikum. Während im Anfängerpraktikum der Schwerpunkt auf der Entwicklung des fachlichen Wissens und Könnens liegt (Abschnitt 5.1), rückt im Fortgeschrittenenpraktikum der Fokus auf eine darauf aufbauende Entwicklung des fachdidaktischen Professionswissens, das im einführenden Seminar aufgefrischt wird und im Hauptteil des Praktikums angewendet wird. Der Hauptteil des Fortgeschrittenenpraktikums gestaltet sich so, dass die Studierenden in den zwei Praktikumssemestern selbst ausgewählte Versuche in verschiedenen Themenbereichen mit schultypischen Experimentiermaterialien zur Funktion bringen und ein didaktisches Konzept zu dessen Durchführung in der Schule entwickeln. Von zentraler Bedeutung sind hierbei konkrete Überlegungen zu den Lernzielen der Versuche und wie diese Ziele im Unterricht durch Einsatz fachdidaktischer Mittel und Umsetzung der Versuche erreicht werden können. Dabei werden die Lehramtsstudierenden unterstützt von den Betreuenden, zu denen unter anderen eine abgeordnete Lehrkraft zählt. Der Versuch wird anschließend von den Studierenden in der Gruppe präsentiert und zusammen mit dem didaktischen Konzept diskutiert. Durch die Konzeption typischer Schulversuche sollen die Studierenden ein breites Spektrum an Versuchen kennenlernen und aufeinander abgestimmtes fachliches und fachdidaktisches Professionswissen erwerben, das ihnen einen souveränen Umgang mit Demonstrations- und Schülerexperimenten in der späteren Unterrichtspraxis ermöglicht. Die hierfür erforderliche solide fachliche Basis sowie die experimentellen Fähigkeiten und Fertigkeiten müssen durch das Anfängerpraktikum geliefert werden.

Eine Zusammenfassung der anhand der Studienpläne (Abb. 4 und 5) erörterten Einbindung des lehramtsspezifischen Anfängerpraktikums ins Curriculum und der wesentlichen Anknüpfungspunkte an andere Lehrveranstaltungen ist in Abbildung 6 dargestellt. Die Säulen stellen von links nach rechts Fach, Fachdidaktik sowie Schulpraxis und Pädagogik dar. Das Anfängerpraktikum basiert zum größten Teil auf den fachphysikalischen Grundlagenvorlesungen der ersten Bachelorsemester und der parallel verlaufenden physikdidaktischen Einführung. Das Anfängerpraktikum selbst ist eine grundlegende Basis für das Fortgeschrittenenpraktikum Lehramt im Master, das wiederum fachliche, fachdidaktische und schulpraktische Aspekte zusammenführt. Außerdem bietet das lehramtsspezifische Praktikum Anknüpfungspunkte in die Vertiefung der Fachdidaktik im Masterstudium. Damit ist das lehramtsspezifische Anfängerpraktikum die zentrale Veranstaltung, die die explizite Verbindung zur Fachdidaktik im Bachelor liefert und die als Grundlage eine weitere Vernetzung von Fach und Fachdidaktik gewährleistet.

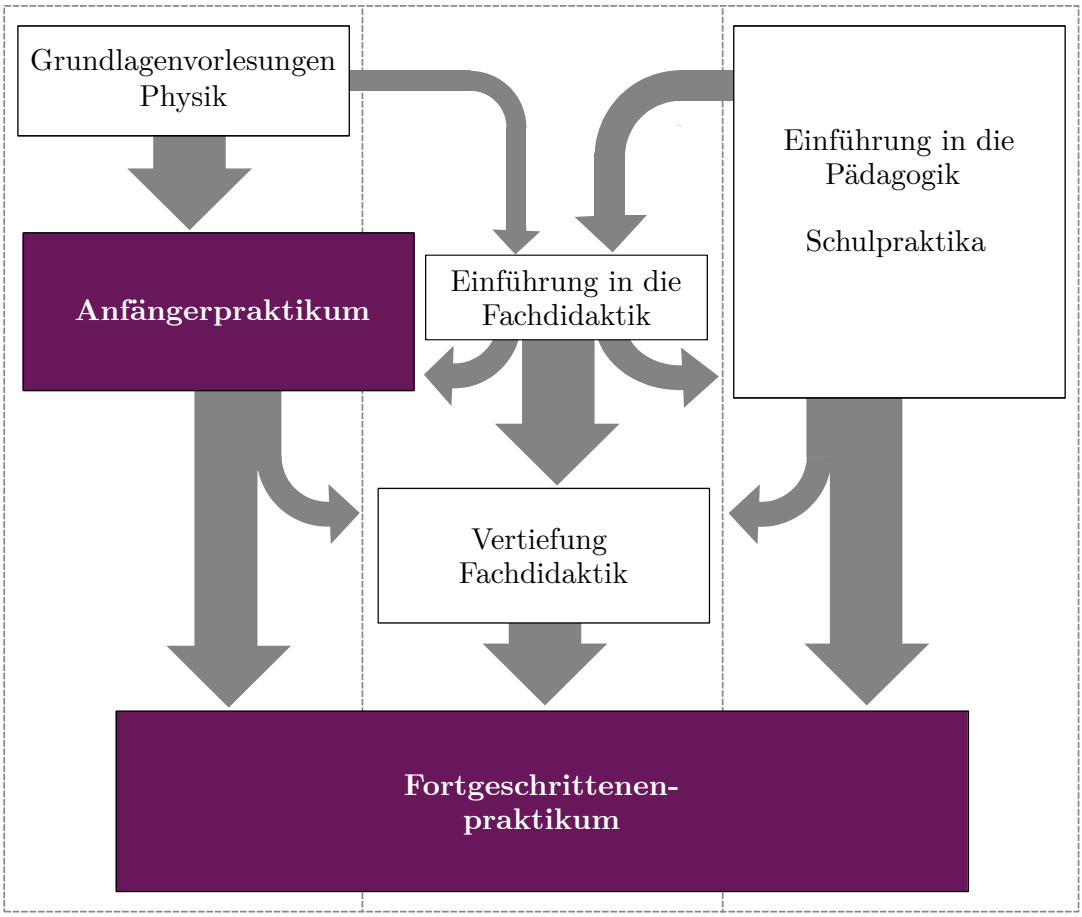


Abbildung 6: *Dieses Schema nach Andersen, Block, Neumann und Wendlandt (2018) verdeutlicht die Vernetzung von Fach (linke Spalte), Fachdidaktik (mittlere Spalte) sowie Pädagogik und Schulpraxis (rechte Spalte) in Bezug auf den laborpraktischen Teil der Lehramtsausbildung in Physik (magenta). Die parallel verlaufenden Veranstaltungen zum Zweitfach wurden in diesem Schema aus Gründen der Übersichtlichkeit nicht dargestellt.*

5.3 Ablauf des Praktikums

Bevor das neue lehramtsspezifische Anfängerpraktikum entwickelt wurde, haben die Lehramtsstudierenden an einem bestehenden klassischen Anfängerpraktikum gemeinsam mit 1-Fach-Bachelor Physik und Studierenden anderer Studienrichtungen wie Materialwissenschaften teilgenommen. Das Praktikum hat sich dabei für die Studierenden des Lehramts lediglich durch eine geringere Anzahl an Versuchen vom Praktikum der 1-Fach-Studierenden unterschieden. Bedingt durch den Zertifizierungsprozess musste das neue Praktikum zunächst formal als Teil dieses alten Praktikums fortgeführt werden. Daraus ergaben sich organisatorische und strukturelle Rahmenbedingungen, die den Spielraum zum Zeitpunkt der Entwicklung in Bezug auf die Organisationsform des neu konzipierten Praktikums erheblich eingeschränkt haben. Damit galt es zwangsläufig, die bestehende Struktur grundsätzlich zu übernehmen, um im Rahmen der Modulbeschreibung zu bleiben und das neue

Praktikum mit dem bisherigen Praktikum kompatibel zu halten. Dadurch entspricht das lehramtsspezifische Anfängerpraktikum in seiner Organisationsform nach wie vor dem klassischen Anfängerpraktikum für Studierende der Physik, wie man es auch an den meisten Hochschulen finden kann (siehe Beschreibung Abschnitt 2.2). Mit der Rezertifizierung der Studiengänge an der CAU und der Einführung der neuen Fachprüfungsordnung (CAU, 2017b) konnten die zwei Teile des Anfängerpraktikums Lehramt als eigenständige Module im Studienplan verstetigt werden. Eine Änderung der Struktur wurde nicht vorgenommen, um die bereits laufenden Untersuchungen nicht zu stören.

Der Ablauf des Anfängerpraktikums, das sich über zwei aufeinanderfolgende Semester erstreckt, ist in Abbildung 7 dargestellt. Beide Praktikumssemester beginnen mit einer Einführungsveranstaltung, bei der vorrangig organisatorische Aspekte geklärt werden und bei der eine Sicherheitsunterweisung in das Labor erfolgt. Im Verlauf des Semesters werden in jedem Praktikumssemester zwölf Praktikumsversuche von den Studierenden bearbeitet, die sich in Themenbereiche mit jeweils vier Versuchen gruppieren. Über das gesamte Praktikum mit zwei Semestern führen die Studierenden somit insgesamt 24 Versuche zu sechs Themenbereichen durch. Die Studierenden führen dabei im Regelfall einen Praktikumsversuch pro Woche an einem Präsenztermin von 3.5 Zeitstunden durch. Gemäß Studienplan beginnen die meisten Studierenden das Praktikum im Wintersemester. Die Reihenfolge der Praktikumsteile und Praktikumsversuche ist aber für die Studierenden grundsätzlich änderbar. Jedes Praktikumssemester endet mit einem Abschlusstermin für die Rückgabe der letzten Protokolle. Für das Lehramtspraktikum wurde ein Zeitaufwand von fünf Semesterwochenstunden und sechs Leistungspunkte errechnet (CAU, 2017b).

Abbildung 7: *Das Anfängerpraktikum Lehramt erstreckt sich über zwei Semester. Das Praktikum verläuft in jedem der beiden Semester nach dem hier dargestellten Schema. Jedes Semester beginnt mit einer Einführungsveranstaltung und endet mit einem Abschlusstermin (orange). Während des Lehramtspraktikums werden in jedem Praktikumssemester jeweils zwölf Praktikumsversuche von den Studierenden durchgeführt (magentafarbene Quadrate), die sich in Themenbereiche mit jeweils vier Versuchen gruppieren.*

Der Versuchstag verläuft gemäß der Beschreibung des klassischen Praktikumstages in Abschnitt 2.2. Das Vorgespräch zur Prüfung der Vorbereitung der Studierenden auf den Praktikumsversuch trägt zu 50 % zur Modulnote bei. Die anderen 50 % der Modulnote bestimmen die Protokolle[2], die die Studierenden anhand ihrer Aufzeichnungen im Laborbuch und einer Auswertung der Messdaten innerhalb einer Woche handschriftlich anfertigen. Für den erfolgreichen Abschluss des Moduls müssen alle zugeteilten Versuche durchgeführt worden sein und die zugehörigen Protokolle ein Testat bekommen haben, also nach maximal einer Verbesserungsmöglichkeit vollständig und mängelfrei sein.

Somit gibt es zwischen dem neuen Anfängerpraktikum für Lehramtsstudierende an der CAU und einem klassischen Anfängerpraktikum wie dem Praktikum für 1-Fach-Studierende des Faches Physik an der CAU kaum Unterschiede in der Struktur. Auch der Ablauf des Praktikumstages inklusive Vorgespräch, Dokumentation der Arbeit im Laborbuch sowie die Nachbereitung in Form eines handschriftlichen Protokolls haben sich im Lehramtspraktikum nicht verändert. Damit ist eine Adaption des neuen Konzepts des Lehramtspraktikums für viele andere Standorte mit bestehendem klassischen Anfängerpraktikum und somit ähnlichen Randbedingungen eine umsetzbare Option.

5.4 Lehramtsstudierende

Den Lehramtsstudierenden kommt im Rahmen dieser Arbeit eine besondere Rolle zu, da sie die Zielgruppe des Praktikums darstellen. Daher müssen sich die Inhalte und der Schwierigkeitsgrad der neu entwickelten Praktikumsversuche an ihnen orientieren. Zudem ist zum Beispiel die Anzahl der Studierenden nicht nur für die Praktikumskonzeption relevant, sondern auch wichtig in Bezug auf die spätere Datenauswertung der begleitenden Beforschung des Praktikums. Daher wird an dieser Stelle ein Überblick über die Lehramtsstudierenden im Praktikum gegeben. Zum Zeitpunkt der Konzeption des Praktikums lagen nur erfahrungsbasierte Anhaltspunkte über die Studierendengruppe des Lehramts Physik aus den vorangegangenen Semestern vor, die sich aber anhand der Beforschung des Praktikums bestätigt haben. Um die Lehramtsstudierenden konkreter charakterisieren und die Passung der Konzeption aufzeigen zu können, wird daher an dieser Stelle auch auf die Beschreibung der Stichprobe der Beforschung des Praktikums vorgegriffen.

Zum Zeitpunkt der Konzeption des Praktikums nahmen üblicherweise etwa 15 bis 20 Lehramtsstudierende pro Semester am herkömmlichen Anfängerpraktikum teil. Die Studierendenzahl pro Semester gibt für die Konzeption des Praktikums den

[2] Die Modulnote wurde mit der neuen Fachprüfungsordnung (CAU, 2017b) ab dem Wintersemester 2017/2018 im Praktikum eingeführt. Für das Vorgespräch können von den Betreuenden maximal 2 Punkte vergeben werden. Im Protokoll bekommt ein direktes Testat 2 Punkte, ein Vortestat mit kleineren Korrekturen 1 Punkt und mit größeren Korrekturen 0 Punkte. Anhand der Gesamtpunktzahl wird die Modulnote errechnet.

Bedarf an Betreuenden, Terminen und Räumen vor. Außerdem bedeutet diese geringe Anzahl an Studierenden, dass im Vorfeld keine repräsentative Erfassung des Vorwissens, der experimentellen Erfahrung oder der Vorstellungen über ein Lehramtspraktikum in einem angemessenen zeitlichen Rahmen vorgenommen werden konnte. In der Pilotierung der in dieser Studie eingesetzten Fragebögen und Tests konnten lediglich wenige Hinweise darauf erkannt werden, dass die Lehramtsstudierenden eher heterogene Vorstellungen über ein Praktikum für Lehramtsstudierende haben und recht unterschiedliche Zweitfächer belegen. Um weitere Informationen zum Vorwissen der Lehramtsstudierenden zu erhalten, wurde daher auf die Analyse des Curriculums (Abschnitt 5.2) zurückgegriffen.

Im Anschluss an die Konzeption wurde das neue Lehramtspraktikum im Zeitraum 2016 bis 2019 über sieben Semester mit einer Vollerhebung aller Teilnehmenden des Lehramtspraktikums beforscht. Dadurch konnten ein paar grundlegende Informationen über die Studierenden erfasst werden, die auch Hinweise aus vorangegangenen Semester bestätigen. Um an dieser Stelle ein konkreteres Bild über die Studierendengruppe zu geben, für die das Praktikum konzipiert wurde, wird daher an dieser Stelle auf die Beschreibung der Stichprobe vorgegriffen. Innerhalb der sieben Semester haben insgesamt 89 Lehramtsstudierende des Faches Physik teilgenommen (Tab. 1). Mit 36 % ist der Anteil der weiblichen Lehramtsstudierenden Physik erfreulich hoch im Vergleich zum 1-Fach-Studium Physik (z. B. 15 % in Sorge et al., 2016). Von den erfassten Studierenden haben 68 (76 %) am gesamten Praktikum, also an beiden Praktikumsteilen, teilgenommen. Die restlichen 24 % haben entweder vorher bereits einen Teil im alten Praktikum absolviert oder im letzten beforschten Semester erst mit dem Praktikum begonnen. Nur in sehr selten Ausnahmefällen ($\approx 3\,\%$) haben die Studierenden nicht beide Praktikumsteile in Folge belegt. Ein Überblick über die Zahl der teilnehmenden Studierenden in den einzelnen Semestern ist in Tabelle 2 dargestellt. Unter Berücksichtigung zweier Semester mit doppelter Jahrgangsstärke[3] ergibt sich damit eine durchschnittliche Anzahl von 18 Lehramtsstudierenden pro Praktikumssemester, wie auch etwa aus den vorangegangenen Jahrgängen geschlossen wurde. Die durchschnittlich 18 Studierenden bzw. neun Zweierteams pro Versuchstermin wurden dabei von zwei Betreuern und zwei wissenschaftlichen Hilfskräften bei der Versuchsdurchführung unterstützt.

Elf verschiedene Zweitfächer wurden für die Lehramtsstudierenden Physik im Forschungszeitraum festgestellt. Dabei liegt der Hauptanteil mit 74% auf Zweitfächern aus dem aus dem MINT-Bereich (Tab. 1, Abb. 8). Mathematik macht mit 42% den Großteil aus und mit Chemie (11%) zusammen etwa die Hälfte der Studierenden. Zu den seltenen Zweitfächern mit jeweils $\leq 2\%$ gehören Geschichte, Philosophie, Englisch und Deutsch. Die unterschiedlichen Zweitfächer zeigen zum einen die

[3] Durch die Rezertifizierung der Studiengänge an der CAU wurde das Anfängerpraktikum vom fünften und sechsten in das dritte und vierte Semester verschoben (CAU, 2007a, 2017b). Durch den Übergang der Studienplanänderung gab es im Wintersemester 2018/2019 und Sommersemester 2019 mit einem doppelten Jahrgang.

unterschiedlichen Interessen der Studierenden auf. Zum anderen können die Zweitfächer auch bedeuten, dass die Lehramtsstudierenden unterschiedliche experimentelle Erfahrungen, mathematisches Wissen oder EDV-Kenntnisse mitbringen. Folglich können Kenntnisse oder Fähigkeiten in diesen Bereichen nicht im Praktikum vorausgesetzt werden. Diese Heterogenität der Studierendengruppe muss bei der Neukonzeption eines Lehramtspraktikums und der Analyse von zum Beispiel (Vor-)Wissen berücksichtigt werden. Eine weiterführende Analyse des Vorwissens ist durch diese lang angelegte Studie möglich geworden (siehe Teil IV) und kann damit für Neukonzeptionen lehramtsspezifischer Praktika an anderen Standorten mit ebenfalls wenig Studierenden genutzt werden.

Lehramtsstudierende (SS16-SS19)	
Gesamtanzahl	89
beide Praktikumsteile^{+}	68 (76%)
weiblicher Anteil	32 (36%)
männlicher Anteil	55 (62%)
2. Fach im MINT-Bereich	66 (74%)
Alter	M=24 (SD=3)

Tabelle 1: *Überblick über die Lehramtsstudierenden im neuen Lehramtspraktikum im Zeitraum Sommersemester 2016 - 2019. Es sind die absolute Anzahl und in Klammern der prozentuale Anteil von der Gesamtanzahl angegeben. Für das Alter der Studierenden bei Studienbeginn ist abweichend ein Mittelwert (M) und in Klammern die Standardabweichung (SD) angegeben. ($^{+}$: Nur ein Teil der Gesamtanzahl an Studierenden hat an beiden Praktikumsteilen im Sommer- und Wintersemester teilgenommen.)*

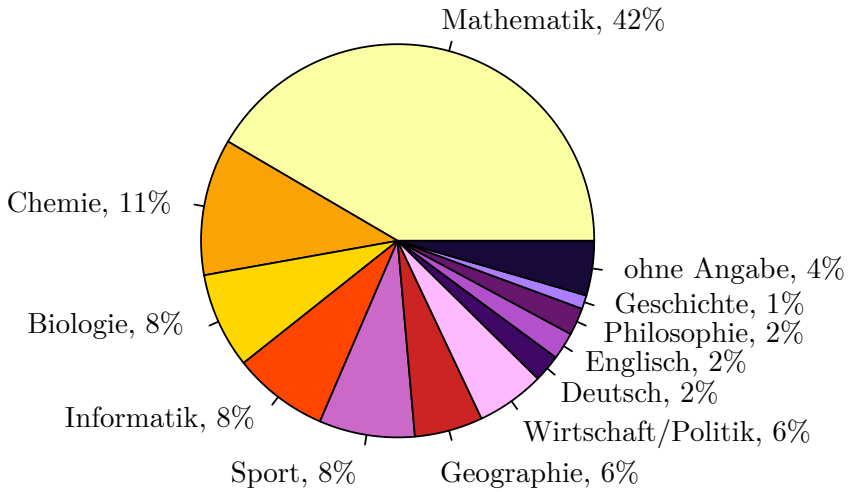


Abbildung 8: *Überblick über den Anteil der verschiedenen Zweitfächer der insgesamt 89 Lehramtsstudierenden im neuen Lehramtspraktikum im Zeitraum Sommersemester 2016 - 2019.*

Praktikumsteil Sommersemester	**Anzahl**	**Praktikumsteil Wintersemester**	**Anzahl**
SS16	14		
		WS1617	18
SS17	16		
		WS1718	24
SS18	22		
		WS1819	33^{+}
SS19	31^{+}		

Tabelle 2: *Überblick über die Anzahl der Studierenden in den Praktikumsteilen in den Sommer- (SS) und Wintersemestern (WS) der Jahre 2016 bis 2019. Ein großer Teil (76 %) der Studierenden hat in diesem Zeitraum am gesamten Praktikum, also an beiden Praktikumsteilen im Sommer- und Wintersemester, teilgenommen, sodass diese Studierenden in dieser Tabelle doppelt geführt sind. ($^{+}$: Im Wintersemester 2018/2019 und Sommersemester 2019 hat aufgrund des Wechsels in die neue Studienordnung (CAU, 2017b) ein doppelter Jahrgang am Praktikum teilgenommen.)*

6 Neukonzeption Lehramtspraktikum

Indem in den vorangegangenen Abschnitten ein Kriterienkatalog für die Gestaltungselemente adressatenspezifischer Praktika erarbeitet wurde und die Rahmen für die Konzeption an der Christian-Albrechts-Universität zu Kiel abgesteckt wurden, wurde die Basis für die nun folgende konkrete Neukonzeption des lehramtsspezifischen Anfängerpraktikums gelegt. Durch die lehramtsspezifische Konzeption der Praktikumsversuche sollen die Studierenden im Praktikum nicht nur auf die Anforderungen des experimentellen Teils des Physikunterrichtens vorbereitet werden. Vielmehr soll auch die wahrgenommene Relevanz der Lehramtsstudierenden in Hinblick auf die Versuche positiv beeinflusst werden, um den Lernerfolg zu steigern. Im Folgenden wird die konkrete Umsetzung des lehramtsspezifischen Anfängerpraktikums an der CAU erläutert.[4] Einzelne Praktikumsversuche werden dabei nur als Beispiele angeführt, können aber anhand von Kurzbeschreibungen im Anhang A.1 nachvollzogen werden. Die vollständigen Versuchsbeschreibungen in der Version aus dem Beforschungszeitraum (Sommersemester 2016-2019) sind im digitalen Anhang B zu finden. Insgesamt wurden 25 Praktikumsversuche innerhalb des ersten Projektjahres auf Basis der Gestaltungsprinzipien (siehe Abschnitt 2.3) entwickelt. Hierbei wurde die Entwicklungsarbeit so gestaltet, das bereits von Anfang an eine parallele Beforschung erfolgen konnte.

Die nächsten Abschnitte strukturieren sich anhand der sechs identifizierten Gestaltungsprinzipien adressatenspezifischer Praktika. Zunächst wird unter Berücksichtigung der Randbedingungen aufgezeigt, wie die **Lernziele** des neuen Lehramtspraktikums auf den Erwerb des fachlichen Professionswissens für die spätere Lehrtätigkeit ausgerichtet sind (Abschnitt 6.1) und an das spezifische **Vorwissen** der Lehramtsstudierenden basierend auf vorangegangenen Lerngelegenheiten anknüpfen (Abschnitt 6.2). Zudem wird dargestellt, wie sich die **Versuchsinhalte** im Lehramtspraktikum an schulrelevanten Themen der schleswig-holsteinischen Fachanforderungen orientiert (Abschnitt 6.3). Weiterhin wird dann verdeutlicht, wie das neue Praktikum schulrelevante Experimentiermaterialien, **Messtechnik** und Software behandelt (Abschnitt 6.4), den Versuch durch entsprechende Versuchsbeschreibungen in einen schulischen **Kontext** situiert und die **Vernetzung** von Fach und Fachdidaktik fördern soll (Abschnitt 6.5). Diesen Abschnitten schließt sich eine Zusammenfassung an (Abschnitt 7).

6.1 Lernziele

Bei der Entwicklung einiger adressatenspezifischer Praktika wurde eine Diskrepanz zwischen den von Experten für die Studierendengruppe formulierten Lernzielen und den angestrebten Zielen im Praktikum erkannt (z. B. Theyßen, 2000; Neumann,

[4] Grundzüge der Konzeption wurden auch in Andersen, Block, Neumann und Wendlandt (2018) veröffentlicht.

2004). Daher ist ein wichtiges Element bei der Entwicklung eines lehramtsspezifischen Praktikums, die Lernziele auf die zukünftige berufliche Tätigkeit der Lehramtsstudierenden auszurichten (vgl. Abschnitt 2.3). Für das Lehramtspraktikum stehen die angestrebten Hauptlernziele physikalischer Praktika weiterhin außer Frage (Abschnitt 5.1). Innerhalb dieses Rahmens bieten sich allerdings auch Spielräume, um auf Detailebene spezifische Lernziele zu adressieren. Diese lehramtsspezifischen Lernziele stellen eine Grundlage für nachfolgende konkrete Maßnahmen zur Konzeption des Lehramtspraktikums dar.

Diese Hauptlernziele im Anfängerpraktikum sind damit für Studierende des Lehramts Physik und für die 1-Fach-Studierenden der Physik gleich. Die Studierenden sollen experimentelle Fähigkeiten erwerben und ihre bestehenden fachlichen Kenntnisse aus den Themenbereichen der Grundlagenvorlesungen wiederholen und vertiefen (Neumann, 2004; CAU, 2017b, Abschnitte 2.1 und 5.1). Das klassische Praktikum für 1-Fach-Studierende der Physik hat über diese Lernziele hinaus allerdings das spezifische Ziel, auf eine spätere Tätigkeit im Forschungslabor vorzubereiten (Neumann, 2004). Lehramtsstudierende des Faches Physik sollen in ihrem Studium hingegen vor allem Kenntnisse und Methoden im Bereich der Vermittlung physikalischer Zusammenhänge sowie Arbeitsweisen erwerben und lernen, den Physikunterricht gleichzeitig für die Schüler*innen interessant und auf dem Stand der Zeit zu gestalten (Großmann & Urban, 2006). Ein Einblick in Forschungsmethoden ist sicher ein wichtiger Aspekt der Lehramtsausbildung, um dies im Physikunterricht adäquat darstellen zu können (Grossman et al., 2005). Dennoch spielt das praktische Forschen für Lehrpersonen nur in seltenen Fällen eine Rolle. 1-Fach-Studierende der Physik sollen beispielsweise im Praktikum Kenntnisse im Umgang mit labortypischen Messgeräten erlangen. Labortypische Messgeräte und damit verbundene Messtechniken unterscheiden sich aber mit ihrer Ausrichtung auf Präzision grundlegend von Messgeräten in der Schulpraxis, die an erster Stelle eine didaktische Funktion erfüllen müssen. Zudem umfasst ein Themenbereich des herkömmlichen Anfängerpraktikums Grundlagenübungen im graphischen Programmiersystem LabVIEW, das zur Datenauswertung im Labor gern genutzt wird, in der Schulpraxis aber keine Rolle spielt. Damit kann das Praktikum in der klassischen Umsetzung die für Lehramtsstudierende im weiteren Verlauf des Studiums, für das Referendariat und für die Schulpraxis notwendigen spezifischen Kompetenzen nur bedingt vermitteln.

Vor dem Hintergrund der Einordnung des Anfängerpraktikums in das Curriculum der Lehramtsausbildung im vorangegangenen Abschnitt 5.2 sollen in diesem Abschnitt detaillierter die Anforderungen der Ausbildungselemente, die im Studienverlaufsplan auf das Anfängerpraktikum folgen, betrachtet werden, um die spezifischen Lernziele des Anfängerpraktikums Lehramt weiter auszurichten. Insbesondere bedeutet dies, dass eine bessere Vernetzung mit dem Fortgeschrittenenpraktikum notwendig ist (Abschnitt 5.2). Der Fokus im Fortgeschrittenenpraktikum liegt auf der Verknüpfung von fachlichem und fachdidaktischem Wissen und setzt damit Fach-

wissen und experimentelle Kompetenzen voraus. Das Anfängerpraktikum muss eine solide fachliche Basis für das fachdidaktisch orientierte Fortgeschrittenenpraktikum Lehramt im 1. und 2. Mastersemester bilden (Abschnitt 5.2). Im Fortgeschrittenenpraktikum standen die Lehramtsstudierenden allerdings bisher oft vor der Herausforderung, die bis dahin meist unbekannte schultypische Messtechnik und Software eigenständig zu nutzen. Auch das theoretische Verständnis fachlicher Grundlagen einiger im Fortgeschrittenenpraktikum behandelter Demonstrations- und Schülerexperimente musste vertieft werden. Dadurch hatte sich der eigentliche Fokus von fachdidaktischen hin zu fachlichen Fragestellungen verschoben, die eigentlich bereits im Anfängerpraktikum hätten vermittelt werden müssen. Diese Problematik beschränkt sich nicht nur auf den Kieler Standort (z. B. Krey & Rabe, 2012). Daher muss eine bessere Abstimmung der Praktika untereinander und eine stärkere Ausrichtung des Anfängerpraktikums Lehramt auf die Vermittlung von Kompetenzen, die notwendig sind für die spätere berufliche Tätigkeit als Physiklehrkraft, angestrebt werden.

Folglich muss ein lehramtsspezifisches Ziel des neuen Praktikums sein, dass die Studierenden experimentelle Fähigkeiten und Kenntnisse speziell im Umgang mit schultypischer Messtechnik und Software erlernen. Zudem ist erforderlich, dass die Studierenden im Lehramtspraktikum insbesondere die fachlichen Kenntnisse auf Hochschulniveau vertiefen, die relevant für typische Experimente des Physikunterrichts sind. Diese Lernziele stellen damit Spezifikationen im Rahmen der Hauptlernziele dar. Nur durch Integration dieser spezifischen Lernziele kann das neue Anfängerpraktikum die Lehramtsstudierenden effektiv auf die spezifischen Anforderungen in der weiteren Lehramtsausbildung und in der späteren Lehrtätigkeit vorbereiten. Details zur Umsetzung der Lernziele innerhalb der gegebenen Rahmenbedingungen werden in den folgenden Abschnitten konkretisiert.

6.2 Vorwissen

Bevor die zuvor formulierten lehramtsspezifischen Lernziele (siehe Abschnitte 5.1 und 6.1) konkret adressiert werden können, muss auf Basis der Inhalte der vorangestellten Lehrveranstaltungen (siehe Abschnitt 5.2) in diesem Abschnitt das Vorwissen der Lehramtsstudierenden berücksichtigt werden.

Anhand der Lehrveranstaltungen in den ersten Semestern des Lehramtsstudiums (siehe Abschnitt 5.2) kann geschlussfolgert werden, dass die Studierenden zu Beginn des Praktikums sowohl über mathematische als auch physikalische Grundlagenkenntnisse verfügen, um sich eigenständig auf die Inhalte eines Versuches und dessen Datenauswertung vorzubereiten. Im Praktikum kann dieses Grundwissen wieder aufgegriffen werden, um es gemäß den Hauptzielen des Praktikums zu vertiefen (Abschnitt 5.1). Eine ausreichende Versuchsvorbereitung wird in Anfängerpraktika von den Betreuenden gewöhnlich in Vorgesprächen abgefragt (Abschnitt 2.2). Im neuen Lehramtspraktikum führen die Betreuenden begleitend zu jedem Versuch in-

tensive, oftmals mehrphasige Gespräche, die über ein Abfragen des vorbereiteten Grundlagenwissens hinaus gehen. In den Gesprächen werden zum Beispiel Theorie und Praxis explizit verbunden, übergreifende Konzepte erarbeitet und Messergebnisse kritisch hinterfragt, um diese Vertiefung des Fachwissens besonders zu fördern. Diese Gespräche bilden damit ein wesentliches Element des Praktikumsversuches und tragen neben den Protokollen zur Modulnote bei (Abschnitt 5.3).

Die Einordnung in das Curriculum (siehe Abschnitt 5.2) hat außerdem verdeutlicht, dass bei den Lehramtsstudierenden, anders als bei den 1-Fach-Bachelorn, keine experimentellen Vorkenntnisse zu Beginn des Anfängerpraktikums vorausgesetzt werden können. Experimentelle Kompetenzen müssen daher im Rahmen des Praktikums systematisch von Grund auf aufgebaut werden. Gleiches gilt für die oft wenig ausgeprägten EDV-Kenntnisse der Studierenden. Auch hier muss eine Basis geschaffen werden, damit diese dann im weiteren Praktikumsverlauf eingefordert werden können. Neben experimentellen Kompetenzen wie Dokumentation der Durchführung im Laborbuch, Datenauswertung, Bestimmung von Messunsicherheiten und Protokollierung müssen daher auch Fähigkeiten im Bereich EDV und Messtechnik vermittelt werden. Daher wird, anders als im bisherigen Praktikum, eine Einheit zum Thema Messtechnik an den Anfang des Praktikums gestellt. In dieser Experimentiereinheit zur Messtechnik sollen den Lehramtsstudierenden Lerngelegenheiten angeboten werden, die in die experimentelle Arbeitsweise und die genutzte Messtechnik mit zugehöriger Auswertesoftware einführen und deren Anwendung einüben (siehe Abschnitt 6.4 für Details zu Messtechnik und Software). Aufgrund des Vorwissens sind im Anfängerpraktikum Lehramt freie Aufgabenstellungen nur eingeschränkt möglich, sodass ein angeleitetes Arbeiten erforderlich ist. Die Versuchsbeschreibungen im Bereich Messtechnik sind daher kleinschrittig mit vielen Unterstützungsmaßnahmen, damit sich die Studierenden zu Beginn des Praktikums leichter in die Arbeit im Praktikum einfinden können. Dies ermöglicht ein eigenständigeres Arbeiten auf einem zum bisherigen Anfängerpraktikum vergleichbaren Schwierigkeitsniveau in den darauffolgenden Praktikumsversuchen, in denen weniger Unterstützungsmaßnahmen gegeben sind. Auch Neumann (2004) beginnt in dem von ihm entwickelten Physikpraktikum für Physiker ebenfalls mit einer Einheit zur Einführung der Messgeräte. In den darauffolgenden Einheiten reduziert Neumann (2004) schrittweise die Unterstützungsmaßnahmen und erhöht damit den Schwierigkeitsgrad, sodass die Studierenden am Ende des Praktikums in der Lage sind, selbstständig ein Projekt zu erarbeiten. Aufgrund organisatorischer Rahmenbedingungen können im neuen Lehramtspraktikum bis auf den vorgeschalteten Teil zur Messtechnik keine aufeinander aufbauenden Versuche angeboten werden. Die Reihenfolge der Durchführung der nachfolgenden Versuche richtet sich nach den verfügbaren Arbeitsplätzen, sodass deren Schwierigkeitsgrad etwa gleich gehalten werden muss. In jedem Versuch wird allerdings ein anderer Schwerpunkt in Bezug auf beispielsweise die Auswertungsmethode oder die Bestimmung der Messunsicherheiten gelegt (siehe Anhang A.1.1 für Details zur Bewertung des Schwierigkeitsgrades). Eine weitere Konsequenz aus den

gering ausgeprägten Experimentierkenntnissen ist eine engmaschige Betreuung im Lehramtspraktikum. Auf acht Praktikumsgruppen a zwei Studierenden kommen im Lehramtspraktikum zwei Betreuende und zwei wissenschaftliche Hilfskräfte. Die Betreuenden führen daher nicht nur die bereits angesprochenen intensiven Vorgespräche, in denen auch messtechnische Aspekte diskutiert werden, sondern unterstützen die Studierenden auch direkt bei experimentellen Fragestellungen in der Versuchsdurchführung, um eventuelle Lücken in der heterogenen Studierendengruppe (siehe Abschnitt5.4) auffangen und den Erwerb experimenteller Kompetenzen so individuell unterstützen zu können.

Während des Anfängerpraktikums hören die Lehramtsstudierenden lediglich parallel eine Einführung in die allgemeine Fachdidaktik (Abschnitt 5.2), sodass im Praktikum daher keine physikdidaktischen Vorkenntnisse vorausgesetzt werden können. Wie dennoch eine Vernetzung zu fachdidaktischen und schulpraktischen Aspekten aus dem Praktikum selbst heraus gelingen kann, wird in Abschnitt 6.5 betrachtet.

Zusammengefasst wurde das spezifische Vorwissen der Lehramtsstudierenden im neuen Lehramtspraktikum insbesondere dadurch berücksichtigt, dass eine Einführungseinheit zur Messtechnik und Software die oft wenig ausgeprägten EDV-Kenntnisse und Experimentiererfahrungen der Lehramtsstudierenden auffängt. Durch kleinschrittigere Unterstützungsmaßnahmen und engmaschige Betreuung wird dabei eigenständigeres Experimentieren auf dem Schwierigkeitsniveau des bisherigen Anfängerpraktikums im weiteren Verlauf des Praktikums vorbereitet. Die Vertiefung der bestehenden Kenntnisse aus den Grundlagenvorlesungen wird dabei nicht nur durch die Bearbeitung der Praktikumsversuche alleine gefördert, sondern vor allem auch durch intensive Gespräche mit den Betreuenden im Verlauf der Versuchsdurchführung.

6.3 Versuchsinhalte

Um das Praktikum auf die Entwicklung des fachlichen Professionswissens der angehenden Lehrkräfte auszurichten, muss ein spezifisches Lernziel des Lehramtspraktikums die Vertiefung der fachlichen Kenntnisse sein, die relevant für typische physikalische Experimente in der Schule sind (Abschnitt 6.1). Zudem soll eine geeignete Auswahl von Versuchsinhalten gemäß der Designkriterien adressatenspezifischer Praktika auch dazu dienen, dass die Lehramtsstudierenden die Praktikumsversuche als relevant für ihre spätere Lehrtätigkeit wahrnehmen (z. B. Theyßen, 2000, Abschnitt 2.3).

Die Rahmenbedingungen geben die fünf fachlichen Themenbereiche Mechanik, Elektrizitätslehre, Wärmelehre, Atomphysik und Optik vor, die im physikalischen Anfängerpraktikum vertieft werden sollen (Abschnitt 5.1). Zudem wurde in Abschnitt 6.2 als sechster Themenbereich Messtechnik im Lehramtspraktikum zur Anknüpfung an das spezifische Vorwissen der Studierenden eingeführt. Innerhalb dieser

gegebenen sechs Themen müssen damit lehramtsspezifische Versuchsinhalte ausgewählt werden.

Die Auswahl der Inhalte der einzelnen Praktikumsversuche basierte dabei auf den schulischen Lernzielen und den Basiskonzepten der schleswig-holsteinischen Fachanforderungen für das Fach Physik in Sekundarstufe I und II (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019). Diese Orientierung an den schulischen Fachanforderungen diente dazu, Versuchsinhalte zu identifizieren, die im Physikunterricht bedeutsam sind. Aus schulischer Perspektive können die fünf klassischen Themenbereiche des Anfängerpraktikums auch als Kernthemen des physikalischen Schulunterrichts identifiziert werden (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019). Eine entsprechende Auswahl der Versuchsinhalte in diesem Rahmen vernachlässigt die universitären Hauptlernziele und daran geknüpfte Themenbereiche als Rahmenbedingung somit nicht. Die Versuchsinhalte, die anhand der schulischen Fachanforderungen ausgewählt wurden, sollen von den Lehramtsstudierenden als relevant für ihre zukünftige Tätigkeit wahrgenommen werden. Dies legen Untersuchungen von Lorentzen et al. (2019) zur wahrgenommenen Relevanz von Fachinhalten von Lehramtsstudierenden im Fach Chemie dar, wonach die Relevanz am häufigsten aus einer schulischen Perspektive begründet wird und dazu der Fachinhalt mit Inhalten aus dem schulischen Unterricht in Verbindung gesetzt wird.

Zudem sollen durch diese Auswahl von lehramtsspezifischen Versuchsinhalten die Lehramtsstudierenden besser auf ihre spätere Lehrtätigkeit vorbereitet werden. Sie sollen zum einen durch die Durchführung von Praktikumsversuchen mit schulrelevanten Themen einen Überblick über die aktuellen Themen des Physikunterrichts erhalten. Zum anderen soll den Studierenden durch Bearbeitung der schulrelevanten Versuche auf Hochschulniveau vertieftes Hintergrundwissen über den Schulstoff hinaus vermittelt werden. Der Schwierigkeitsgrad der Praktikumsversuche liegt damit nach wie vor auf dem universitären Niveau des bisherigen Anfängerpraktikums und wurde durch entsprechende Schwerpunktsetzung für alle Versuche möglichst gleich hoch gehalten (siehe Anhang A.1.1 für Details zur Bewertung des Schwierigkeitsgrades). Dadurch sollen die Studierenden im Anfängerpraktikum, wie bereits in Abschnitt 6.1 formuliert, fachliche Kompetenzen erwerben, wie sie für die souveräne Durchführung der Versuche im Schulunterricht notwendig sind. Beim Franck-Hertz-Versuch beispielsweise, der im Schulunterricht eine wichtige Rolle bei der Entwicklung des Atommodells spielt, sollen die Studierenden nicht nur die diskreten Energieniveaus der Atome ermitteln, sondern auf Basis von gezielten Parametervariationen eine vollständige Erklärung der beobachteten Messsignale geben können.

Somit wurden auf Basis dieser Kriterien die Inhalte der insgesamt 25 neu konzipierten Praktikumsversuche ausgewählt (siehe Details zu den Inhalten in den Kurzfassungen im Anhang A.1). Im Themenbereich Mechanik wurden Versuche zu den Inhalten Newtonsche Axiome, Reibung, Doppler-Effekt und Stöße ausgedehnter Körper entwickelt. Die neuen Versuche aus dem Bereich Elektrizitätslehre sind der Wasserstromkreis, der Feldeffekttransistor, die Lorentzkraft und der Fieldplotter. Im

Bereich Atomphysik sind es die entwickelten Versuche Zerfallsgesetz, Franck-Hertz-Versuch, Photoeffekt und Rutherford'sche Streuung. Dem Themenbereich Optik zugeordnet sind die neuen Versuche Interferenz am Smartphone, Lego-Interferometer, Augenmodell sowie Brechung, Reflexion und Polarisation. Im Themenbereich Wärmelehre sind es die Versuche Thermohaus, Wärmeäquivalent, Wärmepumpe und Schwarzer Strahler. Für den Themenbereich Messtechnik wurden die Versuche Auf- und Entladung eines Kondensators, Oszillierendes Smartphone, Verformung eines Balles, Hören und Roller Coaster entwickelt.

In den Tabellen 4 und 5 sind die Sachgebiete der Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) aufgeführt. Die 25 neu entwickelten Praktikumsversuche decken diese Sachgebiete breit ab, wie aus der Zuordnung in diesen Tabellen hervorgeht. Der quali- und quantitative Energiebegriff im Sachgebiet Energie in der Sekundarstufe I wird beispielsweise im Praktikumsversuch zur Verformung eines Balles behandelt (Tab. 4; Anhang A.1.4). Im Versuch zum Fallenden Doppler-Effekt werden auch explizit Lerninhalte aus der Sekundarstufe I (Freier Fall) und II (Doppler-Effekt) vernetzt (Anhang A.1.21). Nur sehr wenige der Sachgebiete der Fachanforderungen sind dabei unterrepräsentiert (Tab. 4 und 5). Um diese vereinzelten Lücken nach dieser Studie zu schließen, sind im Rahmen des Projektes parallel zur Beforschung einige weitere Praktikumsversuche entstanden wie beispielsweise für das Sachgebiet Teilcheneigenschaften der Materie in der Sekundarstufe II (Tab. 5) der Millikanversuch mithilfe von Videoanalyse (Jacke, 2018). In der Tabelle 3 sind analog die Basiskonzepte der Physik, die in den Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) genannt werden, und Beispiele für Praktikumsversuche gelistet, in denen diese Basiskonzepte aufgegriffen werden. Während die Basiskonzepte Energie, System und Wechselwirkung in mehreren Versuchen auftauchen, ist nur das Basiskonzept Materie in den aktuellen Praktikumsversuchen unterrepräsentiert.

Basiskonzepte	**Versuchsbeispiele**
Energie	Verformung eines Balles
Materie	
System	Feldeffekttransistor
Wechselwirkung	Stöße ausgedehnter Körper

Tabelle 3: *In der linken Spalte dieser Tabelle sind die Basiskonzepte des Physikunterrichts in Schleswig-Holstein anhand der Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) gelistet. In der rechten Spalte sind Beispiele für Praktikumsversuche angegeben, die diese Basiskonzepte zum primären Thema haben. Details zu den Versuchen sind im Anhang A.1 zu finden. Kursiv sind dabei die Basiskonzepte markiert, die nur als Teilaspekt in Praktikumsversuchen enthalten sind.*

Sachgebiete Sekundarstufe I	Praktikumsversuche
Energie	
Quali- und quantitativer Energiebegriff	Verformung eines Balles
	Wärmeäquivalent
	Wärmepumpe
	Stöße ausgedehnter Körper
Herausforderungen der Energieversorgung	-
Elektrizitätslehre	
Einfache elektrische Stromkreise	Wasserstromkreis
Stromstärke und Spannung	Wasserstromkreis
Magnetismus	
Magnetismus	*(Teilaspekt z. B. bei Lorentzkraft)*
Elektromagnetismus	(in Vorbereitung: Transformator)
Optik	
Ausbreitung des Lichts	Augenmodell
Reflexion an ebenen Flächen	Reflexion, Brechung und Polarisation
Lichtbrechung und optische Abbildungen	Augenmodell
	Reflexion, Brechung und Polarisation
Farben	*(Teilaspekt bei Schwarzer Strahler)*
Mechanik	
Geschwindigkeit	Newtonsche Axiome
Statische Kräfte	Verformung eines Balles
Dichte und Druck	*(Teilaspekt beim Wasserstromkreis)*
Beschleunigte Bewegungen	Roller Coaster
	Newtonsche Axiome
	Reibung
	Fallender Doppler-Effekt
Wärme	
Temperatur & Wärmetransport	Thermohaus
	Wärmeäquivalent
	Wärmepumpe
	Schwarzer Strahler
Atom und Kernphysik	
Elementarteilchen	*(Teilaspekt bei Rutherford'sche Streuung)*
Radioaktiver Zerfall	Zerfallsgesetz
Kernenergie	-

Tabelle 4: *In der linken Spalte dieser Tabelle sind die Sachgebiete des Physikunterrichts in Schleswig-Holstein in der Sekundarstufe I anhand der Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) gelistet. In der rechten Spalte sind die Praktikumsversuche angegeben, die diese Sachgebiete zum primären Thema haben. Details zu den Versuchen sind im Anhang A.1 zu finden. Kursiv sind dabei die Sachgebiete markiert, die nur als Teilaspekt in Praktikumsversuchen enthalten sind. In grau sind die Sachgebiete gekennzeichnet, die in den Praktikumsversuchen dieser Studie nicht behandelt werden.*

Sachgebiet Sekundarstufe II	Praktikumsversuche
Mechanik	
Kinematik	Verformung eines Balles
	Newtonsche Axiome
	Fallender Doppler-Effekt
Dynamik	Roller Coaster
	Newtonsche Axiome
	Reibung
	Stöße ausgedehnter Körper
Felder	
Homogenes elektrisches Feld	Auf- und Entladung Kondensator
	Fieldplotter
Bewegungen in radialsym. Feldern	Stöße ausgedehnter Körper
Bewegungen in Magnetfeldern	Lorentzkraft
Elektrodynamik	*(Teilaspekt z. B. bei Roller Coaster)*
Wellen	
Schwingungen und Wellen	Oszillierendes Smartphone
	Hören
Welleneigenschaften des Lichts	Interferenz am Smartphone
	LEGO-Interferometer
	Reflexion, Brechung und Polarisation
Spektren	Hören
	Schwarzer Strahler
	Fallender Doppler-Effekt
Welleneigenschaften der Materie	-
Quanten	
Teilcheneigenschaften der Materie	(in Vorbereitung: Millikanversuch)
Teilcheneigenschaften des Lichts	Photoeffekt
Quantenobjekte	-
Quantenphysikalisches Atommodell	Franck-Hertz-Versuch
	Rutherford'sche Streuung

Tabelle 5: *In der linken Spalte dieser Tabelle sind die Sachgebiete des Physikunterrichts in Schleswig-Holstein in der Sekundarstufe II anhand der Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) gelistet. In der rechten Spalte sind die Praktikumsversuche angegeben, die diese Sachgebiete zum primären Thema haben. Details zu den Versuchen sind im Anhang A.1 zu finden. Kursiv sind dabei die Sachgebiete markiert, die nur als Teilaspekt in Praktikumsversuchen enthalten sind. In grau sind die Sachgebiete gekennzeichnet, die in den Praktikumsversuchen dieser Studie nicht behandelt werden.*

Um sicherzustellen, dass die Studierenden die Relevanz der ausgewählten Versuchsinhalte für den schulischen Physikunterricht erkennen, wird von ihnen in der Versuchsbeschreibung für die Versuchsvorbereitung eine Einordnung des Versuchsinhaltes in die Fachanforderungen gefordert (siehe auch Abschnitt 6.5). Dies ist insbesondere dann wichtig, wenn die Studierenden keine persönliche Erinnerung an das physikalische Thema oder das Experiment aus eigener Schulzeit haben.

Eine Besonderheit bei der Umsetzung der Versuchsthemen in den lehramtsspezifischen Praktikumsversuchen sind quantitative Untersuchungen von Analogie-Experimenten. Die Idee einer Analogie ist, dass man eine Ähnlichkeit zwischen zwei Objekten oder Sachstrukturen ausnutzt, um den Lernprozess in der Schule zu unterstützen (Kircher et al., 2015). Eine Voraussetzung für die fachdidaktische Nutzung von Analogien ist ein tieferes Verständnis der Sachverhalte, sodass der Lehrkraft nicht nur die Übereinstimmungen der Analogie bekannt sind, sondern auch deren Grenzen. Um diese Grenzen experimentell zu erfahren und quantitativ zu analysieren, werden im lehramtsspezifischen Praktikum mehrere Analogie-Versuche eingesetzt. Dadurch werden Fach und Fachdidaktik im Praktikum explizit vernetzt. In der Schulpraxis ist eine der am häufigsten verwendeten Analogien die zwischen Wasserströmungen und elektrischen Strömen, die im Schulunterricht in der Sekundarstufe I zur Einführung der Begriffe Spannung und Stromstärke eingesetzt wird. Im daran angelehnten Analogie-Experiment wird anders als in einem klassischen Praktikumsversuch ein Versuchsthema wie Schaltung von Widerständen nicht an einem elektrischen Stromkreis erarbeitet, sondern an Strömungen von Flüssigkeiten. Durch die quantitative Untersuchung des Wasserstromkreises im Praktikumsversuch können die Lehramtsstudierenden die Wasserstromkreisanalogie vertiefen und so die Chancen und Grenzen der Analogie identifizieren (Anhang A.1.23).

Zu Themen aus der Atom- und Kernphysik können bzw. dürfen Versuche zum Teil aufgrund von Sicherheitsbestimmungen oder gesetzlichen Vorgaben nicht im Schulunterricht durchgeführt werden (vgl. Kultusministerkonferenz, 1994). Aus diesem Grund greift man insbesondere in diesem Themenbereich nicht selten auf Analog-Experimente zurück, um die mikroskopische Welt der Teilchen sichtbar und begreifbar zu machen. So kann beispielsweise der radioaktive Zerfall mit Analogie-Experimenten mittels Würfeln und Bierschaum veranschaulicht werden (Anhang A.1.7). Für den Rutherford'schen Streuversuch, der eine wichtige Rolle bei der Entwicklung des Atommodells spielt, wurde für das lehramtsspezifische Praktikum ein Analogie-Experiment mit Magnetpucks auf einem Luftkissentisch entwickelt, deren Streuung mittels Videoanalyse untersucht wird (Anhang A.1.10). Die experimentell umgesetzten Analogien und die quantitative Untersuchung ihrer Möglichkeiten und Limitationen sind damit ein besonderer Bestandteil des lehramtsspezifischen Praktikums, da sie neben einer fachlichen Vertiefung des Wissens auch eine direkte fachdidaktische Komponente haben.

Zusammengefasst wird den Lehramtsstudierenden im Praktikum durch die Auswahl der Versuchsinhalte anhand der schulischen Fachanforderungen und der Hinzunahme von Analogie-Experimenten eine Lernumgebung geboten, um sich gerade jene fachlichen Kompetenzen aneignen zu können, die wichtig sind für ihre spätere Lehrtätigkeit. Zudem soll den Lehramtsstudierenden durch diese Auswahl der Versuchsinhalte die Relevanz der Praktikumsversuche für den Physikunterricht deutlich werden.

6.4 Messtechnik

Ein weiteres lehramtsspezifisches Lernziel des neuen Praktikums ist, dass die Studierenden experimentelle Fähigkeiten und Kenntnisse speziell im Umgang mit schultypischer Messtechnik erlernen, um das fachliche Professionswissen für ihre spätere Lehrtätigkeit zu entwickeln (Abschnitt 6.1). Die Auswahl relevanter Messtechnik in Hinsicht auf das Berufsfeld der Studierendengruppe zählt außerdem zu den Designkriterien eines adressatenspezifischen Praktikums (z. B. Neumann, 2004, Abschnitt 2.3). Wie auch die schulrelevanten Versuchsinhalte im vorangegangenen Abschnitt 6.3 soll geeignete Messtechnik im Lehramtspraktikum damit auch dazu dienen, dass die Lehramtsstudierenden die Praktikumsversuche als relevant für ihre zukünftige berufliche Tätigkeit als Lehrkraft wahrnehmen.

Um diese Ziele zu erreichen, wurden für das Lehramtspraktikum nicht nur Messtechnik und zugehörige Software gezielt für die Lehramtsstudierenden ausgewählt, sondern die Experimentiermaterialien insgesamt. Im lehramtsspezifischen Anfängerpraktikum werden daher hauptsächlich Experimentiermaterialien eingesetzt, die auch in einer gut ausgestatteten Schulsammlung vorhanden sind und damit gängige Arbeitsmaterialien einer Lehrkraft im Fach Physik sind. Die Studierenden sollen durch diesen gezielten Einsatz schulrelevanter Experimentiermaterialien und Messtechnik mit dem notwendigen Handwerkszeug für den experimentellen Teil des Physikunterrichts ausgerüstet werden. Durch diese Auswahl an Materialien sind alle Praktikumsversuche grundsätzlich im Physikunterricht durchführbar. Dies soll den Studierenden die Relevanz der Praktikumsversuche für den Physikunterricht verdeutlichen. Damit unterscheiden sich die Praktikumsversuche für Lehramtsstudierende durch die Auswahl schultypischer Experimentiermaterialien vom denen für 1-Fach-Studierende Physik, die in der Regel nicht oder nur eingeschränkt in der Schule umgesetzt werden können. Trotz des unterschiedlichen Experimentiermaterialien können die Lehramtsstudierenden das Hauptziel des Anfängerpraktikums, den Erwerb experimenteller Fähigkeiten (Abschnitt 5.1), erreichen. Der Schwierigkeitsgrad der Praktikumsversuche liegt dabei nach wie vor auf dem universitären Niveau des bisherigen Anfängerpraktikums.

Ein Großteil der Experimentiermaterialien sind gängige Produkte der Lehrmittelhersteller wie beispielsweise der komplette Aufbau des Franck-Hertz-Versuches (Anhang A.1.8). Es werden aber auch Materialien eingesetzt, die leicht nachge-

baut, vereinfacht oder durch herkömmliche Gegenstände einfach ersetzt werden können. Beim Versuch zum Michelson-Interferometer beispielsweise wird auf professionelle optische Bauteile verzichtet und stattdessen die kostengünstige Variante aus LEGO®-Bausteinen eingesetzt und diskutiert (myphotonics, o.D. Linke, 2017; Anhang A.1.12). Auch beispielsweise der Lautsprecher im Versuch zum Fallenden Doppler-Effekt (Anhang A.1.21) muss nicht zwangsläufig wie im Praktikumsversuch einen extra dafür konstruierten Turm mit Auslösemechanismus hinunterfallen, sondern kann vereinfacht auch in einem Treppenhaus nach unten in eine gepolsterte Kiste fallen. Um den Studierenden solche Möglichkeiten der schulpraktischen Umsetzung zu verdeutlichen, werden mögliche schulische Umsetzungen mithilfe von fachdidaktischen Reflexionsfragen oder der schulbezogenen Einleitung in der Versuchsbeschreibung angeregt. Weitere Erläuterungen und Beispiele zu den fachdidaktischen Reflexionsfragen und den Versuchsbeschreibungen finden sich in Abschnitt 6.5.

Ebenso wurde bei der Auswahl der Messtechnik und Software auf deren Eignung für den Einsatz in der Schule Wert gelegt. Zu Beginn des lehramtsspezifischen Praktikums wird dazu Messtechnik als eigener Themenbereich angeboten (siehe Abschnitt 6.2). Die Studierenden erlernen dabei in speziell dafür konzipierten Versuchen die Nutzung entsprechender Messmethoden, die in den anschließenden Praktikumsversuchen angewendet und vertieft werden. Dazu zählen Messtechnik und Software, die Lehrmittelhersteller für den Einsatz in der Schule anbieten oder die als Freeware erhältlich sind.

In diesem Zuge gilt es auch zu berücksichtigen, dass die Digitalisierung und damit das computergestützte Experimentieren an Bedeutung gewinnt. Immer mehr Schulen legen sich Messwerterfassungssysteme mit zugehöriger Auswertesoftware zu. Daher wird in den lehramtsspezifischen Versuchen die Anwendung des digitalen Messwerterfassungssystem CASSY (LD-Didactic, o.D. a) mit zugehörigen Messsensoren und Auswertungssoftware vermittelt. Dieses Messwerterfassungssystem wird auch vom Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein (IQSH) im Referendariat eingesetzt. Das CASSY-System ist dabei exemplarisch für die diversen Messsysteme im Lehrmittelbereich zu verstehen, da Kenntnisse eines Messsystems die Einarbeitung in andere, ähnliche Messsysteme erheblich erleichtern. Um sich die Grundlagen des Messsystems anzueignen, führen die Studierenden zu Beginn des Praktikums einen Einführungsversuch zur Strom- und Spannungsmessung und Auswertung der Messungen mit CASSY am Beispiel der Auf- und Entladung eines Kondensators durch (Anhang A.1.2). Auf dieser Basis wird die Nutzung der Vielzahl an Sensoren in weiteren Versuchen erlernt (z. B. Lichtschranken, Bewegungsmesswandler und Kraftsensoren im Versuch Newtonsche Axiome, Anhang A.1.19), und es werden Möglichkeiten der Datendarstellung und Auswertung vertieft.

Ein weiteres Augenmerk wird auf die Anwendung von moderner Messtechnik gelegt, die zunehmend auch im Schulunterricht Beachtung findet. Eingesetzt wird zum Beispiel ein Smartphone, das im Schulunterricht nicht nur Störfaktor sein kann,

sondern auch als Messgerät und Dokumentationsmittel große Vorteile im Physikunterricht bietet (z. B. Müller et al., 2010; Vogt & Kuhn, 2014). Mit frei erhältlichen Apps wie PhysicsToolbox (R. Vieyra et al., 2015; C. Vieyra & Vieyra, o.D.) oder phyphox (Staacks et al., 2018; Staacks, o.D.) können die verschiedenen Messsensoren des Smartphones ausgelesen und weiter ausgewertet oder dargestellt werden. Dieser Einsatz als Messgerät wird in einem Praktikumsversuch anhand der Beschleunigungsmessungen eines schwingenden Smartphones geschult (Anhang A.1.3). Smartphones können zum Beispiel auch einfache konventionelle Kameras für Videoanalysen ersetzen und werden daher auch in Versuchen zur Videoanalyse eingesetzt (Anhang A.1.6). Auch Messtechnik wie die Thermographie mit einer Infrarot-Kamera kann in speziellen Smartphones integriert sein oder über Adapter ergänzt werden. Aber auch alleinstehende Infrarot-Kameras werden für Schulen immer erschwinglicher oder sind zumindest leihbar. Daher wird auch der Umgang mit einer Infrarot-Kamera und die quantitative Analyse von Wärmebildern in einem Praktikumsversuch anhand eines Modellhauses eingeübt (Anhang A.1.15).

Zu der schulrelevanten kostenfreien Software zählt unter anderem die sehr nützliche Software Tracker (D. Brown & Cox, 2009; D. Brown, 2019), die zahlreiche Möglichkeiten der Videoanalyse bietet. Eingeführt wird die Nutzung der Funktionen von Tracker mithilfe einer Hochgeschwindigkeitskamera zur Analyse der Bewegung eines fallenden Balles (Anhang A.1.4). Tracker bietet aber auch die Möglichkeit zum Beispiel Photos von Interferenzmustern auszuwerten (Anhang A.1.11). Ebenso bieten sich der Toneditor Audacity (Audacity, o.D.) und die Tonanalysesoftware Sonic Visualiser (Cannam et al., 2010; SonicVisualiser, o.D.) als freie Software für (zukünftige) Lehrkräfte an. Der Einsatz dieser Software wird an einem Versuch zum Thema Hören (Anhang A.1.5) eingeführt und dann bei der Spektralanalyse des Doppler-Effektes beim freien Fall vertieft (Anhang A.1.21).

Zusammengefasst zielt also der Einsatz von schulrelevanten Experimentiermaterialien (z. B. Produkte der Lehrmittelhersteller oder leicht beschaffbare LEGO®-Bausteine), Messtechnik (z. B. Messwerterfassungssystem CASSY) und Software (z. B. Videoanalyse-Software Tracker) darauf ab, dass die Lehramtsstudierenden im Praktikum das notwendige Handwerkszeug für den experimentellen Teil des Physikunterrichts erlernen. Durch die Auswahl schulrelevanter Materialien und Messtechnik soll den Lehramtsstudierenden außerdem die Relevanz der Praktikumsinhalte für ihre spätere Lehrtätigkeit deutlich werden, da alle Praktikumsversuche grundsätzlich im Physikunterricht durchführbar sind.

6.5 Kontext und Vernetzung

Die letzten zwei Gestaltungselemente des Kriterienkatalogs für adressatenspezifische Praktika sind die Bezugnahme auf Kontexte aus dem Interessenfeld der Studierendengruppe und die Vernetzung physikalischer und studienfachbezogener Fragestellungen (Abschnitt 2.3). Beide Elemente sollen besonders dazu dienen, die Relevanz der Lerninhalte des Praktikums für die Studierenden deutlich machen. Diese beiden Gestaltungselemente werden in diesem Abschnitt zusammen betrachtet, da beide im Lehramtspraktikum in einer lehramtsspezifischen Gestaltung der Versuchsbeschreibung umgesetzt werden und dabei ineinandergreifen. Die Studien zu adressatenspezifischen Praktika zeigen, dass die Versuchsbeschreibung ein zentrales Element zur adressatengerechten Gestaltung eines Praktikumsversuches ist (Abschnitt 2.3). In der Versuchsbeschreibung kann nicht nur die Relevanz des Versuches zum Beispiel in einer Einleitung direkt aufgezeigt werden, sondern auch der Versuch in einen Kontext bezogen auf das Studienfach oder den späteren Versuch eingebettet und physikalische mit studienfachbezogenen Fragestellungen verknüpft werden (z. B. Theyßen, 2000; Plomer et al., 2010).

Der Grundaufbau der Versuchsbeschreibungen der neuen lehramtsspezifischen Versuche wurde an die Form der Beschreibungen von Praktikumsversuchen angelehnt, die im physikalischen Anfängerpraktikum für 1-Fach Bachelor Physik an der CAU eingesetzt werden. Dadurch wurde wieder eine Kompatibilität zum ursprünglichen Anfängerpraktikum für 1-Fach-Studierende hergestellt, auch wenn die Rahmenbedingungen nicht so eng sind wie beim Ablauf des Praktikums (Abschnitt 5.3). Abbildung 9 zeigt auf der linken Seite die Grundelemente einer solchen klassischen Versuchsbeschreibung, die sich im Wesentlichen aus einer Einleitung zur Motivation, Literatur und Aufgaben zur Vorbereitung, Materialien und Versuchsaufbau sowie Messauftrag und einer Anleitung zur Auswertung zusammensetzt. Rehfeldt (2018a) und Brüggemann (2015) haben bei der Entwicklung eines Instruments zur Bewertung der Qualität von Praktikumsskripten die Versuchsanleitungen verschiedener Praktika an Hochschulen gesichtet und fassen die Grundelemente als strukturelle Gemeinsamkeiten der Praktikumsanleitungen ähnlich zusammen. Die Versuchsbeschreibung des neuen lehramtsspezifischen Anfängerpraktikums (Abb. 9 rechts) unterscheidet sich dabei von der klassischen Form durch die Gestaltung der Einleitung. Die klassischen Grundelemente werden in der lehramtsspezifischen Versuchsbeschreibung zudem ergänzt um eine schulbezogene Hausaufgabe zur Vorbereitung des Versuches und um fachdidaktische Reflexionsfragen nach den klassischen fachlichen Versuchsaufgaben.

Die Einleitung in der klassischen Versuchsbeschreibung (Abb. 9 links) ordnet den Versuch üblicherweise in den physikalischen Themenkanon ein. Der Versuch wird dabei ohne weitere Kontextualisierung in einen rein fachlichen Bezug gesetzt. Die neuen lehramtsspezifischen Einleitungen (Abb. 9 rechts) hingegen betten die Versuche in einen Schulkontext ein, um die Bedeutsamkeit der Versuchsinhalte für

die zukünftige Lehrtätigkeit zu verdeutlichen. Während Theyßen (2000) den Bezug zur Medizin in ihrem Praktikum über die Notwendigkeit der Physik zum Verständnis eines medizinischen Zusammenhangs aufzeigt, muss in einem Praktikum für Lehramtsstudierende des Fachs Physik die Relevanz über schulische Kontexte wie Lehrsituationen erzeugt werden. Im Folgenden soll die Herstellung dieses Schulbezugs und die Gestaltung der Einleitung anhand von Beispielen deutlich gemacht werden. Die Einleitungstexte für alle Versuche sind im Anhang A.1 zu finden.

Der Schreibstil in einer klassischen Versuchseinleitung ist üblicherweise angelehnt an ein Physik-Lehrbuch und nutzt sachliche Fachsprache und passive Formulierungen. Für die Einleitung in der schulbezogenen Versuchsbeschreibung wurde ein Schreibstil gewählt, der mit lebendiger Sprache und direkter Anrede in eine journalistische Richtung geht, um den Studierenden den Zugang zu erleichtern und einen persönlichen Bezug herstellen zu können. In manchen schulbezogenen Einleitungen wird zur Anschaulichkeit und zur Schaffung eines zusätzlichen Schulkontexts das Bild einer typischen Situation gezeichnet, in der sich Lehrpersonen befinden können. Dies können zum Beispiel Gespräche von Lehrpersonen mit Schüler*innen im Klassenraum oder auch Interaktionen zwischen Lehrpersonen in Fortbildungssituationen, auf Schulkonferenzen oder im Lehrerzimmer sein. Durch derartige schulbezogene Ausgestaltungen gestalten sich die lehramtsspezifischen Einleitungen mit einer halben bis einer ganzen Seite etwas umfangreicher als die oft knapp gehaltenen klassischen Einleitungen von etwa einer halben Seite.

Ein wesentliches Element am Ende der schulbezogenen Einleitung sind dabei ausformulierte, an die Studierenden gerichtete Fragen in den unterrichtsrelevanten Kontexten. Diese Leitfragen oder auch „Driving-Questions“ sollen ähnlich wie bei der Methode des Problem-Based Learning (z. B. Weber, 2004; English & Kitsantas, 2013) eine Problemstellung als übergeordnetes Versuchsziel zu Beginn des Lernprozesses generieren, die bei den Studierenden eine Lern-Herausforderung und ein „need to know“ hervorruft. Dazu eignen sich besonders authentische, fachdidaktische oder schulpraktische Problemstellungen, denen die Studierenden in ihrem zukünftigen Berufsleben als Lehrperson begegnen können. Da die Lehramtsstudierenden zum Zeitpunkt des Anfängerpraktikums nur sehr geringe fachdidaktische Vorkenntnisse haben (Abschnitt 5.2), sind diese Problemstellungen auf niedrigem Schwierigkeitsniveau und aus dem Praktikumsversuch selbst heraus lösbar. Die Einleitung und die Leitfragen sollen die Relevanz der Problemstellung im Praktikumsversuch für die zukünftigen Lehrkräfte deutlich machen, das Interesse an der Lösung des Problems stimulieren und zum aktiven und selbstständigen Lernen anregen.

Im Versuch zum Wärmeäquivalent (Anhang A.1.16) beispielsweise wird in der schulbezogenen Einleitung die Fehlvorstellung von Schüler*innen zu Temperatur und Wärme aufgegriffen. Schüler*innen fällt es oft schwer, zwischen Temperatur und Wärme zu unterscheiden, zumal der alltägliche Sprachgebrauch mit Aussagen wie „Heute ist es warm.“ dies auch nicht unterstützt. Es wird in der Einleitung eine Unterrichtssituation beschrieben, in der Schüler*innen in wörtlicher Rede ihren Un-

mut und ihr Unverständnis zum Thema deutlich machen. Die Leitfrage für diesen Praktikumsversuch geht entsprechend darauf ein: „Könnten Sie den Schülerinnen und Schülern helfen, die Begriffe zu unterscheiden?“. Der Zusammenhang und der Unterschied der Größen wird den Studierenden im Praktikumsversuch anhand der Bestimmung von Wärmekapazitäten unterschiedlicher Materialien verdeutlicht. Ein weiteres Beispiel für eine eingesetzte schulrelevante Problemstellung ist die schulpraktische Umsetzung von Versuchen. Die Einleitung zum Lego-Interferometer (Anhang A.1.12) beispielsweise wirft die Frage auf, ob und inwieweit auch günstige Alternativen wie Lego-Bausteine für Experimente in der Schule geeignet sind. In den Einleitungen zu Analogieversuchen wie dem Wasserstromkreis (Anhang A.1.23) oder der Rutherford'schen Streuung (Anhang A.1.10) sollen die Möglichkeiten und Grenzen von Analogien (siehe auch Abschnitt 6.3) ausgelotet werden. Für den Franck-Hertz-Versuch wird in der Einleitung ein unstimmiger Lehrbuchtext aufgegriffen (Anhang A.1.8). All diese Problemstellungen sollen den Studierenden die Relevanz der Praktikumsversuche für ihre spätere Lehrtätigkeit deutlich machen.

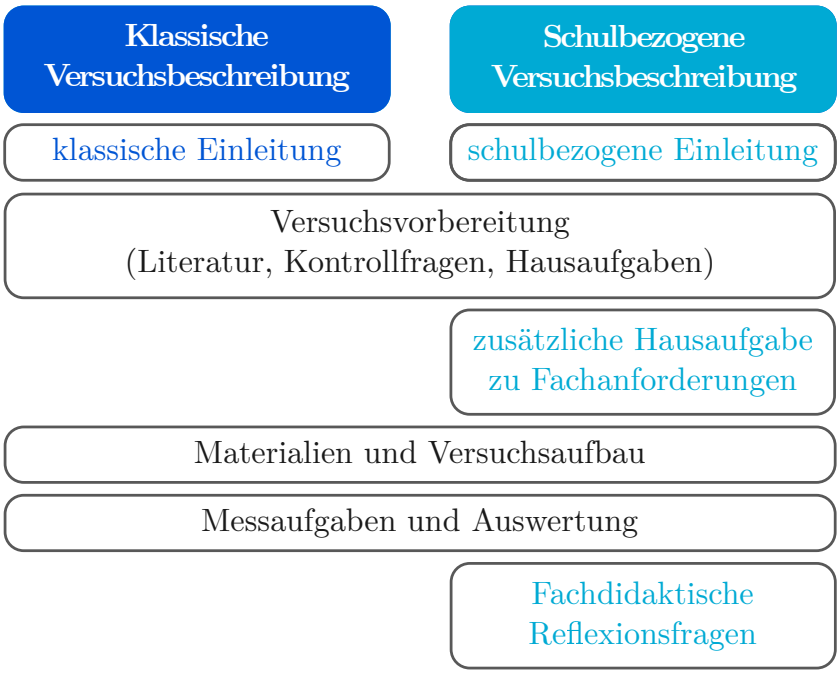


Abbildung 9: *Dieses Schema zeigt von oben nach unten die aufeinanderfolgenden Grundelemente einer Versuchsbeschreibung. Auf der linken Seite (dunkelblau) sind die Elemente der klassischen Versuchsbeschreibung zu finden. Darauf aufbauend ist auf der rechten Seite (hellblau) die in dieser Studie eingesetzte schulbezogene Versuchsbeschreibung schematisch dargestellt. Durchgezogene Kästen (schwarz) zeigen Elemente, die beide Versuchsbeschreibungen gemeinsam haben.*

Die aufgebrachten Leitfragen werden durch zwei fachdidaktische Reflexionsfragen am Ende der Versuchsbeschreibung aufgegriffen (Abb. 9 rechts) und damit ein Bogen zurück zur schulbezogenen Einleitung gespannt. Wie auch für die Leitfragen werden an dieser Stelle keine ausgeprägten fachdidaktischen Kenntnisse vorausgesetzt. Durch die fachdidaktischen Reflexionsfragen soll vor allem ein zusätzlicher Schulbezug in der Versuchsbeschreibung hergestellt werden. Dadurch soll nicht nur die Schulrelevanz weiter verdeutlicht werden, sondern auch die Möglichkeit genutzt werden, die Entwicklung des fachdidaktischen Professionswissens anzubahnen, auch wenn dessen Förderung kein explizites Ziel des lehramtsspezifischen Anfängerpraktikums ist. Zudem werden durch die Verknüpfung von fachwissenschaftlichen Fragestellungen der Versuchsaufgaben mit den fachdidaktischen Reflexionsfragen bereits im Anfängerpraktikum fachliche und fachdidaktische Aspekte vernetzt und fachdidaktische Kenntnisse angebahnt. In der fachdidaktischen Reflexion des angeführten Versuches zum Wärmeäquivalent (Anhang A.1.16) wird zum Beispiel mit der Frage „Wie könnten Sie auf die Schülerfrage ‚Wo ist denn der Unterschied zwischen Temperatur und Wärme?‘ eingehen?“ an die in der Einleitung aufgebrachte Leitfrage zur Fehlvorstellung angeknüpft. Entsprechend haben andere fachdidaktische Fragen beispielsweise die schulpraktische Umsetzung des Experimentes (z. B. Fallender Doppler-Effekt, Anhang A.1.21) oder die Grenzen und Möglichkeiten von Analogien (z. B. Wasserstromkreis, Anhang A.1.23) im Fokus.

Vereinzelt wird in den Reflexionsfragen auch die Eignung einer reduzierten Version des Praktikumsversuches für bestimmte Klassenstufen erfragt. Dies bezieht sich dann auf die zusätzliche erste Hausaufgabe (Abb. 9 rechts) in der Versuchsvorbereitung, die nicht auf die Vorbereitung des Versuchsinhaltes abzielt, sondern spezifisch für Lehramtsstudierende ist. In der schulbezogenen Versuchsbeschreibung wird in der ersten Hausaufgabe eine Einordnung des Versuchsthemas in die schleswig-holsteinischen Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) in Bezug auf das Sachgebiet oder das Basiskonzept und die Sekundarstufe gefordert. Dadurch wird eine erste Auseinandersetzung mit den Fachanforderungen angeregt. Zudem soll damit aber vor allem sichergestellt werden, dass die Studierenden die Relevanz des Versuches für den schulischen Physikunterricht anhand formaler Vorgaben erkennen.

Zusammengefasst sollen die Versuchsbeschreibungen also nicht nur auf die Durchführung des Versuches vorbereiten, sondern durch den Schulkontext in den Einleitungen und die Verknüpfung von fachwissenschaftlichen und fachdidaktischen Fragestellungen mittels Reflexionsfragen die Relevanz der fachlichen Inhalte der Praktikumsversuche für die zukünftigen Lehrkräfte verdeutlichen. Zudem soll die Vernetzung von Fach und Fachdidaktik angebahnt werden.

6.6 Beispiel: Rutherford'scher Streuversuch

Nach Vorstellung der recht theoretischen und abstrakten Designprinzipien der Neukonzeption des Lehramtspraktikums in den vorangegangenen Abschnitten, soll nun exemplarisch der Versuch „Rutherford'sche Streuung" vorgestellt werden, um eine konkrete Vorstellung von der Praktikumspraxis zu ergänzen. Der lehramtsspezifische Praktikumsversuch „Rutherford'sche Streuung" ist ein Analogieversuch (Abschnitt 6.3) zu dem bedeutsamen Experiment, in dem Rutherford (1911) den Atomkern entdeckt hat. Rutherford (1911) schoss α-Teilchen auf eine dünne Goldfolie und konnte anhand deren Streuung auf die positive Ladung und die konzentrierte Masse des Atomkerns schließen (vgl. z. B. Demtröder, 2016a). Da dieses Experiment einen wesentlichen Beitrag zur historischen Entwicklung des Atommodells geleistet hat, ist es auch im Physikunterricht der Oberstufe relevant (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019). Damit ist das Experiment nach den Designkriterien (Abschnitt 6.3) auch interessant für ein lehramtsspezifisches Praktikum. In der Schule ist der Rutherford'sche Streuversuch im Original aufgrund der radioaktiven Strahlung und des präzisen Aufbaus nur theoretisch Thema. Das Spannende des historischen Rutherford'schen Streuversuches ist vor allem auch die Wechselwirkung zwischen α-Teilchen und Goldatomkern, die nicht direkt im Experiment beobachtet werden kann. Um diese mikroskopische Welt der Teilchen sichtbar und (be-)greifbar zu machen, greift man insbesondere in der Schule nicht selten auf Analogie-Experimente zurück. Für den Rutherford'schen Streuversuch gibt es vereinzelt Analogie-Versuche (Kircher et al., 2015; Heptin & Fromme, 2016), die aber schnell an ihre Grenzen stoßen. Daraus entstand die Idee eines neuen Analogie-Versuches, um den Rutherford'schen Streuversuch als wichtiges Experiment zur Entwicklung des Atommodells zu veranschaulichen. Dieser Analogie-Versuch erforderte eine komplette Neuentwicklung im Rahmen dieser Arbeit. Verschiedene Aspekte des Versuches wurden zudem bzgl. ihrer Tauglichkeit für den Praktikumseinsatz im Rahmen der Bachelorarbeiten von Geertz (2016) und Brauer (2017) und Marten (2018) getestet und optimiert.

Die Idee des neuen Analogie-Versuches ist, statt der Abstoßung im elektrischen Feld diese im magnetischen Feld zu betrachten. Beim historischen Rutherford-Experiment werden die positiven α-Teilchen vom positiv geladenen Goldatomkern abgestoßen. Im Analogie-Experiment wird ein Magnetpuck (Analogie: α-Teilchen) auf einem Luftkissentisch mit einer dafür konstruierten Abschussvorrichtung auf einen in der Mitte des Tisches befestigen Magneten („Kernmagnet", Analogie: Goldatom) geschossen (Abb. 10). Durch die gleiche Ausrichtung der Magnete stoßen sich diese ebenfalls ab. Im Gegensatz zum originalen Rutherford-Experiment kann die Abstoßung der Magnete aber auf makroskopischer Ebene betrachtet werden. Sowohl die Coulomb-Kraft, als auch die Kraft zwischen den magnetischen Dipolen nimmt mit dem Abstand r der Ladungen bzw. Magnete ab, aber im unterschiedlichen Maß

(Coulomb-Kraft $\sim 1/r^2$, Kraft zwischen den Magneten $\sim 1/r^4$). Damit ist die Abstoßung in beiden Feldern qualitativ vergleichbar.

Die Studierenden nehmen im Praktikumsversuch mit einer Kamera verschiedene resultierende Streukurven des Magnetpucks auf und analysieren diese anschließend mit der Videoanalyse-Software Tracker (D. Brown & Cox, 2009; D. Brown, 2019) (Abb. 11). Anhand dessen nehmen die Studierenden Energiebetrachtungen vor und bestimmen die Beschleunigung des Pucks bzw. die Kraft auf den Puck und deren Abhängigkeit vom Abstand (siehe auch Versuchsbeschreibung in Anhang A.1.10). Zudem werden geometrische Überlegungen zu den Streuwinkeln angestellt, diese für verschiedene Stoßparameter gemessen (Abb. 12) und die Gültigkeit des von Rutherford aufgestellten Zusammenhangs dieser Größen für das Magnetfeld geprüft. Abschließend sollen die Studierenden den Rutherford'schen Streuversuch im Coulomb-Potential eines Goldatomkerns mit dem durchgeführten Streuversuch im magnetischen Potential in Hinblick auf qualitative und quantitative Entsprechungen und Unterschiede vergleichen. Daraus sollen die studierenden Rückschlüsse auf die Möglichkeiten und Grenzen der Analogie ziehen und damit deren Güte beurteilen.

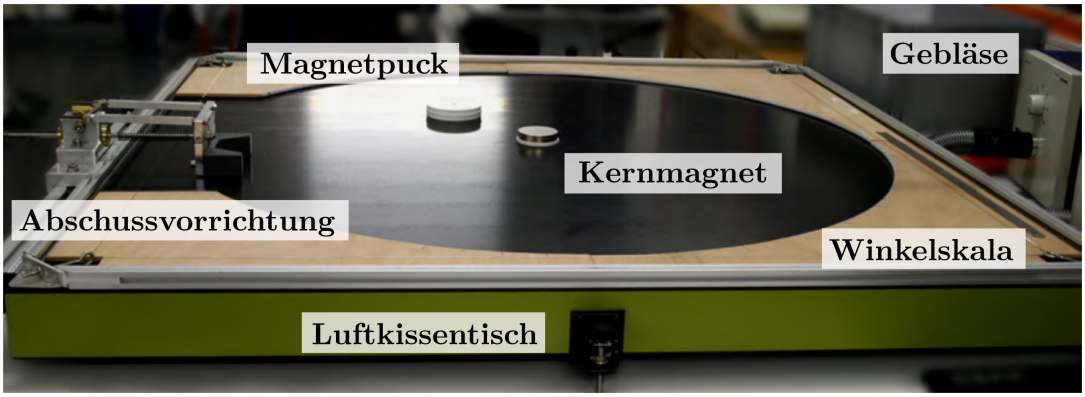


Abbildung 10: *Aufbau zum Analogieversuch Rutherford'sche Streuung bestehend aus einem Luftkissentisch betrieben mit einem Gebläse (rechts). Mit einer Abschussvorrichtung (links) kann ein Magnetpuck (Mitte links) auf einen zentral auf dem Luftkissentisch montierten Kernmagneten (Mitte rechts) geschossen werden. (Photo von Brauer, 2017)*

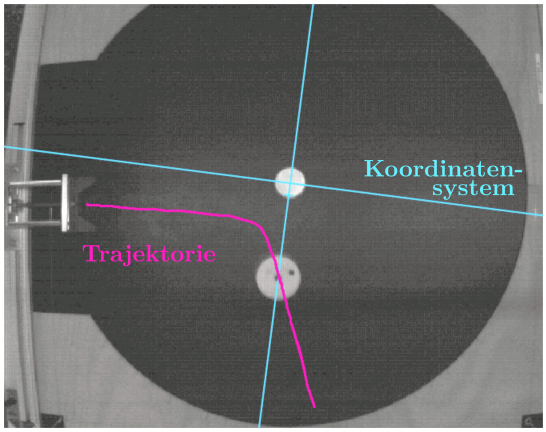


Abbildung 11: *Mithilfe einer Videoanalyse wird im Praktikumsversuch die Trajektorie (magenta) des vom Kernmagneten abgelenkten Magnetpucks bestimmt. Um die radiale Beschleunigung des Magnetpucks zu bestimmen kann in Tracker das Koordinatensystem (blau) mitgedreht werden.*

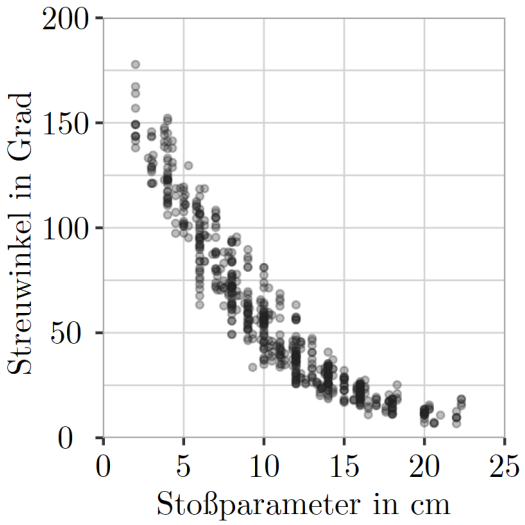


Abbildung 12: *Die Studierenden bestimmen für verschiedene Stoßparameter, also Positionen der Abschussvorrichtung relativ zum Kernmagneten, den Streuwinkel des Magnetpucks und tragen diese wie abgebildet zueinander auf.*

Um diese Praktikumsaufgaben zu bearbeiten, sollen die Lehramtsstudierenden im Sinne der Hauptlernziele (Abschnitte 5.1 und 6.1) Fachwissen aus den Grundlagenvorlesungen zu Stößen, Energie- und Impulserhaltung, elektrischen Feldern und Coulomb-Kraft wiederholen. Diese Kenntnisse werden erweitert und vertieft, indem diese untereinander und mit der Wechselwirkung der magnetischen Dipole in Beziehung gesetzt werden. Dafür spielt der Vergleich der Stöße auf der mikroskopischen und makroskopischen Ebene eine wesentliche Rolle.

Die Studierenden können für den Rutherford-Versuch an den Messtechnik-Versuch (Verformung eines Balles, Anhang A.1.4) zur Einführung der schultypischen, kostenfreien Videoanalyse-Software Tracker anknüpfen (Abschnitte 6.2 und 6.4). Durch den Messtechnik-Versuch haben sich die Studierenden die Grundlagen des Programmes Tracker und der Einstellmöglichkeiten (z. B. Belichtungszeit) der Kamera erarbeitet. Im Rutherford-Versuch können diese experimentellen Kompetenzen gemäß der Hauptlernziele (Abschnitte 5.1 und 6.1) angewendet und vertieft werden, indem speziellere Auswertefunktionen von Tracker wie das automatische Tracken von Trajektorien, Manipulation des Koordinatensystems für ein mitbewegtes Bezugssystem und der Fit von Modellfunktionen eingesetzt werden. Dadurch werden die Studierenden nicht nur sicherer im Umgang mit dieser Software, um diese im späteren Schulunterricht souverän nutzen können. Sie lernen das Programm Tracker auch in seiner vollen Leistungsstärke kennen, um sich den vielfältigen Einsatzmöglichkeiten für den Physikunterricht bewusst zu werden.

Obwohl das Rutherford'sche Atommodell ein Thema des Schulunterrichts der Sekundarstufe II ist und die Videoanalyse-Software Tracker in der Schule eingesetzt wird, wird an diesen Ausführungen deutlich, dass im Rutherford'schen Analogieversuch Fachwissen weit über den Schulstoff hinaus vermittelt und vertieft wird. Damit befindet sich der Schwierigkeitsgrad des Praktikumsversuches trotz der Relevanz für die Schule klar über dem Niveau eines „Schulversuches" und so auf universitärem Niveau. Um den Gesamtschwierigkeitsgrad aller 25 Praktikumsversuche auf einem ähnlichen Bereich auszurichten (siehe auch Abschnitt 6.2), wurde jeder Versuch in

einem Punktesystem bewertet (siehe Kurzbeschreibungen der Versuche in Anhang A.1), das den Schwierigkeitsgrad (leicht: 1 Punkt, mittel: 2 Punkte, schwer: 3 Punkte) für die sechs Kategorien Theoretische Grundidee, Messaufbau, Messtechnik, Versuchsdurchführung, Auswertung und Fehlerrechnung einordnet. Für jeden Praktikumsversuch wurden dabei bestimmte Lernschwerpunkte gesetzt, die einen höheren Schwierigkeitsgrad haben. Der Rutherford-Versuch liegt mit einer Gesamtschwierigkeit von insgesamt 13 Punkten etwas über den angestrebten 12 Punkten und ist damit ein eher schwerer Versuch im Lehramtspraktikum (Tab. 6). Beim Rutherford liegt ein Schwerpunkt auf den theoretischen Grundlagen aufgrund des Vergleiches und der Differenzierung der Stöße im Coulomb-Potential und magnetischen Potential. Der andere Schwerpunkt betrifft die Auswertung mit dem Erlernen spezifischer Auswertefunktionen in Tracker. Als Ausgleich dafür ist der Messaufbau mit der Justierung des Luftkissentisches und der Ausrichtung der Abschussvorrichtung leicht gehalten. Dadurch soll sicher gestellt werden, dass die Studierenden ihren Fokus auf die Lernschwerpunkte ausrichten können.

Auch die intensiven Gespräche und Hilfestellungen als Teil der engmaschigen Betreuung der Studierenden bei der Versuchsdurchführung, auf die in Abschnitt 6.2 hingewiesen wurde, sollen anhand dieses exemplarischen Versuches konkretisiert werden. Idealerweise läuft die Betreuung am Versuch zur Rutherford'schen Streuung in fünf Phasen hab. Wenn die Studierenden am Versuchsplatz angekommen sind, wird ihnen zunächst einmal eine technische Einweisung in den Versuchsaufbau und die Bedienung der Kamera gegeben. Mit diesen Informationen können die Studierenden den Aufbau zunächst justieren und erste Pucks abschießen und aufnehmen. Anschließend wird von den Betreuenden die Umsetzung überprüft und der erste Teil der intensiveren Gespräche geführt. Diese Gespräche gehen über die klassischen Vorgespräche von Physikpraktika hinaus. Die Betreuenden prüfen nicht nur Grundwissen aus der Vorbereitung ab, sondern stellen mit den Studierenden auch gemeinsam Verbindungen zwischen Theorie und Praxis her und erarbeiten übergreifende Konzepte, um das Fachwissen zu vertiefen. Im Fall des Rutherford-Versuches wird in diesem ersten Gesprächsteil dazu der Schwerpunkt auf die Gemeinsamkeiten und Unterschiede des historischen Rutherford-Versuches und der Analogie in Bezug auf Aufbau und Durchführung gelegt. Anschließend bearbeiten die Studierenden die ersten Versuchsaufgaben und erhalten insbesondere beim Einsatz der neuen Auswertefunktionen in Tracker erneut Hilfestellung. Nach Auswertung der Streukurven beginnt eine weitere Phase des Gespräches mit den Betreuenden, in dem die Messergebnisse kritisch diskutiert werden. Nachdem alle Versuchsaufgaben abgeschlossen sind, wird im abschließenden Teil des Betreuenden-Gespräches der Blick nochmal auf Bewertung der Analogie unter Einbezug der experimentellen Ergebnisse gelenkt. Durch diese mehrphasige Betreuung der Studierenden soll das vorhandene Vorwissen optimal aktiviert und vertieft werden, sowie der Erwerb experimenteller Fähigkeiten individuell unterstützt werden.

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	leicht	1
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	schwer	3
Fehlerrechnung	mittel	2
Summe		13

Tabelle 6: *Einordnung des Schwierigkeitsgrades des Versuches „Rutherford'sche Streuung" in den sechs Teilkategorien des Versuches und resultierendem Gesamtschwierigkeitsgrad.*

Nach diesen Beschreibungen des Versuches und dessen Durchführung und Bearbeitung, soll nun der Blick auf die lehramtsspezifische Versuchsbeschreibung gelenkt werden, die ein wichtiges Element zur Verdeutlichung der Relevanz des Versuches für die zukünftigen Lehrkräfte darstellt (Abschnitt 6.5). In einer klassischen Versuchsbeschreibung würde der Rutherford'sche Streuversuch im Einleitungstext üblicherweise in den physikalischen Themenkanon, also in die historische Bedeutung des Rutherford'schen Atommodelles, eingeordnet werden. In diesem lehramtsspezifischen Praktikumsversuch hingegen wird der Versuch in den schulischen Kontext eingebettet (siehe Einleitung in Anhang A.1.10). Dazu wird in der schulbezogenen Einleitung Bezug genommen auf die fachdidaktische Bedeutung der Analogie allgemein (siehe auch Abschnitt 6.3) und einer Analogie des Rutherford'schen Streuversuches mit einem elektrisch geladenen Tischtennisball, die in einem Standardwerk der physikdidaktischen Ausbildung veröffentlicht wurde (Kircher et al., 2015). Dabei wird der Analogiebereich und die Grenzen dieser Analogie kritisch betrachtet und anhand dieses Beispiels anschließend die „Driving-Questions" für den Analogieversuch des Praktikums abgeleitet. Der Versuch steht damit unter der Leitfrage, wo die Grenzen und Möglichkeiten des Rutherford-Analogie-Versuches liegen, wie auch bereits in der Beschreibung der Versuchsaufgaben angeklungen ist. Die Leitfrage wird in den fachdidaktischen Reflexionsfragen am Ende der Versuchsanleitung abschließend wieder aufgegriffen. Eine Voraussetzung für die fachdidaktische Nutzung von Analogien ist ein tieferes Verständnis über die Möglichkeiten und Grenzen dieser Analogie. Für den Vergleich des historischen Versuches und des Analogie-Versuches sind neben den fachphysikalischen Ergebnissen der qualitativen und quantitativen Analyse des Versuches und den fachdidaktischen Informationen der Einleitung keine weiteren Kenntnisse notwendig. Dadurch vernetzt der Versuch aus sich heraus fachliche und fachdidaktische Fragestellungen. Dies soll nicht nur fachdidaktisches Professionswissen anbahnen, sondern auch die Relevanz der Problemstellung für das Lehramt verdeutlichen, Interesse an deren Lösung bei den Studierenden hervorrufen und damit zum aktiven und selbstständigen Lernen anregen.

Der Analogie-Versuch zur Rutherford'schen Streuung steht exemplarisch für die über 25 in dieser Arbeit neu entwickelten Versuche des lehramtsspezifischen Praktikums an der CAU. Einige dieser lehramtsspezifischen Praktikumsversuche, darunter insbesondere auch die Analogieversuche (Abschnitt 6.3), haben eine vollständige Neuentwicklung von Grund auf erfordert (z. B. Interferenz am Smartphone Display (Anhang A.1.11), Fallender Doppler-Effekt (Anhang A.1.21), Wasserstromkreis (Anhang A.1.23)). Bei anderen handelt es sich um bekannte Praktikumsversuche, die anhand der Designprinzipien eine neue lehramtsspezifische Ausrichtung erfahren haben (z. B. Franck-Hertz-Versuch (Anhang A.1.8), Lego-Interferometer (Anhang A.1.11), Newtonsche Axiome (Anhang A.1.19)).

7 Synopse zur Praktikumskonzeption

Im Rahmen dieser Arbeit ist ein adressatenspezifisches Anfängerpraktikum für Lehramtsstudierende des Faches Physik an der Christian-Albrechts-Universität zu Kiel entstanden. Durch die hier vorgestellte Gestaltung der lehramtsspezifischen Praktikumsversuche sollen die Lehramtsstudierenden auf die Anforderungen des experimentellen Teils des Physikunterrichtens vorbereitet werden. Zudem bildet die adressatenspezifische Gestaltung die Basis für die Untersuchung der wahrgenommenen Relevanz der Praktikumsversuche und der Wirkannahmen für adressatenspezifische Praktika in den nachfolgenden Teilen dieser Arbeit.

Die grundlegenden Gestaltungsprinzipien adressatenspezifischer Praktika wurden aus bisheriger Forschung abgeleitet (siehe Abschnitt 2.3) und im Rahmen der Gestaltungsmöglichkeiten auf die lehramtsspezifische Konzeption des Praktikums spezifiziert und ausgestaltet. Die folgende Liste gibt einen Überblick über die Gestaltungsprinzipien der in dieser Arbeit entwickelten lehramtsspezifischen Praktikumsversuche:

- **Lernziele:** Ausrichtung der klassischen Hauptlernziele von Physikpraktika auf den Erwerb des fachlichen Professionswissens für die spätere Lehrtätigkeit
- **Vorwissen:** Berücksichtigung des Vorwissens, insbesondere durch eine kleinschrittige Einführungseinheit zu Messtechnik und Software
- **Versuchsinhalte:** Auswahl der Versuchsthemen anhand der schulischen Fachanforderungen und Hinzunahme von Analogie-Experimenten
- **Messtechnik:** gezielter Einsatz von schulrelevanten Experimentiermaterialien, Software und Messtechniken
- **Kontexte:** Einbettung der Versuche in schulbezogene Kontexte in der Einleitung der Versuchsbeschreibung
- **Vernetzung:** durchgängige Verknüpfung fachlicher und fachdidaktischer Fragestellungen in den Versuchsbeschreibungen

Nach diesen Designprinzipien sind im Laufe des ersten Jahres dieser Entwicklungsarbeit 25 lehramtsspezifische Praktikumsversuche individuell ausgewählt, konzipiert und umgesetzt worden. Darunter waren vollständige Neuentwicklungen, wie der zuvor exemplarisch vorgestellte Analogie-Versuch zur Rutherford'schen Streuung, aber auch lehramtsspezifische Neukonzeptionen bekannter Praktikumsversuche.

Ein besonders interessanter Aspekt dabei ist, dass diese Neukonzeptionen aufgrund der Studienordnung im Rahmen enger Randbedingungen umgesetzt wurden. Trotz der lehramtsspezifischen Konzeption sind die klassischen Hauptlernziele dabei unverändert geblieben. Auch beim Schwierigkeitsgrad mussten keine Abstriche gemacht werden, da die schulrelevanten Versuchsinhalte auf universitärem Niveau vertieft werden. Durch Berücksichtigung dieser Randbedingungen, die nicht nur am

Standort Kiel eine Rolle spielen, konnte damit eine formale Äquivalenz zum bisherigen Anfängerpraktikum hergestellt werden. Ein klassisches Praktikum kann (und wurde) damit im Rahmen der gleichen Studienordnung durch das hier vorgestellte lehramtsspezifische Anfängerpraktikum vollständig ersetzt (werden), was hinsichtlich der Übertragbarkeit auf andere Standorte oder andere Fächer (z. B. Chemie, Biologie) interessante Optionen aufzeigt.

Teil III
Wahrgenommene Relevanz

Nachdem in Teil I die Forschungsfragen dieser Arbeit anhand von Theorie und aktueller Forschung präzisiert wurden, beginnt anhand dessen nun in diesem Teil III die Beforschung des in Teil II vorgestellten neuen Praktikums. Im Zentrum dieses Teils stehen die Forschungsfragen 1 und 2. Das heißt, dass der Fokus dieses Teils der Arbeit auf der Untersuchung der durch die Studierenden wahrgenommenen Relevanz der Praktikumsversuche liegt.

Für diese Untersuchung ergeben sich verschiedene Herausforderungen, da Praktika einen komplexen und vielschichtigen Forschungsgegenstand darstellen. Um die vielfältigen Einflussgrößen auf die wahrgenommene Relevanz identifizieren und damit analysieren zu können, war es notwendig, eine Mischung aus erprobten und innovativen Instrumenten einzusetzen. Daher ist es in dieser Untersuchung zunächst erforderlich, das methodische Vorgehen und dessen Besonderheiten zu betrachten (Abschnitt 8). Auf dieser Basis können anschließend die Analyse (Abschnitt 9) sowie die Bewertung und Deutung der Ergebnisse (Abschnitt 10) vorgenommen werden.

8 Methodisches Vorgehen **70**

8.1 Studiendesign 70

8.2 Erhebungsinstrumente 76

8.3 Statistische Methodik 85

9 Analyse der wahrgenommenen Relevanz **106**

9.1 Prüfung von zeitlichen Effekten 106

9.2 Wahrgenommene Relevanz 112

9.3 Wahrgenommene Relevanz und Schulbezug 117

9.4 Wahrgenommene Relevanz und Versuchseigenschaften 122

10 Synopse zur wahrgenommenen Relevanz **139**

8 Methodisches Vorgehen

Für die Analyse der Forschungsfragen ist es zunächst erforderlich, drei Komponenten zum methodischen Vorgehen zu betrachten. Zunächst wird das Design der Gesamtstudie (Abschnitt 8.1) vorgestellt, um die zentralen Merkmale und Besonderheiten der Studie darzulegen. Ein wichtiges Element der Studie sind die eingesetzten Erhebungsinstrumente und deren Besonderheiten (Abschnitt 8.2). Als drittes ist eine Betrachtung der Datenstruktur und geeigneter statistischer Analyseverfahren notwendig (Abschnitt 8.3). Unter Berücksichtigung dieser drei Komponenten zum methodischen Vorgehen kann dann die eigentliche Analyse der wahrgenommenen Relevanz in Abschnitt 9 erfolgen.

8.1 Studiendesign

An das Studiendesign sind nicht nur Ziele in Form der Beantwortung von Forschungsfragen in Bezug auf die Wirkung des Praktikums (Abschnitte 4.2 und Abschnitt 3.5), sondern, wie an die Praktikumsentwicklung, ebenfalls Randbedingungen geknüpft (Abschnitt 5). Das Studiendesign ist daher stark geprägt vom Ablauf des Lehramtspraktikums als Forschungsgegenstand (siehe Abschnitt 5.3). Zudem erfordert die Beforschung eines so komplexen Untersuchungsgegenstandes wie dem Praktikum die Erhebung von vielfältigen Einflussgrößen. Ein Überblick über das Design mit Messzeitpunkten und Erhebungsinstrumenten ist in Abbildung 13 gegeben.

Das Lehramtspraktikum wurde für diese Teilstudie in einer Längsschnittstudie mit zwei Prä-Post-Messungen und 12 versuchsbegleitenden Messungen untersucht. In Abbildung 13 werden die eingesetzten Erhebungsinstrumente diesen Messzeitpunkten im Überblick zugeordnet. Details zu den einzelnen Instrumenten und deren Auswahl folgen in den Abschnitten 8.2 und 11.1. Da das Praktikum einen vielschichtigen Forschungsgegenstand darstellt, wurde möglichst ökonomisch eine Vollerhebung der Bedingungsfaktoren zur Prüfung der Wirkannahmen vorgenommen. Dazu wurden vor allem etablierte Messinstrumente genutzt, die bereits mehrfach erfolgreich eingesetzt wurden. Um den vielfältigen Untersuchungsgegenstand möglichst vollständig zu erfassen, musste aber auch ein geeignetes Instrument neu entwickelt werden. In den Prä- und Post-Messungen am Einführungs- und Abschlusstermin des Praktikums liegt der Fokus auf den Größen, die den Lernerfolg ausmachen (siehe Abschnitt 3.1). Daher wurden in drei Testheften das Fachwissen Physik in den verschiedenen Themenbereichen sowie in einem Fragebogen die Selbstwirksamkeitserwartungen in Bezug auf Planung und Durchführung von Experimenten im Unterricht, das Interesse an Physik und am Unterrichten von Physik erhoben. Im Prä-Test wurden zusätzlich demographische Daten der Studierenden erhoben. Der Zuwachs an experimenteller Kompetenz und Fachwissen wurde in einem zusätzlichen Fragebogen im Post-Test von den Studierenden eingeschätzt. Zu jedem der 12 lehramtsspezifischen Praktikumsversuche im Verlauf des Praktikumssemesters wur-

den versuchsbegleitende Erhebungen vorgenommen. Das bedeutet, dass nach jedem der durchgeführten Versuche von den Lehramtsstudierenden ein Fragebogen mit Bezug zum Versuch ausgefüllt wurde. In den versuchsbegleitenden Messungen bilden die affektiven Größen wahrgenommene Relevanz und situationales Interesse (siehe Abschnitte 3.2 und 3.3) den Schwerpunkt und werden ergänzt um die Versuchseigenschaften. Das Praktikum erstreckt sich über zwei Semester, sodass sich dieser Studienablauf im zweiten Teil des Praktikums wiederholt. Alle Befragungen wurden dabei mithilfe von Paper-Pencil-Tests mit Einführungstext und geschlossenem Antwortformat durchgeführt, um eine möglichst hohe Objektivität der Messung zu gewährleisten. Die Paper-Pencil-Tests wurden außerdem ausgewählt, um fehlende Daten möglichst zu minimieren. Die versuchsbegleitenden Fragebögen wurden zum jeweiligen Versuch ausgeteilt und mit dem Versuchsprotokoll wieder eingesammelt, um die Rückgabe der Fragebögen bestmöglich zu kontrollieren. Ein pseudonymisierter Kennungscode für jede*n Teilnehmende*n ermöglicht eine Zuordnung der Fragebögen zueinander.

Aus der bisherigen Forschung zu adressatenspezifischen Praktika und theoretischen Überlegungen geht hervor, dass eine Stichprobe von mindestens 100 Studierenden notwendig ist, um die angenommene Wirkung des Lehramtspraktikums nachweisen zu können (Abschnitt 3.6; Cohen, 1992). Am Anfängerpraktikum nehmen jedoch nur rund 20 Lehramtsstudierende pro Jahr teil (Abschnitt 5.4). Aus diesem Grund war es erforderlich, die Beforschung des Lehramtspraktikums über einen möglichst langen Zeitraum durchzuführen. Durch die schnelle Umsetzung des Praktikums konnte ein Beforschungszeitraum von sieben Semestern realisiert werden, sodass eine ausreichend große Stichprobe von fast 90 Studierenden befragt werden konnte. Die zeitliche Herausforderung der Laufzeit der Beforschung brachte auch mit, dass über einen langen Zeitraum keine Veränderungen der Versuche und der Bedingungen im Praktikum vorgenommen wurden, um die Einflussfaktoren konstant zu halten und keine störenden Kohorteneffekte zu generieren. Sowohl die Anleitungen, die Versuchsmaterialien, die betreuenden Personen und auch die Inhalte der Vorgespräche über den Studienzeitraum wurden dazu möglichst unverändert gehalten. Aus organisatorischen Gründen über die sieben Semester war lediglich ein Wechsel der Hilfskräfte nicht ausgeschlossen.

Zur Beantwortung der Forschungsfrage 1, die auf den Einfluss von Kontextualisierung auf die wahrgenommene Relevanz abzielt, wurden die Studierenden zudem in Interventionsgruppen mit unterschiedlichen Versuchsbeschreibungen eingeteilt. Diese Einleitungen sind entweder im klassischen Lehrbuchstil oder mit schulbezogenem Kontext verfasst (siehe Abschnitt 6.5). Während die schulbezogene Versuchsbeschreibung den Praktikumsversuch durch Aufzeigen eines fachdidaktischen Problems in einen Schulkontext einbettet, ordnet die klassische Versuchsbeschreibung den Versuch in den physikalischen Themenkanon ein. Dadurch enthält die klassische Versuchsbeschreibung keine expliziten Bezüge zu schulpraktischen Aspekten im Gegensatz zur schulbezogenen Beschreibung.

Abbildung 13: *Diese Abbildung erweitert den Ablauf eines Praktikumssemesters aus Abbildung 7 um die Messzeitpunkte und die zugehörigen Erhebungsinstrumente. Die Studie umfasst zwei Prä-Post-Messungen (orange Kreuze) und 12 versuchsbegleitenden Messungen (magentafarbene Kreuze), die in die Einführungs- und Abschlusstermin (orange Quadrate) bzw. in die Versuche des Praktikums (magentafarbene Quadrate) integriert wurden. Die zu diesen Messzeitpunkten eingesetzten Erhebungsinstrumente sind entsprechend farblich gelistet.*

An dieses Studiendesign sind weitere Anforderungen geknüpft. An die Interventionsgruppen ist die Bedingung geknüpft, dass keiner der Studierenden bedingt durch die unterschiedlichen Versuchsbeschreibungen der Studie einen Vor- oder Nachteil beim Lernen im Praktikum und daraus resultierend in der bewerteten Leistung haben soll. Im Rahmen der Erhebung der Daten ist es zudem wichtig, systematische Störeinflüsse zu minimieren oder zumindest zu identifizieren. So ist zum Beispiel bei einer immer gleichen Reihenfolge der Praktikumsversuche zu erwarten, dass unerwünschte zeitliche Effekte in den versuchsbegleitenden Messungen auftreten. Die Einteilung der Interventionsgruppen zusammen mit der Versuchsreihenfolge kann Störeinflüsse wie Übertragungseffekte hervorrufen. Aus diesem Grund sollen im Folgenden die Versuchsreihung und darauf aufbauend die Einteilung der Interventionsgruppen näher betrachtet und bewertet werden.

Ein Beispiel für einen möglichen, aber unerwünschten zeitlichen Effekt durch eine feste Versuchsreihenfolge ist, dass die Studierenden sich nach Durchführung der ersten Hälfte der Versuche an das Vorgehen im Praktikum gewöhnt haben, ihre Aufmerksamkeit auf andere Aspekte lenken können und dadurch zu einer anderen Einschätzung eines Versuches am Ende des Praktikums gelangen können. Bei einer festen Reihenfolge der Versuche können Einflüsse durch parallel laufende Lehrveran-

staltungen, Prüfungen oder auch Wetterbedingungen und Feiertage systematische Messfehler verursachen. Zur Vermeidung von systematischen Effekten wurde daher eine zufällige Reihenfolge der Themenbereiche und der einzelnen Praktikumsversuche in diesen Themenbereichen gewählt. Dadurch wird die Erhebung nicht systematisch durch die zeitlichen Störeinflüsse für die einzelnen Versuche verzerrt, sondern der statistische Fehler aller durchgeführten Versuche geringfügig erhöht.

In Abbildung 14 ist beispielhaft die Reihenfolge der Versuchsthemen für die Studierendengruppen im Wintersemester 2016/2017 dargestellt. Das Thema Messtechnik bildet eine Ausnahme, da es für alle Studierenden als Einführungseinheit an den Anfang des Praktikums gestellt ist (siehe Abschnitt 6.2). Alle weiteren Themenbereiche folgen anschließend in unterschiedlicher Reihenfolge. Während maximal vier der Praktikumsgruppen Versuche zu einem Themenbereich, wie in diesem Beispiel Optik, durchführen, arbeiteten maximal vier weitere Gruppen an Versuchen zu einem anderen Themenbereich, wie hier Atomphysik, und anschließend umgekehrt. Innerhalb eines Themenblockes führen die Gruppen an vier Praktikumstagen jeweils einen Praktikumsversuch durch (Abb. 14). Die Einteilung der Praktikumsversuche wurde im Regelfall anhand eines Lateinischen Quadrats vierter Ordnung in Standardform (z. B. Bortz & Schuster, 2010) vorgenommen (Abb. 15). So wird organisatorisch sicher gestellt, dass an jedem Versuchszeitpunkt jeder vorhandene Praktikumsversuch nur einmal durchgeführt wird. Dies reduziert ebenfalls zeitliche Störeffekte. Zudem erhält man durch zyklische Permutation im Lateinischen Quadrat eine gleichbleibende Abfolge der Versuche innerhalb eines Themenbereiches bei variierendem ersten Versuch. Mit Ausnahme der Versuche zur Einführung der Messtechnik am Semesterbeginn wurden die Versuche so konzipiert, dass sie in beliebiger Reihenfolge ohne Erkenntnisse aus vorangegangenen Versuchen durchgeführt werden können. Trotzdem kann die Versuchsreihenfolge insbesondere innerhalb eines Themenbereiches mit verwandten Inhalten einen Einfluss auf den Erwerb der fachlichen Kompetenzen haben. Durch Beibehalten der Versuchsabfolge innerhalb jedes zusammenhängenden Themenbereiches bleibt ein solcher Reihenfolgen-Effekt aber möglichst konstant.

Diese Reihung der Themenbereiche und Versuche erlaubt es, die Versuche quasi-randomisiert bearbeiten zu lassen und gleichzeitig zudem ökonomisch mit personellen, räumlichen und materiellen Kapazitäten zu wirtschaften. Die zum Teil kostspieligen Versuchsmaterialien müssen so beispielsweise nur ein- bis zweimal statt in Jahrgangsstärke vorhanden sein.

Die Einteilung in die Praktikumsteams erfolgte bei der Anmeldung der Studierenden in der Online-Lernplattform Olat. Dabei konnten die Studierenden bei der Einschreibung durch vorherige Absprache ein gemeinsames Team bilden oder sich zufällig zu einem anderen Teammitglied einbuchen. Anhand der Gruppennummer, die von den Studierenden bei der Einschreibung frei ausgewählt wurde, wurden die Praktikumstermine und damit die Versuchsreihenfolge quasi-randomisiert zugeteilt.

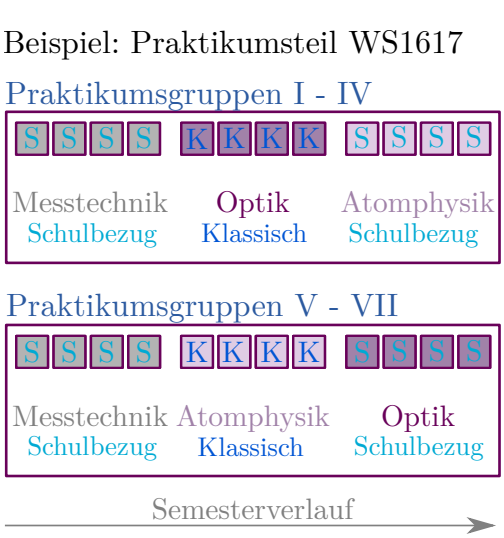


Abbildung 14: *Am Beispiel des Wintersemesters 2016/2017 wird hier die Zuordnung von sieben Praktikumsgruppen (I - VII) zu den Themenbereichen und Versuchsbeschreibungen veranschaulicht. Jedes der farbigen Quadrate stellt einen der Praktikumsversuche, von denen jeweils vier aufeinanderfolgende, farblich zusammengefasste aus einem Themenbereich Messtechnik, Optik und Atomphysik stammen. Zudem ist die Zuordnung zur schulbezogenen (S) und zur klassischen (K) Versuchsbeschreibung gekennzeichnet.*

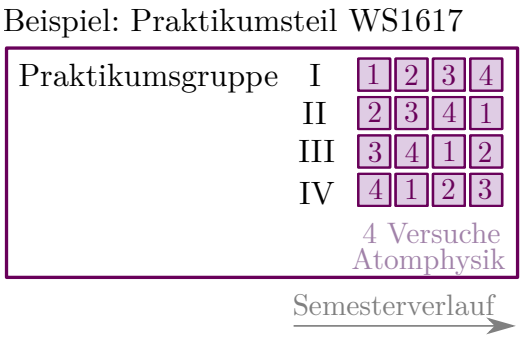


Abbildung 15: *Die Zuordnung von vier Praktikumsgruppen (I - IV) zu den Praktikumsversuchen wird hier gezeigt am Beispiel des Themenbereichs Atomphysik im Wintersemester 2016/2017. Jedes magentafarbende Quadrat stellt einen der Praktikumsversuche Atomphysik 1-4 dar. Die Zuordnung und Reihenfolge der Praktikumsversuche erfolgte anhand eines Lateinischen Quadrats.*

Damit keiner Interventionsgruppe ein Nachteil im Praktikum entsteht, sollten die Studierenden die Praktikumsversuche mit Versuchsbeschreibungen zu etwa gleichen Anteilen im klassischen Lehrbuchstil und mit Schulkontext bearbeiten. Zudem sollten für die Analyse der Praktikumsversuche diese auch in gleichen Anteilen mit klassischer und schulbezogener Versuchsbeschreibung bearbeitet werden, um gleichmäßige Stichproben zu erhalten. Weitere Aspekte, die man bei der Einteilung der Interventionsgruppen bedenken muss, sind mögliche Verzerrungseffekte durch Erkennen der Interventionsmaßnahme oder auch Übertragungs- oder auch Carry-Over-Effekt. Letzteres würde bedeuten, dass sich die Wahrnehmung der durchgeführten Versuche aufgrund deren Versuchsbeschreibung auf die Wahrnehmung nachfolgender Versuche auswirkt. Um diese Anforderungen zu erfüllen und Störeffekte zu minimieren, knüpft die Einteilung der Interventionsgruppen an die quasi-randomisierte Reihenfolge der Praktikumsthemen und Versuche an.

Zur Veranschaulichung ist in Abbildung 14 die Zuordnung der Form der Versuchsbeschreibungen zu den Praktikumsversuchen und Themenbereichen beispielhaft für das Wintersemester 2016/2017 dargestellt. Der Themenblock Messtechnik spielt eine besondere Rolle, da dieser immer als erster einführender Block durchlaufen wird

(siehe Abschnitt 6.2) und ausschließlich mit schulbezogener Versuchsbeschreibung angeboten wird. Dies soll dazu dienen, den Studierenden die gleiche Möglichkeit zu bieten, alle in eine gleiche Ausgangslage in Bezug auf die im Praktikum genutzte Messtechnik und Software sowie den Schulkontext kommen zu können. Die Änderung der klassischen und schulbezogenen Versuchsbeschreibungen in den nachfolgenden Versuchen wird an die Themenblöcke geknüpft, sodass diese Änderung durch den physikalischen Themenwechsel weniger als Erhebungsinstrument als eine inhaltliche Konsequenz empfunden wird, um eine Verzerrung durch den Hawthorne-Effekt (vgl. z. B. Döring & Bortz, 2016) zu minimieren. Die Zuordnung wurde dabei möglichst so vorgenommen, dass die Themenblöcke mit klassischer und schulbezogener Variante im Wechsel und über die zwei Praktikumssemester gleich häufig bearbeitet wurden, sodass die Anteile der Varianten für Studierende und Versuche gleich sind. Während die Praktikumsgruppen im Beispiel also im ersten und dritten Themenblock schulbezogene Versuchsbeschreibungen erhalten, arbeiten sie dazwischen im zweiten Themenblock mit klassischen Versuchsbeschreibungen. Diese blockweise Einteilung ist für die Fragebögen organisatorisch zudem weniger fehleranfällig.

Gleichzeitig hat das genutzte Studiendesign insbesondere bei kleinen Stichproben gegenüber reinen between-subject Designs den Vorteil, dass eine unsystematische Variation durch unterschiedliche Eigenschaften von Vergleichsgruppen reduziert werden kann, indem alle Versuchspersonen beide Untersuchungsbedingungen durchlaufen. Allerdings kann bei diesem Studiendesign nur eingeschränkt von einem within-subjects Design gesprochen werden, da die Untersuchungsbedingungen in Form der unterschiedlichen Versuchsbeschreibungen mit den durchgeführten Praktikumsversuchen konfundiert sind. Die damit einhergehende Komplexität der Datenstruktur wird im nachfolgenden Abschnitt 8.3.1 vorbereitend auf die Auswertung näher beleuchtet.

In Bezug auf den Übertragungseffekt ist es unwahrscheinlich, dass der Effekt der Versuchsbeschreibung innerhalb einer Woche zwischen zwei Versuchstagen ausklingt, da sich die Studierenden in dieser Zeit durch das Protokollieren noch mit dem Versuch beschäftigen. Eine Verlängerung dieser „Wash-Out“-Zeit ist organisatorisch nicht möglich. Allerdings kann an dieser Stelle wie bereits oben zum zeitlichen Effekt auf den Vorteil der quasi-randomisierten Zuordnung verwiesen werden.

Um Verzerrungseffekte nach Rosenthal (1966) durch Erwartungen der Assistenz und Hilfskräfte zu vermeiden, wurde die Gruppenzugehörigkeit nicht mitgeteilt. Die Assistenz und Hilfskräfte im Praktikum waren jedoch darüber informiert, dass es unterschiedliche Formen der Versuchsbeschreibung gab, um Verwirrungen bei Nachfragen der Studierenden mit Bezug auf die Versuchsbeschreibung zu vermeiden. Lediglich die Praktikumsleitung hatte Kenntnis über die Zuteilung in die Vergleichsgruppen, um die Fragebögen entsprechend auszuteilen. Wenn auch somit nicht von einer reinen Doppelblind-Studie gesprochen werden kann, wurden trotzdem Verzerrungseffekte minimiert.

Zusammengenommen bildet dieses Studiendesign eine abgestimmte Basis für die Untersuchung der Forschungsfragen zu den lehramtsspezifischen Praktikumsversuchen und deren Wirkung. Durch die lange Laufzeit konnte eine ausreichend große Stichprobe befragt werden, um mögliche Effekte nachweisen zu können. Die Aufteilung der Studierenden in Interventionsgruppen bezüglich der unterschiedlichen Formen der Versuchsbeschreibung ermöglicht insbesondere eine systematische Untersuchung der Forschungsfrage 1. Das Studiendesign ermöglicht durch die versuchsbegleitenden Messungen die zeitnahe Erfassung der affektiven Faktoren der Studierenden sowie der Eigenschaften aller einzelnen Praktikumsversuche. In Prä-Post-Messungen wurden Größen zum Lernerfolg erfasst. Die Besonderheit dieser Studie ist damit die sehr umfassende Erhebung von Bedingungsfaktoren, die nicht nur integral, sondern auch auf Ebene der insgesamt 24 einzelnen Praktikumsversuche stattfand.

8.2 Erhebungsinstrumente

Nachdem im vorangegangenen Abschnitt 8.1 das Design dieser Studie vorgestellt wurde, soll nun weiter auf die Untersuchung der Forschungsfragen 1 und 2 in diesem Teil III fokussiert werden. Dazu gehört die Operationalisierung der zu prüfenden Annahmen. Die Erhebungsinstrumente sind damit die zweite wesentliche methodische Komponente dieser Studie und bestimmen nicht nur was gemessen wird, sondern auch in welcher Qualität. Abbildung 16 zeigt den für Forschungsfrage 1 und 2 betrachteten Teil des Wirkmodells und ergänzt ihn um die Erhebungsinstrumente, die die entsprechenden Komponenten des Wirkmodells erfassen sollen. Das Kernstück der gesamten Studie ist die versuchsbegleitende Messung der wahrgenommenen Relevanz. In der Arbeit soll der Einfluss der Kontextualisierung und gegebenenfalls weitere Eigenschaften der Praktikumsversuche auf die wahrgenommene Relevanz geprüft werden. Um die lehramtsspezifischen Praktikumsversuche diesbezüglich zu charakterisieren, spielen zum einem die Interventionsgruppen mit den unterschiedlichen Versuchsbeschreibungen ein Rolle (Forschungsfrage 1) und zum anderen die versuchsbegleitende Erhebung der Eigenschaften der Praktikumsversuche (Forschungsfrage 2). Tabelle 7 zeigt eine Übersicht über die gemessenen Konstrukte und Kenngrößen der entsprechenden Skalen, die für diese Teilstudie relevant sind. Die statistischen Verfahren und angenommenen Grenzwerte, die die Grundlage zur Beurteilung der Güte dieser Skalen bilden, werden in Abschnitt 8.2.1 dargestellt. Details zur Güte und Besonderheiten der Skalen selbst werden in den Abschnitten 8.2.2 und 8.2.3 ausgeführt. Ein interessanter Punkt dabei ist, dass es sich um eine Mischung aus bewährten und innovativen Messinstrumenten handelt, da das Instrument zur Erfassung der Versuchseigenschaften für diese Studie neu entwickelt werden musste (Abschnitt 8.2.3).

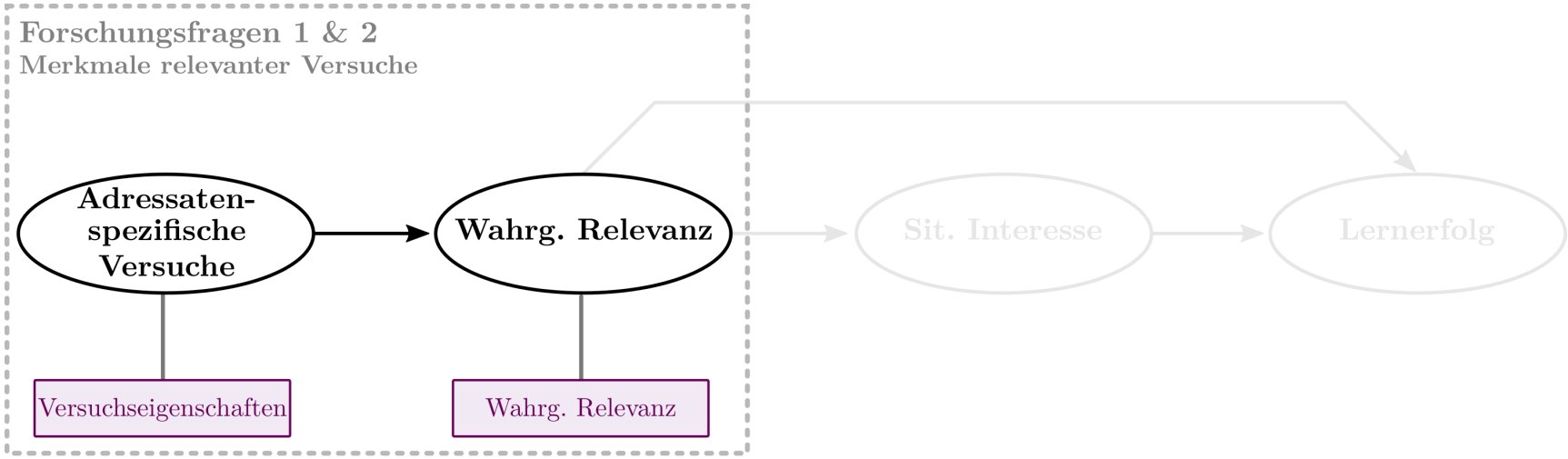


Abbildung 16: *Operationalisierung zur Prüfung der Wirkannahmen zu Forschungsfrage 1 und 2 (Abschnitt 4.2.1) in Teil III dieser Arbeit. Zur Untersuchung des Einflusses der lehramtsspezifischen Praktikumsversuche auf die wahrgenommene Relevanz, werden die Versuchseigenschaften mit einem spezifisch dafür entwickelten Instrument erhoben.*

Konstrukt	MZP	Skala	Items	α	Quelle
Wahrgenommene Relevanz	versuchs-begleitend	5-stufige Likert-Skala	8	.81	Schiefele et al. (2002)
Versuchseigenschaften	versuchs-begleitend	3-stufige Likert-Skala	27	-	Eigenentwicklung

Tabelle 7: *In dieser Tabelle sind zur Übersicht die Konstrukte, die für diese Teilstudie zur wahrgenommenen Relevanz erhoben wurden, mit Messzeitpunkt (MZP), Skala, Anzahl der Items und Cronbachs α als Maß für die Reliabilität gelistet.*

8.2.1 Methoden zur Bewertung der Erhebungsinstrumente

Die Erhebungsinstrumente sind ein wichtiger Faktor für das Gelingen einer Untersuchung. Um die Güte von Erhebungsinstrumenten zu beurteilen, werden üblicherweise Kennwerte wie Reliabilität und Trennschärfe herangezogen. Während deren Definitionen einheitlich in Lehrbüchern zu finden sind (siehe z. B. Moosbrugger & Kelava, 2007; Döring & Bortz, 2016), wird weniger einheitlich angegeben und diskutiert, ab welchen Werten die Güte der Instrumente für ausreichend befunden wird (z. B. Schmitt, 1996). Daher soll im Folgenden zunächst der Blick auf diese Gütekriterien geworfen werden, bevor sie anschließend genutzt werden, um die Erhebungsinstrumente dieser Studie in den nachfolgenden Abschnitten zu bewerten. Während für die Skala der wahrgenommenen Relevanz anhand dieser Gütekriterien unter anderem sichergestellt werden soll, das alle Items das gleiche Merkmal messen (Abschnitt 8.2.2), soll das neu entwickelte Erhebungsinstrument hingegen die vielfältigen Versuchseigenschaften möglichst breit erfassen (Abschnitt 8.2.3). Daher muss an dieser Stelle auch ein dafür geeigneter methodischer Ansatz betrachtet werden.

Alle Analysen dieser Studie wurden mit der freien Programmiersprache R (R Core Team, 2019) und gängigen R-Paketen durchgeführt (u. a. *psych* von Revelle, 2019). Abbildungen wurden in den meisten Fällen mithilfe des zusätzlichen R-Pakets *ggplot2* (Wickham, 2016) erzeugt. Weitere besondere R-Pakete sind in den nachfolgenden Beschreibungen der statistischen Methoden aufgeführt.

Reliabilität

Der erste Schritt der Itemanalyse zur Beurteilung der Güte einer Skala ist üblicherweise die Überprüfung, inwieweit alle Items einer eingesetzten Skala das gleiche Merkmal erfassen. Als gängiges Maß für die interne Konsistenz wird Cronbachs α (Cronbach, 1951) berechnet. Die Idee dieser Schätzung der Reliabilität basiert auf der Korrelation aller Items einer Skala untereinander (siehe z. B. Moosbrugger & Kelava, 2007). Generell sollte die Reliabilität des Tests und damit Cronbachs α möglichst hoch sein, um die Zuverlässigkeit des Erhebungsinstrumentes sicherzustellen. Für die Bewertung der Güte der Instrumente in dieser Studie stellt sich nun

die Frage, wie hoch eine gute Reliabilität sein sollte. Da die Reliabilität von verschiedenen Bedingungen abhängt, ist es schwierig, einen allgemeingültigen Richtwert für Cronbachs α anzugeben (Moosbrugger & Kelava, 2007). Cronbachs α ist stark von der Anzahl der Items der Skala abhängig und unterliegt damit der Kritik, dass es kein eindeutiges Maß für die Eindimensionalität einer Skala darstellt (z. B. Schmitt, 1996; Field et al., 2012). Moosbrugger und Kelava (2007) beurteilen die anzustrebende Höhe der Reliabilität daher für den betrachteten Test auch anhand von weiteren Kriterien wie der Länge des Erhebungsinstrumentes aufgrund der Einsatzbedingungen oder die Art des zu erfassenden Merkmals im Vergleich zu anderen Tests. Somit variieren auch die angegebenen Richtwerte für Cronbachs α für eine gute Reliabilität (vgl. z. B. Schmitt, 1996; Field et al., 2012; Döring & Bortz, 2016). Diese Studie orientiert sich an dem gängigen Richtwert von Cronbachs $\alpha \geq .7$ für gute Reliabilitäten (z. B. Schmitt, 1996) und diskutiert diesen gegebenenfalls zusätzlich hinsichtlich der Merkmale des Tests (vgl. Moosbrugger & Kelava, 2007). In der nachfolgenden Untersuchung der Erhebungsinstrumente wird zusätzlich der Wert für Cronbachs α ohne einzelne Items (α_{drop}) betrachtet (z. B. Field et al., 2012). Da der Fokus dieser Studie nicht auf der Skalenkonstruktion und deren Optimierung liegt, werden diese α_{drop} im Folgenden nicht berichtet, sofern sie nicht wesentlich größer sind als das α des Gesamttest.

Trennschärfe

Der nächste Schritt zur Überprüfung der Güte einer Skala ist die Untersuchung der Itemtrennschärfe. Die Trennschärfe ist ein Maß dafür, wie gut das Gesamtergebnis des Tests durch das beantwortete Item vorhergesagt werden kann und sich somit hohe und niedrige Ausprägungen des untersuchten Merkmals mithilfe des Items unterscheiden lassen (Döring & Bortz, 2016). Die Trennschärfe $r_{i(t-i)} \in [-1, 1]$ eines Items i ergibt sich als die Korrelation des Testwertes des Items mit dem Gesamtwert der Skala ohne dieses Item. Die angestrebten hohen Werte für die Itemtrennschärfe bedeuten folglich, dass Probanden mit einem hohen Gesamttestwert auch einen hohen Wert in der Antwort auf das einzelne Item haben. Richtwerte für die angestrebte Trennschärfe werden dabei recht einheitlich angegeben und werden daher auch in dieser Studie entsprechend für die Bewertung der Erhebungsinstrumente genutzt. Items mit einer Trennschärfe von $.3 \leq r_{i(t-i)} \leq .5$ gelten als mittelmäßig trennscharf und mit Werten $r_{i(t-i)} \geq .5$ als hoch trennscharf (z. B. Döring & Bortz, 2016). Trotzdem weisen Döring und Bortz (2016) darauf hin, dass insbesondere bei breit abgefragten Konstrukten diese Bewertungskriterien nicht mechanisch anwendbar sind und die Maßstäbe angepasst werden sollten. Dies ist insbesondere in Teil IV in dieser Arbeit für die Beurteilung des Wissenstests zum Fachwissen relevant.

Korrelationen

Die zuvor erläuterte Reliabilität und Trennschärfe stellen Gütekriterien für Skalen dar, die mit mehreren Items ein Merkmal erfassen. Ziel des im Rahmen dieser Arbeit entwickelten Erhebungsinstrumentes zur Charakterisierung der Praktikumsversuche ist hingegen eine möglichst breite Erfassung ihrer Eigenschaften bei gleichzeitigem ökonomischen Testeinsatz. Reliabilität und Trennschärfe sind daher nicht geeignet, um die Güte des Instrumentes zur Erfassung der Versuchseigenschaften zu analysieren. Für dieses Instrument ist es stattdessen erforderlich, es auf Redundanzen zu untersuchen. Dazu werden in Abschnitt 8.2.3 paarweise Korrelationen der Merkmale betrachtet. Die Korrelation ist ein typisches Maß für lineare Zusammenhänge zweier metrischer Variablen. Sie wird meist in Form des Pearsonschen Korrelationskoeffizient $r \in [-1, 1]$ angegeben, der sich aus der Kovarianz der zwei standardisierten Variablen ergibt (siehe für Details z. B. Eid et al., 2010). Wenn die Signifikanz des Korrelationskoeffizienten nach Pearson geprüft werden soll, dann ist eine Normalverteilung der Daten Voraussetzung (z. B. Field et al., 2012). Für nicht-parametrische Daten sollten Rangkorrelationen nach Spearman oder nach Kendall berechnet werden, um Fehlschätzungen zu vermeiden (z. B. Field et al., 2012). In dieser Arbeit wird zur besseren Vergleichbarkeit und Interpretation auch für nicht-parametrische Daten bevorzugt die Korrelation nach Pearson angegeben, sofern diese keine großen Unterschiede zur Rangkorrelation nach Spearman aufweist. Dies trifft auch für die nicht-normalverteilten Daten der Versuchseigenschaften zu. Eine hohe, signifikante Korrelation zweier Versuchseigenschaften kann damit einen Hinweis darauf geben, dass die Items sehr ähnliche Eigenschaften abgefragt haben und redundant sind. Dabei orientiert sich die Einschätzung der Höhe nach der groben Klassifizierung von Korrelationen von Cohen (1988): $|r| \approx .1$: schwacher Zusammenhang, $|r| \approx .3$: mittlerer Zusammenhang und $|r| \approx .5$: starker Zusammenhang. Zudem wird auch die inhaltliche Plausibilität betrachtet, da die Zusammenhänge auch mit anderen Einflussgrößen konfundiert sein können. Da die Versuchseigenschaften für alle Personen- und Versuchskombinationen betrachtet werden, um eine möglichst hohe Fallzahl ($n \approx 900$) zu haben, sind die Korrelationen aufgrund der Datenabhängigkeit tendenziell etwas zu hoch geschätzt. Zur Identifikation redundanter Items ist eine Abschätzung der Größenordnung der Korrelationen der Items untereinander mit diesem einfachen Vorgehen jedoch gut geeignet.

8.2.2 Erhebungsinstrument: Wahrgenommene Relevanz

Die wahrgenommene Relevanz der lehramtsspezifischen Praktikumsversuche ist eine zentrale Komponente der Wirkannahmen in dieser Studie (Abschnitt 3.3). Daher ist die Güte der Erfassung der wahrgenommene Relevanz ein wichtiger Faktor für die Qualität der Untersuchung und wird in diesem Abschnitt mithilfe der zuvor vorgestellten Kennwerte beurteilt.

Die wahrgenommene Relevanz der Studierenden wurde versuchsbegleitend für jeden der durchgeführten Praktikumsversuche abgefragt (Abschnitt 8.1). Zur Erfassung der wahrgenommenen Relevanz wurde eine etablierte Skala verwendet, die von den Autoren Schiefele et al. (2002) untersucht und mit einer guten Reliabilität eingesetzt wurde. Die Studierenden bewerten dabei acht Aussagen zu ihrer Wahrnehmung des jeweiligen Versuches bezüglich ihrer Zustimmung anhand einer 5-teiligen Likert-Skala. Der Bezug der originalen Items auf Lehrveranstaltungen im Allgemeinen wurde in der Formulierung leicht auf den Praktikumstag beziehungsweise den Praktikumsversuch angepasst. Ein typisches Beispielitem der Skala lautet: „Das Wissen, das an diesem Praktikumstag vermittelt wurde, ist für meine späteren beruflichen Tätigkeiten ziemlich wichtig.“. Die wahrgenommene Relevanz wurde im originalen Fragebogen mit fünf Items erhoben. Die Skala wurde für diese Studie um drei Items ergänzt, sodass insgesamt mit acht Items gemessen wurde. Diese höhere Messwiederholung der wahrgenommenen Relevanz pro Messzeitpunkt (Praktikumsversuch) soll eine höhere und über die Zeit stabilere Genauigkeit liefern, insbesondere da die Heterogenität der 24 Praktikumsversuche in Bezug auf die Relevanz im Vorfeld schwer abschätzbar war. Da die Items zur wahrgenommenen Relevanz in den versuchsbegleitenden Testheften mit den Items zum situationalen Interesse gemischt wurden, wurde dadurch auch die Gewichtung der Konstrukte im Testheft angeglichen. Die ergänzenden Items wurden nah an den originalen Items formuliert, aber in der Formulierung dabei weniger auf die Bedeutsamkeit für die zukünftige Lehrtätigkeit bezogen sondern mehr auf die Bedeutsamkeit für den Schulunterricht (z. B. „Den heutigen Versuch halte ich für bedeutsam für den Schulunterricht.“). Um einer Testmüdigkeit durch das häufige Bewerten der Items entgegen zu wirken und die Validität zu fördern, wurde eines der neuen Items negativ formuliert. Zu diesem Zweck wurde zudem die Reihenfolge der Items im Fragebogen für jeden Versuch variiert. Die acht Items sind in Anhang A.2.1 aufgeführt.

Die angepasste Skala zur wahrgenommenen Relevanz wurde im Wintersemester 2015/2016 mit Bachelor-Studierenden der Fachrichtungen 1-Fach-Physik, 2-Fach-Physik und Materialwissenschaften im Vorgänger-Praktikum pilotiert. Die Items wurden für die Pilotierung statt auf den Praktikumsversuch auf das Praktikum allgemein bezogen. Die Untersuchung der Güte der Skala basiert auf den üblichen statistischen Verfahren zur Bestimmung von Reliabilität und Trennschärfe (siehe Abschnitt 8.2.1). Die Pilotierung zeigt mit einem Cronbachs $\alpha = .80$ ($n = 69$) eine gute Reliabilität der Skala. Die Trennschärfen der Items liegen für die Pilotierung in einem mittleren bis hohen Bereich ($.31 \leq r_{i(t-i)} \leq .69$). Ein Entfernen einzelner Items hätte die Reliabilität nicht wesentlich verbessert. Aufgrund der guten Kennwerte wurde die Skala zur wahrgenommenen Relevanz in dieser Form beibehalten.

Auch die Analyse der später im Praktikum erhobenen Daten liefert gute Werte für Cronbachs α. Ein Überblick über die Kennwerte der der eigentlichen Datenerhebung im Rahmen dieser Studie gibt Tabelle 8. Für die verschiedenen Messzeitpunkte (Praktikumsversuche) liegen die Reliabilitäten anhand Cronbachs α in einem guten

	M	SE	n	α	$\bar{r}_{i(t-i)}$
Gesamttest	3.65	0.02	1805	.81	.52
Wertebereich für Versuche	3.12 - 4.09	0.06 - 0.09	65 - 89	.74 - .87	.45 - .63

Tabelle 8: *Deskriptive Werte für die wahrgenommene Relevanz. In der ersten Zeile der Tabelle sind die Werte für den Gesamtdatensatz (alle Kombinationen an Versuchen und Personen) angegeben. In der zweiten Zeile ist der Wertebereich der wahrgenommenen Relevanz unter Berücksichtigung der Messzeitpunkte, sprich Praktikumsversuche (ohne Versuch Messtechnik 5), angegeben. Für die Deskription der Datensätze sind Mittelwert (M), Standardfehler (SE), Stichprobengröße (n), Cronbachs α und die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$) angegeben. In Hinblick auf die Übersicht sind weitere Details zu den Reliabilitäten der einzelnen Versuche im Anhang A.2.1 zu finden.*

Bereich mit $.74 \leq \alpha \leq .87$ ($65 \leq n \leq 89$). In Hinblick auf die Übersicht sind weitere Details zu den Reliabilitäten der einzelnen Versuche im Anhang A.2.1 zu finden. Auf Ebene des Gesamtdatensatzes für alle Kombinationen von Studierenden und Praktikumsversuche, der später in der Studie betrachtet wird, erhält man Cronbachs α von $\alpha = .81$ ($n = 1805$). Da die Daten auf Ebene des Gesamtdatensatz teilweise abhängig sind, kann eine Verzerrung allerdings nicht ausgeschlossen werden. Die Trennschärfe der Items liegt auch für die aktuelle Studie mit Werten im Bereich von $.38 \leq r_{i(t-i)} \leq .62$ in einem mittleren bis hohen Bereich (Tab. 8; Anhang A.2.1). Das negativ formulierte Item zeigt dabei für jeden Messzeitpunkt (Praktikumsversuch) im Vergleich zu den anderen Items tendenziell niedrige Itemtrennschärfen, die im Mittel aber vertretbar im mittleren Bereich liegen ($\bar{r}_{i(t-i)} = .38$, Anhang A.2.1) und mit Blick auf die hohen Werte für die Reliabilität ebenfalls keinen Grund bieten, die Güte der Skala anzuzweifeln.

Insgesamt zeigt sich, dass die angepasste Skala zur wahrgenommenen Relevanz eine konsistente und trennscharfe Messung liefert, um die wahrgenommene Relevanz der Praktikumsversuche durch die Studierenden zu untersuchen.

8.2.3 Erhebungsinstrument: Versuchseigenschaften

In der zweiten Forschungsfrage der Arbeit sollen die Merkmale identifiziert werden, die besonders relevant wahrgenommene Praktikumsversuche ausmachen (Abschnitt 4.2.1). Nur anhand einer möglichst vollständigen Einschätzung der Versuchseigenschaften durch die Studierenden kann ein adäquater Zusammenhang zwischen diesen und der wahrgenommenen Relevanz hergestellt werden. Da hierfür kein geeignetes Instrument zur Verfügung stand, wurde für diesen Zweck eine Eigenentwicklung vorgenommen und ab dem Wintersemester 2017/2018 eingesetzt.

Die Entwicklung des Instrumentes zur Erfassung der Versuchseigenschaften wurde anhand von Expertengesprächen in zwei Phasen vorgenommen. Die Expertenrunden bestanden aus Physiker*innen mit Erfahrungen in der Betreuung von Prak-

tikumsversuchen. In der ersten Phase wurden zunächst Eigenschaften von Praktikumsversuchen gesammelt. Dies hatte das Ziel, eine möglichst vollständige Abdeckung aller möglichen Eigenschaften zu erhalten, um keinen potentiellen Einflussfaktor auf die wahrgenommene Relevanz unberücksichtigt zu lassen. Anschließend wurden die gefundenen Eigenschaften in der zweiten Runde durch Entfernung von Redundanzen und Bildung von Kategorien zusammengefasst. Dadurch haben sich 27 Haupteigenschaften herauskristallisiert, die die gesamte Breite an möglichen Merkmalen der Praktikumsversuche abdecken. Die Studierenden bewerten zu jedem durchgeführten Versuch auf dem resultierten einseitigen Fragebogen mittels einer 3-teiligen Likert-Skala das Zutreffen der Eigenschaften auf den Versuch. Ein Beispiele für ein Item im Fragebogen ist: „Die Durchführung des Versuches erfordert praktische Geschicklichkeit." Für die Strukturierung des Fragebogen wurden die Eigenschaften in neun Kategorien sortiert, die sich auf Elemente des Praktikumsversuches wie beispielsweise Anleitung, Material und Durchführung beziehen. Um weitere potentielle Einflussgrößen zu erfassen, wurde die Erhebung der Versuchseigenschaften durch weitere Fragen ergänzt, die sich auf eigene Schulerfahrungen in Bezug auf den Versuch, auf den Zeitaufwand bei der Bearbeitung des Praktikumsversuches und den wahrgenommenen Schwierigkeitsgrad des Versuches beziehen. Der so entstandene Fragebogen wurde anschließend mit wissenschaftlichen Hilfskräften aus der Praktikumsbetreuung mithilfe der Think-Aloud-Methode (vgl. z. B. Häder, 2015) auf seine Verständlichkeit getestet und entsprechend verbessert. Die finalen Items zur Versuchscharakterisierung sowie deren Abkürzungen, die im Folgenden zur Übersicht genutzt werden, sind in Anhang A.2.2 zu finden.

Zwar wurde mithilfe der Expertenrunden eine möglichst große Abdeckung aller Eigenschaften von Praktikumsversuchen erreicht, um alle potentiellen Einflussfaktoren auf die wahrgenommene Relevanz in nachfolgenden Analysen berücksichtigen zu können. Aber um die 27 Praktikumsmerkmale des Erhebungsinstrumentes auf Redundanzen beurteilen zu können, werden im Folgenden paarweise Korrelationen zwischen den Merkmalen betrachtet, wie in Abschnitt 8.2.1 erläutert. Abbildung 17 zeigt die Korrelationsmatrix für die verschiedenen Versuchseigenschaften des Erhebungsinstrumentes. Für den Großteil der Versuchseigenschaften sind die Korrelationen klein ($r < .3$). Diese geringen Zusammenhänge zeigen, dass die Versuchseigenschaften in diesem Satz an Items weitgehend unabhängig sind. Nur für wenige Einzelfälle liegt die Korrelation deutlich höher, was auf Redundanzen hindeutet. Die Merkmale *verständliche Aufgabenstellung* und *klares Versuchsziel* korrelieren hoch ($r > .5$) mit *klar strukturierte Anleitung*. Dieser Zusammenhang der Merkmale lässt sich auch inhaltlich gut nachvollziehen, da alle drei strukturelle Aspekte der Anleitung sind. Daher werden diese drei Items in den folgenden Auswertungsschritten über Mittelwertbildung zusammengefasst, um die Redundanz zu entfernen und damit ohne große inhaltliche Verluste den Umfang des Datensatzes zu reduzieren. Auf weitere Zusammenfassung von Items wird verzichtet, da alle anderen Korrelationen niedrig sind und somit mit inhaltlichen Verlusten zu rechnen wäre.

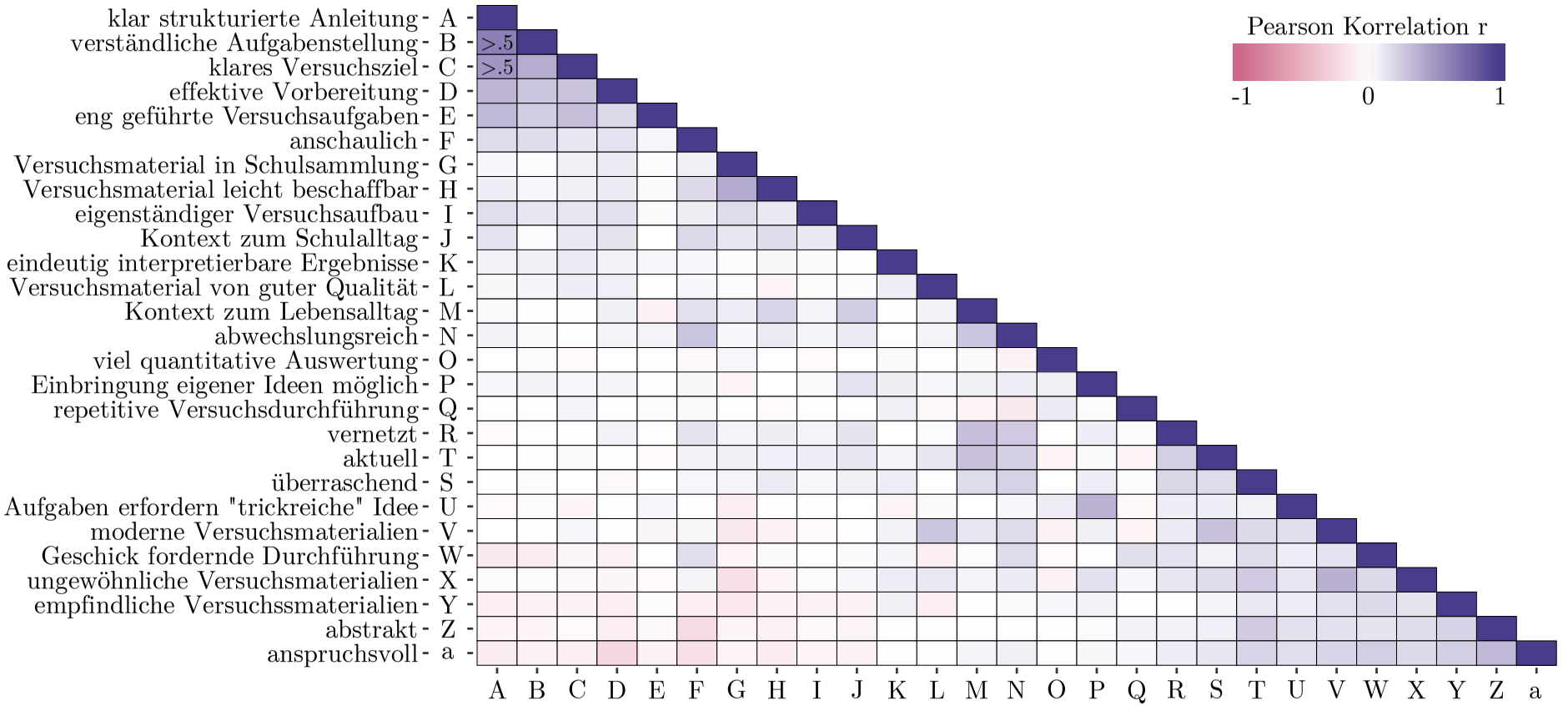


Abbildung 17: *Pearson Korrelation r der Versuchseigenschaften untereinander für den Gesamtdatensatz mit allen Kombinationen an Personen und Versuchen ($n \approx 900$). Zur Übersicht sind die Versuchseigenschaften auf der x-Achse mit Buchstaben A-a abgekürzt, die entsprechend der Beschriftungen an der y-Achse zugeordnet werden können. Hohe Korrelationen von $r > .5$ sind im Korrelationsplot zusätzlich zur Farbkodierung gekennzeichnet.*

Da dieses Erhebungsinstrument speziell für den Einsatz in diesem Praktikum neu entwickelt wurde, steht eine Prüfung der Validität des Instrumentes noch aus. Eine inhaltliche Zuverlässigkeit des Instrumentes ist allerdings Voraussetzung für die Qualität der Erhebung. Inwieweit dieses Instrument tatsächlich die Versuchseigenschaften misst, wird in dieser Arbeit im Rahmen der Analyse der Versuchseigenschaften eingeschätzt (Abschnitt 9.4.1). Da keine vergleichbaren Tests bekannt sind, wird zur Prüfung der Plausibilität der von den Studierenden erkannten Versuchseigenschaften die Perspektive der Entwicklungssicht einbezogen. Ein Vorteil dieses neuen Instruments gegenüber anderen Vorgehen, wie zum Beispiel explorativen Repertory Grid Interviews (z. B. Massolt & Borowski, 2020), ist, dass es die Eigenschaften der Versuche zudem zeitlich ökonomisch erfasst.

Zusammengenommen zeigt sich, dass das entwickelte Instrument die Eigenschaften der Versuche mit wenig zeitlichem Aufwand in großer Bandbreite erfasst. Zusammen mit der Messung der wahrgenommenen Relevanz bildet die Erfassung der Versuchseigenschaften somit eine gute Basis für die Untersuchung der Forschungsfragen zur wahrgenommenen Relevanz der Praktikumsversuche.

8.3 Statistische Methodik

Nachdem im vorangegangenen Abschnitt das Studiendesign und die Erhebungsinstrumente vorgestellt wurden, soll dieser Abschnitt einen kurzen Überblick über die statistischen Verfahren geben und vor allem deren Anwendung zur Analyse der Daten erläutern. Die Untersuchung eines komplexen Forschungsgegenstandes wie eines Praktikums erfordert dabei ein besonderes Vorgehen, wie es aus der vorangestellten Betrachtung der Datenstruktur hervorgeht. Für weiterführende Informationen zu den Verfahren und mathematischen Darstellungen wird an dieser Stelle auf die angegebenen Literaturhinweise verwiesen.

8.3.1 Datenstruktur

Nach den vorangegangenen methodischen Betrachtungen könnte nun direkt die Beschreibung der statistischen Vorgehensweise zur Untersuchung der Forschungsfragen erfolgen. Die Komplexität der Untersuchung des Praktikums spiegelt sich allerdings auch in der Datenstruktur wieder. Diese Datenstruktur hat einen direkten Einfluss auf die Analysen. Daher ist das erste Ziel dieses Abschnittes, eine detaillierte Kenntnis der Datenstruktur zu vermitteln, damit die unterschiedlichen Analyseebenen und die entsprechend gewählten Analysemethoden in Abschnitt 9 und Teil IV überblickt und eingeordnet werden können.

Im Rahmen dieser Arbeit wurde das neu entwickelte Anfängerpraktikum für Lehramtsstudierende über sieben Semester lang begleitend beforscht. Über den Zeitraum der Studie wurden alle Versuche und die Bedingungen, unter denen die Studie-

renden die Versuche durchgeführt haben, möglichst gleich gehalten. Anhand dieser Konzeption der Studie ist plausibel, dass der Zeitpunkt, an dem der von den Studierenden in Hinsicht auf Relevanz eingeschätzte Versuch stattgefunden hat, keinen oder nur einen sehr geringen Einfluss hat. In diesem Fall könnten zeitliche Effekte und Kohorteneffekte vernachlässigt werden, sodass die Daten als Gesamtstichprobe betrachtet werden können. Dies vereinfacht zum einen die komplexe Datenstruktur und damit die Analysen der Forschungsfragen deutlich. Zum anderen lässt dies die Betrachtung der großen Gesamtstichprobe statt mehrerer kleiner Stichproben zu. Dies verbessert die Nachweisbarkeit von Effekten. Für Korrelationen gibt Cohen (1992) als Orientierungswerte für das Standard-Signifikanzlevel von $\alpha = .05$ und die empfohlene statistische Power von $\beta = .8$ an, dass 783 Probanden notwendig sind für den Nachweis kleiner Effekte ($r = .1$) und 85 Probanden für mittlere Effekte ($r = .3$). Für Regressionsanalysen gelten ähnliche Größenordnungen. Anhand dessen ist eine Zusammenlegung der Kohorten zu einer Gesamtstichprobe von insgesamt 89 Probanden entscheidend für den Nachweis von Wirkungen im Praktikum. Die Hypothese, dass keine zeitlichen Effekte und Kohorteneffekte nachweisbar sind, wurde für die jeweils betrachteten Daten im Vorfeld der Analyse geprüft, wie zum Beispiel in der Betrachtung für die wahrgenommene Relevanz in Abschnitt 9.1. In der folgenden Erläuterung der Datenstruktur wird angenommen, dass Kohorteneffekte und zeitliche Effekte vernachlässigbar sind.

Wie auch aus der Beschreibung des Studiendesigns hervorgeht (Abschnitt 8.1), wurden einige Merkmale in Prä- und Post-Tests erfasst und andere begleitend zu jedem der durchgeführten Praktikumsversuche. Für die versuchsbegleitend erfasste wahrgenommene Relevanz würde man damit im Idealfall für jeden der 89 Lehramtsstudierenden 24 versuchsbegleitende Messungen, also insgesamt etwa über 2100 Messungen der wahrgenommenen Relevanz erhalten. In der Realität zeigen sich statt dieser Vollerhebung natürlicherweise trotzdem fehlende Werte. Wie aus der Betrachtung der Stichprobe in Abschnitt 5.4 hervorgeht, liegt dies zu einem Großteil daran, dass rund 20 der Studierenden nicht an beiden Praktikumsteilen teilgenommen haben und somit nur 12 statt 24 Versuche durchgeführt haben. Eine andere Ursache für fehlende Daten ist, dass trotz Kontrollmaßnahmen die versuchsbegleitenden Fragebögen teilweise nicht oder nicht ausgefüllt zurückgegeben wurden. Dadurch ist für keinen der Proband*innen ein vollständiger Datensatz mit versuchsbegleitenden Messungen für alle Praktikumsversuche vorhanden. Dennoch konnten knapp 1800 versuchsbegleitende Messungen vorgenommen werden, was rund 80 % der Vollerhebung entspricht. Daraus folgt für den Datensatz trotzdem, dass er unvollständig ist, was für die Auswahl des statistischen Analyseverfahrens bedeutend ist.

Nach Cohen (1992) bedeutet dies, dass bei einer so hohen Stichprobe auch kleine Effekte nachweisbar werden. Allerdings ist an dieser Stelle zu bedenken, dass die einzelnen versuchsbegleitenden Beobachtungen der wahrgenommenen Relevanz durch die Messwiederholung in Personen und Praktikumsversuche gruppiert und damit abhängig sind. Diese Abhängigkeit der Daten muss in Analysen berücksich-

tigt werden, da andernfalls das Risiko von Fehlschlüssen bei Interpretationen von Zusammenhängen durch Vermischung der Datenebenen und erhöhte Irrtumswahrscheinlichkeiten bei der Absicherung von Zusammenhangsmaßen besteht (siehe z. B. Eid et al., 2010). Man spricht in diesem Fall von einer Mehrebenenstruktur der Daten oder auch von hierarischen Datenstrukturen (siehe z. B. Eid et al., 2010). Ein typisches Beispiel in der fachdidaktischen Forschung ist die Gruppierung, auch Clusterung, Schachtelung oder Nestung, von Schüler*innen in Schulklassen. Die unterste Datenstufe, die im folgenden mit Level 1 bezeichnet wird, wäre in diesem Beispiel die Beobachtung der Schüler*innen. Level 2 würde dann die Gruppierung der Daten in Klassen beschreiben. In der Mehrebenenstruktur in diesem Datensatz kann jede einzelne Beobachtung (Level 1) zwei unabhängigen Gruppen, also Person (Level 2a) und Versuch (Level 2b), zugeordnet werden. Bei Berücksichtigung beider Gruppen spricht man von *cross-classification* anstatt einer hierarischen Datenstruktur (z. B. Snijders & Bosker, 1999). Anhand der Gruppenzugehörigkeit kann der Datensatz auf unterschiedlichen Ebenen betrachtet werden. Dies ist insofern interessant, da je nach betrachteter Ebene unterschiedliche Mittelungen und Datenanalysemethoden erforderlich sind.

In Abbildung 18 sind die Datenebenen (Versuche und Personen) für den Datensatz schematisch veranschaulicht, um in den nachfolgenden Abschnitten darauf zurückgreifen zu können. Abbildung 18A zeigt den *Gesamtdatensatz* bestehend aus allen einzelnen Beobachtungen, also allen vorhandenen Kombinationen von Personen- und Versuchsbewertungen. Abbildung 18B zeigt die Ebene *Zwischen Personen*, für die die versuchsbezogenen Einschätzungen für jede Person aggregiert werden. Ein Beispiel dafür wäre der Vergleich der Personenmittelwerte der wahrgenommenen Relevanz über alle Versuchsbewertungen als ein Maß für das grundsätzliche Relevanzempfinden im Praktikum. Ebenso kann auf Ebene *Zwischen Versuchen* zum Beispiel die mittlere wahrgenommene Relevanz für jeden Versuch bestimmt und als Versuchsmaß verglichen werden (Abb. 18C). Die Ebene *Innerhalb Personen* bezieht sich auf den Vergleich der Beobachtungen einer Person und kann damit auch als Spezialfall der Ebene *Zwischen Versuchen* bezogen auf eine Person gesehen werden (Abb. 18D). Analog verhält es sich für die Ebene *Innerhalb Versuche* als Betrachtung aller Beobachtungen in Bezug auf einen einzelnen Versuch (Abb. 18E).

Diese Systematisierung der Analyseebenen der versuchsbegleitenden Daten dient nicht nur dazu, die Datenebene im Folgenden mit einem prägnanten Begriff beschreiben zu können und dadurch Orientierung zu bieten. Diese Datenebenen sind nützlich, um beim statistischen Vorgehen die beschriebenen Abhängigkeiten der Daten neben der Unvollständigkeit der versuchsbegleitenden Messungen zu berücksichtigen.

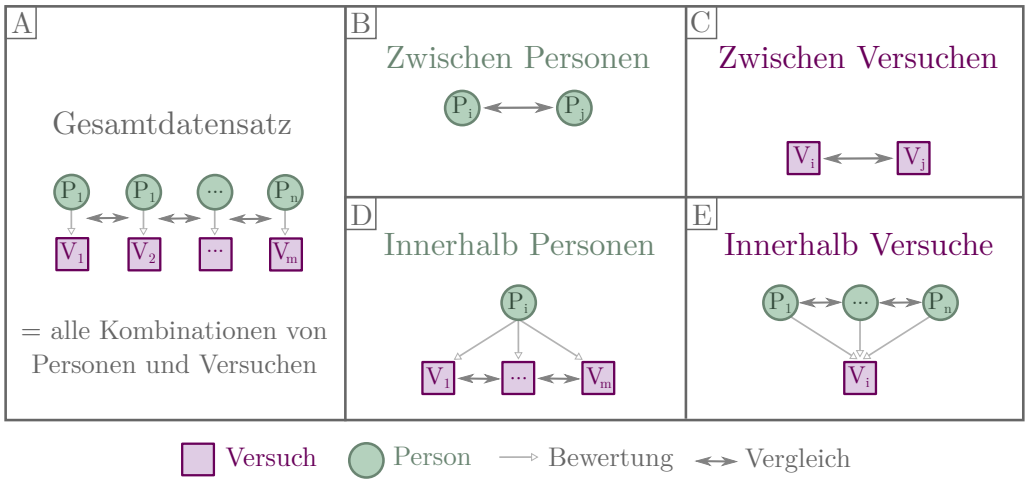


Abbildung 18: *Darstellung von fünf Betrachtungsebenen (A: Gesamtdatensatz, B: Zwischen Personen, C: Zwischen Versuchen, D: Innerhalb Personen, E: Innerhalb Versuche) der Struktur des Datensatzes mit den Clustern Personen (grüne Kreise, P) und Versuche (magentafarbene Vierecke, V). Die Indizes $i|j \in \{1, ..., n|m\}$ nummerieren die verschiedenen Personen bzw. Versuche durch, wobei sich n bzw. m auf die Anzahl aller Personen bzw. Versuche bezieht. Die hellgrauen Pfeile stehen für die Bewertung eines Versuches durch eine Person. Die dunkleren Doppelpfeile stehen für den Vergleich von Daten.*

8.3.2 Zentrierung

Aus der vorangegangenen Betrachtung des Datensatzes zeigt sich, dass bei den statistischen Analysen die Abhängigkeit der versuchsbegleitenden Daten durch die Gruppierung in Personen und Versuche berücksichtigt werden muss. Andernfalls besteht die Gefahr von Fehlschlüssen unter anderem durch verzerrte Standardfehler. Field et al. (2012) zeigt auf, dass für Visualisierungen zum Vergleich von Unterschieden bei Messwiederholungen die Standardfehler adjustiert werden müssen, um korrekte Fehlerbalken darzustellen. Field et al. (2012) schlägt dazu eine Zentrierung der Daten um den Personenmittelwert vor. Dadurch werden die Daten um die Varianz zwischen den Personen bereinigt, die Abhängigkeit der Daten berücksichtigt und die Standardfehler möglichst gut geschätzt. Zentrierungen an Gruppenmittelwerten werden auch bei Mehrebenenanalysen genutzt, um Prädiktoren auf der Datenebene innerhalb der Gruppe zu erhalten (z. B. Eid et al., 2010, siehe auch Abschnitt 8.3.5). Die Zentrierung von Daten ist in dieser Arbeit daher nicht nur für die Visualisierung von Gruppenunterschieden relevant, wie in diesem Abschnitt nachfolgend weiter dargestellt. Auch in den folgenden Abschnitten wird das Vorgehen der Zentrierung von Daten wieder aufgegriffen, um die besondere Struktur des Datensatzes zu berücksichtigen.

Die Zentrierung ist nachfolgend beispielsweise relevant für die Betrachtung der wahrgenommenen Relevanz gemittelt für die einzelnen Praktikumsversuche (Abb. 22), also auf der Datenebene *Zwischen Versuche* (siehe Abb. 18C). Die wahrgenom-

mene Relevanz wird für die unterschiedlichen Praktikumsversuche erhoben, sodass es sich nicht um eine Wiederholung der gleichen Messung handelt. Da aber die gleichen Personen die Relevanz der zu vergleichenden Versuche bewertet haben, sind die Daten in Bezug auf die Personen abhängig. Bei der Erfassung der wahrgenommenen Relevanz setzt jede*r Proband*in seine subjektive Skala als Maßstab für die Bewertung der unterschiedlichen Versuche an. Varianz zwischen den Personen entsteht dann dadurch, dass es Personen gibt, die alle Versuche grundsätzlich als relevanter ansehen als andere Personen. Zur Bereinigung der Daten um diese Varianz zwischen den Personen wird der Personenmittelwert gebildet. Der Personenmittelwert der wahrgenommenen Relevanz über alle Versuchsbewertungen kann als ein Maß für die grundlegende Relevanzwahrnehmung der Person im Praktikum insgesamt interpretiert werden. Voraussetzung dafür ist, dass eine zuverlässige Schätzung dieser Grundeinstellung vorliegt. Da die Personenmittelwerte in dieser Arbeit aber anhand einer Vielzahl von Versuchsbewertungen (bis zu 24) pro Person bestimmt werden, wird angenommen, dass die Grundeinstellung der Person zur Relevanz der Praktikumsversuche gut mit dem Personenmittelwert erfasst wird.

Für die Zentrierung wird von den gemessenen Werten der Relevanz $x_{p,v}$ für die Kombination von Person p und Versuch v der zugehörige Personenmittel $\bar{x}_p$ abgezogen. Dadurch wird die Varianz der Mittelwerte auf Null gesetzt. Gemäß der Interpretation des Mittelwertes als Grundeinstellung werden alle Messwerte um die Varianz zwischen Personen bedingt durch Persönlichkeitsunterschiede und die dadurch zustande kommenden Abhängigkeit der Daten bereinigt. Das Zentrieren reduziert die Standardfehler, die bei Betrachtung der Daten als unabhängig voneinander in der Regel zu hoch geschätzt würden (siehe Field et al., 2012). Auf die so von der Personenvarianz bereinigten Werte wird anschließend der Mittelwert des Gesamtdatensatzes $\bar{X}$ addiert, um der Verteilung der resultierenden zentrierten Werten ein sinnvoll interpretierbares Zentrum zu geben. Dieser Schritt ist mathematisch nicht zwingend notwendig, aber er vermeidet Fehlinterpretationen, die beispielsweise aus der Visualisierung von Balkendiagrammen der zentrierten Daten mit einer mittleren Relevanz von Null und negativen Werten der Relevanz resultieren kann. Seien n die Anzahl der Personen, m die Anzahl der Versuche, $k \in \{1, ..., n\}$ der Personenindex und $i \in \{1, ..., m\}$ der Versuchsindex für den Fall vollständiger Daten, dann ergeben sich die zentrierten Werte $z_{p,v}$ wie folgt:

$$\begin{aligned} z_{p,v} &= x_{p,v} - \frac{1}{m}\sum_{i=1}^{m} x_{p,i} + \frac{1}{n \cdot m}\sum_{k=1}^{n}\sum_{i=1}^{m} x_{k,i} \\ &= x_{p,v} - \bar{x}_p + \bar{X}. \end{aligned}$$

Das angesprochene Balkendiagramm der zentrierten wahrgenommenen Relevanz aufgeschlüsselt auf die verschiedenen Praktikumsversuche (Abb. 22) zeigt im Vergleich zum analogen Diagramm für die nicht zentrierten, gemessenen Daten (Abb. 86 im Anhang), dass die Reihung der Versuche absteigend nach der Relevanz sowie deren Mittelwerte nahezu gleich sind und lediglich die Standardfehler gemäß der theo-

retischen Überlegungen zur Zentrierung merklich geringer sind (vgl. auch Field et al., 2012). Betrachtet man zusätzlich die Varianz der gemessenen wahrgenommenen Relevanz für alle Kombinationen von Personen und Versuchen (*Gesamtdatensatz*, siehe Abb. 18A), so zeigt sich, dass 25 % der Gesamtvarianz auf Unterschieden zwischen Personen beruhen (siehe auch Abschnitte 8.3.5 und 9.2). Für die zentrierten Daten ergibt sich dieser Varianzanteil durch Personenunterschiede zu annähernd Null. Dies zeigt, dass die Zentrierung der versuchsbezogenen Daten um den Personenmittelwert eine einfache Methode darstellt, um die Personenvarianz aus den Daten zu entfernen.

8.3.3 Prüfung von Testvoraussetzungen

An die Anwendung statistischer Verfahren sind oftmals verschiedene Voraussetzungen geknüpft, die zu Beginn der statistischen Auswertung zu prüfen sind (vgl z. B. Eid et al., 2010). Für die in dieser Arbeit eingesetzten Verfahren gehören zu diesen Bedingungen die Unabhängigkeit, Normalverteilung, Varianzhomogenität und Intervallskalierung der Daten. Genügen die Daten nicht den entsprechenden Anforderungen, müssen Tests für nicht-parametrische Daten herangezogen werden. Nachfolgend wird kurz aufgezeigt, wie diese Testvoraussetzungen in dieser Arbeit geprüft werden, um entsprechend geeignete statistische Tests auswählen zu können.

- **Unabhängigkeit der Daten:** Wenn sich die Stichproben hinsichtlich der Personen und Messungen nicht gegenseitig bedingen, dann sind diese unabhängig (z. B. Eid et al., 2010). Unabhängige Stichproben liegen daher beispielsweise vor, wenn getrennte Gruppen von unterschiedlichen Personen (z. B. durch unterschiedliche Untersuchungsbedingungen) zum gleichen Merkmal befragt werden. Werden hingegen gleiche Personen mithilfe des gleichen Erhebungsinstrumentes an zwei unterschiedlichen Zeitpunkten (z. B. zwei verschiedene Praktikumsversuche) befragt, dann liegt eine abhängige Stichprobe vor. Die Unabhängigkeit bzw. Abhängigkeit der zu analysierten Daten ergibt sich damit für die jeweilige Betrachtung aus dem Design dieser Studie.

- **Normalverteilung:** Um zu prüfen, ob eine Normalverteilung von Daten angenommen werden kann, wurde in dieser Arbeit der Shapiro-Wilk-Test, kurz **Shapiro-Test**, herangezogen, wie es von Field et al. (2012) empfohlen wird. Die Idee des Tests ist, die erwartete Varianz der Daten bei Normalverteilung und die tatsächliche Varianz der Daten zu vergleichen (siehe z. B. Hedderich & Sachs, 2016). Der Shapiro-Test prüft die Null-Hypothese, dass die Verteilung normal ist (siehe z. B. Field et al., 2012). Ein signifikanter Test bedeutet, dass die Verteilung der Daten signifikant unterschiedlich zu einer Normalverteilung ist. Berichtet wird die Teststatistik W und das Signifikanzniveau p. Zusätzlich wurden die Daten auch zur Absicherung graphisch betrachtet, da der Test bei großen Stichproben an seine Grenze stößt und auch bei geringen Abweichungen von der Normalverteilung signifikant wird (siehe z. B. Field et al., 2012).

Dargestellt sind dazu neben Häufigkeitsverteilungen in Histogrammen auch Box-Whisker-Diagramme (kurz: **Box-Plots**; nach Tukey, 1977, z. B. Abb. 19, siehe dazu z. B. Bortz und Schuster, 2010). In dieser Darstellungsform umfasst die Box die mittleren 50 % der Verteilung (Interquartilabstand). Zusätzlich zum darin markierten Median wurden in dieser Arbeit auch Mittelwerte mit Standardfehlern hinzugefügt, da diese in weiterführenden Analysen betrachtet werden. Striche (Whisker) markieren die Ränder der Verteilung. In dieser Arbeit wurde die Variante der Darstellung gewählt, in der die Whisker auf die 1.5-fache Länge des Interquartilabstandes beschränkt sind und Daten darüber hinaus als Punkte markiert sind.

- **Varianzhomogenität:** Die Homogenität der Varianz von Stichproben wird in dieser Arbeit mit dem **Levene-Test** aus dem R-Paket *car* (Fox & Weisberg, 2019) überprüft, wie es zum Beispiel von Field et al. (2012) empfohlen wird. Der Levene-Test prüft die Null-Hypothese, dass der Unterschied zwischen den Varianzen verschiedener Gruppen Null ist (siehe z. B. Levene, 1960; Eid et al., 2010; Field et al., 2012). Ist der Test daher nicht signifikant, bedeutet dies, dass die Varianzen als homogen angenommen werden können. Die Teststatistik ist formal identisch zur F-verteilten Teststatistik der nachfolgend beschriebenen einfaktoriellen Varianzanalyse angewendet auf Abweichungen vom Gruppenmittelwert der Variablen statt auf die Variable selbst (siehe z. B. Field et al., 2012). Berichtet wird die Teststatistik mit den Freiheitsgraden $F(df_1, df_2)$ und das Signifikanzniveau p.

- **Intervallskalierung:** Für einige Tests müssen die Variablen intervallskaliert sein. Das bedeutet, dass der Abstand zweier aufeinanderfolgender Skalenwerte konstant ist und sinnvoll interpretiert werden kann (z. B. Döring & Bortz, 2016). In der vorliegenden Arbeit werden hauptsächlich Likert-Skalen zur Einschätzung von Merkmalsausprägungen und später in Teil IV auch intervallskalierte Testscores aus dichotomen Skalen genutzt. Für diese Skalen kann eine Intervallskalierung angenommen werden (vgl. z. B. Döring & Bortz, 2016), sodass diese Testvoraussetzung in der Arbeit als vorhanden gilt und nicht weiter geprüft wird.

8.3.4 Prüfung von Stichprobenunterschieden

In diesem Teil III sowie in Teil IV der Arbeit sind für mehrere Fragestellungen Stichproben auf Unterschiede zu prüfen. In dieser Teilanalyse wird die wahrgenommene Relevanz in Bezug auf Unterschiede zwischen den Kohorten, verschiedenen Praktikumsversuchen und unterschiedlichen Versuchsbeschreibungen geprüft. Bevor auf die methodischen Besonderheiten bezüglich dieser Analysen näher eingegangen wird, müssen zunächst die Auswahlkriterien der Tests sowie die Voraussetzungen und grundlegenden Ideen der Test als Basis der Analyse kurz umrissen werden. Für

weiterführende Informationen sei auch an dieser Stelle auf die Literatur verwiesen (z. B. Bortz & Schuster, 2010; Eid et al., 2010; Field et al., 2012).

Wenn Mittelwertvergleiche vorgenommen werden sollen, dann sind für die korrekte Wahl des Tests drei Überlegungen vorzunehmen. Im ersten Schritt muss festgestellt werden, ob zwei oder mehr Stichproben hinsichtlich der Mittelwerte verglichen werden sollen. Anschließend ist zum einen zu schauen, ob die Daten abhängig sind. Zum anderen ist zu prüfen, ob die weiteren Voraussetzungen des Tests auf die Daten zutreffen und ggf. auf einen nicht-parametrischen Test zurückgegriffen werden muss. Im Folgenden sind die vorgestellten Tests und deren Voraussetzungen in Bezug auf die erste Überlegung zur Testauswahl anhand der Anzahl der zu vergleichenden Stichproben gegliedert.

Unterschiede zweier Stichproben

Ein bewährtes parametrisches Testverfahren zur Prüfung, ob sich zwei Stichproben systematisch unterscheiden, ist der **t-Test** (z. B. Bortz & Schuster, 2010; Eid et al., 2010; Field et al., 2012). Die Null-Hypothese des t-Tests ist dabei, dass ein beobachteter Unterschied zufällig zustande gekommen ist. Ist der t-Test signifikant bedeutet dies folglich, dass sich die Stichproben signifikant voneinander unterscheiden und der Unterschied nicht zufällig ist. Man unterscheidet dabei den t-Test für unabhängige und abhängige Stichproben. Unabhängige t-Tests werden beispielsweise eingesetzt, wenn zwei Gruppen mit unterschiedlichen Probanden bezüglich derselben abhängigen Variable befragt wurden und diese auf Unterschiede geprüft werden sollen. Die Teststatistik des unabhängigen t-Tests beruht in diesem Fall auf dem Unterschied der Stichprobenmittelwerte. Für den unabhängigen t-Test ist zu prüfen, ob für beide Stichproben Normalverteilungen und Varianzhomogenität vorliegen (z. B. Field et al., 2012). Der unabhängige t-Test gilt als robust gegenüber Verletzungen dieser Voraussetzungen, sofern die Stichproben etwa vergleichbar groß sind (vgl. Bortz & Schuster, 2010). Bei Verletzung der Varianzhomogenität wird zudem in R automatisch eine Korrektur vorgenommen, bei der man auch vom Welch's t-Test spricht (z. B. Field et al., 2012). Werden gleiche Personen mithilfe des gleichen Erhebungsinstrumentes an zwei unterschiedlichen Zeitpunkten befragt, dann liegt beispielsweise eine abhängige Stichprobe vor. In diesem Fall beruht die Teststatistik auf der mittleren Differenz aller Messwertpaare. Unvollständige Datensätze werden beim abhängigen t-Test daher nicht berücksichtigt. Für den abhängigen t-Test ist zu prüfen, ob die Differenz der Messwertpaare normalverteilt ist (z. B. Field et al., 2012). Auch der abhängige t-Test gilt als robust gegenüber Verletzungen, sofern die Messwertreihen nicht negativ miteinander korrelieren (vgl. Bortz & Schuster, 2010). Berichtet wird in dieser Arbeit für den t-Test die Teststatistik mit Freiheitsgrad $t(df)$ und das Signifikanzniveau p.

Auch wenn der t-Test unter bestimmten zusätzlichen Voraussetzungen als robust gegenüber einer Verletzung der Testvoraussetzung gilt, empfehlen Bortz und Schuster (2010) insbesondere bei kleinen Stichprobenumfängen die Anwendung nicht-parametrischer Tests statt des t-Tests. Daher wird in Hinblick auf die zum Teil kleinen Stichproben bei Verletzung der Voraussetzungen des t-Tests in dieser Arbeit zur Absicherung der Ergebnisse zusätzlich ein nicht-parametrischer Test berechnet.

Das nicht-parametrische Pendant zum t-Test ist der **Wilcoxon-Rangsummen-Test** für unabhängige Daten und der **Wilcoxon-Vorzeichen-Rangtest** für abhängige Daten (z. B. Bortz & Schuster, 2010; Eid et al., 2010; Field et al., 2012). Beide Tests nehmen die Null-Hypothese an, dass ein beobachteter Unterschied zufällig zustande gekommen ist. Ein signifikanter Test bedeutet somit, dass sich die Stichproben signifikant voneinander unterscheiden. Die Idee der Wilcoxon-Tests ist weitreichender als für einen t-Test. Beim Wilcoxon-Rangsummen-Test für unabhängige Daten werden die Daten beider Stichproben hinsichtlich des Merkmals aufsteigend sortiert und durch Durchnummerieren ein Rang vergeben. Die Summen der Ränge der beiden Stichproben sollten gleich sein, wenn die Stichproben keinen Unterschied aufweisen. Entsprechend basiert die Teststatistik auf den Rangsummen. Durch dieses Vorgehen muss keine Normalverteilung der Daten vorausgesetzt werden. Beim Wilcoxon-Vorzeichen-Rangtest für abhängige Daten wird nach der gleichen Idee dem Betrag der Differenz der Messwertpaare ein Rang zugeordnet. Anschließend werden Summen über die Ränge der positiven und negativen Differenzen gebildet. Diese sollten gleich groß sein, wenn die Stichproben keinen Unterschied aufweisen, und bilden damit die Grundlage der Teststatistik. Berichtet wird die Teststatistik W und das Signifikanzniveau p. Sofern t-Test und zugehöriger Wilcoxon-Test keine wesentlich unterschiedlichen Ergebnisse zeigen, werden in dieser Arbeit die Ergebnisse des gängigeren t-Test wegen der einfacheren Interpretierbarkeit und Vergleichbarkeit dargestellt.

Zur Einschätzung der Signifikanz eines Tests wird der p-Wert angegeben. Dabei wird auf die Konvention zurückgegriffen, die p-Werte mit hochgestellten Sternen zu kodieren: signifikant: $^{*}p < .050$, hoch signifikant: $^{**}p < .010$ und höchst signifikant: $^{***}p < .001$. Zur Einschätzung der Größe des gefundenen Unterschiedes zwischen zwei Stichproben werden zusätzlich standardisierte Effektgrößen betrachtet. In der vorliegenden Arbeit wird dafür **Cohens d** angegeben, da es im Vergleich zum Korrelationskoeffizient nach Pearson auch bei sehr unterschiedlichen Gruppengrößen ein zuverlässiges Maß bildet (vgl. Field et al., 2012). Cohens d ergibt sich aus dem Quotienten aus Mittelwertdifferenz und der (gepoolten) Standardabweichung (Cohen, 1988, vgl. auch Olejnik und Algina, 2000, Torchiano, 2020). Dabei orientiert sich die Einschätzung der Höhe der Effektgröße in dieser Arbeit nach der Klassifizierung von Cohen (1988): $|d| \approx 0.2$: kleiner Effekt, $|d| \approx 0.5$: mittlerer Effekt und $|d| \approx 0.8$: großer Effekt. Im Detail wird die tatsächliche Bedeutsamkeit des Effektes zudem in Bezug auf die Studie interpretiert.

Unterschiede mehrerer Stichproben

Bei mehr als zwei Stichproben würde die mehrfache Anwendung des t-Tests zu einer α-Fehler-Kumulierung führen (vgl. z. B. Field et al., 2012). Das bedeutet, dass sich der Fehler erster Ordnung erhöht, eine Nullhypothese fälschlicherweise zurückzuweisen. Die Wahrscheinlichkeit, keinen α-Fehler zu machen, würde sich für jeden der unabhängig voneinander durchgeführten Vergleichstests multiplizieren. Bei einem Signifikanzniveau von 5 % und fünf Vergleichen läge die Wahrscheinlichkeit, einen α-Fehler zu begehen, dann bei etwa 23 % ($1 - 0.95^5$) statt der angestrebten 5 %. Aus diesem Grund wird für die Prüfung mehrerer unabhängiger Stichproben auf Mittelwertunterschiede die Varianzanalyse (kurz: ANOVA für ANalysis Of VAriance) eingesetzt (z. B. Eid et al., 2010; Bortz & Schuster, 2010; Field et al., 2012). Man spricht von einer einfaktoriellen Varianzanalyse, wenn die Stichproben durch eine kategoriale unabhängige Variable (z. B. mehrere Kohorten) gruppiert werden. Die Idee der einfaktoriellen Varianzanalyse ist der Vergleich der systematischen Varianz durch die Gruppenzugehörigkeit und der unsystematischen Varianz durch andere Faktoren anhand einer F-verteilten Teststatistik. Die Null-Hypothese der unabhängigen Varianzanalyse ist, dass alle Mittelwerte der abhängigen Variable für die verschiedenen Gruppen gleich sind. Ein signifikant ausfallender Test bedeutet also, dass sich mindestens ein Mittelwert von den anderen unterscheidet. Berichtet wird die Teststatistik mit den Freiheitsgraden $F(df_1, df_2)$ und das Signifikanzniveau p. Die Varianzanalyse ist an folgende Testvoraussetzungen geknüpft: Unabhängigkeit der Stichproben, Normalverteilung innerhalb der Stichproben, Varianzhomogenität zwischen den Stichproben und Intervallskalierung der abhängigen Variable (z. B. Field et al., 2012). Die Varianzanalyse gilt im Allgemeinen als robust gegenüber Verletzungen der Normalverteilung und der Varianzhomogenität (vgl. Eid et al., 2010; Field et al., 2012).

Trotz der Robustheit der einfaktoriellen Varianzanalyse gegenüber Verletzungen der Testvoraussetzungen, empfehlen Bortz und Schuster (2010) insbesondere bei kleinen und ungleichen Stichprobenumfängen die Anwendung nicht-parametrischer Tests. Daher werden in diesem Fall wie beim t-Test auch bei der Varianzanalyse die Ergebnisse mit einem nicht-parametrischen Test abgesichert. Als nicht-parametrische Entsprechung der einfaktoriellen unabhängigen Varianzanalyse wird von Field et al. (2012) zum Beispiel der **Kruskal-Wallis-Test** vorgeschlagen. Der Kruskal-Wallis-Test beruht wie der Wilcoxon-Rangsummen-Test auf Rangplatzsummen allerdings für mehrere Stichproben (siehe z. B. Eid et al., 2010; Field et al., 2012). Die Null-Hypothese des Tests ist, dass die Mediane aller Gruppen gleich sind. Ein signifikanter Test bedeutet somit, dass signifikante Unterschiede zwischen den Stichproben bestehen. Berichtet wird die Teststatistik mit Freiheitsgrad $H(df)$ und das Signifikanzniveau p.

Für Varianzanalysen werden verschiedene Maße für die Effektgröße betrachtet, die alle Vor- und Nachteile aufweisen (siehe z. B. Eid et al., 2010; Olejnik & Algina, 2000). In dieser Arbeit wird **Eta-Quadrat (η^2)** betrachtet, das zwar bei sehr

kleinen Stichproben den Effekt etwas überschätzen kann (Okada, 2013), aber ein etabliertes und leicht zu interpretierendes Maß ist. η^2 ergibt sich als Verhältnis der durch das Modell erklärten Quadratsumme durch die gesamte Quadratsumme (siehe z. B. Field et al., 2012). Damit kann η^2 wie das Bestimmtheitsmaß R^2 als Anteil der aufgeklärten Varianz interpretiert werden. Bei mehreren Variablen im Modell wird das partielle η^2 berechnet, indem die erklärte Quadratsumme durch die Quadratsumme, die durch keine andere Variable erklärt wird, dividiert wird (siehe z. B. Field et al., 2012). Auch für die Effektstärke η^2 wird sich in dieser Arbeit auf Kategorien von Cohen (1988) bezogen: $\eta^2 \approx .01$: kleiner Effekt, $\eta^2 \approx .06$: mittlerer Effekt und $\eta^2 \approx .14$: großer Effekt. Wie bei allen Angaben von Grenzwerten, werden auch diese Kategorien diskutiert (z. B. Olejnik & Algina, 2000) und sind mit Bezug auf die Studie zu interpretieren.

Bei mehreren unabhängigen Variablen spricht man von mehrfaktoriellen oder spezifischer zum Beispiel **zweifaktoriellen Varianzanalysen**. Bei einer zweifaktoriellen Varianzanalyse werden insgesamt drei Hypothesen in Bezug auf die Haupteffekte der zwei Faktoren alleine und den Effekt durch die Kombination der beiden Faktoren (auch: Interaktion) geprüft. Für die zweifaktorielle Varianzanalyse wird die Idee der einfaktoriellen Varianzanalyse erweitert, indem die gesamte Quadratsumme entsprechend dieser varianzgenerierenden Quellen in drei Anteile aufgeteilt wird (siehe z. B. Bortz & Schuster, 2010). Eine Erweiterung des Kruskal-Wallis-Tests für zweifaktorielle nicht-parametrische Analysen stellt der Scheirer–Ray–Hare-Test dar (Scheirer et al., 1976; Mangiafico, 2020).

Bei abhängigen und vollständigen Daten kann auf die **Varianzanalyse mit Messwiederholung** zurückgegriffen werden (siehe z. B. Eid et al., 2010; Field et al., 2012). Eine einfaktorielle Varianzanalyse mit Messwiederholung kann als zweifaktorielle Varianzanalyse mit den Personen als zufälligen Faktor aufgefasst werden, sodass die gesamte Quadratsumme in drei Anteile bezüglich des zu untersuchenden Faktors, der Personen und der Residuen aufgeteilt wird. Für die Varianzanalyse mit Messwiederholung wird als Effektstärke das generalisierte η^2_G mit der Einordung $\eta^2_G \approx .02$. kleiner Effekt, $\eta^2_G \approx .13$: mittlerer Effekt und $\eta^2_G \approx .26$: großer Effekt angegeben (Olejnik & Algina, 2003; Bakeman, 2005). Aufgrund der Verletzung der Unabhängigkeitsannahme gilt für die Varianzanalyse mit Messwiederholung neben der Normalverteilung die Annahme, dass die Daten sphärisch sind. Unter **Sphärizität** versteht man, dass die Varianzen der Differenzen zwischen allen möglichen Faktorstufen homogen sind. In diesem Fall wird nicht mit dem Levene-Test, sondern mit dem **Mauchly-Test** gearbeitet, der bei Signifikanz zeigt, dass die Annahme der Sphärizität nicht erfüllt ist (siehe z. B. Eid et al., 2010; Field et al., 2012). Zudem wird das Greenhouse-Geisser-Epsilon betrachtet (siehe z. B. Eid et al., 2010; Field et al., 2012), das bei p Faktorstufen im Bereich $1/(p-1) \leq \epsilon_{GG} \leq 1$ liegt. Je näher das ϵ_{GG} am Wert 1 liegt, desto eher ist die Annahme der Sphärizität erfüllt. Das ϵ_{GG} dient zudem als Korrekturfaktor zur Bestimmung des p-Wertes (p_{GG}). Die Varianzanalyse mit Messwiederholung wird mit dem R-Paket *ez* berechnet (Lawrence, 2016). Die

nicht-parametrische Entsprechung der Varianzanalyse mit Messwiederholung ist der Friedman-Test oder Rangvarianzanalyse nach Friedman, der auf Rangzuordnungen der Messzeitpunkte beruht und die Mediane auf Unterschiede prüft (z. B. Eid et al., 2010; Field et al., 2012).

Bei der Varianzanalyse und den zugehörigen nicht-parametrischen Tests handelt es sich um Omnibus-Tests, mit denen festgestellt werden kann, ob sich die abhängige Variable zwischen den Gruppen signifikant unterscheidet. Um herauszufinden, für welche der Gruppen Unterschiede bestehen, werden **Post-Hoc-Tests** eingesetzt, mit denen paarweise Vergleiche für die verschiedenen Gruppen unter Berücksichtigung der α-Fehler-Kumulierung vorgenommen werden. Es gibt eine Vielzahl von Verfahren für Post-Hoc-Tests, die alle verschiedene Vor- und Nachteile bezüglich Stichprobenumfang, Anzahl der Vergleichsgruppen, Testvoraussetzungen, etc. aufweisen (vgl. z. B. Eid et al., 2010; Field et al., 2012). Am einfachsten nachvollziehbar ist die Idee dahinter anhand der **Adjustierung des p-Wertes nach Bonferroni** bei paarweiser Anwendung der zuvor vorgestellten t-Tests und Wilcoxon-Tests auf mehrere Stichproben (vgl. z. B. Eid et al., 2010; Field et al., 2012). Um für die α-Fehler-Kumulierung zu korrigieren wird eine spezifische Irrtumswahrscheinlichkeit für jeden der paarweisen Vergleichstests angesetzt. Bonferroni teilt dazu das angestrebte Signifikanzniveau (5 %) durch die Anzahl der Vergleiche, sodass bei fünf Vergleichen die Irrtumswahrscheinlichkeit bei .01 liegt. Über alle Vergleichstests hinweg erhält man dadurch wieder etwa das angestrebte Signifikanzniveau (5 %). In dieser Arbeit wird für Post-Hoc-Untersuchungen das Ergebnis der paarweisen t-Tests mit Bonferroni-Adjustierung des p-Wertes berichtet. Falls die Testvoraussetzungen für den t-Test nicht erfüllt sind, wird das Ergebnis mit dem entsprechenden paarweisen Wilcoxon-Tests mit Bonferroni-Adjustierung abgesichert. Aus der Vorgehensweise der Adjustierung nach Bonferroni folgt, dass diese Methode insbesondere bei vielen Stichprobenvergleichen recht konservativ ist (vgl. Eid et al., 2010). Daher wurde zur Absicherung der Ergebnisse zusätzlich bei unabhängigen Daten auch der **Tukey-Test** betrachtet, der Vorteile bei einer großen Zahl an Vergleichen bietet, aber ähnliche Stichprobengrößen und Varianzhomogenität voraussetzt (vgl. Tukey, 1959; Field et al., 2012).

Besonderheiten der Analysen aufgrund der Datenstruktur

Aus der Betrachtung der Datenstruktur in Abschnitt 8.3.1 haben sich zwei Herausforderungen für die Analyse der versuchsbegleitenden Messungen herauskristallisiert: Die Abhängigkeit der Daten durch die Gruppierung in Personen und Versuche sowie die starke Unvollständigkeit der Daten. Diese Eigenschaften des Datensatzes haben eine große Auswirkung auf die Auswahl geeigneter Verfahren zur Prüfung von Stichprobenunterschieden. Die Unabhängigkeit der Daten ist eine Voraussetzung für einige der Tests zu Stichprobenvergleichen, wie vorangehend dargestellt.

Im Abschnitt 9 tritt diese Herausforderung der Testauswahl in vier Fällen auf. Zunächst soll der erste dieser Fälle als Beispiel dienen, um die Überlegungen zum Einsatz der Testverfahren zu verdeutlichen. Für die Analyse der wahrgenommenen Relevanz muss in einem ersten Schritt, wie in Abschnitt 8.3.1 erörtert, die Annahme zur Vernachlässigbarkeit von Kohorteneffekten geprüft werden (Abschnitt 9.1). Die versuchsbegleitend gemessene wahrgenommene Relevanz soll dazu auf Unterschiede in den insgesamt sieben Semestern geprüft werden. Die Daten sind aufgrund der Wiederholungen von Messungen für die unterschiedlichen Versuche allerdings abhängig von den Personen und den Versuchen. Da mehrere Gruppen verglichen werden sollen, liegt die Wahl der Varianzanalyse nahe. Die Varianzanalyse gilt gemeinhin als robust bei Verletzungen der Testvoraussetzungen. Dies gilt allerdings nur für die Normalverteilung und die Varianzhomogenität der Daten und weniger für die Unabhängigkeit der Daten, da letzteres das Risiko eines verzerrten α-Fehlers birgt (vgl. z. B. Bortz & Schuster, 2010). Der Abhängigkeit der Daten würde man mit einer Varianzanalyse mit Messwiederholung begegnen können. Da Varianzanalysen mit Messwiederholung wie auch der abhängige t-Test durch listenweisen Fallausschluss nur auf vollständige Datensätze angewendet werden können, kommt an dieser Stelle allerdings die starke Unvollständigkeit der Daten mit mehr als 20 % fehlenden Werten ins Spiel. Im hier untersuchten Datensatz sind die Daten für jeden der Probanden unvollständig, sodass die Varianzanalysen mit Messwiederholung keine Option ist. Um diesen Herausforderungen zu begegnen, wird in dieser Arbeit auf die Zentrierung von Daten zurückgegriffen. Durch die Zentrierung der wahrgenommenen Relevanz um den Personenmittelwert wird die Idee verfolgt, die Grundeinstellung der Personen zur Relevanz und damit die Varianz zwischen den Personen zu entfernen, wie in Abschnitt 8.3.2 erörtert. Damit könnten die zentrierten Daten als unabhängig in Bezug auf die Personen angenähert werden. Eine wichtige Voraussetzung dafür ist aber, dass diese Grundeinstellung zuverlässig geschätzt wird. Aufgrund der Vielzahl an bis zu 24 Messungen kann der Personenmittelwert grundsätzlich als gute Schätzung für die Grundeinstellung angenommen werden. Jedoch ist die Zentrierung keine vollständig reliable Schätzung der personenunabhängigen Komponente der Daten, da Abhängigkeiten der Versuchsbewertungen durch die Personen jenseits der Grundeinstellung nicht erfasst werden. Wenn die Grundeinstellung aber einen Großteil der Personenvarianz ausmacht, kann angenommen werden, dass die Zentrierung hinreichend gut ist, um die Unabhängigkeit der Daten anzunehmen. Dies kann durch die Berechnung der Intraklassen-Korrelation (siehe Abschnitt 8.3.5) in Abschnitt 8.3.2 bestätigt werden, die für die gemessene wahrgenommene Relevanz einen $ICC = .25$ und für die zentrierte wahrgenommene Relevanz einen $ICC \approx 0$ ergibt. Dies bedeutet, dass nach einer Zentrierung keine Varianz zwischen Personen mehr vorliegt und damit die Testvoraussetzung zur Unabhängigkeit nicht verletzt wird (z. B. Muthén, 1994). Folglich kann durch eine Zentrierung eine unabhängige Varianzanalyse für die Analyse der wahrgenommenen Relevanz angewendet werden, ohne eine starke Verzerrung der Irrtumswahrscheinlichkeit α befürchten zu müssen. Nach Prüfung

der Testvoraussetzungen kann nun entschieden werden, inwieweit ein unabhängiger nicht-parametrischer Test hinzugezogen werden muss, um die Ergebnisse zusätzlich abzusichern. Für den konkreten Fall der Analyse der Kohorteneffekte bietet sich eine unabhängige zweifaktorielle Varianzanalyse mit der zentrierten wahrgenommenen Relevanz als abhängige Variable und den Semestern und Versuchen als unabhängige Variablen an, um den Effekt der Kohorten insgesamt und in Bezug auf die Versuche zu prüfen.

Die Herausforderungen sind auch für weitere Stichprobenvergleiche in dieser Arbeit sehr ähnlich, sodass in diesen Fällen ebenfalls auf die Zentrierung zurückgegriffen wird. Zum Beispiel wird so die wahrgenommene Relevanz hinsichtlich der Unterschiede in Bezug auf die 25 einzelnen Praktikumsversuche untersucht (siehe Abschnitt 9.2). Auch für diesen Vergleich mehrerer Mittelwerte werden die Daten zentriert, um eine unabhängige und in diesem Fall einfaktorielle Varianzanalyse mit dem Faktor Versuche anwenden zu können. Für die Betrachtung des Effektes der unterschiedlichen Versuchsbeschreibungen (Abschnitt 9.3) wird durch Zentrierung der wahrgenommenen Relevanz die Unabhängigkeit der Daten angenähert. Da es zwei Formen unterschiedlicher Versuchsbeschreibungen gibt, in die sich die Stichprobe aufteilt, wird in diesem Fall ein unabhängiger t-Test für den Vergleich ausgewählt. In einem zweiten Schritt der Analyse werden dann auch die Versuche mit in diese Betrachtung eingeschlossen und wie für die Untersuchung der Kohorteneffekte eine unabhängige zweifaktorielle Varianzanalyse für die zentrierte wahrgenommene Relevanz und die Gruppierung in zwei Versuchsbeschreibungen und 25 Versuche angewendet.

Die Zentrierung ist in dieser Arbeit folglich eine wichtige Komponente bei der Untersuchung von Mittelwertunterschieden, um die Struktur dieses Datensatzes möglichst gut zu berücksichtigen und Fehleinschätzungen zu vermeiden.

8.3.5 Regressions- und Mehrebenenanalysen

In der empirischen Forschung spielt neben der Prüfung auf Unterschiede von Stichproben auch die Analyse von Zusammenhängen mehrerer Größen eine wichtige Rolle. In der Forschungsfrage 2 dieser Arbeit wird zum Beispiel nach einem Zusammenhang der wahrgenommenen Relevanz und der Eigenschaften der Praktikumsversuche gefragt (siehe Abschnitt 4.2.1). Bei Zusammenhangsanalysen wird in der Regel zunächst die Korrelation r von zwei Variablen betrachtet. Die Korrelation ist ein typisches Maß für lineare Zusammenhänge zweier metrischer Variablen (siehe auch Abschnitt 8.2.1, Details z. B. in Eid et al., 2010). Daher werden auch in der Analyse von Forschungsfrage 2 in Abschnitt 9.4.2 paarweise Korrelationen der wahrgenommenen Relevanz mit den einzelnen Versuchseigenschaften genutzt, um einen ersten Eindruck über die Zusammenhänge der Größen zu erhalten. Dabei wird die wahrgenommene Relevanz zentriert, um die Abhängigkeit der Daten auf Personenebene zu berücksichtigen, um die Datenebene *Innerhalb Personen* zu betrachten und um damit Fehlschlüsse zu vermeiden . Die Datenebene *Innerhalb Personen* ist dabei für

die Fragestellung von besonderem Interesse, da damit die versuchsbedingten Einschätzungen jenseits der stabilen Grundeinstellungen der Personen betrachtet werden, die im Gegensatz zu den Grundeinstellungen durch die Praktikumsgestaltung beeinflusst werden können. Bei der Untersuchung der Zusammenhänge mittels Korrelationen werden allerdings Zusammenhänge zwischen den Versuchseigenschaften untereinander und Aspekte der Mehrebenenstruktur vernachlässigt. Um die Ergebnisse der Korrelationen abzusichern und die Datenebene *Zwischen Personen* in die Betrachtung einbeziehen zu können, sind daher weitere Zusammenhangsanalysen der einzelnen Versuchseigenschaften mit der wahrgenommenen Relevanz unter Kontrolle der anderen Eigenschaften und unter Berücksichtigung der Mehrebenenstruktur der Daten erforderlich. Weitere Informationen über Zusammenhänge können mittels Regressionsanalysen gewonnen werden.

Regressionsanalysen

Das Prinzip der **Regressionsanalysen** wird im Folgenden vorgestellt, da es nicht nur relevant für die Analyse der Forschungsfrage 2, sondern auch in Teil IV wieder aufgegriffen wird. Bei einer **einfachen linearen Regression** geht man davon aus, dass zwischen einer abhängigen Variable y und einer unabhängigen Variable x, auch Prädiktor, ein linearer Zusammenhang der Form $y = b \cdot x + b_0$ besteht (siehe z. B. Eid et al., 2010). Mithilfe der Kleinste-Quadrate-Methode werden Regressionsgewicht b und Achsenabschnitt b_0 dieser Regressionsgerade geschätzt, die die Daten bestmöglich annähert. Wurden die abhängigen und unabhängigen Variablen vor der Regression für Vergleiche z-standardisiert, sodass sie Mittelwert 0 und Standardabweichung 1 haben und damit unabhängig von der Skala sind, so wird das Regressionsgewicht im Folgenden mit dem griechischen Buchstaben (β) bezeichnet. Das standardisierte Regressionsgewicht entspricht bei der einfachen linearen Regression dem Korrelationskoeffizienten und gibt an, um wieviele Standardabweichungen sich die abhängige Variable ändert, wenn die unabhängige Variable um eine Standardabweichung erhöht wird. Bei Betrachtung und Vergleich der standardisierten Regressionsgewichte β kann daher auch von Korrelationsanalysen gesprochen werden. Das **Bestimmtheitsmaß** R^2 der einfachen Regression ist der Anteil der Varianz der abhängigen Variable die durch die Varianz der unabhängigen Variable aufgeklärt wird. Damit entspricht R^2 dem quadrierten Korrelationskoeffizienten r.

Werden in einer linearen Regressionsanalyse mehrere (m) unabhängige Variablen x_i mit $i \in \{1, ..., m\}$ berücksichtigt, dann spricht man von einer **multiplen linearen Regression**. Der angenommene lineare Zusammenhang erweitert sich zu $y = \sum_{i=1}^{m}(b_i \cdot x_i) + b_0$. Das Regressionsgewicht b_i kann interpretiert werden als die Änderung der abhängigen Variable y, wenn die unabhängige Variable x_i um den Wert 1 erhöht wird und alle anderen unabhängigen Variablen konstant gehalten werden. Das standardisierte Regressionsgewicht β_i kann bei der multiplen Regression nur dann direkt als Korrelationskoeffizient interpretiert werden, wenn keine Korrelationen der unabhängigen Variablen untereinander vorhanden sind. Für die multiple Regression

wird das Bestimmtheitsmaß R^2 interpretiert als der Anteil der Varianz der abhängigen Variable, der durch alle unabhängigen Variablen gemeinsam erklärt wird. Durch die Aufnahme mehrerer unabhängiger Variablen werden somit Redundanzen durch Zusammenhänge der unabhängigen Variablen berücksichtigt und Störvariablen kontrolliert. Damit bietet sich in Bezug auf Forschungsfrage 2 eine multiple Regressionsanalyse an, um die Zusammenhänge der verschiedenen Versuchseigenschaften bei der Analyse zu berücksichtigen. Mit Verzerrungen der Regressionskoeffizienten ist allerdings bei der Fehlspezifikationen des Modelles zum Beispiel durch Auslassen relevanter Variablen zu rechnen (z. B. Eid et al., 2010). Um dem zu begegnen wurden in dieser Arbeit die Versuchseigenschaften möglichst vollständig erfasst (siehe Abschnitt 8.2.3), damit keine relevanzstiftende Eigenschaft unberücksichtigt bleibt. Im Sinne eines formativen Messmodels (Döring & Bortz, 2016) wird dazu für diese Untersuchung davon ausgegangen, dass die wahrgenommene Relevanz zu einem Großteil durch die Eigenschaften des Praktikumsversuches bestimmt ist. Mit der inferenzstatistischen Absicherung linearer Regressionen, also ob die für die Stichprobe gefundenen Zusammenhänge auf die Population übertragbar sind, sind neben der Linearität des Zusammenhanges drei weitere Hauptannahmen verbunden: Varianzhomogenität der Residuen (Homoskedastizität), Normalverteilung der Residuen und Unabhängigkeit der Residuen (z. B. Eid et al., 2010). Damit verbunden ist die Forderung der Unabhängigkeit der abhängigen Variable für die lineare Regression (z. B. Field et al., 2012), die anhand der Gruppierung in Personen und Versuche nicht gegeben ist (siehe Abschnitt 8.3.1), sodass von einer Verzerrung der Standardfehler bei einer Analyse der Daten mit multipler Regression auszugehen ist (z. B. Eid et al., 2010).

Mehrebenenanalysen

Um das zuvor angesprochene Problem der Abhängigkeit der Daten zu lösen, werden **Mehrebenenanalysen**, oder auch hierarchische lineare Modelle genutzt. Diese werden für die weiterführende Analyse der Forschungsfrage 2 in Abschnitt 9.4.2 benötigt. Da über die Theorie von Mehrebenenanalysen ganze Bücher geschrieben wurden (z. B. Snijders & Bosker, 1999), können im Rahmen dieser Arbeit nur knapp die Ideen der Analyse umrissen, wichtige Punkte für diese Arbeit angesprochen werden und für weitere Details nur auf entsprechende Literatur verwiesen werden (z. B. auch Eid et al., 2010; Field et al., 2012). Hierarchische lineare Modelle sind im Grunde ebenfalls Regressionsanalysen, in denen lineare Zusammenhänge zwischen abhängigen und unabhängigen Variablen analysiert werden. Der Unterschied zu den vorhergehend beschriebenen Regressionsanalysen besteht darin, dass Datenabhängigkeiten durch Mehrebenenstrukturen (siehe Abschnitt 8.3.1) berücksichtigt werden. Beachtet man die Mehrebenenstruktur in den Regressionen nicht, kann sich zum einen die Irrtumswahrscheinlichkeit α erhöhen, da die Standardfehler der Parameter bei Vernachlässigung der Abhängigkeiten der Daten im Modell zu gering geschätzt werden. Zudem besteht auch das Risiko, falsche Rückschlüsse zu ziehen,

indem fälschlicherweise gefundene Zusammenhängen auf der Gruppenebene als Zusammenhänge auf Ebene der einzelnen Beobachtungen interpretiert werden. In einer Mehrebenenanalyse werden daher die Zusammenhänge der Variablen auf den verschiedenen Ebenen des Datensatzes betrachtet. Neben der zuverlässigeren Schätzung der Modellparameter ermöglicht eine Mehrebenenanalyse im Gegensatz zur Regressionsanalyse somit auch die Beantwortung anspruchsvollerer Fragestellungen auf den verschiedenen Datenebenen. Für jede der Datenebenen werden in der Mehrebenenanalyse lineare Modellgleichungen formuliert. Die Regressionskoeffizienten auf Level 1 (z. B. einzelne Beobachtungen der Relevanz, siehe auch Abschnitt 8.3.1) werden dabei durch Modellgleichungen auf Level 2 (z. B. Gruppierung in Personen) ausgedrückt. Diese Schachtelung der Modellgleichungen der verschiedenen Ebenen wird anhand der nachfolgenden Erläuterung der konkreten Modelle dieser Analyse deutlicher.

Die Modelle können entsprechend der Fragestellung und Anforderung unterschiedlich spezifiziert werden. Die Mehrebenenanalyse beginnt gewöhnlich mit der Betrachtung des Nullmodells oder *Intercept-Only-Modells*, in dem die Varianz der abhängigen Variable ohne Hinzunahme eines Prädiktors auf den unterschiedlichen Datenebenen betrachtet wird. Dieses Modell gibt Aufschluss darüber, wie groß die Variation auf Level 1 und Level 2 ist. In diesem Zusammenhang wird auch die **Intraklassen-Korrelation** ($ICC \in [0, 1]$) betrachtet, die den Anteil der Gesamtvarianz angibt, der auf Unterschiede zwischen Level-2-Einheiten zurückzuführen ist (z. B. Eid et al., 2010). Die ICC ist folglich auch ein Maß für die Abhängigkeit der Daten. Ein $ICC = 0$ bedeutet, dass die gesamte Varianz auf Level-1-Einheiten zurückzuführen ist, somit keine Unterschiede zwischen den Gruppierungen in Level 2 bestehen und damit die Annahme von unabhängigen Beobachtungen nicht verletzt ist (z. B. Muthén, 1994). Auf die Frage, ob Mehrebenenstrukturen berücksichtigt werden müssen, wird als Daumenregel oft angeführt, dass bei einem $ICC > .1$ der Level-2-Cluster auf eine Mehrebenenanalyse zurückgegriffen werden sollte (z. B. Hartig & Rakoczy, 2010). Auch die Grenze von $ICC = .05$ wird angegeben, die als kleiner bis mittlerer Effekt eingeordnet wird (vgl. z. B. LeBreton & Senter, 2008). Häufig wird auch der **Designeffekt** nach Kish (1965), der neben der ICC noch die Größe der Cluster mit einbezieht, für diese Entscheidung hinzugezogen. Dazu wird empfohlen, bei einem Designeffekt deff$>$ 2, die Mehrebenenanalyse zu nutzen (z. B. Lai & Kwok, 2015). Daumenregeln und Empfehlungen wie diese werden aus statistischer Sicht oft kritisch diskutiert (z. B. Lai & Kwok, 2015), sollen in dieser Arbeit dennoch als Orientierung gelten. Zudem wird für die Entscheidung zur Mehrebenenanalyse in dieser Arbeit zusätzlich auch die Fragestellung und die Komplexität des resultierenden Modelles einbezogen.

Nach Betrachtung des Nullmodells werden die Prädiktoren im Mehrebenenmodell ergänzt. Dabei können die Prädiktoren feste Effekte haben, bei denen der Prädiktor über alle Gruppen gleich ist. Zufällige Effekte in diesem Kontext bedeuten, dass der Effekt der Prädiktoren zwischen den Gruppen als unterschiedlich ange-

nommen wird. Zufällige Effekte der Regressionsgewichte werden als *Random-Slopes* bezeichnet. Entsprechend bezeichnen *Random-Intercepts* die zufälligen Effekte der Achsenabschnitte. Von *Random-Coefficients* spricht man entsprechend bei zufälligen Effekten beider Regressionsparameter. Zudem kann die Korrelation zwischen zufälligen Achsenabschnitten und zufälligen Regressionsgewichten bestimmt werden. Im Modell können auf der Beobachtungsebene (Level 1) und Gruppenebene (Level 2) unterschiedliche oder die gleichen Prädiktoren relevant sein. Wenn der Effekt der gleichen unabhängigen Variable auf beiden Ebenen betrachtet werden soll, dann wird diese als Level-1-Prädiktor um den Gruppenmittelwert zentriert und für den Level-2-Prädiktor ein Gruppenmittelwert gebildet (z. B. Eid et al., 2010). Dies erinnert an das beschriebene Vorgehen der Zentrierung in den Abschnitten 8.3.2 und 8.3.4. Die Frage nach der aufgeklärten Varianz ist bei Mehrebenenmodellen nicht so leicht zu klären, da es keine allgemein akzeptierte Definition für R^2 gibt (z. B. Snijders & Bosker, 1999). In dieser Arbeit wird, wie von Eid et al. (2010) vorgeschlagen, ein Pseudo-R^2 betrachtet, das die aufgeklärte Varianz des Modells inklusive Prädiktor mit der des Nullmodells ohne Prädiktor vergleicht. Dadurch erhält man für jedes Level ein R^2, das angibt, welchen Effekt der Prädiktor relativ zum Nullmodell hat.

Die Regressionsparameter und Varianzkomponenten werden mittels Maximum-Likelihood-Methode geschätzt, bei dem die Wahrscheinlichkeit der beobachteten Daten unter dem gegebenen Modell iterativ maximiert wird (vgl. z. B. Eid et al., 2010). Die Mehrebenenanalysen wurden mit den R-Paketen *lme4* (Bates et al., 2015) und *lmerTest* (Kuznetsova et al., 2017) berechnet. In R steht die Maximum-Likelihood-Methode (ML) und die Restricted-Maximum-Likelihood-Methode (REML) zur Verfügung. Beide Verfahren haben unterschiedliche Vor- und Nachteile, schätzen aber sehr ähnliche Modellparameter (vgl. Field et al., 2012). Während das REML-Verfahren zum Beispiel eine Stärke bei Schätzung der Varianzkomponenten hat, wird das ML-Verfahren beim Schwerpunkt der Analyse auf Regressionsparameter bevorzugt. Entsprechend wird auch in dieser Arbeit das Schätzverfahren anhand des gelegten Schwerpunktes der Modellparameter gewählt. Daran schließt sich die Frage nach der Modellpassung an. Der Koeffizient, der zur Beurteilung der Passung betrachtet wird, ist die Devianz, die auf der maximalen Likelihood-Funktion basiert (z. B. Eid et al., 2010). Je größer dieser Wert für das betrachtete Modell in Relation zu einem anderen geschachtelten Modell ist, desto schlechter wird die Passung des Modells erachtet. Für Mehrebenenanalysen treffen im Grunde bis auf die Unabhängigkeit die gleichen Annahmen wie für Regressionsanalysen zu. Mehrebenenanalysen gelten dabei als recht robust (vgl. Eid et al., 2010). Zudem bietet die Mehrebenenanalyse den großen Vorteil insbesondere auch für diese Arbeit, dass fehlende Daten kein Problem darstellen.

Modelle zur Untersuchung von Forschungsfrage 2

In der weiterführenden Analyse von Forschungsfrage 2 in Abschnitt 9.4.2 wird eine Mehrebenenanalyse berechnet, um die Zusammenhänge der einzelnen Versuchsei-

genschaften mit der wahrgenommenen Relevanz unter Kontrolle der anderen Eigenschaften zu betrachten, um die Ergebnisse der Korrelationen damit abzusichern und um die Datenebenen der Mehrebenenstruktur in die Betrachtung mit einzubeziehen. Die dafür zugrundeliegende Spezifikation der Modelle soll an dieser Stelle erläutert werden. Damit werden auch die zuvor beschriebenen abstrakten Modellbegriffe anhand des konkreten Anwendungsbezuges deutlich.

Zur Untersuchung der Mehrebenenstruktur wird für diese Fragestellung, wie zuvor erläutert, zunächst das Nullmodell für die wahrgenommene Relevanz als abhängige Variable betrachtet. Die gemessene wahrgenommene Relevanz y_{ij} wird aufgrund unterschiedlicher Skalen der Größen im Modell z-standardisiert, sodass der Gesamtmittelwert null ist und die Varianz eins ist. In diesem Fall handelt es sich um zwei Nullmodelle, da die Beobachtungen sowohl in Personen (Model 0P), als auch Versuche (Model 0V) gruppiert sind. In der Tabelle 9 sind die Gleichungen des Modells auf beiden Datenebenen angeben. Hier zeigt sich die Verschachtelung der Gleichungen aus Level 1 und 2, die in Abschnitt 8.3.5 genannt wurde. Gemäß dem Namen Intercept-Only-Modell beinhaltet das Modell keinen Prädiktor und schätzt nur den gruppenunspezifischen Achsenabschnitt γ_{00} und die Residuen e und u auf Level 1 und 2.

Modell 0	Mehrebenenanalyse - Intercept-Only-Modell	
	Datenebenen	Level 1: Beobachtungen (i)
		Level 2: Gruppe (j) (Personen (P) oder Versuche (V))
	AV	wahrgenommene Relevanz (y_{ij}, z-standardisiert)
	Level 1	$y_{ij} = \beta_{0j} + e_{ij}$
	Level 2	$\beta_{0j} = \gamma_{00} + u_{0j}$
Modell 1	Mehrebenenanalyse - Random-Intercept-Modell	
	Datenebenen	Level 1: Beobachtungen (i)
		Level 2: Personen (j)
	UV - Level 1	$k \in \{1, ..., 28\}$ Versuchseigenschaften um den Personenmittelwert zentriert (x_{kij}, z-standardisiert)
	UV - Level 2	Personenmittelwerte der $k \in \{1, ..., 28\}$ Versuchseigenschaften ($\overline{x}_{ki}$, z-standardisiert)
	AV	wahrgenommene Relevanz (y_{ij}, z-standardisiert)
	Level 1	$y_{ij} = \beta_{0j} + \sum_k(\beta_{kj} \cdot x_{kij}) + e_{ij}$
	Level 2	$\beta_{0j} = \gamma_{00} + \sum_k(\gamma_{0k} \cdot \overline{x}_{kj}) + u_{0j}$
		$\beta_{kj} = \gamma_{k0}$

Tabelle 9: *Kurzbeschreibung der Modelle 0 und 1 zur Analyse der Zusammenhänge zwischen Versuchseigenschaften und der wahrgenommenen Relevanz. (UV: Unabhängige Variable, AV: abhängige Variable, γ_{00}: gruppenunspezifischer Achsenabschnitt, γ_{k0}: gruppenunspezifische Regressionsgewichte, γ_{0k}: gruppenspezifische Regressionsgewichte, e: Level 1 Residuum, u: Level 2 Residuen)*

Basierend auf den Ergebnissen der Nullmodelle wird in Abschnitt 9.4.2 die Spezifikation von Modell 1 mit Beobachtungen auf Level 1, der Gruppierung in Personen in Level 2 und der Vernachlässigung der Gruppierung in Versuche geschlussfolgert. Im Vergleich zum Nullmodell (0P) kommen im Random-Intercept-Modell (Modell 1) Prädiktoren hinzu. Zur Beantwortung von Forschungsfrage 2 soll der Effekt von 28 Versuchseigenschaften auf die wahrgenommene Relevanz untersucht werden. Diese sollen sowohl innerhalb der Personen (Level 1) und zwischen Personen (Level 2) betrachtet werden. Dabei sind Level 2 Prädiktoren im Modell 1 die Personenmittelwerte der $k \in \{1, ..., 28\}$ Versuchseigenschaften $\bar{x}_{ki}$ und Level 1 Prädiktoren die um den Personenmittelwert zentrierten Personenwerte x_{kij} (Tab. 9). Neben den Residuen (e, u) und dem gruppenunspezifischen Achsenabschnitt γ_{00} werden die gruppenunspezifischen Regressionsgewichte γ_{k0} (Level 1) und die gruppenspezifischen Regressionsgewichte γ_{0k} (Level 2) geschätzt. Insgesamt werden in diesem Modell also 56 Regressionsgewichte geschätzt. Die Varianz des Residualwertes u steht für die Varianz des zufälligen Effekte der gruppenspezifischen Achsenabschnitte (Random-Intercepts). Die Modelle werden mit dem ML-Verfahren geschätzt, da der Fokus der Analyse auf den festen Effekten liegt. Dieses Modell der Mehrebenenanalyse erfüllt damit alle Anforderungen, um die Forschungsfrage 2 unter Einbezug der Mehrebenenstruktur und der Beziehungen der Versuchseigenschaften als Prädiktoren untereinander zu analysieren.

8.3.6 Wachstumskurven-Modelle

Das Prinzip der Mehrebenenanalyse kommt noch an einer weiteren Stelle in der Untersuchung der wahrgenommenen Relevanz und auch später in Teil IV zum Tragen. Bei der Beschreibung der Datenstruktur in Abschnitt 8.3.1 wurden Annahmen zur Vernachlässigbarkeit von zeitlichen Effekten getroffen, die es vor der eigentlichen Analyse zu prüfen gilt. In Abschnitt 8.3.4 wurde bereits das Vorgehen zur Prüfung von Kohorteneffekten mittels Varianzanalyse vorgestellt. Da darüber hinaus eine Zunahme der wahrgenommenen Relevanz über den Zeitraum plausibel ist (siehe Abschnitt 9.1), sollen auch zeitliche Effekte auf die versuchsbegleitenden Messungen im Verlauf des Praktikums von einem bis zwei Semestern geprüft werden. Für diese Untersuchung bietet sich ein Wachstumskurven-Modell an. Das Ziel von Wachstumskurven-Modellen ist, die (lineare) Änderung einer wiederholt gemessenen Variable mit der Zeit zu betrachten (z. B. Field et al., 2012). Wachstumsmodelle sind Mehrebenenanalysen (Abschnitt 8.3.5), bei denen sich Beobachtungen (Level 1) aufgrund der Messwiederholungen in Probanden (Level 2) gruppieren und die Zeit als Prädiktor gewählt wird. Dafür wird ein kontinuierlicher Zeitverlauf angenommen. In das Modell können auch Kovariaten aufgenommen werden. Der besondere Vorteil dieses Vorgehens für die Analysen in dieser Arbeit ist neben der Berücksichtigung der Mehrebenenstruktur, dass fehlende Daten kein Problem darstellen. Die in dieser Arbeit zur Untersuchung der zeitlichen Änderung genutzten Modelle sind in Tabelle 10 dargestellt. Wie bei der Mehrebenenanalyse zuvor wird zunächst ein

Nullmodell (Modell 0) betrachtet, um die Mehrebenenstruktur der abhängigen Variable ohne Prädiktor zu untersuchen. In das Wachstumskurven-Modell (Modell A) wird der Prädiktor t_{ij} als Zeitabschnitt aufgenommen und das gruppenunspezifische Regressionsgewicht γ_{10} als zeitliche Veränderungsrate und der Achsenabschnitt γ_{00} als Ausgangswert zum Zeitpunkt $t_{ij} = 0$ mit dem ML-Verfahren geschätzt. Zusätzlich wird im Modell über die Residualwerte u_{0j} und u_{1j} die Varianz der zufälligen Effekte der gruppenspezifischen Achsenabschnitte (Random-Intercepts) und gruppenspezifischen Änderungsraten geschätzt (Random-Slopes). Zudem wird bei der Analyse auch die Korrelation von Ausgangswert und Änderungsrate bestimmt. Mithilfe dieser Analyse kann insbesondere durch Betrachtung des festen Effektes γ_{10} die zeitliche Änderung der wahrgenommenen Relevanz innerhalb von Personen über die Zeit geprüft werden.

Modell 0	Mehrebenenanalyse - Intercept-Only-Modell	
	Datenebenen	Level 1: Beobachtungen (i)
		Level 2: Gruppe (j) (Personen (P) oder Versuche (V))
	AV	Merkmal (y_{ij}, z-standardisiert)
	Level 1	$y_{ij} = \beta_{0j} + e_{ij}$
	Level 2	$\beta_{0j} = \gamma_{00} + u_{0j}$
Modell A	Wachstumskurven-Modell	
	(Mehrebenenanalyse - Random-Coefficients-Modell)	
	Datenebenen	Level 1: Beobachtungen (i)
		Level 2: Personen (j)
	UV	Messzeitpunkte ($t_{ij} \in [0, 5]$)
	AV	wahrgenommene Relevanz (y_{ij}, z-standardisiert)
	Level 1	$y_{ij} = \beta_{0j} + \beta_{1j} \cdot t_{ij} + e_{ij}$
	Level 2	$\beta_{0j} = \gamma_{00} + u_{0j}$
		$\beta_{1j} = \gamma_{10} + u_{1j}$

Tabelle 10: *Kurzbeschreibung der Modelle 0 und A des Wachstumskurvenmodells. Modell 0 dient als Nullmodell zur Untersuchung der Mehrebenenstruktur. Modell A dient zur Mehrebenenanalyse des zeitlichen Verlaufs eines Merkmals y_{ij}. (UV: Unabhängige Variable, AV: Abhängige Variable, γ_{00}: gruppenunspezifischer Achsenabschnitt, γ_{10}: gruppenunspezifisches Regressionsgewicht, γ_{01}: gruppenspezifisches Regressionsgewicht, e: Level 1 Residuum, u: Level 2 Residuen)*

9 Analyse der wahrgenommenen Relevanz

Nachdem im vorigen Abschnitt 8 das methodische Vorgehen für die Untersuchung zur wahrgenommenen Relevanz erläutert und in die Struktur des Datensatzes eingeführt wurde, kann auf dieser Grundlage nun die Analyse der Relevanz der Praktikumsversuche selbst folgen. Grundlage der Analyse sind die Forschungsfragen, die in den Abschnitten 1 und 4 entwickelt wurden. Die folgende Analyse der wahrgenommenen Relevanz gliedert sich dazu in vier Teilanalysen. Zunächst wird geprüft, ob für die wahrgenommene Relevanz Kohorteneffekte oder zeitliche Effekte vernachlässigbar sind und damit die Vereinfachungen des Datensatzes bezüglich der wahrgenommenen Relevanz zulässig sind (Abschnitt 9.1). Anschließend kann untersucht werden, in welchem Maß die Praktikumsversuche von den Studierenden überhaupt als relevant wahrgenommen werden und ob sich die Versuche bezüglich der eingeschätzten wahrgenommenen Relevanz unterscheiden (Abschnitt 9.2). Auf dieser Basis schließt sich die Untersuchung von Forschungsfrage 1 an, das heißt, in welchem Maß sich durch das Aufzeigen eines Bezuges zum Lehrberuf die von den Studierenden wahrgenommene Relevanz von physikalischen Versuchen im Anfängerpraktikum Lehramt erhöhen lässt (Abschnitt 9.3). In der vierten Analyse wird die Erfassung der Versuchseigenschaften hinzugenommen, um herauszufinden, welche Merkmale der Praktikumsversuche dazu führen, dass die Versuche von den Lehramtsstudierenden als relevanter wahrgenommen werden (Abschnitt 9.4). Vor diesem Hintergrund kann abschließend die Bewertung und Deutung der Ergebnisse ins Zentrum gestellt werden (Abschnitt 10).

9.1 Prüfung von zeitlichen Effekten

Die Datenstruktur wurde in Abschnitt 8.3.1 unter der Annahme vereinfacht, dass der Zeitpunkt, an dem der Praktikumsversuch von den Studierenden durchgeführt und bewertet wurde, keinen Effekt auf die Messung der Merkmale hat. Da die Bedingungen und Versuche im Praktikum möglichst gleich gehalten wurden, ist diese Annahme zwar plausibel, muss aber vor der eigentlichen Analyse bezüglich der Forschungsfragen geprüft werden. In diesem Abschnitt soll daher zunächst geprüft werden, ob die wahrgenommene Relevanz über die verschiedenen Semester hinweg vergleichbar ist und die Vereinfachung der Datenstruktur damit zulässig ist.

Um diese Hypothese zu prüfen wird zunächst die mittlere wahrgenommene Relevanz pro Person über die Semester hinweg auf Unterschiede hin verglichen. Da in den Semestern unterschiedliche Themenbereiche und Versuche durchgeführt werden, wurden die Winter- und Sommersemester (WS und SS) dafür getrennt voneinander untersucht. Dadurch sind die Daten unabhängig voneinander. Die Abbildung 19 zeigt Box-Plots der mittleren wahrgenommenen Relevanz pro Person jeweils für die Winter- und Sommersemester. Die Visualisierung der wahrgenommenen Relevanz deutet auf keine Unterschiede der Mittelwerte über die Semester hin. Dies bestätigt

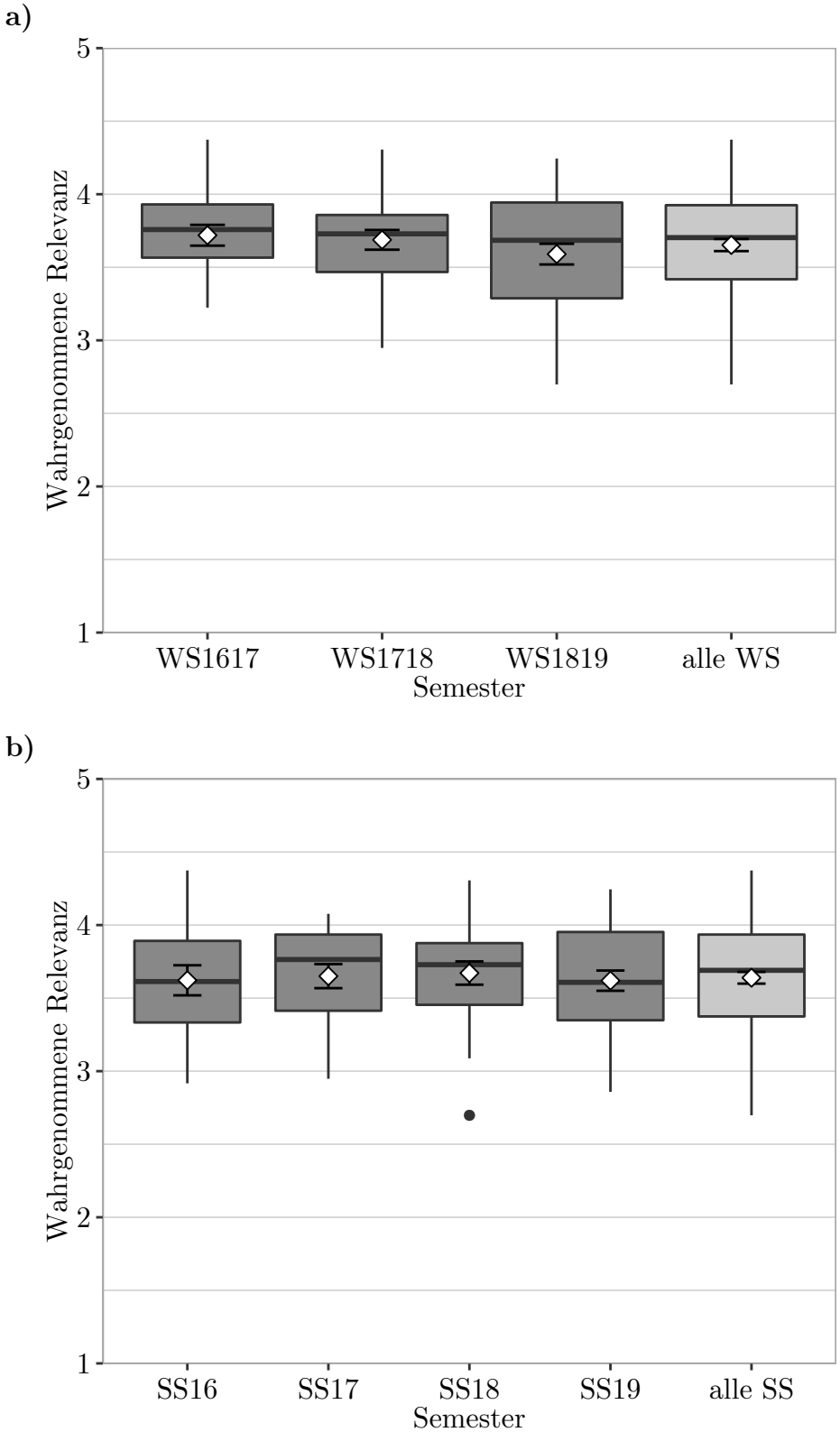


Abbildung 19: *Darstellung der wahrgenommenen Relevanz (5-teilige Likert-Skala, Mittel pro Person) für a) die Wintersemester und b) die Sommersemester ($14 \leq n \leq 33$). Der Boxplot zeigt neben Minimum, unterem Quartil, Median, oberem Quartil und Maximum zusätzlich die Mittelwerte (weiße Rauten) mit Standardfehlern.*

auch eine quantitative Analyse. Da mehrere Semester auf Mittelwertunterschiede geprüft werden müssen, wird eine unabhängige einfaktorielle Varianzanalyse durchgeführt (Abschnitt 8.3.4). Die Testvoraussetzungen für die Varianzanalyse sind erfüllt, da die Daten innerhalb der einzelnen Semester normalverteilt sind (Shapiro-Test: $0.95 \leq W \leq 0.99, p > .05$) und Varianzhomogenität aufweisen (Levene-Test: WS: $F(2, 72) = 1.57, p = .215$; SS: $F(3, 80) = 0.37, p = .771$). Das Ergebnis der Analyse zeigt keine signifikanten Unterschiede zwischen den Winter- und Sommersemestern (WS: $F(2, 72) = 0.915, p = .405, \eta^2 = .00$; SS: $F(3, 80) = 0.098, p = .961, \eta^2 = .02$). Damit ist die mittlere wahrgenommene Relevanz der Studierenden in allen Semestern als gleich anzusehen.

Durch Aggregation der wahrgenommenen Relevanz über Personen wurde in der Betrachtung allerdings nicht untersucht, inwieweit sich Kohorteneffekte für einzelne Praktikumsversuche zeigen. Um dies zu prüfen, wurde eine unabhängige Varianzanalyse mit den zentrierten Daten der wahrgenommenen Relevanz als abhängige Variable und den zwei unabhängigen Variablen Semester und Praktikumsversuch berechnet. Diese Analyse der Haupteffekte zeigt, dass keine Unterschiede der Effekte der Semester auf die wahrgenommene Relevanz für die einzelnen Versuche nachweisbar sind (Semester: $F(6, 1693) = 1.11, p = .357, \eta^2 = .00$, Versuche: $F(24, 1693) = 10.41, p < .001, \eta^2 = .13$, Interaktion Semester und Versuche: $F(81, 1693) = 1.07, p = .326, \eta^2 = .05$). Da die Testvoraussetzungen anhand von Shapiro- und Levene-Tests nicht für alle Semester und Versuche erfüllt sind, wurde die Analyse ergänzend mit dem Scheirer–Ray–Hare-Test durchgeführt, der die Ergebnisse bestätigt. Dieser Befund ergibt sich auch für die Betrachtung der nicht zentrierten Daten.

Die Ergebnisse aller Untersuchungen zusammengenommen liefern ein konsistentes Bild. Demnach sind keine Unterschiede der wahrgenommenen Relevanz durch die verschiedenen Semester nachweisbar. Dieses Ergebnis ist ein wichtiges Indiz dafür, dass die Studierenden die Versuche über die Semester unter konstant gehaltenen Bedingungen bearbeitet haben. Das Bestreben, Einflussfaktoren wie Anleitungen, Versuchsmaterialien und Betreuung konstant zu halten (Abschnitt 8.1), ist offenbar erfolgreich gewesen. Die Homogenität der wahrgenommenen Relevanz über die Semester zeigt außerdem, dass die Neuheit des Praktikums im Sinne der Redewendung „Neue Besen kehren gut.“ im ersten Jahr nach der Einführung des Moduls nicht zu einem zeitlichen Effekt in Form einer signifikant höheren Relevanz geführt hat.

Neben den Kohorteneffekten muss noch ein zweiter zeitlicher Effekt auf Ebene der einzelnen Studierenden in Bezug auf die Vereinfachung der Datenstruktur geprüft werden. Jede*r der Lehramtsstudierenden wurde in dieser Studie über einen langen Zeitraum von ein bis zwei Praktikumssemestern zur wahrgenommenen Relevanz von 12 bis 24 Praktikumsversuchen befragt (Abschnitt 8.1). Über einen Zeitraum von einem Jahr mit 24 Befragungssituationen ist trotz entgegenwirkender Gestaltung des Studiendesigns (Abschnitt 8.1) nicht auszuschließen, dass zeitliche Effekte auf dieser

Ebene eine Rolle spielen. Ein Beispiel für einen zeitlichen Entwicklungseffekt könnte dabei sein, dass die Studierenden sich nach Durchführung der ersten Versuche an das Vorgehen im Praktikum gewöhnt haben, ihren Fokus dadurch auf andere Aspekte lenken können und somit zu einer anderen Einschätzungen bezüglich der Versuche gelangen können. Ein anderer zeitlicher Effekt könnte sein, dass die wahrgenommene Relevanz im Laufe des Semesters abnimmt, da der Prüfungszeitraum am Ende des Semesters näher rückt und andere Prioritäten fordert. Auch ein zeitlicher Übertragungseffekt wäre möglich, bei dem die Studierenden nach Durchführung der ersten lehramtsspezifischen Versuche eine grundsätzlich höhere wahrgenommene Relevanz dem Lehramtspraktikum gegenüber entwickeln und sich diese auf die Wahrnehmung nachfolgender Versuche überträgt. Wenn zeitliche Aspekte bei der wahrgenommenen Relevanz vorhanden sind, dann hat dies nicht nur Einfluss auf weitere Analysen der Daten, sondern auch auf die optimale Anordnung der Praktikumsversuche im Praktikum. Aus diesem Grund wird in diesem Abschnitt vor der eigentlichen Analyse der Forschungsfragen untersucht, ob der Praktikumsverlauf einen zeitlichen Effekt auf die wahrgenommene Relevanz hat.

Im Folgenden werden unter Berücksichtigung der in Abschnitt 8.3.1 beschriebenen Mehrebenenstruktur der Daten ein Wachstumskurven-Modell zur Untersuchung der Änderung der wahrgenommenen Relevanz über den Zeitraum der Befragung berechnet (siehe Abschnitt 8.3.6 für Details zu den einzelnen Modellen). Zur Voruntersuchung der Mehrebenenstruktur der wahrgenommenen Relevanz wurden zwei Nullmodelle mit Level 1 als Beobachtungen (=Versuch als Messsituation) und den Level 2 Gruppierungen Personen (Modell 0P) und Versuche (Modell 0V) berechnet, deren Parameterschätzungen in Tabelle 11 dargestellt sind. Der Designeffekt liegt in beiden Fälle über dem Grenzwert von deff= 2, sodass demnach beide Gruppierungen in der Mehrebenenstruktur berücksichtigt werden sollten (siehe Abschnitt 8.3.5). Anhand der Intraklassen-Korrelation (ICC) (siehe Abschnitt 8.3.5) sind 26 % der Varianz auf Unterschiede zwischen den Personen ($n(pid) = 81$) zurückzuführen und 7 % auf Unterschiede zwischen Versuchen ($n(VS) = 25$). Da der Varianzanteil für Versuche klein und unterhalb des Grenzwertes der ICC von 10 % liegt, die Hauptvarianz damit in den Personen liegt und die Fragestellung sich auf die zeitliche Entwicklung der Personeneinstellung bezieht, wird nur die Personengruppierung in der anschließenden Analyse berücksichtigt, um auch die Komplexität des Modelles zu reduzieren.

Zur Untersuchung der zeitlichen Änderung der wahrgenommenen Relevanz wurden die Praktikumssemester jeweils in drei Zeiträume aufgeteilt, in denen die Studierenden jeweils vier Versuche zu einem Themenbereich durchgeführt und hinsichtlich der wahrgenommenen Relevanz bewertet haben. Es ist nur die Reihenfolge der durchgeführten Themenbereiche bekannt, da die Reihenfolge der Versuche nicht erfasst wurde, um für die Studierenden die Anonymität zu erhalten. Für das Wachstumskurvenmodell wird der Zeitabschnitt (0-5), in dem die wahrgenommene Relevanz in den Praktikumssemestern beobachtet wurde, als Prädiktor hinzugenommen (Mo-

AV: wahrgenommene Relevanz, UV: Messzeitpunkt

	Modell 0P Personen	**Modell 0V** Versuche	**Modell A**
feste Effekte			
zeitliche Änderung (γ_{10})			$\gamma_{10} = .01(.02)$ $g_{10} = 0.01(0.01)$
zufällige Effekte			
Varianz zwischen Gruppen (Var(u_{0j}))	0.26	0.07	0.27
Varianz innerhalb Gruppen (Var(e))	0.73	0.93	0.70
Varianz der Steigungen (Var(u_{1j}))			0.01
Korrelation Ausgangswert und Änderung			-0.26
ICC	.26	.07	.29
Designeffekt	6.6	5.6	
Anzahl Beobachtungen	1650	1650	1650
Anzahl Gruppen	81	25	81
Devianz	4333	4607	4316

Modell 0: Intercept-Only-Modell; **Modell A:** Random-Coefficients-Modell

Signifikanzniveau für feste Effekte: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

Tabelle 11: *Ergebnisse der Mehrebenenanalysen (Intercept-Only-Modell 0 und Random-Coefficients-Modell A, Tab. 10) des zeitlichen Verlaufs der wahrgenommenen Relevanz. (γ: standardisiertes Regressionsgewicht, g: nicht standardisiertes Regressionsgewicht, in Klammern: Standardfehler; Intercepts nicht dargestellt, da bei standardisierten Daten nicht sinnvoll interpretierbar)*

dell A in Tab. 11). Die zufälligen Effekte des Modells zeigen, dass 27 % der Gesamtvarianz zurück geht auf Unterschiede zwischen Personen am ersten Messzeitpunkt, 1 % auf Unterschiede zwischen den Änderungsraten der Personen und 71 % auf Unterschiede innerhalb der Personen. Anhand der resultierenden festen Effekte ist die mittlere Veränderungsrate ($\gamma_{10} = .01$, $g_{1o} = 0.01$) nicht signifikant von Null verschieden. Damit ist keine zeitliche Veränderung der wahrgenommenen Relevanz über den Praktikumsverlauf feststellbar. Zur Veranschaulichung sind in Abbildung 20 die Entwicklungsverläufe ohne Standardisierung für die einzelnen Personen und der mittlere Verlauf dargestellt. Während entsprechend dem Modell im Mittel keine Veränderung der wahrgenommenen Relevanz über die Zeit festzustellen ist, können für die einzelnen Personen Verläufe mit unterschiedlichen Geradensteigungen (Range: $-0.12 < b_{1j} < 0.15$, 95 %-Konfidenzintervall: $-0.02 < b_{1j} < 0.03$) erkannt werden. Die Korrelation von $r = -.26$ zwischen Ausgangswert und Veränderungsrate zeigt, dass für Studierende, die eine hohe wahrgenommene Relevanz zu Beginn des Praktikums haben, die wahrgenommene Relevanz über die Zeit weniger ansteigt als für Studierende mit einer geringeren Relevanz als Ausgangswert. Die Veränderungen der wahrgenommenen Relevanz über die Zeit für die einzelnen Studierenden sind vermutlich auf die Konfundierung mit den Versuchen und andere Einflüsse zurückzuführen. Da die Versuche von fast allen Studierende in unterschiedlichen Reihenfolgen durchgeführt wurden, mitteln sich diese Effekte über den Praktikumsverlauf raus.

Damit zeigt diese Analyse der Wachstumskurven, dass keine zeitlichen Effekte durch Entwicklungen der Studierenden im Verlaufe des Praktikums oder durch Übertragungen von Versuchseinschätzungen auf andere Versuche in Bezug auf die wahrgenommene Relevanz nachweisbar sind. Eine mögliche Ausnutzung eines zeitlichen Effekts für die Praktikumsgestaltung in Form einer optimierten Reihung der Versuche mit höherer und niedriger Relevanz fällt damit weg.

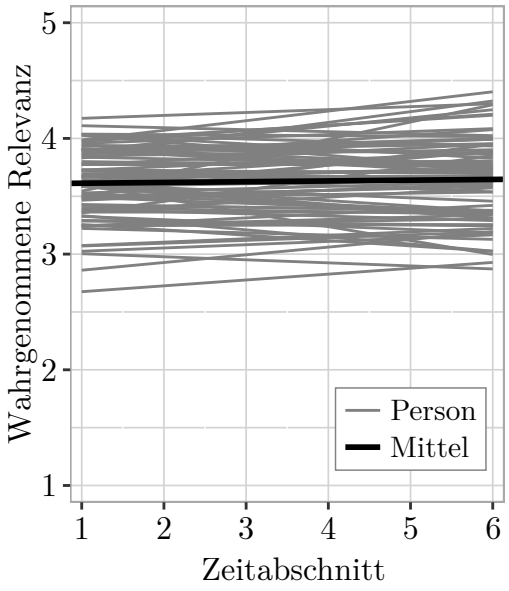


Abbildung 20: *Entwicklungsverläufe aus der Mehrebenenanalyse (Modell A, Tabelle 11) für die wahrgenommene Relevanz (5-teilige Likert-Skala, ohne Standardisierung) abhängig vom Zeitabschnitt der Beobachtung. Die grauen Linien zeigen die Verläufe für alle einzelnen Personen (Range:* $-0.12 < b_{1j} < 0.15$, *95 %-Konfidenzintervall:* $-0.02 < b_{1j} < 0.03$*). Die dickere schwarze Linie ist der Verlauf über alle Personen (Ausgangswert im ersten Zeitabschnitt:* $g_{oo} = 3.61$, *Steigung:* $g_{1o} = 0.01$*).*

Insgesamt bedeuten die Ergebnisse dieses Abschnittes, dass sich für die weitere Auswertung die Datenstruktur durch Vernachlässigung der Kohorten vereinfacht. Es wird zudem geschlussfolgert, dass auch der Zeitpunkt, in dem die wahrgenommene Relevanz im Praktikumsverlauf erfasst wird, in folgenden Untersuchungen vernachlässigt werden kann und damit Modelle ebenfalls an Komplexität verlieren.

9.2 Wahrgenommene Relevanz

Nachdem die vereinfachenden Annahmen zur Datenstruktur für die wahrgenommene Relevanz bestätigt wurden, soll nun zunächst die wahrgenommene Relevanz an sich betrachtet werden. Dabei soll geklärt werden, in welchem Maß das Praktikum bzw. die einzelnen Praktikumsversuche von den Studierenden als relevant wahrgenommen werden und ob sich die Versuche bezüglich der eingeschätzten wahrgenommenen Relevanz unterscheiden.

Um einen ersten Eindruck der wahrgenommenen Relevanz zu erhalten, sollen die Daten auf der obersten Datenebene des *Gesamtdatensatzes* (siehe Abb. 18A), also für alle vorhandenen Kombinationen von 89 Personen und bis zu 24 Versuchen ($n = 1805$), beschrieben werden. Das Histogramm in Abbildung 21 und die deskriptiven Werte im oberen Teil der Tabelle 12 geben einen ersten Überblick über die Daten zur wahrgenommenen Relevanz. Der Mittelwert der wahrgenommenen Relevanz liegt über alle Versuche und Personen hinweg bei $M = 3.65 \pm 0.02$ und damit deutlich über dem mittleren Maß von 3 auf der 5-teiligen Likert-Skala. Der Messbereich ist mit einem minimalen Wert von $MIN = 1.25$ und einem maximalen Wert von $MAX = 5.00$ beinahe voll ausgeschöpft. Die Standardabweichung liegt bei $SD = 0.66$. Damit sind Unterschiede der wahrgenommenen Relevanz für die einzelnen Bewertungen der Versuche messbar. Da die Daten in Bezug auf Personen und Versuche bei dieser Betrachtung abhängig sind, ist jedoch zu bedenken, dass die Standardabweichung und Standardfehler in dieser ersten Beschreibung zu hoch geschätzt sind (siehe Abschnitt 8.3.2).

Versuch	**n**	**M**	**SD**	**SE**	**MD**	**MIN**	**MAX**
alle	1805	3.65	0.66	0.02	3.75	1.25	5.00
Augenmodell (Z)	72	4.09	0.52	0.06	4.09	3.00	5.01
Feldeffekttransistor (Z)	76	3.13	0.72	0.08	3.15	0.94	4.68

Tabelle 12: *Deskriptive Statistik für die wahrgenommene Relevanz. Im oberen Teil sind die Kennwerte für die unzentrierten Daten über alle Kombinationen von Personen und Versuchen (Datenebene Gesamtdatensatz, Abb. 18A) tabelliert. Eine Visualisierung befindet sich in Abbildung 21. Im unteren Teil der Tabelle sind die deskriptiven Werte für die Versuche mit jeweils maximaler und minimaler zentrierter (Z) wahrgenommener Relevanz gelistet (Datenebene Zwischen Versuchen, Abb. 18C). Eine Darstellung für alle Versuche ist in Abbildung 22 gezeigt. Die Details für deskriptive Werte weiterer Versuche sind in Tabelle im Anhang A.2.1 zu finden.*

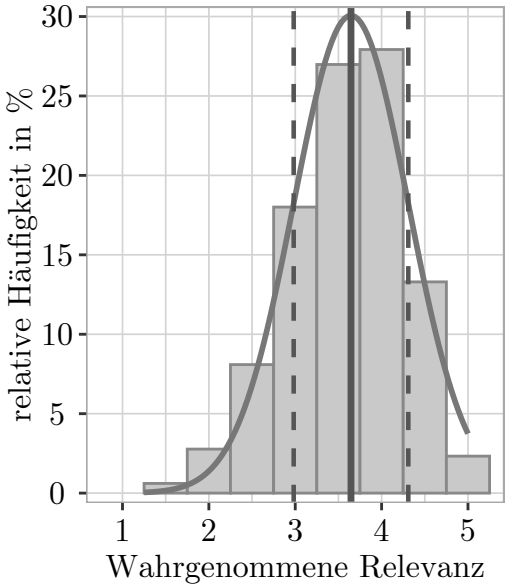


Abbildung 21: *Dieses Histogramm zeigt die relative Häufigkeit der wahrgenommenen Relevanz (5-teilige Likert-Skala) für die Daten aller Kombinationen von 89 Probanden für jeweils bis zu 24 verschiedene Versuche ($n = 1805$, Datenebene Gesamtdatensatz Abb. 18A). Die senkrechten Linien kennzeichnen den Mittelwert ($M = 3.65$) und die Standardabweichung (gestrichelt, $SD = 0.66$).*

Um nun weiter ins Detail einzutauchen und die Einschätzung der wahrgenommenen Relevanz der einzelnen Praktikumsversuche zu untersuchen, wird die Datenebene gewechselt auf *Zwischen Versuchen* (Abb. 18C). Das heißt, dass die wahrgenommene Relevanz im Folgenden über Personen gemittelt und bezüglich der Versuche aufgeschlüsselt wird. Um die Abhängigkeit der Daten in Bezug auf Personen beim Vergleich der Versuche zu berücksichtigen, werden die Personendaten bzgl. des Personenmittelwertes zentriert (siehe Abschnitt 8.3.2), um die Daten um die Grundeinstellung der Personen zu bereinigen. Im unteren Teil der Tabelle 12 ist für den Überblick nur die deskriptive Statistik für die Versuche Augenmodell (Anhang A.1.13) und Feldeffekttransistor (Anhang A.1.24) gelistet, da deren Mittelwerte die maximale und minimale wahrgenommene Relevanz auf der 5-stufigen Likert-Skala aufweisen. In Abbildung 22 ist für alle Versuche die zentrierte wahrgenommene Relevanz in abnehmender Reihenfolge in Form eines Balkendiagramms dargestellt. Deskriptive Werte aller Versuche sind im Anhang A.2.1 zu finden. Die Mittelwerte der wahrgenommenen Relevanz liegen für alle Versuche oberhalb der Mitte der 5-teiligen Skala. Der Versuch zum Augenmodell hat dabei den höchsten Mittelwert mit $M_{Auge} = 4.09 \pm 0.06$. Der Versuch zum Feldeffekttransistor wird mit $M_{FET} = 3.13 \pm 0.08$ im Mittel am wenigsten relevant eingestuft. Der Unterschied der maximalen und minimalen mittleren wahrgenommenen Relevanz beträgt mit $\Delta = 0.96 \pm 0.14$ rund eine Skaleneinheit auf der 5-teiligen Likert-Skala. Damit liegt der Hauptmessbereich etwa innerhalb einer Skaleneinheit oberhalb der Skalenmitte, sodass Decken- oder Bodeneffekte ausgeschlossen werden können. In Hinblick auf die Streuung der einzelnen Versuche zeigt die Deskription, dass die Standardabweichungen für die zentrierte wahrgenommene Relevanz im Bereich von $SD = 0.38$ bis $SD = 0.72$ liegen, wobei der Versuch zum Roller Coaster dabei mit einer Stichprobe von nur $n = 9$ nicht berücksichtigt ist. Für den Versuch zum Feldeffekttransistor ist die Standardabweichung im Vergleich zu den anderen Versuchen auffällig hoch. Ein

Levene-Test bestätigt, dass die Varianz für die einzelnen Versuche nicht homogen ist ($F(24, 1780) = 2.70, p < .001$). Die Studierenden bewerten folglich manche Versuche wie zum Beispiel den Feldeffekttransistor in Hinblick auf die wahrgenommene Relevanz weniger einheitlicher als andere Versuche.

Anhand dieser deskriptiven Statistik der wahrgenommenen Relevanz für die einzelnen Praktikumsversuche stellt sich die Frage, wie genau sich die Versuche hinsichtlich der eingeschätzten Relevanz unterscheiden. Aus Abbildung 22 wird ersichtlich, dass ein Großteil der Versuche in der Mitte des Wertebereiches der mittleren Relevanz liegen und sich anhand des visuellen Eindrucks im Rahmen des Standardfehlers wenig unterscheiden. Um dazu eine definitive Aussage treffen zu können, ist es erforderlich im Folgenden Mittelwertvergleiche für die Praktikumsversuche vorzunehmen, wobei die Abhängigkeitsstrukturen des Datensatzes durch Zentrierung der wahrgenommenen Relevanz des *Gesamtdatensatzes* berücksichtigt werden (siehe Abschnitt 8.3.4). Die Ergebnisse einer einfaktoriellen Varianzanalyse ($F(24, 1780) = 10.54, p < .001, \eta^2 = .12$) zeigen, dass es mit mittlerem Effekt signifikante Unterschiede in der wahrgenommenen Relevanz zwischen den Versuchen gibt. Da die Daten anhand von Shapiro-Tests nur für 80 % der Versuche normalverteilt sind und zudem die Varianzhomogenität, wie oben betrachtet, nicht gegeben ist und somit die Voraussetzungen für die Varianzanalyse nicht erfüllt sind, wurde zusätzlich ein Kruskal-Wallis-Test berechnet (siehe Abschnitt 8.3.3). Die zuvor gefundenen Unterschiede in der Varianzanalyse werden durch den Kruskal-Wallis-Test ($H(24) = 206.62, p < .001$) für die nicht-parametrischen Daten bestätigt.

Um die sich unterscheidenden Versuche zu identifizieren, wurden daraufhin paarweise t-Tests und Wilcoxon-Rangsummentests mit Bonferroni-Korrektur als PostHoc-Tests gerechnet (siehe Abschnitt 8.3.4). Beide Tests weisen konsistente Ergebnisse auf. Abbildung 23 zeigt in einer Matrix paarweise für die Versuche die Effektstärke (Cohens d) und das Signifikanzniveau des t-Tests. Die Reihung der Versuche entspricht der absteigenden wahrgenommenen Relevanz aus Abbildung 22. Tendenziell sind für diese mehrfachen Vergleiche in dieser Stichprobe Unterschiede ab mittlerer Effektstärke von Cohens $d \approx 0.5$ signifikant nachweisbar. Die Versuche Augenmodell und Feldeffekttransistor, die am oberen und unteren Rand der wahrgenommenen Relevanz liegen, unterscheiden sich signifikant mit mittleren bis hohen Effekten von fast allen anderen Versuchen hinsichtlich ihrer Relevanz. Der Versuch Newtonsche Axiome mit der zweithöchsten wahrgenommenen Relevanz unterscheidet sich von den Versuchen in der unteren Hälfte des Wertebereiches der wahrgenommenen Relevanz. Etwa fünf weitere Versuche im oberen und drei Versuche im unteren Wertebereich der wahrgenommenen Relevanz unterscheiden sich ebenfalls signifikant. Für alle anderen Versuchskombinationen ist kein signifikanter Unterschied in der wahrgenommenen Relevanz festzustellen. Mit Rückblick auf die im Rahmen des Standardfehlers ähnlichen Mittelwerte für die Versuche im mittleren Wertebereich der wahrgenommenen Relevanz in Abbildung 22 ist dieses Ergebnis plausibel.

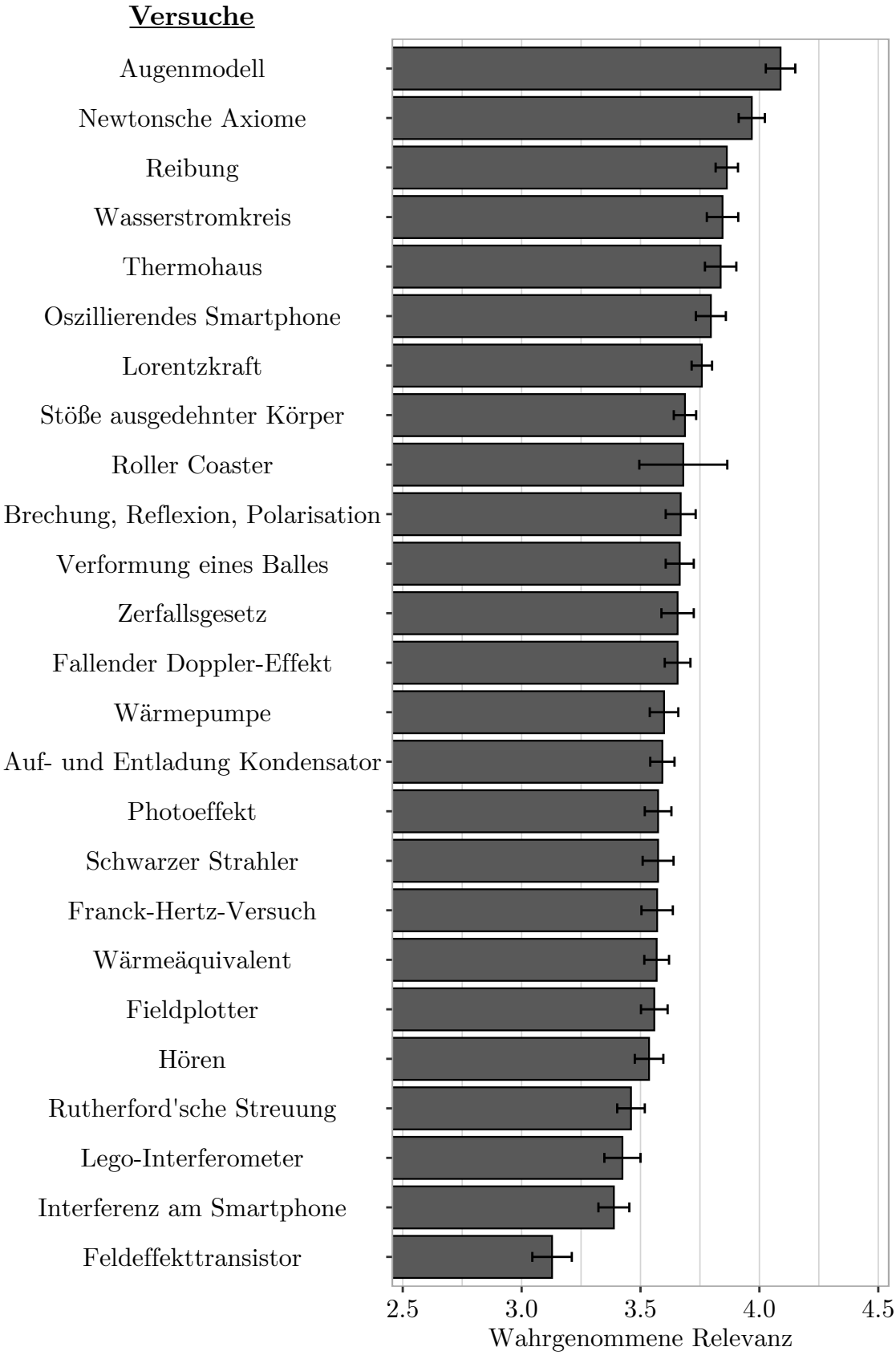


Abbildung 22: *Darstellung der zentrierten wahrgenommenen Relevanz (5-teilige Likert-Skala) absteigend für die einzelnen Praktikumsversuche* ($65 \leq n \leq 89$, $n_{RollerCoaster} = 9$)*. Das Balkendiagramm zeigt Mittelwerte und Standardfehler.*

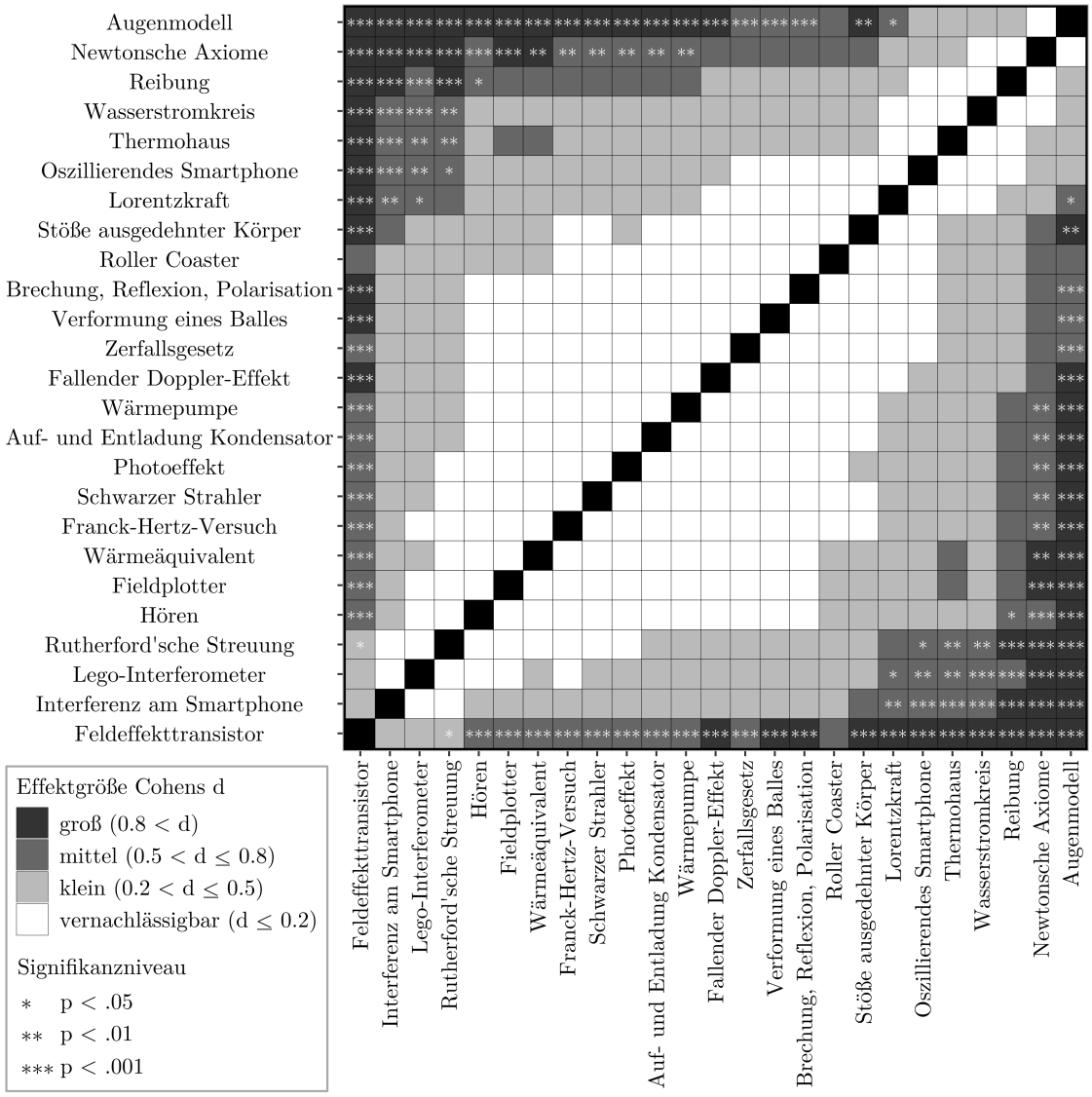


Abbildung 23: *Für die 25 Praktikumsversuche wurden paarweise Mittelwertvergleiche für die zentrierte wahrgenommene Relevanz mittels unabhängigen t-Test mit Bonferroni-Korrektur berechnet ($60 \leq n \leq 89$, $n_{RollerCoaster} = 9$). Angegeben sind für die Versuchskombinationen in Graustufen kodiert die Effektgröße Cohens d (Kategorien orientiert an Cohen (1988), siehe Abschnitt 8.3.4) und symbolisch kodiert das Signifikanzniveau p des t-Tests.*

Damit zeigt dieser erste Blick auf die Daten der wahrgenommenen Relevanz in Bezug auf die Praktikumsversuche, dass alle neu entwickelten Praktikumsversuche von den Studierenden als eher relevant wahrgenommen werden und damit ein wesentliches Ziel der Neuentwicklung des lehramtsspezifischen Praktikums erreicht wurde. Die Versuche werden bzgl. ihrer wahrgenommenen Relevanz von den Studierenden zum großen Teil signifikant unterschiedlich bewertet. Während die Intraklassen-Korrelation für die wahrgenommene Relevanz zeigt, dass 7 % der Gesamtvarianz auf Unterschiede zwischen Versuchen zurückzuführen sind, resultieren sogar 25 % der Gesamtvarianz aus Unterschieden zwischen Personen (siehe Abschnitt 9.1). Somit sind weiterführende Analysen auf Ebene der Personeneinschätzungen besonders vielversprechend.

9.3 Wahrgenommene Relevanz und Schulbezug

Nachdem im vorangegangenen Abschnitt gezeigt wurde, dass die Daten zur wahrgenommenen Relevanz eine gute Grundlage für die nachfolgenden Analysen der Forschungsfragen bilden, kann sich dieser Abschnitt der ersten Forschungsfrage zur Analyse der wahrgenommenen Relevanz widmen. Sie lautet:

In welchem Maße lässt sich durch das Aufzeigen eines Bezuges zum Lehrberuf die von den Studierenden wahrgenommene Relevanz von physikalischen Versuchen im Anfängerpraktikum Lehramt erhöhen?

Zur Herstellung des Bezuges zum Lehrberuf wurden entsprechende Versuchsbeschreibungen konzipiert. Zur Prüfung des Einflusses auf die wahrgenommene Relevanz wurden Versuchsbeschreibungen in unterschiedlichen Versionen im Praktikum eingesetzt (siehe Abschnitt 8.1). Während die klassischen Einleitungen in den Versuchsbeschreibungen eine Einordnung in das physikalische Thema vornehmen, wird der Versuch in den adressatenspezifischen schulbezogenen Einleitungen in einen Schulkontext eingebettet, um den Studierenden die schulische Relevanz des Versuches transparent darzulegen. Der Schulkontext in der Einleitung der Versuchsbeschreibung zeigt dabei den Bezug des Praktikumsversuches zum Lehrberuf auf. Dadurch soll den Studierenden die Wichtigkeit des Versuches für die spätere Lehrtätigkeit deutlich gemacht werden, um die wahrgenommene Relevanz der Lehramtsstudierenden zu erhöhen (siehe Abschnitt 3.4). Durch den Vergleich der Bewertungen für die verschiedenen Versionen der Versuchsbeschreibungen soll geprüft werden, ob gemäß dieser Hypothese eine schulbezogene Ausrichtung der Versuchseinleitung einen Effekt auf die wahrgenommene Relevanz im Praktikum hat.

Es soll in einem ersten Schritt geschaut werden, ob ein Gesamteffekt des Kontextes im Mittel über alle Versuche und Personen feststellbar ist. Der mögliche Einfluss der Versuche wird an dieser Stelle zunächst vernachlässigt, da sich in der vorangegangenen Analyse gezeigt hat, dass die Varianz durch Unterschiede zwischen Versuchen klein ist. Die Unabhängigkeit der Daten in Bezug auf die Personen wird durch Zentrierung angenähert (siehe Abschnitt 8.3.2). In Abbildung 24 sind die Mittelwerte mit Standardfehler der zentrierten wahrgenommenen Relevanz für die Versuchsbeschreibungen klassisch und schulbezogen dargestellt. Hypothesenkonform ist die wahrgenommene Relevanz für die schulbezogenen Versuchsbeschreibungen ($M_S = 3.68 \pm 0.02, n_S = 1125$) im Mittel größer als für die klassischen ($M_K = 3.59 \pm 0.02, n_K = 680$). Der absolute Unterschied ist auf einer 5-teiligen Likert-Skala aber klein. Um den Unterschied auf Signifikanz zu prüfen, ist ein Vergleich zweier Stichproben erforderlich (siehe Abschnitt 8.3.4). Die Prüfung der Testvoraussetzungen zeigt, dass Varianzhomogenität gegeben ist (Levene-Test: $F(1, 1803) = 0.38, p = .539$). Die Daten sind für beide Anleitungsformen anhand des Shapiro-Tests aber nicht normalverteilt (klassisch: $W = 0.98, p < .001$; schulbezogen: $W = 0.99, p < .001$). Da die Testvoraussetzungen somit nicht vollständig erfüllt sind, wurde in diesem Fall ergänzend zum unabhängigen t-Test für zwei Stichproben ein Wilcoxon-Rangsummentest für die Daten der zentrierten wahrgenommenen Relevanz angewendet. Der t-Test ($t(1410.1) = 3.53, p < .001$, Cohens $d = 0.17$) und der Wilcoxon-Test ($W = 420696, p < .001$) zeigen konsistent, dass die wahrgenommene Relevanz für die schulbezogenen Versuchsbeschreibungen signifikant größer ist als für die klassischen.

Aufgrund des kleinen Effektes stellt sich die Frage, ob der Schulkontext, der gemäß Theorie die wahrgenommene Relevanz erhöhen soll, in den schulbezogenen Versuchsbeschreibungen von den Studierenden tatsächlich erkannt wurde. Dies wird daher an dieser Stelle überprüft, um den festgestellten Effekt der Versuchsbeschreibungen auf die wahrgenommene Relevanz besser interpretieren zu können. Dazu wird das Item „Es wird ein Kontext zum Schulalltag hergestellt.“ aus dem Instrument zur Erfassung der Eigenschaften der Versuche (Abschnitt 8.2.3) hinzugezogen. Für diese Überprüfung dienen alle Versuchs- und Personenkombinationen (*Gesamtdatensatz*, Abb. 18A) als Datenbasis, um den Gesamteffekt zu betrachten. Abbildung 25 zeigt dafür die relativen Häufigkeiten der Antworten auf dieses Item für Versuche mit klassischen ($n_K = 386$) und die schulbezogenen Versuchsbeschreibungen ($n_S = 617$). Während für 40 % der Fälle mit schulbezogenen Versuchsbeschreibungen der Schulkontext erkannt wurde, wurde dies für klassische Versuchsbeschreibungen mit einem geringeren Anteil von 26 %. Dies zeigt, dass der Schulbezug bei schulbezogenen Versuchsbeschreibungen erwartungsgemäß besser erkannt wurde als bei klassischen Beschreibungen. Dennoch wurde bei 24 % der Bewertungen von Versuchen mit schulbezogenen Versuchsbeschreibungen kein Schulkontext erkannt. Das bedeutet, dass die Einleitung allein nicht in allen Fällen den Schulkontext verdeutlichen konnte.

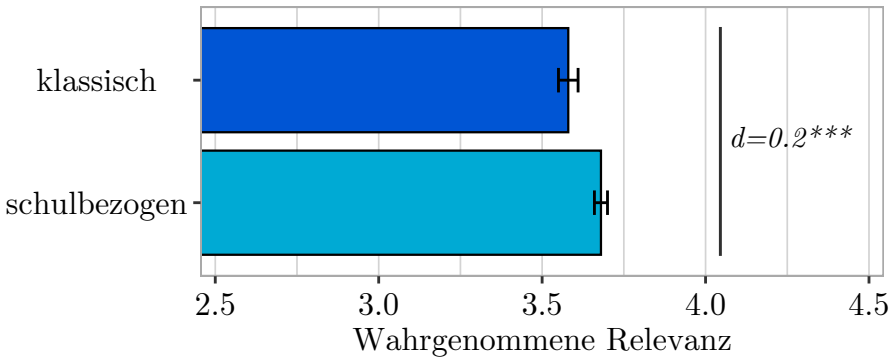


Abbildung 24: *Darstellung der zentrierten wahrgenommenen Relevanz (5-teilige Likert-Skala) über alle Praktikumsversuche aufgeteilt auf die klassische (dunkelblau) bzw. schulbezogene (hellblau) Versuchsbeschreibung ($n_S = 1125, n_K = 680$). Das Balkendiagramm zeigt neben den Mittelwerten mit Standardfehlern auch das Signifikanzniveau des t-Tests ($^{***}p < .001$) und Cohens d als Effektgröße.*

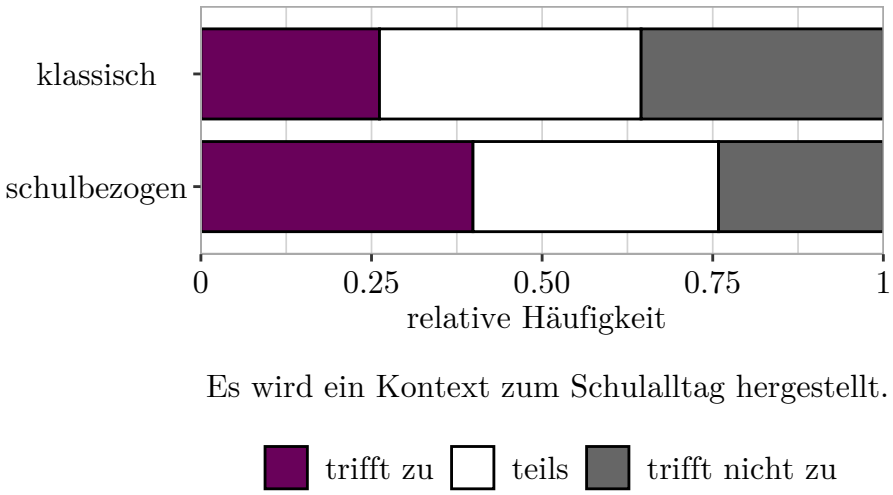


Abbildung 25: *Darstellung der relativen Häufigkeiten der Antworten auf das Item „Es wird ein Kontext zum Schulalltag hergestellt.“ aus dem Instrument zur Erfassung der Versuchseigenschaften (3-teilige Likert-Skala) aufgeteilt auf die klassische bzw. schulbezogene Versuchsbeschreibung ($n_S = 617, n_K = 386$).*

Da jeder Versuch inhaltlich passend zugeschnittene Einleitungen der klassischen und schulbezogenen Form hat, wären unterschiedliche Effekte für einzelne Versuche aufgrund der individuellen Merkmale der Einleitung naheliegend. Der Effekt der Versuchsanleitungen auf die wahrgenommene Relevanz wurde zuvor aber über alle Versuche hinweg festgestellt. Um eine Antwort zu erhalten, warum der Schulkontext nicht in allen Fällen erkannt wurde, ist es im nächsten Schritt somit erforderlich, eine Analyse der einzelnen Versuche hinsichtlich Unterschieden bezüglich des Effektes der Versuchsbeschreibungen auf die wahrgenommene Relevanz vorzunehmen. Abbildung 26 zeigt Mittelwerte der zentrierten wahrgenommenen Relevanz für die

25 verschiedenen Versuche in den zwei unterschiedlichen Versuchseinleitungen. Die Versuche des Themenbereiches Messtechnik hatten dabei nur schulbezogene Einleitungen (siehe Abschnitt 8.1). Für die meisten Versuche liegt der Mittelwert der wahrgenommenen Relevanz für die schulbezogene Einleitung über dem für die klassische Einleitung. Bei Einbezug der Standardfehler unterscheiden sich aber lediglich die Versuche zum Photoeffekt (Anhang A.1.9) und zur Rutherford'schen Streuung (Anhang A.1.10) deutlicher hinsichtlich der Versuchsbeschreibung. Eine Aussage zur Signifikanz ist damit aber nicht möglich. Um die Signifikanz der Unterschiede zu prüfen, wurde eine zweifaktorielle Varianzanalyse für die zentrierte wahrgenommene Relevanz des Gesamtdatensatzes ($n = 1779$) als abhängige Variable und die unabhängigen Variablen Versuch und Versuchsbeschreibung berechnet (siehe Abschnitt 8.3.4). Die Analyse bestätigt zum einen, dass die wahrgenommene Relevanz zwischen den Einleitungsversionen ($F(1, 1779) = 14.13, p < .001, \eta^2 = .01$) und den Versuchen ($F(24, 1779) = 10.64, p < .001, \eta^2 = .13$) signifikante Unterschiede aufweist. Der Effekt für die Einleitungen ist dabei wie in der ersten Untersuchung klein, aber nachweisbar. Zudem sind keine Unterschiede zwischen den Effekten der Versuchsbeschreibungen auf die wahrgenommene Relevanz für die einzelnen Versuche feststellbar ($F(19, 1779) = 0.74, p = .776, \eta^2 = .01$). Shapiro-Tests und ein Levene-Test ($F(44, 1760) = 2.02, p < .001$) zeigen, dass die Testvoraussetzungen bezüglich Normalverteilung und Varianzhomogenität für die Varianzanalyse nicht für alle Gruppen gegeben sind. Ein Scheirer-Ray-Hare-Test für nicht-parametrische Daten bestätigt aber die Ergebnisse der als robust geltenden Varianzanalyse (Versuchsanleitung: $H(1) = 12.7, p < .001$, Versuche: $H(24) = 207.6, p < .001$, Interaktion Anleitung und Versuche: $H(19) = 14.9, p = .731$). Damit sind mögliche Unterschiede in den Effekten der Versuchsbeschreibungen für die einzelnen Versuche zu klein, um sie für diese Stichprobe statistisch signifikant vom zufälligen Fehler zu trennen. Der Ursachenforschung auf Ebene der individuellen Versuchsbeschreibungen sind damit leider an dieser Stelle statistische Grenzen gesetzt.

Eine andere Ursache dafür, dass der Schulkontext nicht für alle schulbezogenen Versuchsbeschreibungen erkannt wurde, könnte auch der Umfang der Intervention in den gegebenen Rahmenbedingungen sein, in dessen Relation der nachgewiesene Effekt betrachtet werden muss. Daher muss auch die Kontextualisierung in Bezug auf die Länge der Versuchsanleitungen und die Bearbeitungszeit in die Beurteilung des Ergebnisses einbezogen werden. Die Einleitungen nehmen in den aktuellen Versuchsbeschreibungen etwa eine halbe bis ganze Seite von insgesamt sieben bis zwölf Seiten, also 5 % bis 15 % des Skriptumfangs ein. Die Einleitung nimmt damit also nur einen kleinen Teil der Versuchsbeschreibung ein. Zudem ist die zeitliche Beschäftigung mit der Einleitung ein noch kleinerer Anteil, wenn man den Gesamtumfang der Praktikumsversuche inklusive Durchführung und Nachbereitung betrachtet. Eine mögliche Ursache für die Ergebnisse ist damit, dass die Einleitung aufgrund des geringen Umfangs in der Vorbereitung die Relevanzwahrnehmung nicht über die

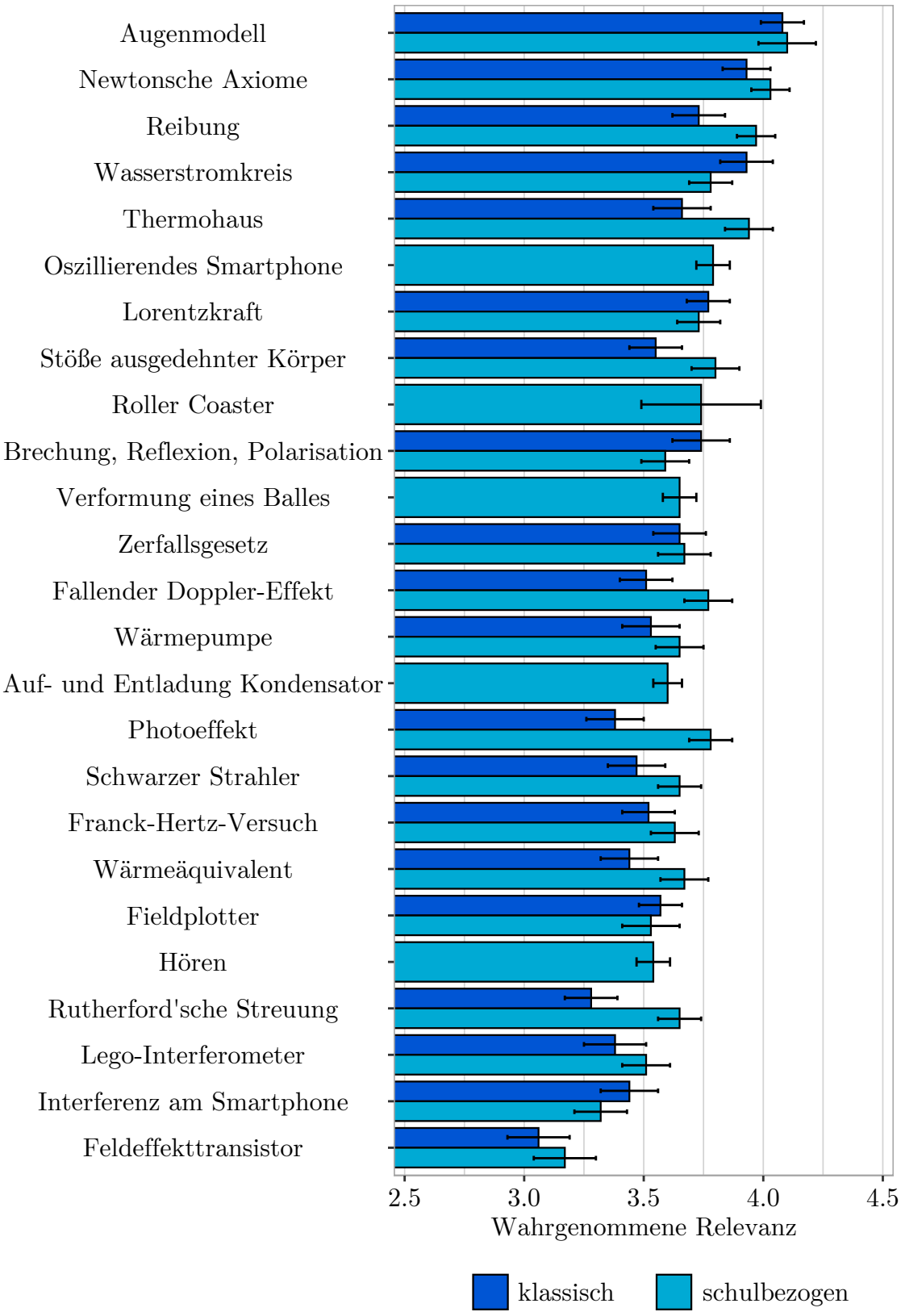


Abbildung 26: *Darstellung der wahrgenommenen Relevanz (5-teilige Likert-Skala) für die verschiedenen Praktikumsversuche aufgeteilt auf die klassische (dunkelblau) bzw. schulbezogene (hellblau) Versuchsbeschreibung* ($16 \leq n \leq 57$). *Das Balkendiagramm zeigt die Mittelwerte mit Standardfehlern.*

zeitintensive Durchführung und Nachbereitung des Versuches aufrechterhalten kann. Eine weitere mögliche Ursache könnte sein, dass die Versuchseinleitungen im Rahmen der Versuchsvorbereitung von einigen Studierenden für die Vorbereitung zum Beispiel aufgrund von Zeitmangel nicht oder kaum gelesen wurden. Auch die fachdidaktischen Fragen und die Frage nach der Schuleignung des Versuches wurden nicht von allen Studierenden im Protokoll aufgegriffen. Vor diesem Hintergrund des geringen Umfangs der Intervention muss geschlussfolgert werden, dass der nachgewiesene positive Effekt der schulbezogenen Versuchsbeschreibungen auf die wahrgenommene Relevanz bereits ein Erfolg ist.

Somit bleibt für Forschungsfrage 1 anhand dieser Analyse an dieser Stelle zunächst festzuhalten, dass die kontextbezogenen Versuchsanleitungen im Mittel mit einer höheren wahrgenommen Relevanz verknüpft sind. Durch die Kontextualisierung der Einleitungen wird der Schulkontext der Versuche dabei häufiger erkannt. Der festgestellte Effekt der Einleitungen auf die wahrgenommene Relevanz anhand des Wertes von Cohens $d = 0.17$ ist klein. Daher kann in vielen Fällen in den folgenden Analysen die Unterscheidung in schulbezogene und klassische Versuchsbeschreibung vernachlässigt werden. Aus dem kleinen Effekt kann aber nicht grundsätzlich auf eine geringe Wirksamkeit geschlossen werden. Zum einen ist der Effekt in Relation zum geringen Umfang der Intervention zu sehen. Trotz geringer Prominenz der Kontextualisierung und einem geringen Gesamtumfang im Rahmen der Versuchsdurchführung konnte bereits ein Effekt nachgewiesen werden. Zum anderen weisen die Versuche mit klassischen Versuchsbeschreibungen, die trotzdem als relevant und als schulbezogen erkannt wurden, darauf hin, dass gemäß der Konzeption der lehramtsspezifischen Praktikumsversuche (siehe Abschnitt 6) noch weitere Versuchseigenschaften in Zusammenhang mit der wahrgenommenen Relevanz der Versuche stehen und die Einleitung offenbar nicht allein ausschlaggebend ist. Bevor also das Maß des Effektes der Versuchsbeschreibungen auf die wahrgenommene Relevanz und die daraus resultierenden Implikationen für die Gestaltung eines lehramtspezifischen Versuches abschließend diskutiert und beurteilt werden können, muss die wahrgenommene Relevanz zunächst weiterführend analysiert werden.

9.4 Wahrgenommene Relevanz und Versuchseigenschaften

Die Ergebnisse in Abschnitt 9.3 haben gezeigt, dass weitere Versuchseigenschaften auf einen Zusammenhang mit der wahrgenommenen Relevanz geprüft werden müssen, um die wahrgenommene Relevanz der lehramtsspezifischen Versuche besser zu verstehen. Für die lehramtsspezifische Gestaltung der Praktikumsversuche sollen daher weitere Einflussfaktoren identifiziert werden, um den Gesamteffekt auf die wahrgenommene Relevanz durch entsprechende Konzeption der Praktikumsversuche weiter verstärken zu können. Forschungsfrage 2 (Abschnitt 4.2.1) lautet daher:

Welche Merkmale charakterisieren die Praktikumsversuche, die von den Lehramtsstudierenden als besonders relevant wahrgenommen werden?

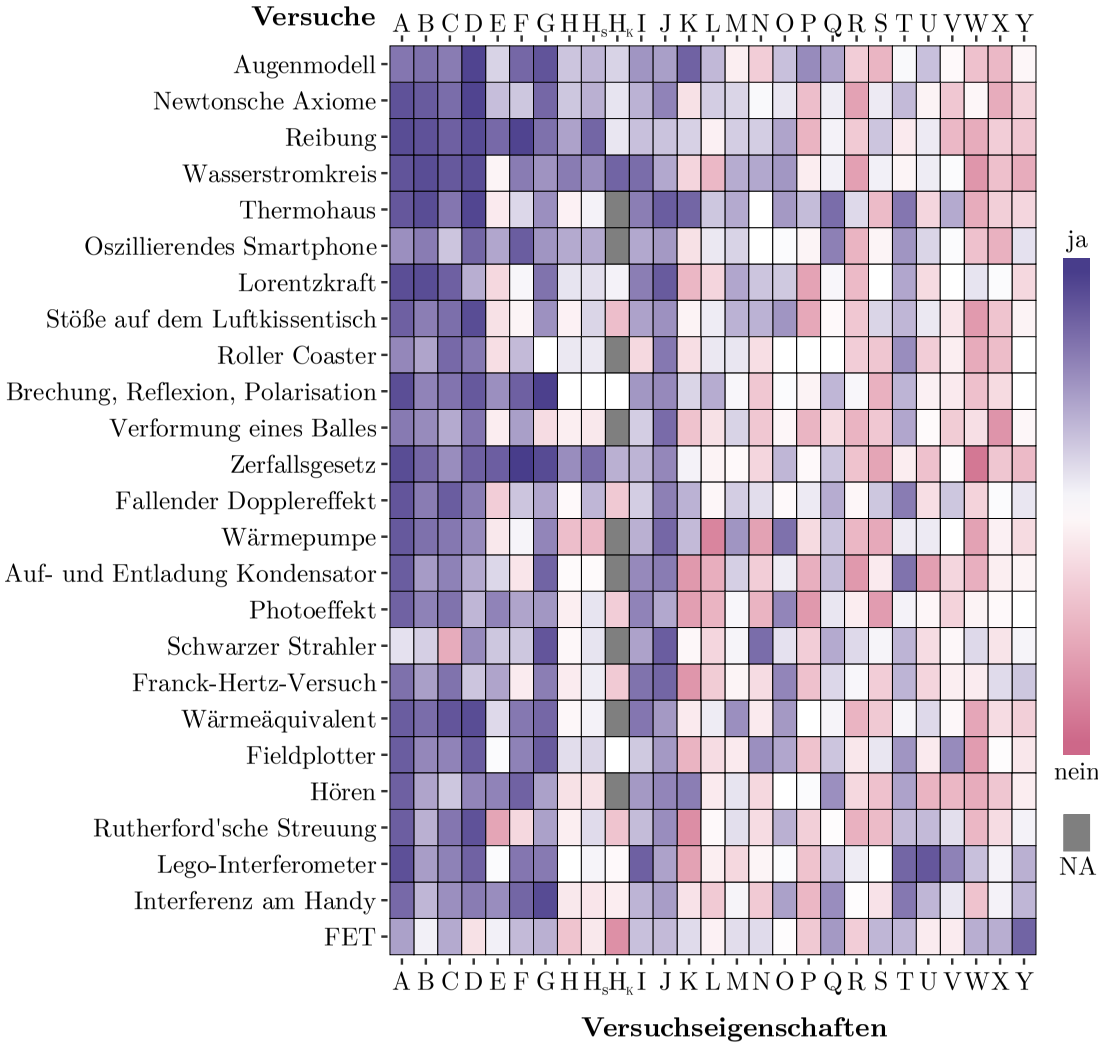
Hierzu ist in einem ersten Schritt erforderlich, die Versuche hinsichtlich der Versuchseigenschaften zu charakterisieren.

9.4.1 Eigenschaften der Praktikumsversuche

Um die Praktikumsversuche hinsichtlich ihrer Merkmale zu charakterisieren, werden die erhobenen Versuchseigenschaften betrachtet. Da das Erhebungsinstrument dazu speziell für dieses Praktikum neu entwickelt wurde (siehe Abschnitt 8.2.3), ist dabei von Bedeutung, die Validität des Instrumentes einzuschätzen. Eine inhaltliche Zuverlässigkeit ist notwendig, um Fehlinterpretationen zu vermeiden. Zur Prüfung der Plausibilität der von den Studierenden erkannten Versuchseigenschaften wird auch die Perspektive der Entwicklungssicht einbezogen. Außerdem soll in dieser Analyse auch ein Augenmerk darauf gelegt werden, ob die entwickelten Praktikumsversuche die Versuchseigenschaften in einer großen Breite abdecken.

Abbildung 27 zeigt dazu farblich kodiert, in welchem Maß auf einer 3-teiligen Likert-Skala die Studierenden im Mittel die Versuchseigenschaften den einzelnen Versuchen zuordnen. Ähnliche Versuchseigenschaften sind dabei anhand der Ergebnisse aus Abschnitt 8.2.3 zur Übersicht möglichst nebeneinander angeordnet. Die Versuche sind in der Abbildung absteigend sortiert nach ihrer eingeschätzten Relevanz (vgl. Abb. 22), sodass im nächsten Abschnitt Muster in den Versuchseigenschaften gegebenenfalls mit der Relevanz der Versuche in Verbindung gebracht werden können. Für Details zu den Praktikumsversuchen und zu den Itemwortlauten der Erhebung der Versuchseigenschaften stehen die Anhänge A.1 und A.2.2 zur Verfügung. In Abbildung 27 zeigt die eher einheitliche Färbung für die Versuchseigenschaften auf der linken und rechten Seite, dass einige Eigenschaften als ähnlich vorhanden (violett) bzw. nicht vorhanden (rot) eingeschätzt werden. Die Versuchseigenschaften dazwischen sind uneinheitlich bewertet worden.

Beginnend auf der linken Seite mit den Eigenschaften, die als vorhanden eingeschätzt wurden, fallen die strukturellen Aspekte *klar strukturierte Anleitung (A)*, *effektive Vorbereitung anhand der Anleitung (B)* und *eng geführte Versuchsaufgaben (C)* auf. Ausnahmen bilden der Versuch Schwarzer Strahler (Anhang A.1.18) und tendenziell auch das Oszillierende Smartphone (Anhang A.1.3). Dies ist konsistent mit der Entwicklung der Versuche. Der Schwarze Strahler ist bewusst sowohl in der Durchführung als auch in der Auswertung des Versuches sehr offen gehalten. Beim Oszillierenden Smartphone sind zwei Teilaufgaben des Versuches eher offen formuliert. Systematische Manipulation der anderen strukturellen Aspekte der Versuchsbeschreibungen (A und B) sind in großer Bandbreite jedoch auch nicht ohne Benachteiligung von Studierenden realisierbar. Ebenso werden fast alle Versuche als *anschaulich (D)* angesehen. Einzig der Versuch Feldeffekttransistor (FET) (Anhang A.1.24), der auch aus Sicht der Entwicklung eher abstrakt ist, wird im



	Versuchseigenschaften
A	klare Anleitung
B	effektive Vorbereitung
C	eng geführte Versuchsaufgaben
D	anschaulich
E	Versuchsmaterial in Schulsammlung
F	Versuchsmaterial leicht beschaffbar
G	eigenständiger Versuchsaufbau
H	Kontext zum Schulalltag
H_S	Kontext zum Schulalltag schulbezogene Einleitung
H_K	Kontext zum Schulalltag klassische Einleitung
I	eindeutig interpretierbare Ergebnisse
J	Versuchsmaterial von guter Qualität
K	Kontext zum Lebensalltag
L	abwechslungsreich
M	viel quantitative Auswertung
N	Einbringung eigener Ideen möglich
O	repetitive Versuchsdurchführung
P	vernetzt
Q	aktuell
R	überraschend
S	Aufgaben erfordern trickreiche Idee
T	moderne Versuchsmaterialien
U	Geschick fordernde Durchführung
V	ungewöhnliche Versuchsmaterialien
W	empfindliche Versuchsmaterialien
X	abstrakt
Y	anspruchsvoll

Abbildung 27: *Eigenschaften der einzelnen Versuche anhand der Einschätzung der Studierenden. Die Eigenschaften sind sortiert anhand der Ähnlichkeit der Bewertung (vgl. Abb. 17), die Versuche nach absteigender wahrgenommener Relevanz vgl. Abb. 22). Zur Übersicht sind die Eigenschaften auf der x-Achse mit Buchstaben A-Y abgekürzt (siehe Box rechts und Details in Anhang A.2.2). Die Farbkodierung gibt den Mittelwert über alle Einschätzungen für die jeweilige Eigenschaft (3-teilige Likert-Skala) für den Versuch an ($9 \leq n \leq 51$; NA: $n < 9$).*

Mittel als weniger anschaulich befunden. Die Studierenden geben zum Item bzgl. des *Versuchsaufbaus (G)* recht einheitlich an, dass die Versuche mit vorhandenen Materialien eigenständig aufgebaut wurden. Allein beim Versuch Verformung eines Balles (Anhang A.1.4), bei dem der Aufbau mit Fallröhre, Beleuchtung und Kamera komplett installiert vorzufinden waren, geben die Studierenden an, den Versuch eher nicht selbstständig aufgebaut zu haben. Obwohl die Versuche sich durchaus im Umfang der erforderlichen Arbeiten unterscheiden, sind die Bewertungen der Studierenden bzgl. des Aufbaus überraschend einheitlich. Dies liegt daran, dass die Studierenden hier offenbar aufgrund der Formulierung den Anteil der benötigten Hilfestellung beim Aufbau bewertet haben. Überraschend ist auch die Bewertung der Studierenden ihrer Ergebnisse für alle Versuche, bis auf den nur selten eingesetzten Roller Coaster (Anhang A.1.6), im Mittel als *eindeutig interpretierbar (I)*. Dies lässt sich bei Durchsicht der Protokolle nicht objektivieren. Diese Diskrepanz erklärt sich aus der Tatsache, dass die Studierenden die Fragebögen zusammen mit dem jeweiligen Protokoll abgeben. Zu diesem Zeitpunkt sind sie offenbar von ihrer Auswertung bzgl. Richtigkeit und damit auch Eindeutigkeit überzeugt. Eindeutigkeit ist hier also auch ein eindeutig subjektiv geprägter Begriff. In Anbetracht der Neuentwicklung der Versuche weniger überraschend ist, dass die Versuchsmaterialien im Mittel für alle Versuche als von *guter Qualität (J)* bewertet wurden. Die Versuche werden tendenziell eher als *aktuell (Q)* bewertet. Dabei fallen der Versuch Thermohaus mit der Infrarot-Kamera und die Versuche mit den Smartphones (Oszillierendes Smartphone und Interferenz am Handy (Anhang A.1.11)) mit besonders positiven Wertungen auf, da sie alle moderne Messtechnik einsetzen. Entsprechend fallen die Einschätzungen des Merkmals *moderne Versuchsmaterialien (T)* ähnlich aus. Neben den Versuchen mit Infrarotkamera und Smartphone stechen hier auch der Versuch Lego-Interferometer (Anhang A.1.12) und die Einführung des Messwerterfassungssystems CASSY anhand der Auf- und Entladung eines Kondensators (Anhang A.1.2) ins Auge.

Mehr unterschiedliche Einschätzungen von Versuchen sind in Hinsicht auf die Eigenschaften *vorhandene (E) oder leicht beschaffbare (F) Versuchsmaterialien* zu finden. Ein Luftkissentisch beispielsweise, der in den Versuchen Stöße auf dem Luftkissentisch (Anhang A.1.22) und Rutherford'sche Streuung (Anhang A.1.10) genutzt wird, ist, wie die Studierenden erkannt haben, eher selten in einer Schulsammlung anzutreffen und auch nicht leicht beschaffbar. Diese schulische Verfügbarkeit der Versuchsmaterialien wurde in den meisten Fällen von den Studierenden sinnvoll eingeschätzt. Nachvollziehbare abweichende Einschätzungen sind zum Beispiel für den Versuch Schwarzer Strahler zu finden, da das USB Spektrometer in der vorliegenden Version zwar von einem Lehrmittelhersteller vertrieben wird, aber ein teures Sonderteil darstellt. Da die Studierenden über keine Vorstellungen von Preisen der Materialien und typischen Budgets einer Schule verfügen können, ist ihre Einschätzung hinsichtlich der Verfügbarkeit somit wenig verwunderlich. Recht unterschiedlich sind die Bewertungen der Versuchseigenschaft *Kontext zum Schulalltag (H)*. Eine Auf-

teilung für die unterschiedlichen Versuchsbeschreibungen (Details siehe Abschnitt 8.1, Abb. 14) zeigt, wie auch Abbildung 25 in Abschnitt 9.3, dass der Schulkontext für die *schulbezogene (H_S) Versuchsbeschreibung* nicht für alle Versuche deutlich höher ist als für die *klassische (H_K)* und damit der Schulkontext aus der Versuchsbeschreibung nicht immer erkannt wurde. Die Einschätzungen zum *Kontext zum Lebensalltag (K)* sind ebenfalls unterschiedlich für die Versuche, aber entsprechen den Erwartungen mit zum Beispiel mehr Erkennen von Kontext in den lebensnahen Versuchen Augenmodell (Anhang A.1.13), Thermohaus (Anhang A.1.15) und Hören (Anhang A.1.5) und weniger in den abstrakteren Versuchen Rutherford'sche Streuung und Franck-Hertz-Versuch (Anhang A.1.8). Ebenfalls konsistent zeigen sich die Unterschiede in den Versuchen bezüglich der Einschätzungen der Eigenschaften *abwechslungsreich (L)*, *viel quantitative Auswertung (M)*, *Einbringung eigener Ideen (N)* und *repetitive Versuchsdurchführung (O)*. Die Bewertungen *abwechslungsreich (L)* und *repetitive Versuchsdurchführung (O)* sind erwartungsgemäß tendenziell gegensätzlich. Der Versuch Schwarzer Strahler fällt mit der Einbringung eigener Ideen bei der Versuchsdurchführung (N) auf, wie anhand der offenen Aufgabenstellung auch zu erwarten ist. Im Kontrast dazu hat besonders der Versuch Wärmepumpe (Anhang A.1.17) keinen Spielraum in der Durchführung.

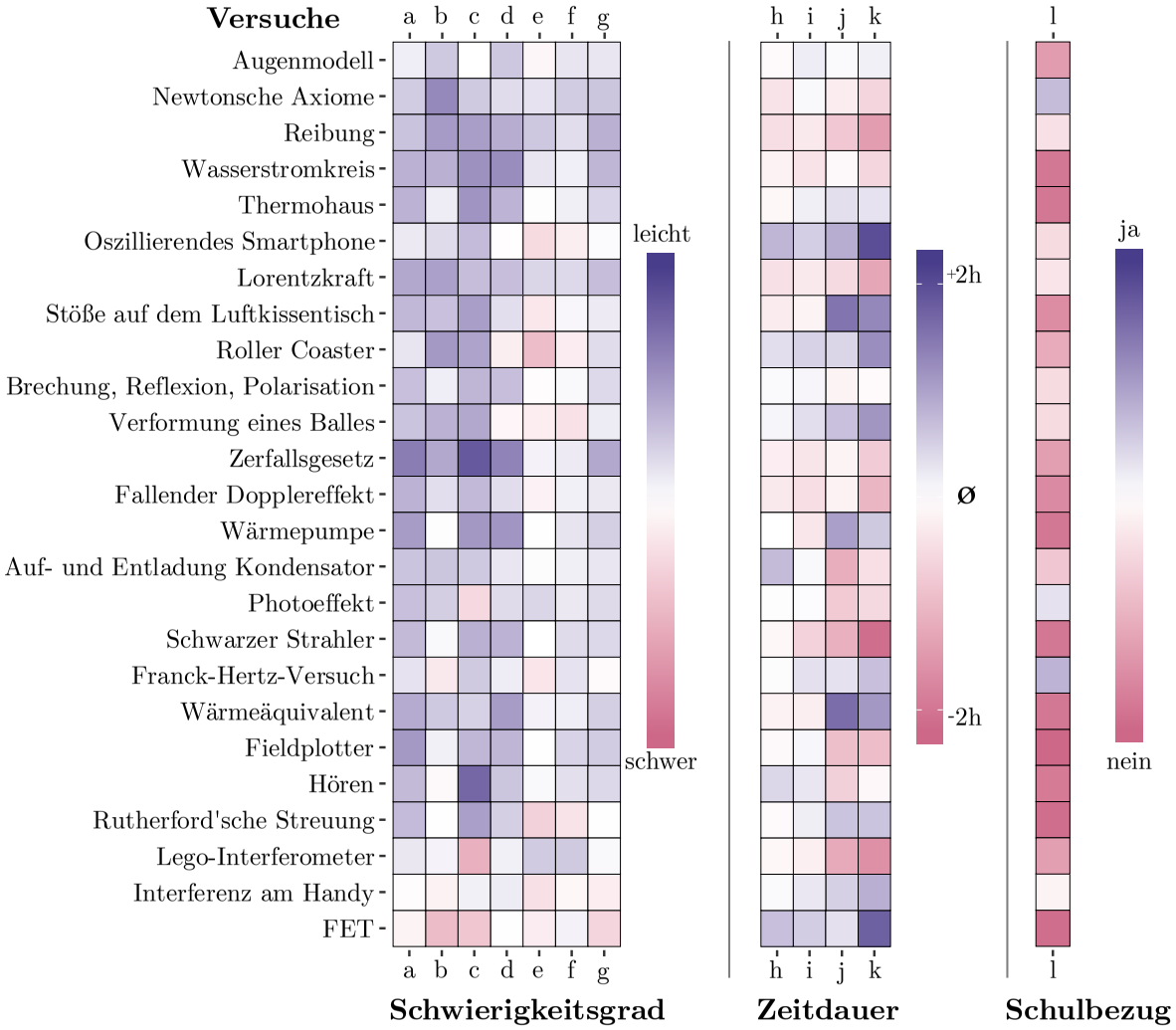
Andere Versuchseigenschaften werden von den Studierenden eher wenigen Versuchen zugeordnet. So gibt es nur wenige Versuche, die als klar *interdisziplinär* eingeschätzt wurden bzw. *unterschiedliche Fachbereiche vernetzen (P)* wie der Versuch zum Augenmodell, der in der Entwicklung mit diesem Ziel angelegt wurde. Auch *ungewöhnliche Versuchsmaterialien (V)* werden nur für wenige Versuche identifiziert. Neben den Versuchen mit Lego und der Infrarot-Kamera wird der Versuch Fieldplotter (Anhang A.1.26) mit den Materialien Leitlackstiften und Widerstandspapier am ungewöhnlichsten empfunden. Weitere Versuchseigenschaften, die eher wenigen Versuchen zugeordnet werden, sind *überraschend (R)* und *Aufgaben erfordern trickreiche Ideen (S)*. Die Einschätzung der *trickreichen Ideen in den Versuchsaufgaben (S)* unterscheidet sich dabei von der *Einbringung eigener Ideen in der Versuchsdurchführung (N)*, was auch zeigt, dass die Studierenden die Versuchsaufgaben und die Versuchsdurchführung als unterschiedliche Aspekte betrachten. Nur für den Versuch FET sehen die Studierenden den Bedarf einer trickreichen Idee, der aus Entwicklungssicht eher mit dem höheren Schwierigkeitsgrades des Versuches zusammenhängen könnte. Die Merkmale *Geschick fordernde Durchführung (U)* und *empfindliche Versuchsmaterialien (W)* werden ebenso nur einzelnen Versuchen zugeordnet. Der Aufbau des Lego-Interferometers ist empfindlich und die Justage benötigt Geschick und Fingerspitzengefühl. Da der Feldeffekttransistor (FET) bei einigen Praktikumsgruppen durch unsachgemäße Nutzung beschädigt wurde, wurde das Versuchsmaterial hier entsprechend als empfindlich bewertet. Der Versuch FET, der bereits als wenig *anschaulicher Versuch (D)* bewertet wurde, wird konsistenter Weise auch als *abstrakter Versuch (X)* eingestuft. Der FET Versuch fällt neben

dem Franck-Hertz-Versuch, dem Lego-Interferometer und der Interferenz am Handy außerdem bei der Versuchseigenschaft als besonders *anspruchsvoll (Y)* auf.

Abbildung 28 erweitert die Eigenschaften aus Abbildung 27 um ergänzende Einschätzungen des Schwierigkeitsgrades und der Bearbeitungsdauer in Bezug auf die einzelnen Versuche sowie eine zusätzliche Frage zum Schulbezug. Versuche, die als eher schwieriger (a-g) (z. B. in der Durchführung (d)) eingestuft werden, werden entsprechend auch als eher zeitintensiv (h-k) (in der Durchführung (i)) eingestuft (Abb. 28). Damit zeigt sich, dass sich der Schwierigkeitsgrad der Versuche auch im Zeitaufwand niederschlägt. Überraschend von Entwicklungsseite ist, dass die Studierenden nur wenige *Versuche aus aus der eigenen Schulzeit kennen (l)*. Ausnahmen bilden nur die Versuche Newtonsche Axiome und Franck-Hertz, die sehr klassisch im Physikunterricht sind.

Insgesamt zeigt die Untersuchung, dass die Einschätzung der Versuchseigenschaften untereinander und größtenteils auch in Bezug auf die Versuchsentwicklung konsistent und plausibel sind und damit die Versuchseigenschaften sinnvoll erkannt wurden. Für das neu entwickelte Erhebungsinstrument zur Erfassung der Versuchseigenschaften bedeutet dies, dass die Versuchseigenschaften valide gemessen wurden und damit eine zuverlässige Basis für die weitere Analyse bilden.

Die Betrachtungen zeigen auch, dass jeder Versuch eine eigene Kombination aus Eigenschaften aufweist. Manche Versuche unterscheiden sich dabei deutlicher von anderen, wie der Feldeffekttransistor als im Vergleich sehr abstrakter und anspruchsvoller Versuch. Auch wenn manche Eigenschaften für den Großteil der Versuche eher einheitlich und nur für wenige Versuche unterschiedlich bewertet wurden, sind die Eigenschaften damit insgesamt breit abgedeckt. Dieses breite Spektrum an Eigenschaften in den Praktikumsversuchen ist eine wichtige Voraussetzung, um den Zusammenhang zwischen den Eigenschaften und der wahrgenommenen Relevanz unverzerrt abzubilden.



	Schwierigkeitsgrad
a	Theorie
b	Aufbau
c	Messtechnik
d	Durchführung
e	Auswertung
f	Fehlerrechnung
g	Gesamtschwierigkeit
	Zeitdauer
h	Vorbereitung
i	Durchführung
j	Nachbereitung
k	Gesamtdauer
	Schulbezug
l	Schulerinnerung

Abbildung 28: *Weitere Eigenschaften der einzelnen Versuche anhand der Einschätzung der Studierenden. Die Versuche sind sortiert nach wahrgenommener Relevanz (Abb. 22). Zur Übersicht sind die Eigenschaften auf der x-Achse mit Buchstaben a-l abgekürzt (siehe Box rechts und Details in Anhang A.2.2). Die Farbkodierung gibt den Mittelwert über alle Einschätzungen für den Schwierigkeitsgrad (5-teilige Likert-Skala, $9 \leq n \leq 891$), die Zeitdauer (Stundenzahl zentriert um jeweiligen Mittelwert ($M_d = 2.4h$, $M_e = 3.4h$, $M_f = 7.3h$, $M_g = 13.2h$), $9 \leq n \leq 50$) und Schulerinnerung (l, 2-teilige Likert-Skala, $9 \leq n \leq 50$) an.*

9.4.2 Wahrgenommene Relevanz und Versuchseigenschaften

Nachdem vorangehend die Eigenschaften der Versuche betrachtet wurden, können diese nun im nächsten Schritt für die Analyse der Forschungsfrage in Verbindung mit der wahrgenommenen Relevanz gebracht werden. Da die Versuche in den Abbildungen 27 und 28 absteigend nach ihrer mittleren wahrgenommenen Relevanz sortiert sind, können daran erste Beobachtungen gemacht werden, in welchen Merkmalen sich die Versuche mit hoher und niedriger Relevanz unterscheiden. Aus Abbildung 27 zeichnet sich die Tendenz ab, dass für Versuche mit hoher wahrgenommener Relevanz (z. B. Augenmodell (Anhang A.1.13) und Newtonsche Axiome (Anhang A.1.19)) ein *Kontext zur Schule (H)* im Versuch erkannt wurde und bei Versuchen mit geringerer wahrgenommener Relevanz (z. B. FET (Anhang A.1.24) und Interferenz am Handy (Anhang A.1.11)) der Kontext zur Schule weniger oder nicht erkannt wurde. Dies stimmt mit den theoretischen Überlegungen zur Wirkung von Kontexten (Abschnitt 3.4) und den Ergebnissen zu Forschungsfrage 1 (Abschnitt 9.3) überein. Außerdem zeigt sich die Tendenz, dass die als *anspruchsvoller (Y)* empfundenen Versuche (z. B. FET und Interferenz am Handy) eine niedrigere wahrgenommene Relevanz haben als weniger anspruchsvolle Versuche (z. B. Newtonsche Axiome, Reibung (Anhang A.1.20), Wasserstromkreis (Anhang A.1.23)). Die Tendenz, dass anspruchsvollere Versuche als weniger relevant wahrgenommen werden, wird ebenfalls anhand der Einschätzung des *Schwierigkeitsgrades (a-g)* für die einzelnen Teilbereiche der Versuche in Abbildung 28 deutlich. Ein höherer Schwierigkeitsgrad ist zudem oft von einer höheren Bearbeitungsdauer (h-k) begleitet. Eine plausible Erklärung dafür ist, dass die Studierenden Versuche, die sie selbst als schwierig und aufwendig empfinden, aus diesem Grund wahrscheinlich auch für zu schwierig für Schüler*innen halten und damit als weniger relevant für den Physikunterricht erachten. Zudem lässt sich den Abbildungen entnehmen, dass jeder Versuch eine eigene Kombination aus verschiedenen Eigenschaften aufweist. Damit sind weitere Zusammenhänge der wahrgenommenen Relevanz und der Versuchseigenschaften nicht eindeutig aus den Darstellungen identifizierbar.

Um weitere Zusammenhänge aufzudecken, werden im Folgenden paarweise Korrelationen zwischen der wahrgenommenen Relevanz und den einzelnen Versuchseigenschaften als Zusammenhangsmaß betrachtet. Die Korrelationen wurden für den Gesamtdatensatz mit allen Kombinationen aus Personen und Versuchen (siehe Abb. 18) berechnet. Da die Daten damit vor allem in Bezug auf die Personen abhängig sind und versuchsbedingte Zusammenhänge auf der Datenebene *Innerhalb Personen* von größerem Interesse sind als Zusammenhänge durch personenbedingte Unterschiede auf der Datenebene *Zwischen Personen*, wurde die wahrgenommene Relevanz um den Personenmittelwert zentriert, um die grundlegende Einstellung zur Relevanz der Personen zu entfernen (siehe Abschnitt 8.3.2). Da die Korrelationen nach Pearson und Spearman ähnliche Ergebnisse liefern, wird hier die Pearson-Korrelation berichtet (siehe Abschnitt 8.2.1). Die resultierenden Korrelationen sind in Abbildung

29 in absteigender Reihenfolge für die einzelnen Versuchseigenschaften farblich kodiert dargestellt. Signifikante positive Korrelationen $r \geq .15$ ergeben sich dabei für Versuchseigenschaften, die auf schulische Versuche bezogen werden können (*Kontext zum Schulalltag, anschaulich, Kontext zum Lebensalltag, abwechslungsreich, Schulerinnerung, vernetzt*) und die strukturelle Aspekte in Bezug auf die Arbeit im Praktikum (*effektive Vorbereitung, klare Anleitung*) abdecken. Die Eigenschaften *anschaulich, Kontext zum Lebensalltag, abwechslungsreich und vernetzt* haben keinen direkten Schulbezug, können aber als Merkmale von typischen Experimenten im Physikunterricht betrachtet werden, weshalb Versuche mit diesen Eigenschaften nachvollziehbar als relevanter wahrgenommen werden. Erinnern sich die Studierenden an den Versuch aus der Schulzeit (*Schulerinnerung*), so wird dieser Versuch als relevanter bewertet, was übertragen die Wichtigkeit der Auswahl von schulrelevanten Versuchsthemen verdeutlicht. Eine klare Anleitung der Versuche und eine anhand dessen mögliche effektive Vorbereitung kann den Studierenden die Ziele des Versuches verdeutlichen, sodass zum einen die Bearbeitung des Versuches erleichtert wird und auch die Relevanz des Versuches klarer wird. Insgesamt sind die Korrelationen zwischen der wahrgenommenen Relevanz und den Versuchseigenschaften eher im mittleren bis kleinen Bereich ($|r| < .3$). Für die zwei Merkmale *Kontext zum Schulalltag* und *anschaulich* ergeben sich dabei die höchsten Korrelationen von mittlerer Größenordnung ($|r| \approx .3$). Erkennen die Studierenden demnach also einen Kontext zur Schule im Praktikumsversuch oder erachten den Versuch als besonders anschaulich, nehmen sie den Versuch als relevanter wahr.

Negative Korrelationen von $r < -.15$ ergeben sich dabei vor allem für die Schwierigkeit des Versuches (*Gesamtschwierigkeit, Schwierigkeit von Theorie und Auswertung*). Nimmt man die signifikanten Korrelationen $-.15 \leq r \leq 0$ hinzu, kommt zu den Merkmalen in Bezug auf die Schwierigkeit der Versuche (*anspruchsvoll, Schwierigkeit von Aufbau, Durchführung, Messtechnik und Fehlerrechnung, abstrakt, empfindliche Versuchsmaterialien*) der Aspekt der Zeitdauer hinzu (*Gesamtdauer, Dauer von Durchführung, Vorbereitung und Nachbereitung*). Für diese negativen Zusammenhänge von Schwierigkeit und Zeitdauer mit der wahrgenommenen Relevanz haben sich auch bereits bei der Betrachtung der Versuchseigenschaften zuvor Hinweise gezeigt.

Für die Versuchseigenschaften in der Mitte der Darstellung zeigen sich die Korrelationen als sehr gering und damit nicht signifikant. Dies kann zum einen daran liegen, dass tatsächlich kein Zusammenhang zwischen Relevanz und Versuchseigenschaft besteht. Zum anderen können hier aber auch sehr individuell unterschiedliche Wahrnehmungen der Studierenden eine Rolle spielen, die im Mittel zu keiner signifikanten Korrelation führen. Letzteres wäre für die Versuchseigenschaften *Geschick fordernde Durchführung* durchaus nachvollziehbar, da während der Betreuung des Versuches zum Lego-Interferometer (Anhang A.1.12) beobachtet werden konnte, dass Studierende mit Geschick und Geduld die Justage und Durchführung des Versuches sehr viel positiver erlebt haben als andere, was sich auf die wahrgenommene

Abbildung 29: *Pearson Korrelation r der Versuchseigenschaften und der zentrierten wahrgenommenen Relevanz für den Gesamtdatensatz über alle Personen und Versuche* ($878 \leq n \leq 1798$). *Die Versuchseigenschaften sind nach absteigender Korrelation sortiert. Zudem ist das Signifikanzniveau gekennzeichnet mit* $^{***}p < 0.001$, $^{**}p < .010$, $^{*}p < .050$. *Nicht signifikante Korrelationen* ($p > .05$) *wurden zur Übersicht ausgeblendet.*

Relevanz auswirken kann. Im Mittel könnten sich diese unterschiedlichen Wahrnehmungen aufheben.

Damit stützen die untersuchten Zusammenhänge der wahrgenommenen Relevanz und der Versuchseigenschaften anhand der Korrelationen die zuvor gesammelten Hinweise aus der Betrachtung der Eigenschaften der Versuche. Die Korrelationen ermöglichen darüber hinaus eine sinnvolle erste Quantifizierung der Zusammenhänge. Die Zentrierung berücksichtigt dabei die Abhängigkeit der Daten auf Personenebene durch Zentrierung und Betrachtung der Datenebene *Innerhalb Personen*. Strukturelle und schulbezogene Versuchseigenschaften zeigen anhand der Korrelationen einen positiven Zusammenhang mit der wahrgenommenen Relevanz, Schwierigkeit und Bearbeitungsdauer einen negativen. Die Eigenschaften *Kontext zum Schulalltag und Anschaulichkeit* weisen dabei signifikante positive Korrelationen zur wahrgenommenen Relevanz mit mittlerer Größenordnung ($r \approx .3$) auf.

Bei der Betrachtung der Korrelationen werden allerdings Zusammenhänge der Versuchseigenschaften untereinander und manche Aspekte der Mehrebenenstruktur vernachlässigt. Da die Korrelationen der Versuchseigenschaften in Abschnitt 8.2.3 als klein und mittelstark befunden wurden und die Gruppierung der Personen durch Zentrierung berücksichtigt wurde, dürften die Abweichungen nicht sehr stark sein. Dennoch sollen zur Absicherung im nächsten Schritt der Analyse die Zusammenhänge der einzelnen Versuchseigenschaften mit der wahrgenommenen Relevanz unter Kontrolle der anderen Eigenschaften und unter Berücksichtigung der Mehrebenenstruktur der Daten analysiert werden. Dabei soll die Fragestellung von der Ebene *Innerhalb Personen*, auf der die Einschätzung von der Situation (=Praktikumsversuch) abhängt, auf die zusätzliche Datenebene *Zwischen Personen* erweitert werden, die den Effekt der konstanten Grundeinstellung der Personen über alle Praktikumsversuche erfasst, um die Effekte auf beiden Ebenen zu vergleichen. Zur Schätzung dieser gesuchten Zusammenhänge wird eine Mehrebenenanalyse durchgeführt (siehe Abschnitt 8.3.5).

Zur Untersuchung der Mehrebenenstruktur und für die Spezifikation des finalen Modells wurden in einem ersten Schritt der Analyse zwei Nullmodelle (für Details siehe Tab. 9 und Abschnitt 8.3.5) mit Level 1 als Beobachtungen und Level 2 als Gruppierungen in Personen (Modell 0P) bzw. Versuche (Modell 0V) für die Daten aller Personen- und Versuchskombinationen zur z-standardisierten wahrgenommenen Relevanz berechnet und in Tabelle 13 dargestellt. Die Intraklassen-Korrelation (ICC) zeigt, dass für diesen Teildatensatz 25 % der Varianz auf Unterschiede zwischen den Personen ($n_{pid} = 57$) zurückzuführen sind und 10 % auf Unterschiede zwischen Versuchen ($n_{VS} = 25$). Damit ist eine Berücksichtigung beider Gruppierungen empfehlenswert, um Verzerrungen der Standardfehler zu vermeiden, wie auch der Designeffekt (deff > 2) zeigt. Um die Komplexität des Modelles in Grenzen zu halten, wurde entschieden nur die Gruppierung der Personen in das Mehrebenenmodell aufzunehmen. Dafür sprechen die Intraklassen-Korrelation für die Versuche von $ICC < .1$ und die Fragestellung mit Bezug auf die Einschätzung der Personen.

Im nächsten Schritt wird dieses Mehrebenenmodell mit den Versuchseigenschaften als Prädiktoren spezifiziert (siehe Modell 1 in Tab. 9 und Abschnitt 8.3.5). Zur Reduktion der Prädiktoren und somit zur Vereinfachung des Modells werden die Gesamtschwierigkeit und die Gesamtdauer genutzt, da die Einzelkomponenten ähnliche Korrelationen mit der wahrgenommenen Relevanz gezeigt haben und somit nicht extra aufgeschlüsselt werden müssen. Somit werden die Praktikumsversuche mit 28 Versuchseigenschaften in diesem Modell breit abdeckend beschrieben (Abschnitt 8.2.3). Um die Fragestellung auf den verschiedenen Datenebenen zu betrachten, gehen als Level 2 Prädiktoren die Personenmittelwerte der Versuchseigenschaften und als Level 1 Prädiktoren die um den Personenmittelwert zentrierten Personenwerte ein. Da die Versuchseigenschaften erst ab dem Wintersemester 2017/2018 erhoben wurden, gehen insgesamt $n = 738$ versuchsbezogene Daten für 57 Probanden in das Modell ein.

Tabelle 13 zeigt die festen und zufälligen Effekte der berechneten Mehrebenenanalyse (Modell 1). Der Fokus liegt dabei auf der Analyse der festen Effekte des Modells. Die Effekte auf Level 1 und 2 sind zur Übersicht nebeneinander in zwei Spalten dargestellt. Die Versuchseigenschaften sind anhand der festen Effekte auf der Ebene *Innerhalb Personen* sortiert. Eine übersichtlichere Darstellung dieser festen Effekte auf Level 1 ist angelehnt an ein Pfaddiagramm (siehe Abschnitt 11.2.3) in Abbildung 30 zu finden. Da das Hauptinteresse anhand der Fragestellung an diesen personenunspezifischen Effekten besteht, werden diese als erstes betrachtet. Auch wenn die Signifikanzniveaus für Mehrebenenanalysen kritisch diskutiert werden (Luke, 2017), wurden diese zur Orientierung hinzugenommen. Bei nicht signifikanten Effekten wurden die Versuchseigenschaften in Abbildung 30 ausgegraut, um den Fokus auf die Versuchseigenschaften mit größeren Effekten zu legen. Die größten Effekte ($\gamma_{k0} > .1$) auf die wahrgenommene Relevanz haben der *Kontext zum Schulalltag, Schulerinnerung* und die *Anschaulichkeit*. Kleinere positive signifikante Effekte ($\gamma_{k0} \leq .1$) haben außerdem die Eigenschaften *Kontext zum Lebensalltag, effektive Vorbereitung, abwechslungsreich, aktuell, Versuchsmaterial von guter Qualität, viel quantitative Auswertung* und *vernetzt*. Die meisten dieser Eigenschaften lassen sich wie bei der Betrachtung der Korrelationen in schulbezogene (*Kontext zum Schulalltag, Schulerinnerung, anschaulich, Kontext zum Lebensalltag, abwechslungsreich, aktuell, vernetzt*) oder strukturelle Kategorien (*effektive Vorbereitung, klare Anleitung*) einordnen. Interessant ist, dass die Eigenschaft *viel quantitative Auswertung* einen positiven Effekt auf die wahrgenommene Relevanz aufweist, da man in den ersten Jahren des Physikunterrichts als Lehrkraft auch einen Großteil der Experimente qualitativ betrachtet. Dass *Versuchsmaterial guter Qualität* einen positiven Effekt auf die wahrgenommene Relevanz in Bezug auf eine spätere Lehrtätigkeit hat, ist eher unerwartet, zumal an Schulen aufgrund des eingeschränkten Budgets selten die hochwertigsten Versuchsmaterialien vorzufinden sind. Eine mögliche Erklärung dafür ist, dass mit Versuchsmaterialien weniger guter Qualität direkt (mess-)technische Schwierigkeiten verbunden werden. Dies würde wiederum

AV: wahrgenommene Relevanz, UV: Versuchseigenschaften

	Modell 0P	**Modell 0V**	**Modell 1**	
	Personen	Versuche	Level 1	Level 2
feste Effekte			γ_{k0}	γ_{0k}
Kontext zum Schulalltag			$.20^{***}$	.08
Schulerinnerung			$.15^{***}$	$-.03$
anschaulich			$.13^{***}$	.06
Kontext zum Lebensalltag			$.10^{***}$	$.23^{+}$
effektive Vorbereitung			$.10^{**}$	.08
abwechslungsreich			$.09^{**}$	$.21^{*}$
aktuell			$.09^{**}$	.00
Versuchsmaterial von guter Qualität			$.08^{**}$	$-.02$
viel quantitative Auswertung			$.08^{**}$	$-.09$
vernetzt			$.06^{*}$	.13
Aufgaben erfordern „trickreiche“ Idee			.05	.06
Geschick fordernde Durchführung			.05	$-.14$
eindeutig interpretierbare Ergebnisse			.02	$-.05$
Einbringung eigener Ideen möglich			.02	$-.06$
eigenständiger Versuchsaufbau			.02	$-.06$
überraschend			.01	$-.03$
abstrakt			.01	$-.13$
repetitive Versuchsdurchführung			.00	.04
empfindliche Versuchsmaterialien			.00	$-.05$
klare Anleitung			$-.01$	.02
Versuchsmaterial leicht beschaffbar			$-.02$	$-.03$
anspruchsvoll			$-.02$	$-.05$
Gesamtdauer			$-.02$	$-.14$
Versuchsmaterial in Schulsammlung			$-.03$	.02
eng geführte Versuchsaufgaben			$-.03$	$-.10$
moderne Versuchsmaterialien			$-.04$	$-.02$
ungewöhnliches Versuchsmaterial			$-.05$	$-.00$
Gesamtschwierigkeit			$-.10^{**}$	$.22^{+}$
zufällige Effekte				
Varianz zwischen Gruppen ($Var(u_{0j})$)	0.25	0.10	0.14	
Varianz innerhalb Gruppen ($Var(e)$)	0.76	0.90	0.52	
ICC	.25	.10	0.21	
Designeffekt	4.4	4.1		
Anzahl Beobachtungen	738	738	738	
Anzahl Gruppen	57	25	57	
Pseudo-R^2			0.32	0.42
Devianz	1980	2056	1880	

Model 0: Intercept-Only-Modell; **Model 1:** Random-Intercept-Model

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$,$^{+}p < 0.10$

Tabelle 13: *Mehrebenenmodelle für Versuchseigenschaften (unabhängige Variable) und wahrgenommenen Relevanz (abhängige Variable) (Modellspezifikationen: Tab. 9, γ standardisierte Regressionsgewichte; Intercepts wegen Standardisierung nicht dargestellt).*

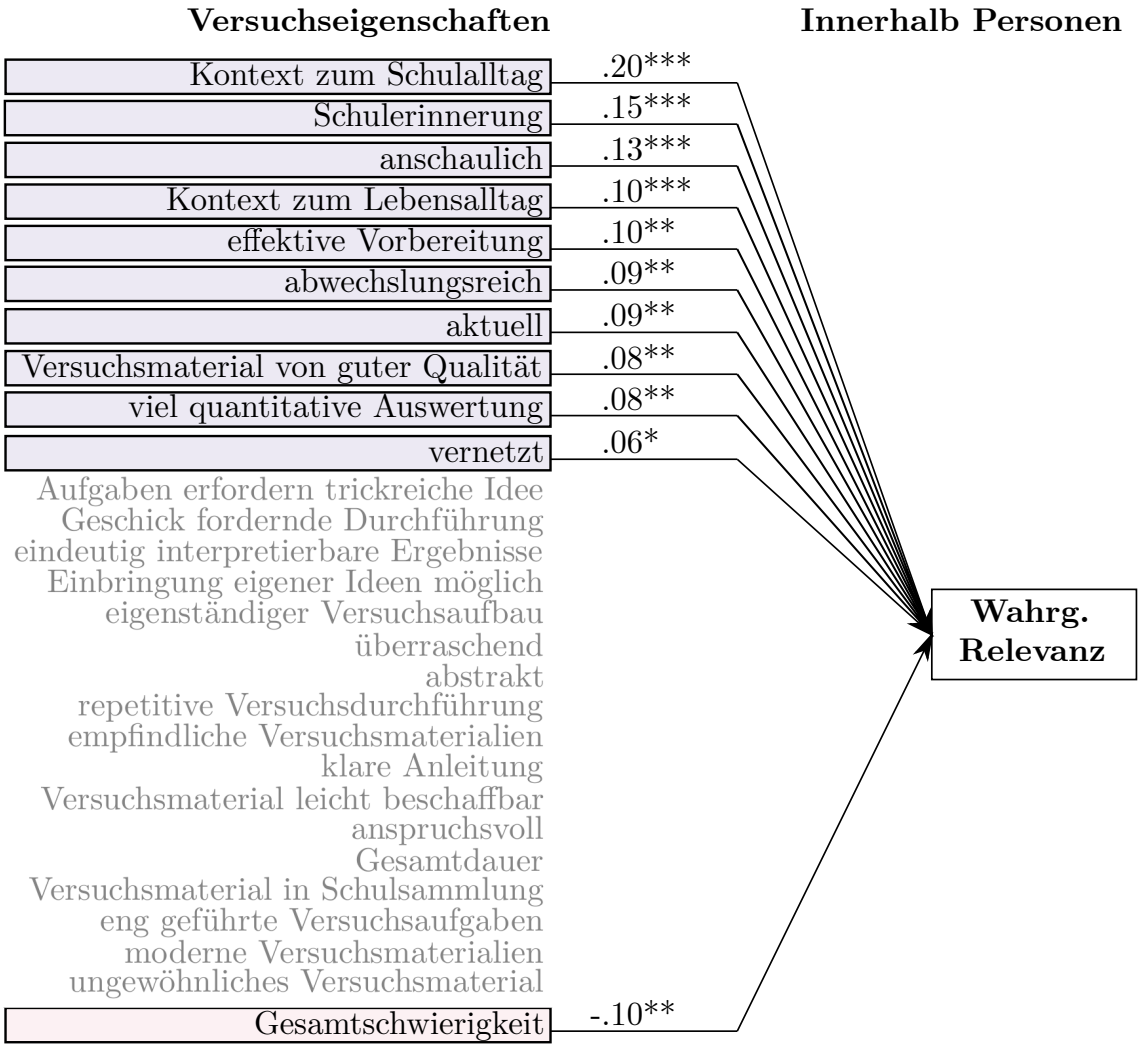


Abbildung 30: *Die Darstellung veranschaulicht die standardisierten Regressionsgewichte der Mehrebenenanalyse von Versuchseigenschaften (UV) und wahrgenommener Relevanz (AV) auf der Datenebene Innerhalb Personen. Versuchseigenschaften, für die das Regressionsgewicht nicht signifikant ist, wurden ausgegraut. Positive Effekte sind violett und negative Effekte rot hinterlegt.*

mit dem negativen Effekt ($\gamma_{k0} = -.10, p = .004$) auf die wahrgenommene Relevanz zusammenpassen, der von der *Gesamtschwierigkeit* des Versuches ausgeht.

Zudem zeigt Tabelle 13 in der rechten Spalte von Modell 1 die festen Effekte auf der Datenebene Zwischen Personen (Level 2), die auf Grundeinstellungen der Personen zurückgeführt werden können. Da nur 57 Personen in diesem Teildatensatz befragt wurden, ist es auf dieser Ebene schwierig, Effekte signifikant nachzuweisen. Daher sollen an dieser Stelle nur kurz die drei Effekte mit Signifikanzniveaus $p < .100$ betrachtet werden und mit den Effekten Innerhalb der Personen verglichen werden. Studierende, die die Praktikumsversuche als generell eher abwechslungsreich einstufen, nehmen die Versuche als relevanter wahr ($\gamma_{0k} = .21, p = .028$). Dies kann ein Hinweis darauf sein, dass diese Studierenden das Praktikum als persönliche Abwechslung von den bis dahin recht theoretischen und inhaltlich fokussierten Lehrveranstaltungen ansehen und/oder das Erlernen experimenteller Kompetenzen im Kontext wechselnder Lerninhalte und -medien als Merkmal von begeisternden Experimenten im Schulunterricht erachten und damit als relevant einstufen. Studierende, die in den Praktikumsversuchen grundsätzlich einen starken Kontext zum Lebensalltag erkennen, nehmen eine größere Relevanz der Versuche wahr ($\gamma_{0k} = .23, p = .093$). Das könnte man damit erklären, dass diese Studierenden insgesamt mehr Zusammenhänge zwischen der Physik und ihrem Lebensalltag ziehen und deshalb eine größere Relevanz von physikalischen Phänomen insgesamt erkennen. Die Versuchseigenschaften *abwechslungsreich* und *Kontext zum Lebensalltag* spielen damit folglich nicht nur eine Rolle auf der Datenebene Innerhalb Personen, sondern auch Zwischen Personen. Auch der Aspekt der Schwierigkeit hat auf dieser Ebene einen Effekt, aber mit umgekehrtem Vorzeichen. Studierende, die alle Praktikumsversuche insgesamt als schwerer empfunden haben, nehmen eine größere Relevanz der Versuche wahr ($\gamma_{0k} = .22, p = .080$). Dies ist nicht widersprüchlich zu dem negativen Effekt auf der Datenebene *Innerhalb Personen* (Level 1), bei denen einzelne Versuche als schwierig und damit als weniger relevant als andere eingestuft werden. Auf dieser Ebene *Zwischen Personen* (Level 2) geht es um Studierende, die die Praktikumsversuche insgesamt für schwierig halten. Ein mögliche Erklärung dafür ist, dass diese Studierenden Schwierigkeiten mit den physikalischen Inhalten haben, ihre Schwächen aber erkennen, gegebenenfalls auch mehr im Praktikum lernen und daher die Notwendigkeit und Relevanz des Praktikums für ihre Zukunft als Lehrkraft erkennen. Um diese Effekte auf Ebene *Zwischen Personen* besser einordnen und interpretieren zu können, müssten weitere Betrachtungen der Daten für die einzelnen Personen angestellt werden. Im Rahmen der Forschungsfragen wird darauf in dieser Arbeit verzichtet, da diese Level 2 Effekte weniger durch entsprechende Konzeption von Versuchen beeinflusst werden können und somit aus Sicht der Praktikumsentwicklung weniger interessant sind.

Im Vergleich zum Nullmodell (Modell 0P) werden 32 % der Varianz auf Level 1 und 42 % der Varianz auf Level 2 mehr durch die Aufnahme der Prädiktoren in Modell 1 aufgeklärt (Abb. 13). Aufgrund der breit angelegten Abfrage der Versuchs-

eigenschaften wird davon ausgegangen, dass diese vollständig erfasst wurden. Mit Verzerrungen der Modellparameter durch Auslassen relevanter Versuchseigenschaften ist daher nicht zu rechnen. Eine Vielzahl anderer Bedingungsfaktoren auf die wahrgenommene Relevanz der Praktikumsversuche, die nicht mit den Eigenschaften der Versuche zusammenhängen, sondern zum Beispiel die Vorerfahrungen oder Einstellungen der Studierenden betreffen, ist aber wahrscheinlich. Vor diesem Hintergrund ist die Änderung des aufgeklärten Varianzanteils durch die Versuchseigenschaften alleine als hoch zu bewerten. Zudem ist bei der Vielzahl an möglichen Einflussfaktoren nicht verwunderlich, dass die Effekte der einzelnen Versuchseigenschaften auf die Relevanz tendenziell eher klein sind.

Ein weiterer Diskussionspunkt ist die Stabilität der Modellschätzung hinsichtlich der bereits zuvor vermuteten individuellen Wahrnehmung relevanter Versuche. Die Gruppe der Lehramtsstudierenden im Fach Physik ist aufgrund ihrer Erfahrungen im Zweitfach sehr heterogen, wie in Abschnitt 5.4 beschrieben. Geht man davon aus, dass die wahrgenommene Relevanz der Versuche nicht nur von den Eigenschaften der Versuche abhängt, sondern auch von Faktoren die Studierenden betreffend wie unterschiedliche Interessens-Orientierungen, dann schließt sich schnell die Hypothese an, dass sich Studierendentypen identifizieren lassen, die unterschiedliche Ansprüche an für sie relevante Versuche stellen. Während in dieser Studie von einer homogenen Studierendengruppe ausgegangen wird und dementsprechend nach der optimalen Merkmalskombination gesucht wird, könnte das Problem komplexer als angenommen sein und soll daher kurz diskutiert werden. Um die Stabilität der Ergebnisse dieser Teilanalyse zu prüfen, wurde die Mehrebenenanalyse für verschiedene Teildatensätze durchgeführt. Dabei wurden zum einen zufällige Teildatensätze betrachtet. Zum anderen wurden die Teildatensätze systematisch ausgewählt, um zu prüfen, ob sich die Ergebnisse für bestimmte Personengruppen unterschieden.

Die Untersuchungen der zufälligen Teildatensätze haben gezeigt, dass die Ergebnisse stabil sind und keine grundlegend anderen Versuchseigenschaften als relevanzschaffend identifiziert werden. Die Betrachtungen zeigen allerdings auch, dass zum Beispiel bei Halbierung des Datensatzes der Stichprobenumfang zu gering ist und kleine Effekte entsprechend nicht mehr signifikant nachgewiesen werden können. Eine naheliegende Vermutung zu systematischen Teilgruppen von Studierenden ist, dass es Unterschiede in der Versuchswahrnehmung zwischen Studierende gibt, die besonders gut im Studium abschneiden, und anderen, die noch viele Lücken zu schließen haben. Ein plausibler Grund dafür wäre, dass die Relevanz sehr abstrakter Versuche schwieriger zu erfassen sein dürfte und damit von den Studierenden mit höherem Fachwissen besser erkannt werden müsste. Die Mehrebenenanalyse liefert für Teildatensätze, die mithilfe eines Mediansplits anhand des Post-Fachwissenstests (siehe Abschnitt 8.1) gefunden wurden, jedoch keine Hinweise darauf, dass deutlich unterschiedliche relevanzstiftende Versuchsmerkmale für Studierende mit hohem und niedrigem Fachwissen zu finden sind. Auch ein exploratives Vorgehen, bei dem mittels Distanzplots und Clusteranalysen (z. B. Backhaus et al., 2016) Gruppen von

Studierenden anhand Ähnlichkeiten oder Abweichungen in den Einschätzungen der wahrgenommenen Relevanz identifiziert wurden, zeigt im Rahmen der Stichprobengröße keine Hinweise darauf, dass wesentliche andere Versuchseigenschaften die besonders relevanten Versuche charakterisieren.

Damit zeigt die Untersuchung der Forschungsfrage bzgl. Studierendengruppen mit unterschiedlichen Interessens-Orientierungen, dass mit kleinen Effekten zu rechnen ist und damit in dieser Arbeit statistische Grenzen gesetzt sind. Zum anderen kann die Stabilität der Ergebnisse der Analyse dieses Abschnittes geschlussfolgert werden.

Die Mehrebenenanalyse und die Betrachtung der Korrelationen zeigen grundsätzlich konsistente Ergebnisse. In beiden Analysen konnten ähnliche Haupteinflussfaktoren auf die wahrgenommene Relevanz identifiziert werden wie bei der Betrachtung der Korrelationen zuvor. Ein direkter quantitativer Vergleich der Regressionsgewichte und der Korrelationen ist allerdings nicht möglich (siehe Abschnitt 8.3.5). Qualitativ kann aber die Reihung bzw. Gewichtung der Versuchseigenschaften nach der Größe der Effekte verglichen werden. Dabei zeigen sich große Ähnlichkeiten, da zum Beispiel in beiden Analysen die Versuchseigenschaften *Kontext zum Schulalltag* und *anschaulich* mit die größten positiven Zusammenhänge aufzeigen und *Gesamtschwierigkeit* die größten negativen Zusammenhänge. Auffällig ist, dass die *Erinnerung der Versuche aus der Schule* durch die Mehrebenenanalyse an die zweithöchste Stelle im Ranking der Versuchseigenschaften steigt. Zudem zeigt sich, dass die Bearbeitungsdauer anhand der Mehrebenenanalyse im Gegensatz zur Korrelation keinen nachweisbaren Effekt auf die wahrgenommene Relevanz hat. Dies kann an dem Zusammenhang von Gesamtdauer und Schwierigkeitsgrad liegen, der zu Beginn des Abschnittes erkannt wurde. Die wesentlichen Erkenntnisse der Mehrebenenanalyse bestehen damit vor allem in einer belastbareren Schätzung der Effekte der einzelnen Versuchseigenschaften auf die wahrgenommene Relevanz.

Zusammengefasst zeigt diese Analyse, dass oberflächliche Versuchsmerkmale (z. B. empfindliches oder modernes Versuchsmaterial) keinen signifikanten Effekt auf die wahrgenommene Relevanz haben. Es lassen sich aber übergeordnete Versuchseigenschaften identifizieren, die einen positiven Effekt auf die wahrgenommene Relevanz haben. Diesen Versuchseigenschaften (u. a. Schulkontext, schulrelevante Versuchsauswahl und Anschaulichkeit) lässt sich ein schulischer Bezug zuordnen.

10 Synopse zur wahrgenommenen Relevanz

Die Forschungsfragen dieses Teils der Arbeit zielen übergeordnet auf die Fragestellung ab, wie ein lehramtsspezifischer Praktikumsversuch gestaltet werden muss, um von den Studierenden als möglichst relevant wahrgenommen zu werden. Die Analysen der Versuchseigenschaften und der wahrgenommenen Relevanz haben gezeigt, dass es nicht die eine Stellschraube gibt, die eine hohe Relevanzwahrnehmung des Praktikumsversuches garantiert. Ein Praktikumsversuch ist ein komplexes und multidimensionales System. In dieser Arbeit konnten daher mehrere kleinere Stellschrauben für die wahrgenommene Relevanz identifiziert werden. Zu den Versuchseigenschaften, die die größten Effekte auf die wahrgenommene Relevanz haben, zählen Kontext zur Schule, Erinnerung der Versuche aus der eigenen Schulzeit, Anschaulichkeit, Schwierigkeitsgrad, Kontext zum Lebensalltag, Abwechslung, Aktualität und Vernetzung. Übergeordnet lassen sich die Versuchseigenschaften, die die größten Effekte auf die wahrgenommene Relevanz haben, somit als schulnah kategorisieren. Dies deutet außerdem darauf hin, dass die Studierenden die Relevanz der Versuche zu einem großen Teil anhand der Eignung dieser Versuche für den Einsatz im Schulunterricht, also aus einer schulpraktischen Sicht beurteilen.

Dies deckt sich mit den Ergebnissen der Studie von Lorentzen et al. (2019), die für Lehramtsstudierende im Fach Chemie gezeigt haben, dass die Relevanz von Fachinhalten am häufigsten aus einer schulischen Perspektive begründet wird (siehe auch Lorentzen, 2020). Zudem bestehen Überschneidungen der gefundenen Eigenschaften relevant wahrgenommener Praktikumsversuche mit denen relevant wahrgenommener Physik-Übungsaufgaben für Lehramtsstudierende. Massolt und Borowski (2020) haben diese unter anderem identifiziert als konzeptuelle Aufgabenstellung, nah am Lebensalltag, geringe mathematische Anforderungen, schulbezogener Kontext und Situationsbezug (z. B. Fehlvorstellung). Damit sind die Ergebnisse dieser Analyse zu den relevanzfördernden Eigenschaften von Praktikumsversuchen konsistent mit den Ergebnissen verwandter aktueller Forschungen aus der Lehramtsausbildung.

Die drei gefundenen Hauptmerkmale der als besonders relevant empfundenen Versuche sind der Schulkontext, die Erinnerung der Versuche aus der eigenen Schulzeit und die Anschaulichkeit von Versuchen. Diese stehen in direktem Zusammenhang mit den Designkriterien des neu entwickelten Lehramtspraktikums (siehe Abschnitt 6). Im Lehramtspraktikum wird der Schulbezug der Praktikumsversuche durch Kontextualisierung der Versuchsbeschreibung und der Vernetzung von fachlichen und fachdidaktischen Fragestellungen hergestellt. Durch diese Analyse wird die konzeptionelle Grundidee aller adressatenspezifischer Praktika, durch Kontextualisierung der Versuchsbeschreibung Relevanz zu schaffen (siehe Abschnitte 2.3 und 3.4), für das Lehramtspraktikum bestätigt. Der Aspekt der Schulerinnerung spielt insofern eine besondere Rolle bei den Versuchseigenschaften, da eine Erinnerung nicht direkt durch die Gestaltung der Versuche erzeugt werden kann. Hinter dem

Aspekt der Schulerinnerung verbirgt sich aber die Wichtigkeit der Auswahl schulrelevanter Versuchsthemen für die wahrgenommene Relevanz. Damit wird die Auswahl der Versuchsthemen anhand der schulischen Fachanforderungen für ein Lehramtspraktikum bestätigt, da durch diese Auswahl die Wahrscheinlichkeit gesteigert wird, dass die Studierenden die Versuche aus eigener Schulzeit kennen. Zudem wurde festgestellt, dass die Anschaulichkeit der Versuche ebenfalls einen positiven Effekt auf die wahrgenommene Relevanz der Versuche hat. Eine plausible Erklärung dafür ist, dass die Studierenden Versuche, die sie selbst als wenig anschaulich und ggf. infolge dessen auch als abstrakter und schwieriger empfinden, aus diesem Grund wahrscheinlich auch als unzureichend anschaulich für Schüler*innen erachten und damit als weniger relevant wahrnehmen. Dies spricht für den Einsatz der Analogie-Versuche im neuen Lehramtspraktikum. Die Analogie-Versuche werden gerade dazu genutzt, um Prozesse, die auf mikroskopischer Ebene ablaufen und damit von sich aus wenig anschaulich sind, in einem Praktikumsversuch auf makroskopischer Ebene anschaulich zu gestalten. Damit kann diese Analyse einige der Designkriterien, die dem neu entwickelten Lehramtspraktikums zugrunde liegen, als wichtige Stellschrauben bestätigen.

Zudem wird hier noch einmal deutlich, dass einige Studierende die Relevanz der Praktikumsversuche wahrscheinlich aus einer schulpraktischen Perspektive beurteilen. Daraus würde folgen, dass Studierende einen Praktikumsversuch, der sich zum Beispiel aufgrund des Themas nicht direkt für den Einsatz im Schulunterricht eignet, als weniger relevant einschätzen, auch wenn dieser Versuch den angehenden Lehrkräften wichtige Kompetenzen und Kenntnisse vermittelt kann. Dies würde wiederum bedeuten, dass den Studierenden in diesem Fall die Relevanz der Lerninhalte für die zukünftige Lehrtätigkeit im besonderen Maße deutlich gemacht werden muss. Dies kann über den Schulbezug in der Versuchsbeschreibung gelingen.

Der schulbezogene Kontext spielt eine Schlüsselrolle für die wahrgenommene Relevanz. In der Analyse hat sich gezeigt, dass eine Kontextualisierung der Einleitung der Versuchsbeschreibung allein bereits schon ausreichend ist, damit der Schulkontext des Versuches besser erkannt wird und die wahrgenommene Relevanz des Versuches erhöht wird. Dies lässt vermuten, dass ein deutlich größerer Effekt durch eine Optimierung der Kontextualisierung erreicht werden kann. Da der Gestaltungsspielraum der Versuchsbeschreibung durch den Schulkontext in der Einleitung alleine noch nicht voll ausgeschöpft ist, stehen sogar noch einige Möglichkeiten offen, den Schulkontext stärker in der Versuchsanleitung zu verankern. Wenn die Rahmenbedingungen bei der Gestaltung lehramtsspezifischer Praktikumsversuche es zulassen, dann sollte sich die Kontextualisierung idealerweise durch die gesamte Versuchsbeschreibung ziehen und an prominenter Stelle der Beschreibung besonders hervorgehoben werden. Der Schulkontext sollte auch über das Rezipieren hinaus stärker zum Beispiel in Form fachdidaktischer Fragen in die Versuchsaufgaben integriert werden, deren Beantwortung im Protokoll eingefordert wird. Eine gezielte Diskussi-

on des Schulkontextes in den Prüfungsgesprächen könnte sogar gezielt sicherstellen, dass die Schulrelevanz erkannt wird. Das Ergebnis zur Kontextualisierung zeigt folglich aber auch, dass eine lehramtsspezifische Ausrichtung eines Physikpraktikums nicht zwangsläufig eine Neukonzeption erfordert, sondern auch in engeren Rahmenbedingungen bei bestehenden Versuchen zielführend umgesetzt werden kann. Die Stellschraube der Kontextualisierung der Einleitung von Versuchsbeschreibungen zeigt attraktive Möglichkeiten für Laborpraktika für Lehramtsstudierende an anderen Standorten und anderer naturwissenschaftlicher Fachrichtungen auf.

Teil IV
Wirkung

Nach der Beforschung der neuen lehramtsspezifischen Praktikumsversuche in Bezug auf die wahrgenommene Relevanz im Teil III, soll in diesem Teil noch ein Schritt über diese Analyse hinaus gewagt werden. Da vorangehend Versuchseigenschaften identifiziert werden konnten, die gemäß den in Teil I aufgestellten Wirkannahmen einen positiven Effekt auf die wahrgenommene Relevanz der Praktikumsversuche haben, soll nun die zweite Komponente des postulierten Wirkmodells im Rahmen von Forschungsfrage 3 geprüft werden. Das heißt, es sollen Erkenntnisse über die Wirkung der wahrgenommenen Relevanz auf den Lernerfolg gewonnen werden.

Auch hier spielt die Vielschichtigkeit des Lehramtspraktikum als Forschungsgegenstand eine Rolle, die sich in diesem Teil besonders in der Operationalisierung des Lernerfolgs im Wirkmodell widerspiegelt. Neben der bereits dargelegten Mehrebenenstruktur der Daten, stellt in diesem Analyseteil auch die kleine Stichprobe eine Herausforderung dar. Daher ist es für diese Analyse erforderlich, zunächst die Erfassung des Lernerfolgs im Praktikum und das methodische Vorgehen präzise zu beschreiben (Abschnitt 11). Erst vor diesem Hintergrund kann dann die Analyse des Wirkmodells selbst erfolgen (Abschnitt 12). Abschließend wird eine Bewertung und Deutung der Ergebnisse in Abschnitt 13 vorgenommen.

11 Methodisches Vorgehen **144**

11.1 Erhebungsinstrumente 144

11.2 Statistische Methodik 173

12 Prüfung des Wirkmodells **183**

12.1 Interesse 183

12.2 Experimentelle Kompetenz 205

12.3 Fachwissen 229

13 Synopse zum Wirkmodell **248**

11 Methodisches Vorgehen

Nachdem das Design der Studie bereits ausführlich im vorangegangenen Teil dargestellt wurde, kann die Betrachtung des methodischen Vorgehens für die weiterführende Analyse des Wirkmodells darauf aufbauen. Dazu ist es notwendig, drei Komponenten zu betrachten. Zentral ist dabei die Operationalisierung und die Güte der Erfassung der zentralen Größen des Wirkmodells zum Lernerfolg im Praktikum (Abschnitt 11.1). Um die Stichprobengröße und die mehrschichtige Datenstruktur zu berücksichtigen, ist nicht zuletzt die Auswahl einer Reihe von unterschiedlichen statistischen Analyseverfahren notwendig (Abschnitt 11.2). Erst nach Betrachtung dieser drei Aspekte zum methodischen Vorgehen kann anschließend die Analyse des Wirkmodells selbst in Abschnitt 12 vorgenommen werden.

11.1 Erhebungsinstrumente

An dieser Stelle wird zunächst ein Überblick über die Operationalisierung des zu prüfenden Wirkmodells und damit über die Wahl der Erhebungsinstrumente gegeben, bevor Details zu den Instrumenten in den nachfolgenden Abschnitten ausgeführt werden.

Im Wirkmodell des Praktikums wird ein positiver Wirkzusammenhang zwischen wahrgenommener Relevanz, situationalem Interesse und Lernerfolg angenommen (siehe Abb. 31). In der vorangegangenen Analyse in Teil III wurde die Erfassung der wahrgenommenen Relevanz als zentrale Ausgangsgröße des Wirkmodells bereits ausführlich betrachtet. Im angenommenen Wirkmodell wird der Effekt der wahrgenommenen Relevanz auf den Lernerfolg über das situationale Interesse Catch mediierte, das als Proxy für die im Praktikumsversuch angeregte Lernmotivation betrachtet wird (siehe Abschnitt 3.2). Zielgröße ist die Verbesserung des Lernerfolgs. Basierend auf den Lernzielen wurden als wichtigste Elemente des Lernerfolgs im Praktikum die Zunahme des individuellen Interesses, der experimentellen Kompetenz und des physikalischen Fachwissens identifiziert und spezifiziert (siehe Abschnitte 3.1 und 3.2). Das individuelle Interesse wurde in dieser Arbeit als Interesse an Physik und am Unterrichten von Physik erfasst. Zudem wurde in Abbildung 31 das erfasste situationale Interesse Hold zum Lernerfolg bzgl. des individuellen Interesses zugeordnet, da dieses als Vorstufe und damit als Prädiktor für individuelles Interesse betrachtet werden kann, wie in Abschnitt 3.2 erläutert. Die Zunahme der experimentellen Kompetenz wurde mithilfe einer Selbsteinschätzung erfasst. Zusätzlich wurden Selbstwirksamkeitserwartungen in Bezug auf das Experimentieren im Unterricht erhoben, wie in Abschnitt 11.1.2 näher erläutert wird. Auch für das Fachwissen wurden mit einem Wissenstest und einer Selbsteinschätzung zwei sich ergänzende Instrumente genutzt, wie in Abschnitt 11.1.3 ausgeführt wird. Auch wenn neben dem fachlichen Wissen auch das fachdidaktische Wissen eine wichtige Komponente des Professionswissens von Lehrkräften bildet, spielt der Erwerb von fachdidaktischem Wissen eine unter-

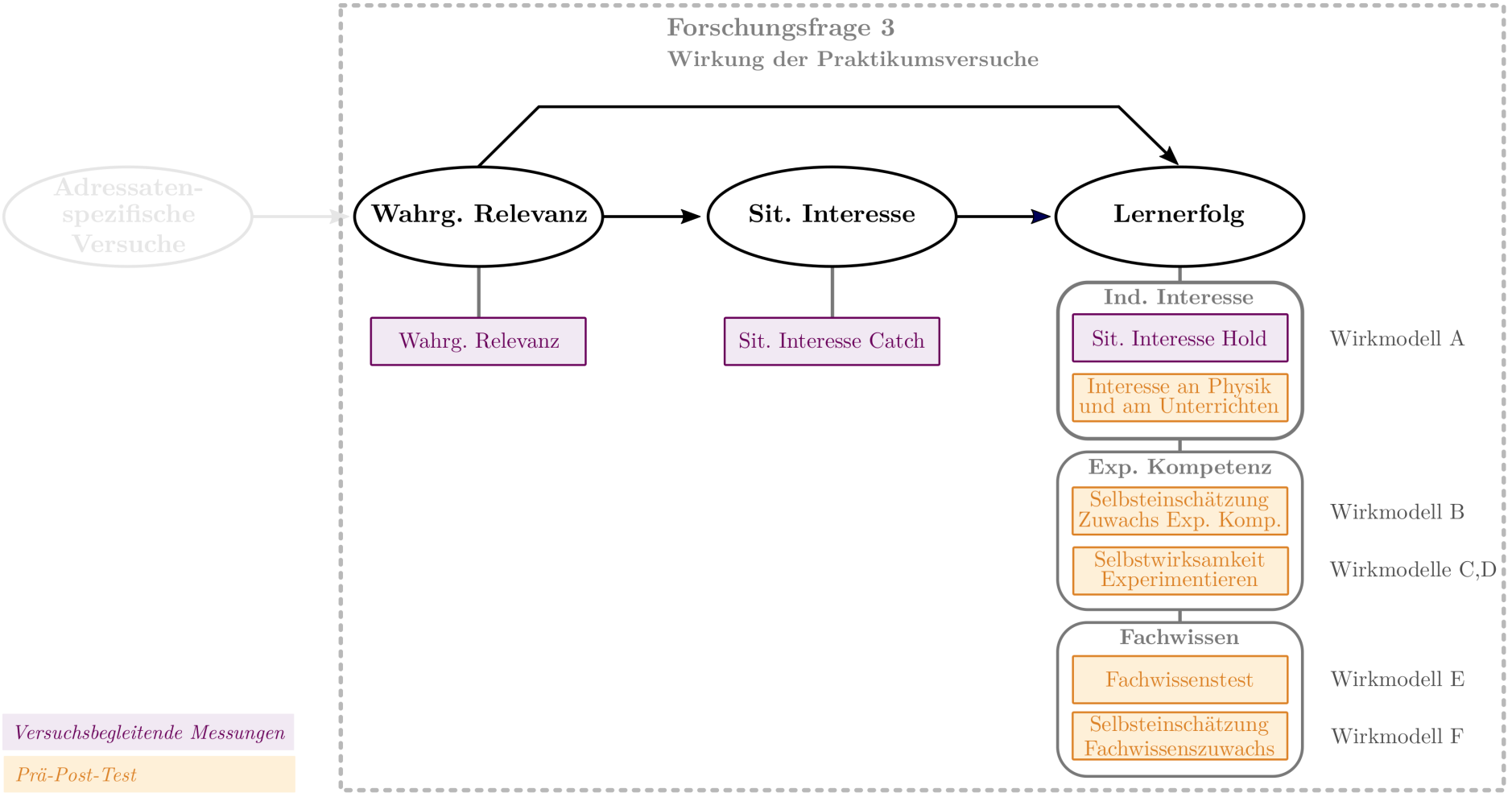


Abbildung 31: *Operationalisierung des Wirkmodells zur Untersuchung von Forschungsfrage 3. Eine höhere wahrgenommene Relevanz fördert demnach das situationale Interesse und damit den Lernerfolg. Der Lernerfolg teilt sich in die drei Komponenten individuelles Interesse, experimentelle Kompetenz und Fachwissen auf. Die farbliche Kodierung zeigt die zugeordneten Erhebungsinstrumente, die versuchsbegleitend (magenta) oder Prä-Post (orange) eingesetzt wurden. Außerdem sind rechts die Wirkmodelle A bis F zugeordnet, die später in Abschnitt 12 geprüft werden.*

Konstrukt	**Subskala**	**MZP**	**Skala**	**Items**	α	**Quelle**
Wahrgenommene Relevanz		versuchs-begleitend	5-stufige Likert-Skala	8	.81	Schiefele et al. (2002)
Situationales Interesse	Catch	versuchs-begleitend	5-stufige Likert-Skala	8	.93	Lewalter und Knogler (2014),
	Hold			8	.85	Knogler (2014)
Interesse an Physik		Prä-Post	5-stufige Likert-Skala	6	.63 – .79	Keller (u.a. 2011)
Interesse am Unterrichten von Physik		Prä-Post	5-stufige Likert-Skala	6	.83 – .92	Keller (u.a. 2011)
Selbstwirksamkeit in Bezug auf Experimente im Unterricht	Planung	Prä-Post	5-stufige Likert-Skala	8	.74 – .84	Rabe et al. (2012)
	Durchführung			8	.79 – .86	Meinhardt (2018)
Fachwissen	6	Prä-Post	Multiple-Choice-Aufgaben	48	Prä: .75, Post: .74	KiL (Kröger, 2019), FCI (Hestenes et al., 1992), (...)
Lernzuwachs Fachwissen	Theorie und Praxis	Post	6-stufige Likert-Skala	8	.87	Rehfeldt (2018a)
Lernzuwachs Experimentelle Kompetenz	9	Post	6-stufige Likert-Skala	40	.97	Rehfeldt (2018a)

Tabelle 14: *In dieser Tabelle sind zur Übersicht die Konstrukte, die für diese Teilstudie zum Wirkmodell erhoben wurden, mit Messzeitpunkt (MZP), Skala, Anzahl der Items und Cronbachs α als Maß für die Reliabilität (Abschnitt 8.2.1) gelistet.*

geordnete Rolle im Anfängerpraktikum Lehramt (siehe Abschnitt 6) und wird daher im Wirkmodell mit Bezug auf das Lehramtspraktikum nicht als Lernerfolg erfasst.

Die auf das Wirkmodell abgestimmten Erhebungsinstrumente bestimmen nicht nur was gemessen wird, sondern auch in welcher Qualität. Da die Erfassung des Lernerfolgs nicht trivial ist und die Güte der Messinstrumente ein wesentlicher Faktor für die Qualität der Untersuchung ist, ist im Folgenden die Bewertung der Erhebungsinstrumente und die Betrachtung deren Besonderheiten besonders wichtig. In Tabelle 14 ist eine Übersicht über die verwendeten Konstrukte und deren Kenngrößen dargestellt. Sowohl in Abbildung 31 als auch in Tabelle 14 ist erkenntlich, welche der Instrumente versuchsbegleitend bzw. Prä-Post erhoben wurden, da dies für die Struktur und damit für die Analyse der Daten bedeutend ist. Die Skalen wurden anhand der in Abschnitt 8.2.1 vorgestellten statistischen Verfahren und angenommenen Grenzwerte hinsichtlich ihrer Güte beurteilt. Im Detail wird auf Güte und Besonderheiten dieser Skalen in den folgenden Abschnitten 11.1.1 bis 11.1.3 eingegangen. Einzelne Items und deren Kenngrößen sind im Anhang A.2 tabelliert.

11.1.1 Erhebungsinstrumente: Interesse

Das Interesse der Lehramtsstudierenden spielt im Wirkmodell eine Doppelrolle (Abb. 31, Abschnitt 3.2). Das situationale Interesse als Proxy für die Lernmotivation bzgl. der Praktikumsversuche wird im Wirkmodell als Mediator aufgefasst. Eine Steigerung des individuellen Interesses wird als Lernerfolg im Praktikum bewertet. Aufgrund dieser zwei Rollen des Interesses im Wirkmodell wird an dieser Stelle als Erstes auf die Messungen der verschiedenen Phasen des Interesses und deren Besonderheiten und Güte eingegangen.

Interesse an Physik und am Unterrichten

Das Interesse am Fach Physik und am Unterrichten von Physik stehen für das als individuelle Interesse. Diese Aspekte des Interesses der Lehramtsstudierenden wurden vor und nach dem Praktikumssemester als Prä-Post-Test erhoben, da sie als recht stabil angenommen werden. Die Skalen zum Interesse wurden größtenteils basierend auf bestehenden Items passend für die Lehramtsstudierenden des Faches Physik zusammengestellt und entsprechend in den Formulierungen angepasst. Grundlagen waren dazu Interessens-Skalen aus den Testheften des Projektes KiL (Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen) am Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik in Kiel (IPN) (z. B. Sorge, Keller et al., 2019), die wiederum auf den Arbeiten von Schiefele, Krapp, Wild et al. (1993) und Kunter et al. (2011) beruhen. Als weitere Quelle für die Items dienten die Skalen aus der Doktorarbeit von Keller (2011) (siehe auch Keller et al., 2014), die ebenfalls auf dem Fragebogen zum Studieninteresse (FSI) von Schiefele, Krapp, Wild et al. (1993) basieren. Zudem wurden Items aus dem Fragebogen von Neumann und Borowski (2011), der auf den Arbei-

ten von Rammert und Webler-Pijahn (2001) basiert, entnommen und angepasst. Die Studierenden bewerten im Prä- sowie im Post-Test jeweils sechs Aussagen zu ihrem Interesse an Physik und zu ihrem Interesse am Unterrichten von Physik anhand einer 5-teiligen Likert-Skala.

Ein typisches Item für das Interesse an Physik ist: „Ich kann mich für Physik begeistern.“. Auch in dieser Skala wurden negativ formulierte Items eingesetzt, um die Aufmerksamkeit der Studierenden beim Lesen der Items nicht zu verlieren: „Außerhalb des Physikstudiums befasse ich mich nicht mit Physik.“. Das Beispielitem „Ich erkläre gerne physikalische Zusammenhänge.“ für das Interesse am Unterrichten von Physik zeigt, dass Items bevorzugt ausgewählt und formuliert wurden, die keinen expliziten Bezug auf den Physikunterricht haben. Da die Lehramtsstudierenden bis auf vernachlässigbare Ausnahmen noch nicht in der Schule tätig sind bzw. waren, wurde für andere Items zudem im Einführungstext des Fragebogens der Bezug zum zukünftigen Unterrichten hergestellt. Eine Übersicht über alle Items der Skalen sind in Anhang A.2.3 gegeben.

Bei der Untersuchung der Güte der beiden Skalen wird auf die statistischen Verfahren aus Abschnitt 8.2.1 zurückgegriffen. Im Wintersemester 2015/2016 wurde mit Bachelor-Studierenden der Fachrichtungen 1-Fach-Physik, 2-Fach-Physik und Materialwissenschaften im alten Anfängerpraktikum eine Pilotierung einiger Skalen dieser Studie durchgeführt. Für die Skala zum Interesse an Physik hat die Pilotierung eine gute Reliabilität der Skala mit einem Cronbachs $\alpha = .79$ ($n = 69$) ergeben. Für das Interesse am Unterrichten von Physik lieferte die Pilotierung ebenfalls Hinweise auf eine gute Reliabilität ($\alpha = .94$), allerdings für eine sehr kleinen Stichprobe ($n = 9$). Die mittleren Trennschärfen der Items liegen für die Pilotierung in einem mittleren bis hohen Bereich (Interesse Physik: $\bar{r}_{i(t-i)} = .54$; Unterrichten: $\bar{r}_{i(t-i)} = .83$). Durch Entfernen einzelner Items wäre eine wesentliche Verbesserung der Reliabilität nicht gelungen. Daher wurden aufgrund der guten Kennwerte die Skalen in dieser Form beibehalten.

Auch im Einsatz lassen sich für alle vier Messzeitpunkte vor und nach dem ersten und zweiten Praktikumssemester (kurz: Prä 1, Post 1, Prä 2, Post 2) größtenteils gute Reliabilitäten für die beiden Skalen feststellen (siehe Tab. 15 und Details in Anhang A.2.3). Cronbachs α für das Interesse an Physik liegt im Bereich von $.63 \leq \alpha \leq .79$ und damit in zwei Fällen knapp unterhalb des angenommenen Grenzwertes (siehe Abschnitt 8.2.1). Für das Interesse am Unterrichten von Physik sind die Reliabilitäten gut mit $.83 \leq \alpha \leq .92$.

Die mittleren Trennschärfen der Items liegen für die vorliegende Arbeit in einem mittleren bis hohen Bereich (Interesse Physik: $.37 \leq \bar{r}_{i(t-i)} \leq .54$; Unterrichten: $.62 \leq \bar{r}_{i(t-i)} \leq .78$) (Tab. 15, Details in Anhang A.2.3). Für das Interesse an Physik stechen bei Betrachtung der einzelnen Itemtrennschärfen die negativ formulierten Items mit geringen Werten ($\bar{r}_{i(t-i)} \approx .30$) hervor. Die zum Teil geringere Reliabilität des Instruments für das Interesse an Physik könnte sich in der Nutzung dieser

zwei negativ formulierten Items von insgesamt sechs Items begründen. Eine differenziertere Analyse der Items zeigt aber, dass ein Entfernen von Items die Reliabilität nicht wesentlich verbessern würde (siehe auch Anhang A.2.3). Da die etwas geringeren Reliabilitäten nur für zwei der Messzeitpunkte auftreten und die anderen Werte deutlich besser sind, wird die Güte der Skala insgesamt nicht angezweifelt. Für das Interesse am Unterrichten von Physik fallen anhand der Deskription in Tabelle 15 hohe Mittelwerte von $M \approx 4.3$ auf einer 5-teiligen Likert-Skala auf, die auf einen Deckeneffekt hinweisen.

Zusammengenommen zeigen diese Betrachtungen, dass die zusammengestellte Skala zum Interesse an Physik und am Unterrichten von Physik eine konsistente und trennscharfe Messung für die Untersuchungen des Lernerfolgs liefert. Für das Interesse am Unterrichten von Physik muss aufgrund der hohen Mittelwerte ein Deckeneffekt und damit mögliche Verzerrungen für die weitere Analyse bedacht werden.

MZP	**Deskription**				**Itemtrennschärfe**		
	M	SE	n	α	$\bar{r}_{i(t-i)}$	min	max
Interesse an Physik							
Prä 1	3.75	0.05	87	.63	.37	.19	.51
Post 1	3.75	0.05	88	.65	.40	.28	.55
Prä 2	3.90	0.07	66	.79	.55	.39	.68
Post 2	3.88	0.07	67	.74	.50	.37	.68
Interesse am Unterrichten von Physik							
Prä 1	4.42	0.05	89	.83	.62	.50	.71
Post 1	4.33	0.06	88	.88	.69	.60	.82
Prä 2	4.31	0.08	68	.89	.72	.60	.86
Post 2	4.32	0.07	68	.92	.78	.74	.81

Tabelle 15: *Kennwerte für das Interesse an Physik (oben) und am Unterrichten von Physik (unten). Dabei wird unterschieden zwischen dem Messzeitpunkt (MZP) vor und nach dem ersten (Prä 1, Post 1) und dem zweiten Praktikumssemester (Prä 2, Post 2). Für die Deskription des Datensatzes sind Mittelwert (M, 5-teilige Likert-Skala), Standardfehler (SE) und Stichprobengröße (n) angegeben. Als Kenngrößen für die Analyse des Erhebungsinstrumentes ist neben der Reliabilität (Cronbachs α) auch die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$), sowie deren Extremwerte (min, max) angegeben.*

Situationales Interesse

Zusätzlich zum individuellen Interesse der Studierenden wurde das situationale Interesse der Studierenden versuchsbegleitend für jeden der Praktikumsversuche erhoben. Das situationale Interesse wurde mit der Skala von Lewalter und Knogler (2014) (siehe auch Knogler, 2014; Knogler et al., 2015) gemessen, die an diese Studie angepasst wurde. Dieser Fragebogen bietet den Vorteil, dass er bereits umfassend von den Autoren eingesetzt und untersucht wurde (z. B. Lewalter & Geyer, 2009; Lewalter & Willems, 2009; Knogler et al., 2015). Das situationale Interesse teilt sich in zwei Komponenten: Die Catch- und die Hold-Komponente (siehe auch Abschnitt 3.2). Die Catch-Komponente deckt dabei mit zwei Subskalen aufmerksamkeits- und gefühlsbezogene Aspekte ab. Die Hold-Komponente umfasst ebenfalls zwei Subskalen, welche den wahrgenommenen Wert und epistemische Orientierung messen. Die Studierenden bewerten insgesamt 16 Aussagen auf einer 5-teiligen Likert-Skala zum situationalen Interesse am jeweiligen Praktikumsversuch.

Im Folgenden soll jeweils ein Beispielitem zu den Subskalen einen Eindruck über den Fragebogen geben. Der Situationsbezug der originalen Items wurde in den Formulierungen angepasst auf den Praktikumstag beziehungsweise den Praktikumsversuch. Ein typisches Beispielitem für die Subskala Aufmerksamkeit der Catch-Komponente lautet somit: „Der Versuch konnte meine Aufmerksamkeit fesseln.“. Die Assoziation des Versuches mit positiven Gefühlen (Catch-Komponente) wird beispielsweise mit dem Item „Der Versuch hat mir Spaß gemacht.“ abgefragt. Die subjektive Bedeutung (Hold-Komponente) des Praktikumsversuches wird mit Items wie „Die Inhalte des Versuches erachte ich für mich persönlich als wichtig.“ erhoben. Die epistemische Orientierung (Hold-Komponente) geht mit Items wie „Ich bin während des Praktikumstages auf Themen gestoßen, zu denen ich gerne mehr Informationen hätte.“ der Frage nach, ob die Studierenden Inhalte des Versuches weiter verfolgen möchten. Die Items sind vollständig in Anhang A.2.4 aufgeführt.

Beide Komponenten des situationalen Interesses werden im originalen Fragebogen von Lewalter und Knogler (2014) mit je sechs Items erhoben. Die Items verteilen sich dabei mit jeweils drei Items zu gleichen Teilen auf die Subskalen. Dem Originalfragebogen wurde für diese Studie je ein Item für die Subskalen hinzugefügt, sodass jede Komponente des situationalen Interesses mit acht Items gemessen wurde. Zum einen wurde dies in Bezug auf die Pilotierung des Bogens und ggf. ausfallende Items gemacht. Da die Heterogenität der 24 Versuchssituationen in Bezug auf das situationale Interesse im Vorfeld nicht abgeschätzt werden konnte, soll die höhere Messwiederholung pro Messzeitpunkt (Praktikumsversuch) zum anderen eine höhere und über die Zeit stabilere Genauigkeit liefern. Die zusätzlichen Items wurden eng an den originalen Items formuliert. Um einer Testmüdigkeit durch das häufige Bewerten der Items entgegen zu wirken, wurden je ein negativ formuliertes Item pro Komponente wie beispielsweise „Der Versuch war langweilig.“ (Aufmerksamkeit, Catch-Komponente) ergänzt. Da die Items zum situationalen Interesse in den versuchsbegleitenden Testheften mit acht Items zur wahrgenommenen Relevanz

gemischt wurden, sollte die gleiche Anzahl Items pro Komponente des situationalen Interesses ebenfalls für eine gleiche Gewichtung der Konstrukte im Testheft sorgen.

Basierend auf den üblichen statistischen Verfahren (Abschnitt 8.2.1) wurde die Güte des Erhebungsinstrumentes untersucht. Im Wintersemester 2015/2016 wurde mit Bachelor-Studierenden der Fachrichtungen 1-Fach-Physik, 2-Fach-Physik und Materialwissenschaften im Vorgänger-Praktikum der angepasste Fragebogen pilotiert. Der Bezug der Items wurde für die Pilotierung statt auf den Versuch auf das Praktikum an sich bezogen. Die Skala zeigte in der Pilotierung ($n = 69$) gute Reliabilitäten mit Werten für Cronbachs α von $\alpha_{Catch} = .92$ für die Catch-Komponente und $\alpha_{Hold} = .79$ für die Hold-Komponente. Die Trennschärfen der Skala zum situationalen Interesse liegen in der Pilotierung für alle Items mit $.39 \leq \bar{r}_{i(t-i)} \leq .83$ in einem mittleren bis hohen Bereich. Die Ergebnisse decken sich aber mit denen aus verschiedenen Studien der Entwickler*innen der Skala, bei denen die Reliabilität der Hold-Komponente auch etwas geringer ist als für die Catch-Komponente (z. B. Lewalter & Geyer, 2009; Lewalter & Willems, 2009; Knogler et al., 2015). Vor dem Hintergrund dieser Untersuchung des Erhebungsinstrumentes hätten alle Items in der Skala unverändert beibehalten werden können. Nach der Pilotierung wurde dennoch die Formulierung einzelner Items der Hold-Komponente zum wahrgenommenen Wert ausgeschärft, um die persönliche Relevanz besser von der beruflichen Relevanz (Abschnitt 8.2.2) abzugrenzen.

Einen Überblick über die erhaltenen Kennwerte der Untersuchung dieser Studie gibt Tabelle 16. Auch für die verschiedenen Messsituationen (Praktikumsversuche) in dieser Arbeit liegen die Reliabilitäten der Komponenten des situationalen Interesses mit $.89 \leq \alpha_{Catch} \leq .96$ und $.77 \leq \alpha_{Hold} \leq .92$ im guten Bereich (Tab. 16). In Hinblick auf die Übersicht sind die α-Werte der einzelnen Versuche im Anhang A.2.4 zu finden. Die Reliabilität wurde auch auf dem Gesamtdatensatz gerechnet, auch wenn diese aufgrund der Abhängigkeit der Daten verzerrt sein kann. Auch für den Gesamtdatensatz ergeben sich gute Werte für Cronbachs α mit $\alpha_{Catch} = .93$ und $\alpha_{Hold} = .85$. Für die verschiedenen Praktikumsversuche der Studie finden sich mittlere bis hohe Trennschärfen im Wertebereich von $.38 < \bar{r}_{i(t-i)} < .81$ (Tab. 16; siehe auch Anhang A.2.4). Aufgrund der negativen Formulierung hat eines der Items eine niedrigere Itemtrennschärfe, die für den Gesamttest aber vertretbar im mittleren Bereich liegt ($\bar{r}_{i(t-i)} = .38$, Anhang A.2.4). Da die Skala insgesamt aber hohe Werte für die Reliabilität zeigt, wird dieses Item beibehalten.

Zusammen zeigt sich somit, dass der Fragebogen von Lewalter und Knogler (2014) in der modifizierten Form eine konsistente und trennscharfe Messung des situationalen Interesses in dieser Studie liefert.

	Subskala	M	SE	n	α	$\bar{r}_{i(t-i)}$
Gesamttest	Catch	3.50	0.02	1805	.93	.77
	Hold	3.30	0.02	1805	.85	.58
Wertebereich	Catch	3.07 - 3.80	0.08 - 0.12	65 - 89	.89 - .96	.73 - .81
für Versuche	Hold	3.11 - 3.55	0.06 - 0.11	65 - 89	.77 - .92	.38 - .70

Tabelle 16: *Deskriptive Werte für das situationale Interesse. Im oberen Teil der Tabelle sind die Werte für den Gesamttest angegeben. Im unteren Teil ist der Wertebereich des situationalen Interesses unter Berücksichtigung der Messzeitpunkte, sprich Praktikumsversuche (ohne Versuch Messtechnik 5), angegeben. Für die Deskription der Datensätze sind Mittelwert (M), Standardfehler (SE), Stichprobengröße (n), Cronbachs α und die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$) angegeben. In Hinblick auf die Übersicht sind weitere Details zu den Reliabilitäten der einzelnen Versuche im Anhang A.2.4 zu finden.*

11.1.2 Erhebungsinstrumente: Experimentelle Kompetenz

Ein zentrales Maß für den Lernerfolg im Praktikum und damit eine wesentliche Komponente der Untersuchung des Wirkmodells in Bezug auf Forschungsfrage 3 ist der Erwerb experimenteller Kompetenzen (siehe Abschnitt 3.1). Die Erfassung der experimentellen Kompetenz im Praktikum stellt eine Herausforderung dar, da der Prozess des Experimentierens komplex ist. Die Auswahl eines geeigneten Messverfahrens ist somit ein Abwägen von Vor- und Nachteilen. Daher muss im Folgenden die Erfassung der experimentellen Kompetenz im Rahmen dieser Studie diskutiert und bewertet werden. Zusätzlich wird in diesem Abschnitt ein Instrument zur Erhebung der Selbstwirksamkeitserwartungen in Bezug auf das Experimentieren im Unterricht betrachtet. Die Selbstwirksamkeitserwartungen werden sowohl als Hinweis auf die experimentelle Kompetenz aufgefasst, als auch als motivationaler Faktor und damit insgesamt als Lernerfolg.

Selbsteinschätzung experimentelle Kompetenz

Der Prozess des Experimentierens ist vielschichtig, sodass die experimentelle Kompetenz ebenfalls breit gefächert ist. Daher stellt die Messung experimenteller Kompetenzen im Praktikum eine Herausforderung dar und bedarf bei der Auswahl eines Instrumentes ein Abwägen von Vor- und Nachteilen. Daher soll im Folgenden vor der Analyse des Erhebungsinstrumentes kurz erläutert werden, wie experimentelle Kompetenz gemessen wird und wie sich die Auswahl des hier eingesetzten Messinstrumentes begründet.

Die experimentelle Kompetenz wird oft mit schriftlichen Wissenstests (z. B. Straube, 2016; Wellnitz et al., 2017) oder mit Experimentiertests an realen Experimenten (z. B. Heidrich, 2016) erfasst. Da in den Wissenstests nur über das Experimentieren nachgedacht wird, somit keine Experimente durchgeführt werden und dadurch die Erfassung von prozessbezogenen Aspekten des Experimentierens fehlt, stehen die Wissenstests in der Kritik nicht ausreichend valide zu sein (z. B. Ruiz-Primo &

Shavelson, 1996; Shavelson et al., 1999; Schreiber et al., 2014). Zur Berücksichtigung von handlungsbezogenen Aspekten und eine dadurch höhere Validität eignen sich Experimentaltests mit Realexperimenten. Diese haben allerdings den Nachteil, dass sie sehr aufwändig sind in Bezug auf die Bearbeitungszeit und das Experimentiermaterial (z. B. Ruiz-Primo & Shavelson, 1996). Auch die Auswertung kann bei prozessbezogenen Analysen von videographierten Tests nochmal deutlich umfangreicher sein, als bei produktbezogenen Analysen z. B. von schriftlichen Protokollen des Experimentiervorgangs. Zudem ist die Erfassung der experimentellen Kompetenz auch mit dem Experiment und der Aufgabenstellung konfundiert und dadurch weniger reliabel (Shavelson et al., 1993; Ruiz-Primo & Shavelson, 1996). Während es einige Experimentiertests für schulische Anwendungen gibt, ist die Erfassung von experimenteller Kompetenz anhand von Handlungsanalysen auf Hochschulniveau und insbesondere für Physikpraktika Thema aktueller Entwicklungen (z. B. Heidrich, 2016; A. Bauer et al., 2018).

Aufgrund des hohen Aufwandes von Experimentiertests mit realen Experimenten werden als Alternative auch Tests mit Simulationen von Experimenten durchgeführt, um den Aufwand zu verringern und die prozessbezogenen Aspekte trotzdem analysieren zu können. Da bisherige Untersuchungen zeigen, dass die Ergebnisse von Experimentiertests mit Realexperimenten und Simulationen nicht hinreichend übereinstimmen, ist die Validität für Tests mit Simulationen Gegenstand von Diskussionen (z. B. Schreiber, 2012; Shavelson et al., 1999). Auch Selbsteinschätzungen der experimentellen Kompetenz als zeitlich ökonomische Alternative kommen in Betracht, sind aber in Hinblick auf die Zuverlässigkeit diskussionswürdig (z. B. Rehfeldt, 2018a). Aus unterschiedlichen Studien folgt, dass eine Selbsteinschätzung einen Experimentiertest nicht ersetzen kann, aber dennoch zuverlässige Anhaltspunkte liefern kann (z. B. Schreiber & Theyßen, 2019; Schreiber et al., 2016; Stefani, 1994).

Diese Ausführungen zeigen, dass die Erhebung von experimentellen Kompetenzen keine triviale Aufgabe ist und jedes Instrument Vorteile, aber auch Nachteile bietet. In Bezug auf die Beforschung des Praktikums wurde entschieden, eine Selbsteinschätzung der Zunahme der experimentellen Kompetenz zur Abschätzung dieser Komponente des Lernerfolgs zu nutzen, da Zeit und Betreuungskapazität im Praktikum keine aufwendigen Tests zulassen. Zudem wurde der dafür genutzte Fragebogen zur Selbsteinschätzung der experimentellen Kompetenz von Rehfeldt (2018a) spezifisch für die Evaluation von Laborpraktika entwickelt. Des Weiteren wurde entschieden, Selbstwirksamkeitserwartungen als ein spezifisches Maß für den Lernerfolg im Lehramtspraktikum zu erfassen, um somit zusätzliche Hinweise auf die experimentellen Kompetenzen sowie motivationale Faktoren zu erhalten.

Im Folgenden soll zunächst das Erhebungsinstrument zur Selbsteinschätzung des Lernzuwachses im Bereich experimentelle Kompetenz nun beschrieben und bewertet werden. Das Instrument ist Teil eines umfassenderen Instrumentes zur Messung der Qualität von naturwissenschaftlichen Experimentalpraktika (kurz: PraQ-Test)

von Rehfeldt (2018a)[5]. Der PraQ-Test ist so konzipiert, dass die Studierenden am Praktikumsende rückblickend die Aussagen zu verschiedenen Aspekten der Praktikumsqualität auf einer 6-teiligen Likert-Skala bezüglich der Zustimmung bewerten. Daher wurde der Test in dieser Studie in der Abschlussbefragung am letzten Termin des Praktikumssemesters eingesetzt.

Für die Operationalisierung des Konstrukts experimentelle Kompetenz hat sich Rehfeldt (2018b) an der Arbeit von Schreiber (2012) orientiert (siehe auch Abschnitt 3.1). Die vielen Facetten der experimentellen Kompetenz werden in der Selbsteinschätzung mit neun Subskalen mit je drei bis sieben von Rehfeldt neu entwickelten Items abgebildet. Zu den verschiedenen Teilkompetenzen im Umgang mit den Experimenten im Praktikum zählen unter anderem die Gerätekenntnis, die Versuchsplanung, die Durchführung von Messungen sowie die Darstellung und Interpretation von Daten. Ein typisches Item der Subskala „Messungen durchführen“ lautet zum Beispiel „Aufgrund dieses Praktikums kann ich besser mit Messgeräten umgehen.“. Eine Übersicht zu den Subskalen und den zugehörigen Items ist in Anhang A.2.6 gegeben. Dieser Fragebogen ist im Gegensatz zu Experimentiertests damit unabhängig von den konkreten Versuchsinhalten und erfasst so den Zuwachs an experimenteller Kompetenz übergeordnet.

Da der PraQ-Test in der Originalformulierung eingesetzt wird und in dieser Form bereits umfangreich in verschiedenen Praktika evaluiert wurde (Rehfeldt, 2018a), wurde der Test im Vorfeld dieser Studie nicht zusätzlich pilotiert. Die anschließende Untersuchung der Güte des Tests basiert auf den in Abschnitt 8.2.1 dargestellten statistischen Verfahren. Ein Überblick über die Kennwerte der Untersuchung der Skala ist in Tabelle 17 gegeben (für Details siehe auch Anhang A.2.6). Die Werte für Cronbachs α als Maß für die Reliabilität sind für die Selbsteinschätzung auf der Gesamtskala zu beiden Messzeitpunkten hoch mit $\alpha_{Post1} = .97$ und $\alpha_{Post2} = .96$. Die Werte für die Subskalen liegen niedriger im Bereich von $.76 \leq \alpha \leq .93$ (Tab. 17). Die Subskalen weisen hohe Trennschärfen im Bereich $.48 \leq r_{i(t-i)} \leq .70$ auf. Damit misst das Instrument die experimentelle Kompetenz reliabel und trennscharf.

Auch wenn diese Selbsteinschätzung des Zuwachses an experimenteller Kompetenz somit ein reliables und trennscharfes Erhebungsinstrument ist, unterliegen Selbsteinschätzungen, wie zuvor bereits bemerkt, der Kritik, eine geringere Validität im Vergleich zu objektiven Leistungstests aufzuweisen. Dieser Aspekt soll zur umfassenden Bewertung des Erhebungsinstrumentes im Folgenden diskutiert werden.

In mehreren Metastudien (Zell & Krizan, 2014; G. Brown et al., 2013; Dochy et al., 1999; Falchikov & Boud, 1989) wurde untersucht, inwieweit Lernende den eigenen Wissensstand korrekt einschätzen können. Den Studien zufolge sind die Lernenden bedingt in der Lage, ihre Leistung akkurat einzuschätzen, da die Qualität der Selbsteinschätzung von einer Vielzahl an Faktoren beeinflusst wird. Demnach spielen neben psychologischen und situativen Faktoren wie zum Beispiel Selbstkonzept,

[5] Version des eingesetzten Tests: PraQ4A vom 18.05.2015 (Rehfeldt, Persönliche Kommunikation, Januar 2016)

Messzeitpunkt	Subskala	Deskription				Itemtrennschärfe		
	der exp. Kompetenz	M	SE	n	α	$\bar{r}_{i(t-i)}$	min	max
Post 1	Gesamtskala	4.64	0.07	82	.97	.63	.46	.78
	Erwartungen formulieren	4.67	0.08	87	.83	.63	.53	.76
	Versuchsplanung und Geräte zusammenstellen	4.45	0.08	88	.83	.62	.56	.67
	Gerätekenntnis (deklarativ)	4.62	0.09	88	.84	.70	.67	.74
	Versuchsanordnung aufbauen	4.50	0.08	87	.84	.69	.55	.75
	Messung durchführen	4.98	0.08	85	.87	.65	.48	.76
	Messungen dokumentieren	5.00	0.09	87	.80	.65	.55	.74
	mit Problemen und Fehlern umgehen	4.84	0.09	88	.87	.72	.69	.79
	Rohdaten graphisch / tabellarisch darstellen	4.49	0.11	86	.90	.74	.63	.86
	Daten interpretieren	4.60	0.10	88	.93	.79	.70	.89
Post 2	Gesamtskala	4.83	0.07	67	.96	.61	.50	.73
	Erwartungen formulieren	4.78	0.09	68	.83	.63	.53	.74
	Versuchsplanung und Geräte zusammenstellen	4.59	0.09	68	.76	.53	.48	.57
	Gerätekenntnis (deklarativ)	4.66	0.10	68	.78	.62	.59	.65
	Versuchsanordnung aufbauen	4.74	0.09	68	.83	.66	.56	.77
	Messung durchführen	5.09	0.08	68	.87	.65	.59	.74
	Messungen dokumentieren	5.16	0.08	68	.80	.65	.60	.73
	mit Problemen und Fehlern umgehen	5.02	0.09	68	.88	.75	.70	.83
	Rohdaten graphisch / tabellarisch darstellen	4.82	0.11	67	.89	.72	.57	.81
	Daten interpretieren	4.86	0.08	68	.85	.64	.57	.71

Tabelle 17: *Kennwerte für die Selbsteinschätzung des Zuwachses an experimenteller Kompetenz auf der Gesamtskala und den Subskalen. Dabei wird unterschieden zwischen dem Messzeitpunkt nach dem ersten (Post 1) und nach dem zweiten (Post 1) Praktikumssemester. Für die Deskription des Datensatzes sind Mittelwert (M, 6-teilige Likert-Skala), Standardfehler (SE) und Stichprobengröße (n) angegeben. Als Kenngrößen für die Analyse des Erhebungsinstrumentes ist neben der Reliabilität (Cronbachs α) auch die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$), sowie deren Extremwerte (min, max) angegeben.*

Lernerfahrung und die Konkretheit der Formulierung der Items, auch die herangezogenen Vergleichsmaßstäbe der Befragten eine Rolle bei der Selbsteinschätzung (z. B. Moritz, 1996; Higgins et al., 1986). Die befragten Personen setzen ihren Wissensstand in Relation zu einem maximal möglichen Wissen, das sich an der eigenen Lernentwicklung, Vorbildern oder sozialen Gruppen orientieren kann. Dadurch bewertet jeder der Probanden seinen Wissensstand auf einer eigenen Messskala, die untereinander nicht unbedingt vergleichbar sind. Bei der Befragung dieser Studie wird davon ausgegangen, dass die Studierenden ihre Bewertung im aktuellen Kontext Laborpraktikum vornehmen. Die hohen Werte der Selbsteinschätzung (Tab. 17) sprechen dafür, dass die Einschätzung stark am maximal erreichbaren Wissen durch die Lerngelegenheit Praktikum orientiert war. Für die Genauigkeit der Selbsteinschätzung experimenteller Kompetenz gibt es wenig Studien. Für Schüler*innen zeigt sich, dass die Fähigkeit zur Selbsteinschätzung der experimentellen Kompetenzen im Mittel vergleichsweise hoch ist, allerdings auf individueller Ebene unterschiedlich stark ausgeprägt ist (Schreiber et al., 2016; Schreiber & Theyßen, 2019). Für ein biologisches Anfängerpraktikum wurde eine hohe Korrelation zwischen der Selbstbeurteilung der Studierenden und der Fremdbeurteilung durch die Lehrenden in Hinblick auf die Versuchsprotokolle (ohne Berücksichtigung der experimentellen Handlungen) sowie eine Tendenz zur Unterschätzung der Leistungsträger gezeigt (Stefani, 1994).

Auch Rehfeldt (2018a) diskutiert den Vorteil der ökonomischen und zudem versuchsunabhängigen Anwendbarkeit der Selbsteinschätzung gegenüber dem Nachteil der geringeren Validität der Selbsteinschätzung im Vergleich zu objektiven Tests. Die Erfassung von Kompetenzen mittels Selbsteinschätzung wird kritisch in Bezug darauf gesehen, wieweit die Einschätzung mit der tatsächlichen Kompetenz zusammenhängt und weniger Bestandteile der Selbstwirksamkeitserwartung (z. B. Cramer, 2010) misst. Um die Kritik auszuräumen und Verzerrungen zu vermeiden, nehmen Rehfeldt (2018a) und Kratschmar (2017) in der Evaluation des PraQ-Tests einen Vergleich der Selbsteinschätzung mit einem etablierten Kompetenztest (Ko-WADiS-Test nach Straube, 2016) und einer Messung der Selbstwirksamkeitserwartung (Skala nach Schroedter & Körner, 2011) in Bezug auf experimentelle Kompetenz vor. Die Korrelationen der Subskalen der PraQ-Selbsteinschätzung mit der Selbstwirksamkeitserwartung fielen zum Teil gering bis mittelgroß aus ($r = .3$) und zum Teil nicht signifikant, was auch am Fokus des Tests zur Selbstwirksamkeitserwartung auf bestimmte Themenbereiche lag. Die Korrelation mit den Ergebnissen aus dem Kompetenztest (Ko-WADiS) waren größtenteils nicht signifikant, was darauf hindeuten kann, dass der PraQ-Test nicht den Zuwachs an experimenteller Kompetenz erfasst. Rehfeldt (2018a) schließt in seiner Diskussion hingegen, dass der Kompetenztest (Ko-WADiS) nicht im ausreichenden Umfang eingesetzt wurde, um die experimentelle Kompetenz belastbar schätzen zu können, und empfiehlt daher weitere Untersuchungen der Selbsteinschätzung des PraQ-Tests.

Alles in allem zeigt die Untersuchung und Diskussion der Erhebung des Zuwachses an experimenteller Kompetenz, dass die Selbsteinschätzung einen Experimentiertest zwar nicht ersetzen kann, aber für diese Studie zeitlich sehr ökonomisch zuverlässige Anhaltspunkte für die experimentelle Kompetenz liefert.

Selbstwirksamkeitserwartungen

Im vorangegangenen Abschnitt hat sich gezeigt, dass die Erhebung der experimentellen Kompetenz keine triviale Aufgabe ist. In dieser Studie wird davon ausgegangen, dass sich der Erwerb experimenteller Kompetenzen auch positiv auf die zugehörigen Selbstwirksamkeitserwartungen der Studierenden auswirken kann. Zudem werden Selbstwirksamkeitserwartungen als eine einflussreiche motivationale Größe beim kompetenten Handeln von Lehrkräften betrachtet (z. B. Tschannen-Moran et al., 1998; Baumert & Kunter, 2006; Klassen et al., 2011), wie später in diesem Abschnitt weiter ausgeführt wird. Daher wurden Selbstwirksamkeitserwartungen in Bezug auf Experimentieren in dieser Arbeit als ein spezifisches Maß für den Lernerfolg von Lehramtsstudierenden im Praktikum erfasst (Abb. 31), um zusätzliche Hinweise auf die experimentellen Kompetenzen aber auch motivationale Faktoren zu erhalten. Da die Selbstwirksamkeitserwartungen im Gegensatz zu den anderen Konstrukten bisher noch nicht beleuchtet wurden, sollen zu Beginn dieses Abschnittes zunächst ein paar Worte zum Konstrukt selbst und dessen Bedeutung für Lehrkräfte gesagt werden, bevor die Erfassung des Konstruktes bewertet wird.

Das Konstrukt der Selbstwirksamkeitserwartungen, oder auch (wahrgenommene) Selbstwirksamkeit, geht zurück auf die sozial-kognitive Theorie von Bandura (1977) (siehe auch Bandura, 2001), wonach die Motivation und die Handlung einer Person stark von der Selbstwahrnehmung abhängen. Selbstwirksamkeitserwartungen werden von Schwarzer und Jerusalem (2002) definiert als „die subjektive Gewissheit, neue oder schwierige Anforderungssituationen auf Grund eigener Kompetenz bewältigen zu können“. Aus dieser Definition geht zum einen die Kontextspezifität der Selbstwirksamkeitserwartungen hervor und zum anderen die Subjektivität der Wahrnehmung der eigenen Kompetenz. Demnach sind Selbstwirksamkeitserwartungen kein direktes Maß für tatsächliche Fähigkeiten. Vielmehr sind Selbstwirksamkeitserwartungen von Bedeutung, weil sie Einfluss darauf nehmen, ob eine schwierige Aufgabe überhaupt angegangen wird und mit welcher Anstrengung und Ausdauer die Handlungsziele dabei angestrebt werden (Schmitz & Schwarzer, 2000). Dadurch wird indirekt auch der Erfolg der Handlung durch die Selbstwirksamkeitserwartungen beeinflusst (Schmitz & Schwarzer, 2000).

Daraus kann geschlussfolgert werden, dass die Selbstwirksamkeitserwartungen auch eine wichtige Rolle für das erfolgreiche Handeln von Lehrkräften spielen (z. B. Baumert & Kunter, 2006). Überträgt man das Konzept der Selbstwirksamkeitserwartungen auf Lehrpersonen, so entsprechen die Selbstwirksamkeitserwartungen den Überzeugungen einer Lehrperson, wie gut es ihr gelingt, den Unterricht so zu

konzipieren und durchzuführen, dass die gesetzten Lehrziele in einem spezifischen Kontext erfolgreich erreicht werden (Tschannen-Moran et al., 1998). Tschannen-Moran et al. (1998) und Klassen et al. (2011) fassen in ihren Reviews zahlreiche Forschungsresultate zu Selbstwirksamkeitserwartungen von Lehrkräften zusammen. Die Ergebnisse weisen dabei unter anderem darauf hin, dass Lehrkräfte mit höheren Selbstwirksamkeitserwartungen größeren Enthusiasmus für den Unterricht aufweisen, Schüler*innen mit Verständnisschwierigkeiten mehr unterstützen und eher im Beruf verbleiben (vgl. Tschannen-Moran et al., 1998). Zudem investieren Lehrpersonen mit hohen Selbstwirksamkeitserwartungen mehr Zeit in zusätzlichen Unterricht und weisen eine geringere Berufsbelastung auf (Schmitz & Schwarzer, 2000). Darüber hinaus deuten Zusammenhänge darauf hin, dass die Selbstwirksamkeitserwartungen der Lehrenden einen positiven Einfluss auf die Motivation, die Selbstwirksamkeitserwartungen und den Wissenserwerb der Schüler*innen haben (vgl. Tschannen-Moran et al., 1998). Demnach stellen die Selbstwirksamkeitserwartungen eine einflussreiche Größe beim kompetenten Handeln von Lehrkräften dar. Daher werden in Modellierungen der professionellen Handlungskompetenz von Lehrkräften nicht nur kognitive Aspekte wie das fachliche und fachdidaktische Wissen als wichtige Komponenten des Kompetenzbegriffs verstanden, sondern auch motivationale Aspekte wie die Selbstwirksamkeitserwartungen einbezogen (Baumert & Kunter, 2006). Rabe et al. (2012) verorten die Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern sogar in der Schnittmenge zwischen dem Bereich des fachdidaktischen Professionswissens und den motivationalen Orientierungen.

Anhand der von Bandura (1977) postulierten Theorien zu Ursachen für Selbstwirksamkeitserwartungen und empirischen Forschungsergebnissen haben Tschannen-Moran et al. (1998) ein zyklischen Modell für die Entwicklung der Selbstwirksamkeitserwartungen von aktiven Lehrkräften entwickelt. Hauptquellen der Selbstwirksamkeitserwartungen sind demnach die eigenen Erfahrungen bei der Unterrichtsdurchführung. Aber auch verbale Rückmeldungen im Sinne von Feedback, Beobachtungen von Handlungen anderer gewonnener Erfahrungen und physiologische und emotionale Reaktionen haben einen Einfluss auf die Selbstwirksamkeitserwartungen. Die Entwicklung der Selbstwirksamkeitserwartungen von angehenden Lehrkräften während des Studiums ist ein eher jüngeres Forschungsfeld. Erste Ergebnisse zeigen, dass Praxisphasen wie Lehr-Lern-Labore, in denen die Lehramtsstudierenden erste eigene Unterrichtserfahrungen gewinnen, positive signifikante Effekte auf die Selbstwirksamkeitserwartungen im MINT-Bereich haben (Weß et al., 2020).

Die Selbstwirksamkeitserwartungen für Lehrpersonen werden im Idealfall noch weiter spezifiziert, da jede Domäne spezifische Aufgaben und Anforderungen zur Konzeption und Durchführung von Unterricht erfordert (z. B. Rabe et al., 2012). Für Physiklehrkräfte sind damit unter anderem die spezifischen Selbstwirksamkeitserwartungen in Bezug auf Planung und Durchführung von Experimenten im Unterricht in dieser Studie relevant, da das Lehramtspraktikum als erste Lerngelegenheit im Studium den Erwerb experimenteller Fähigkeiten als zentrales Ziel hat (siehe

Abschnitt 2.1). Damit ist das Lehramtspraktikum die erste Gelegenheit, experimentbezogene Selbstwirksamkeitserwartungen aufzubauen, um sich eigenständiges Experimentieren überhaupt erstmal zuzutrauen. Obwohl das Praktikum keine Praxisphase in dem Sinne ist, dass Schüler*innen in die Lehrveranstaltung eingebunden sind und erste Lehrerfahrungen gesammelt werden können, wird davon ausgegangen, dass sich ein Erwerb experimenteller Kompetenzen auch positiv auf die zugehörigen Selbstwirksamkeitserwartungen der Studierenden auswirken kann. Der Zusammenhang zwischen Selbstwirksamkeitserwartungen in Bezug auf Experimentieren und den tatsächlichen experimentellen Fähigkeiten wurde allerdings bisher kaum untersucht, da eine valide Erfassung von experimentellen Fähigkeiten bisher nur eingeschränkt möglich ist. Heidrich et al. (2015), die zur Validierung ihres entwickelten Experimentiertests auch die Skala zu Selbstwirksamkeitserwartungen von Rabe et al. (2012) eingesetzt haben, konnten keinen Zusammenhang der Größen feststellen, auch wenn man diesen theoretisch erwarten würde. Eine direkte Schätzung der tatsächlichen experimentellen Fähigkeiten ist somit aus den Selbstwirksamkeitserwartungen nicht unverzerrt möglich, zumal die Selbstwirksamkeitserwartungen durch weitere Faktoren bedingt sind. Daher können die Selbstwirksamkeitserwartungen nur als Hinweis auf die experimentelle Kompetenz und Motivation dienen. Die Selbstwirksamkeitserwartungen werden aber auch als eigenständiges Konstrukt als Lernerfolg aufgefasst. Hohe Selbstwirksamkeitserwartungen bzgl. des Experimentierens können als Teil der professionellen Handlungskompetenz unterstützen, den beobachteten mangelhaften Einsatz von Experimenten im Physikunterricht (z. B. Harlen, 1999; Tesch & Duit, 2004) entgegenzuwirken. Niedrige Selbstwirksamkeitserwartungen von Physiklehrkräften in Bezug auf das Experimentieren hingegen können im schlimmsten Fall dazu führen, dass das Demonstrieren von Experimenten und die Durchführung von Experimenten im Physikunterricht sogar ganz vermieden werden. Damit ist es umso wichtiger, dass die Lehramtsstudierenden nicht nur die experimentellen Kompetenzen, sondern auch das Zutrauen entwickeln, im Unterricht Experimente umzusetzen.

Nachdem nun deutlich geworden ist, was sich hinter dem Konstrukt Selbstwirksamkeitserwartungen verbirgt und welche Rolle es für Lehrkräfte spielt, soll darauf aufbauend das entsprechende Erhebungsinstrument zu Selbstwirksamkeitserwartungen der Lehramtsstudierenden in Bezug auf Experimentieren im Unterricht betrachtet werden. Die erfassten Selbstwirksamkeitserwartungen gliedern sich in zwei Subskalen zur Planung und Durchführung von Experimenten im Unterricht. Beide Komponenten wurden vor und nach dem Praktikumssemester als Prä-Post-Test erhoben. Als Erhebungsinstrument der Selbstwirksamkeitserwartungen wurde eine etablierte Skala verwendet, die von den Autoren Rabe et al. (2012) entwickelt, eingesetzt und untersucht wurde (siehe auch Meinhardt, 2018). Die Items wurden in dieser Studie in der Originalformulierung eingesetzt. Für jede Subskala haben die Studierenden jeweils acht Aussagen auf einer 5-teiligen Likert-Skala bewertet.

Ein Beispielitem für die Planung von Experimenten im Unterricht lautet: „Ich kann in meiner Unterrichtsvorbereitung ein Experiment planen, welches meine Schülerinnen und Schüler dazu anregt, physikalische Fragestellungen zu entwickeln, auch wenn ich dieses Experiment neu entwickeln muss.“. Für die Durchführung von Experimenten werden Items wie „Ich kann physikalische Experimente an interessante Impulse meiner Schülerinnen und Schüler anpassen, auch wenn ich das vorher nicht geplant hatte.“ in der Skala eingesetzt. Die Skala enthält keine negativ formulierten Items. Die Wortlaute weiterer Items sind in Anhang A.2.5 nachzulesen.

Im alten Anfängerpraktikum wurden im Wintersemester 2015/2016 die beiden Skalen mit Lehramtsstudierenden pilotiert. Daran nahmen allerdings nur neun Lehramtsstudierende teil, sodass die Pilotierung nur erste Hinweise liefern konnte. Für die Skala zu Selbstwirksamkeitserwartungen hat die Pilotierung gute Reliabilitäten für die Planung (Cronbachs $\alpha = .86$) und die Durchführung ($\alpha = .77$) von Experimenten im Unterricht gezeigt. Die Trennschärfe der Items ist in der Pilotierung in Einzelfällen sehr niedrig (Planung: $.31 \leq r_{i(t-i)} \leq .91$, Durchführung: $.05 \leq r_{i(t-i)} \leq .78$), was an der geringen Anzahl der befragten Studierenden ($n = 9$) liegen kann. Die Pilotierung hat aber vor allem anhand von mehreren Nachfragen verdeutlicht, dass es den Studierenden ohne Unterrichtserfahrung nicht sinnvoll erscheint und auch schwer fällt, die Items zu Selbstwirksamkeitserwartungen zu beantworten. Aus diesem Grund wurde der Einführungstext des Fragebogens zu Selbstwirksamkeitserwartungen nach der Pilotierung nochmal deutlich ausgeschärft, um den Blickwinkel auf die hypothetische Situation zu richten, dass die Studierenden morgen beginnen würden aktiv Physik zu unterrichten.

Die Reliabilität für diese Studie ist für alle vier Messzeitpunkte vor und nach dem ersten und zweiten Praktikumssemester (Prä 1, Post 1, Prä 2, Post 2) für beide Skalen gut (Tab. 18, siehe auch Anhang A.2.5). Cronbachs α für die Selbstwirksamkeitserwartungen in Bezug auf Planung von Experimenten liegt bei $.74 \leq \alpha \leq .84$ und für die Durchführung bei $.79 \leq \alpha \leq .86$. In dieser Studie liegen die mittleren Trennschärfen für die Planung im mittleren bis hohen Bereich $.43 \leq \bar{r}_{i(t-i)} \leq .57$ und für die Durchführung im hohen Bereich bei $.53 \leq \bar{r}_{i(t-i)} \leq .61$ (Tab. 18). Einzelne Items weisen eine eher geringe Trennschärfe auf (Anhang A.2.5). Dies zeigt sich vermutlich deshalb insbesondere für die Planung von Experimenten, da diese im Vergleich zur Durchführung von Experimenten im Lehramtspraktikum nicht Hauptlernziel waren, die Studierenden in diesem Bereich auch weniger Erfahrungen mitbringen und sich ggf. dadurch auch nicht immer in die Situation des Unterrichtens mit Experimenten hineinversetzen können. Auf Basis dieser Untersuchung würde ein Entfernen einzelner Items die Reliabilität aber nicht verbessern (Anhang A.2.5). Die Originalskala wurde zudem aus dem Grund beibehalten, um gegebenenfalls in Zukunft Vergleiche mit anderen Studien zu ermöglichen.

MZP	Deskription				Itemtrennschärfe		
	M	SE	n	α	$\bar{r}_{i(t-i)}$	min	max
Selbstwirksamkeit - Planung							
Prä 1	2.92	0.07	88	.84	.57	.44	.68
Post 1	3.47	0.05	88	.74	.43	.33	.56
Prä 2	3.42	0.06	67	.83	.55	.39	.64
Post 2	3.60	0.06	61	.78	.48	.40	.57
Selbstwirksamkeit - Durchführung							
Prä 1	3.20	0.07	85	.81	.53	.50	.57
Post 1	3.52	0.06	86	.84	.58	.50	.66
Prä 2	3.51	0.07	65	.86	.61	.51	.65
Post 2	3.69	0.06	64	.79	.50	.39	.62

Tabelle 18: *Kennwerte für die Selbstwirksamkeitserwartungen in Bezug auf die Planung (oben) und Durchführung (unten) von Experimenten im Schulunterricht. Dabei wird unterschieden zwischen dem Messzeitpunkt (MZP) vor und nach dem ersten (Prä 1, Post 1) und dem zweiten Praktikumssemester (Prä 2, Post 2). Für die Deskription des Datensatzes sind Mittelwert (M, 5-teilige Likert-Skala), Standardfehler (SE) und Stichprobengröße (n) angegeben. Als Kenngrößen für die Analyse des Erhebungsinstrumentes ist neben der Reliabilität (Cronbachs α) auch die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$), sowie deren Extremwerte (min, max) angegeben.*

Zusammengenommen zeigt diese Analyse der Skalen zur Selbstwirksamkeitserwartung in Bezug auf Experimentieren im Unterricht, dass die Messung durchaus konsistent und im mittleren Maß trennscharf ist. Damit werden die Selbstwirksamkeitserwartungen als lehramtsspezifischer Lernerfolg zuverlässig erfasst.

11.1.3 Erhebungsinstrumente: Fachwissen

Eines der zentralen Lernziele im Praktikum ist die Wiederholung und die Vertiefung von Fachwissen. Die Messung des physikalischen Fachwissenszuwachses als Lernerfolg ist damit eine wesentliche Komponente der Untersuchung des Wirkmodells in Bezug auf Forschungsfrage 3. Das Fachwissen wird dabei mithilfe eines gängigen Wissenstests und einer Selbsteinschätzung erfasst (Abb. 31). Während der Fachwissenstest physikalisches Wissen in einzelnen Themenbereichen misst, wird der Fachwissenszuwachs mithilfe der Selbsteinschätzung auf einer globalen Dimension erfasst. In diesem Abschnitt wird die Güte dieser eingesetzten Instrumente untersucht und diskutiert, wie diese Instrumente sich gegenseitig ergänzen.

Fachwissenstest

Als ein Maß für den Lernerfolg der Studierenden im Praktikum wurde das Fachwissen Physik vor und nach dem Praktikumssemester in den sechs Themenbereichen des

Praktikums (Messtechnik, Atomphysik, Optik, Wärmelehre, Mechanik und Elektrizitätslehre) erhoben. Für jeden der vier Versuche pro Themenbereich enthält das Testheft zwei Aufgaben. Somit umfasst der Test acht Aufgaben pro Themenbereich und 48 Aufgaben insgesamt. Da im Prä- und Post-Test die gleichen Fachwissensaufgaben eingesetzt wurden, kann ein Übungseffekt durch Vertrautheit der Aufgaben im Post-Test nicht ausgeschlossen werden. Um das Wiedererkennen der Testhefte und damit den Störeffekt zumindest zu reduzieren, wurde die Reihenfolge der zu bearbeitenden Aufgaben in den Testheften variiert. In den Fachwissensaufgaben werden grundlegende physikalische Themen abgefragt, die in den Praktikumsversuchen auf spezifische Weise behandelt werden. Zu den Fachwissensaufgaben zum Themenbereich Messtechnik sei der Hinweis gegeben, dass diese keine messtechnischen Aspekte abfragen, sondern sich an den fachwissenschaftlichen Inhalten in den physikalischen Themenbereichen Mechanik und Elektrizitätslehre orientieren, anhand denen in den Versuchen messtechnischer Inhalt vermittelt wird.

Mehr als die Hälfte der Aufgaben wurden aus etablierten Aufgabensammlungen ausgewählt. Die Nutzung bewährter Aufgaben bietet den Vorteil, dass diese bereits umfassend eingesetzt und untersucht wurden, sodass von einem validen Instrument ausgegangen werden kann. Die Hauptquelle der Aufgaben ist die Aufgabensammlung des Projektes KiL (Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen) (Kröger, 2019), die umfassend die Fachgebiete Mechanik, Elektrizitätslehre, Optik, Thermodynamik, Festkörperphysik, Atom- und Kernphysik, Relativitätstheorie und Quantenmechanik abdeckt. Die KiL Aufgabensammlung beruht neben Neuentwicklungen ebenfalls auf vorhandenen Aufgabensammlungen. Dazu zählen unter anderem auch die Tests *Force Concept Inventory* (FCI) und *Brief Electricity and Magnetism Assessment* (BEMA), die ebenfalls für die Zusammenstellung des Fachwissenstests dieser Studie ergänzend herangezogen wurden. Der FCI ist ein bekannter Test, der Kenntnisse und Anwendungen des Kraftkonzepts abfragt (Hestenes et al., 1992). BEMA enthält Fragen zu den Themen Elektrizitätslehre, Magnetismus und deren Wechselwirkungen (Ding et al., 2006). Für die Aufgabenauswahl in diesem Themengebiet wurde zudem der Test *Conceptual Survey of Electricity and Magnetism* (CSEM) hinzugezogen (Maloney et al., 2001). Für den Themenbereich Mechanik wurde zusätzlich der Test of Understanding Graphs-Kinematics (TUG-K) zum Verständnis von Graphen zu Bewegungsabläufen betrachtet (Beichner, 1994). Im Bereich Optik wurde mehrfach auf den Test *Basic Wave Optics Survey* (BWOS) zurückgegriffen, den Mešić et al. (2016) für Studien zum Unterricht von Wellenoptik entwickelt haben. Zwei der Fachwissensaufgaben sind aus der Aufgabensammlung des Projekts *The German National Educational Panel Study* (NEPS) (Blossfeld, 2019). Die Aufgaben sind im digitalen Anhang B mit Angabe der jeweiligen Quelle dokumentiert.

Für die Aufgabenauswahl wurde die Passung der Aufgaben auf die Praktikumsversuche in einer Expertenrunde diskutiert. Wenn keine Aufgaben für bestimmte Inhaltsbereiche der Praktikumsversuche zur Verfügung standen, wurden Eigenent-

wicklungen ergänzt, um die Abdeckung aller Inhaltsbereiche zu erreichen. Insgesamt wurden 21 der 48 Aufgaben neu entwickelt, da insbesondere in den Themenbereichen Atomphysik und Wärmelehre die Auswahl an Fachwissensaufgaben gering und wenig passend war. Die neuen Aufgaben wurden in der Arbeitsgruppe auf ihre Verständlichkeit getestet. Diese Eigenentwicklungen wurden in ihrer Konzeption den Aufgaben der anderen Tests nachempfunden, sodass sie ein vergleichbares Format haben. Das Fachwissen Physik wurde mit Aufgaben mit gebundenem Aufgabenformat erhoben. Dabei wurden hauptsächlich Entscheidungsfragen mit dichotomem Antwortformat (z. B. „Ist folgende Aussage richtig oder falsch?“) und Auswahlaufgaben mit mehr als zwei Antwortmöglichkeiten (Multiple-Choice) genutzt. Die Auswahlaufgaben haben dabei genau eine korrekte Lösung. Statt einer „Weiß nicht“-Option wurde mündlich der Hinweis gegeben, keine Antwort auszuwählen, wenn die Antwort nicht beantwortet werden kann.

Im Fachwissenstest wurde eine korrekte Antwort mit 1 kodiert und eine falsche Antwort mit 0. Im Fall von Entscheidungsfragen beinhaltet eine Fachwissensaufgabe mehrere Aussagen, die auf Korrektheit geprüft werden sollen. Für Aufgaben dieser Art wird ein Mittelwert über die einzelnen Items gebildet. Dies soll einer verzerrenden Gewichtung auf bestimmte fachliche Inhalte im Gesamttestwert entgegenwirken. Gleiches Vorgehen gilt für Multiple-Choice-Aufgaben, die aus mehreren Teilaufgaben bestehen. Somit kann jede Aufgabe einen Testwert zwischen 0 und 1 erhalten, der dem Anteil der korrekten Antwort für die Aufgabe entspricht. Ein Testwert von 0.5 bedeutet folglich, dass die einzelnen Teilitems der Aufgabe zur Hälfte korrekt beantwortet wurden.

Die nachfolgende Analyse der Testgüte basiert auf den statistischen Verfahren aus Abschnitt 8.2.1. Tabelle 19 gibt einen Überblick über die Kennwerte der Untersuchung des Fachwissenstests dieser Studie. Detailliertere Ergebnisse der Analysen sind im Anhang A.2.7 zu finden. Im Test wurde das physikalische Fachwissen mit 48 Aufgaben zu sechs Themenbereichen bzw. zu 24 Versuchsinhalten breit abgefragt. Nach dem von Gramzow (2015) beschriebenen Reliabilitäts-Validitäts-Dilemma kann eine breite Erfassung von physikalischen Inhalten mit einer höheren Validität des Fachwissenstests einhergehen, aber auch mit einer niedrigeren inneren Konsistenz des Konstrukts. Für den Gesamttest an den beiden Messzeitpunkten Prä und Post der Studie liegen die Reliabilitäten der Inhalte mit Cronbachs $\alpha_{Pre} = .75$ und $\alpha_{Post} = .74$ jedoch im guten Bereich. Eine differenziertere Analyse der Aufgaben zeigt, dass ein Entfernen von Aufgaben die Reliabilität nicht wesentlich verbessern würde (Anhang A.2.7). Trotz der inhaltlichen Breite zeigt der Gesamttest für das Fachwissen somit eine zufriedenstellende Reliabilität, sodass der Fachwissenstest als Ganzes als belastbar anzusehen ist.

Die Reliabilitäten für die sechs Themenbereiche liegen im Prä-Test im Bereich von $.22 \leq \alpha_{Pre,Thema} \leq .53$ und im Post-Test im Bereich von $.14 \leq \alpha_{Post,Thema} \leq .49$ (Anhang A.2.7) und damit unter dem angenommenen Grenzwert von $\alpha_G = .7$ (siehe Abschnitt 8.2.1). Aufgrund der Kombination aus begrenzter Testlänge durch die

Messzeitpunkt	**Deskription**				**Itemtrennschärfe**			**Lösungsrate**			**Bearbeitungsgrad**		
	M	SE	n	α	$\bar{r}_{i(t-i)}$	min	max	$\bar{p}_i$	min	max	$\bar{bg}_i$	min	max
Prä	.46	.01	88	.75	.22	-.17	.57	.52	.13	.87	.71	.24	.98
Post	.59	.01	88	.74	.21	-.18	.60	.63	.06	.97	.77	.41	.99

Tabelle 19: *Kennwerte für den Fachwissenstest. Dabei wird unterschieden zwischen dem Messzeitpunkt vor (Prä) und nach (Post) dem Praktikumssemester. Für die Deskription des Datensatzes sind Mittelwert (M) in anteiligem Testwert, Standardfehler (SE) und Stichprobengröße (n) angegeben. Als Kenngrößen für die Analyse des Testinstrumentes ist neben der Reliabilität (Cronbachs α) auch die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$), die mittlere Lösungswahrscheinlichkeit ($\bar{p}_i$) und der mittlere Bearbeitungsgrad ($\bar{bg}_i$) der Items sowie deren Extremwerte (min, max) angegeben.*

Einsatzbedingungen des Tests im Praktikum und inhaltlicher Breite der untersuchten Praktikumsversuche, ist diese geringe Reliabilität auf Ebene der Subskalen nicht unerwartet. Zum einen liegt der Grund für die geringen Werte für Cronbachs α in der geringen Anzahl von nur acht Items auf den Subskalen. Kröger (2019) erhält im Projekt KiL für die Subskalen seines Testinstruments mit einer vergleichbaren Anzahl von Items ebenfalls Cronbachs $\alpha < .6$ und zeigt, dass eine Erhöhung der Itemanzahl in den Subskalen auf 20 bei gleicher inhaltlicher Breite zu akzeptablen Reliabilitäten führen würde. Zum anderen werden die einzelnen Themenbereiche inhaltlich breit erfasst, was nach dem Reliabilitäts-Validitäts-Dilemma mit einer geringeren Reliabilität einhergehen kann (Gramzow, 2015). Die Fachwissensaufgaben dieser Studie wurden bewusst so gewählt, dass sie abgestimmt auf die Praktikumsversuche mehrere Inhaltsbereiche eines Themenbereiches abdecken. Obwohl beispielsweise die Newtonschen Gesetze und der Doppler-Effekt beide in den Themenbereich Mechanik fallen, bedingt das Wissen über einen der Inhaltsbereiche nicht zwangsläufig das des anderen. Verschiedene Autoren sprechen sich mit Blick auf ebensolche Aspekte dafür aus, die allgemein angenommenen Grenzwerte für die Reliabilität mit Hinsicht auf die Bedingungen und Ziele des Tests weniger streng zu beurteilen (Moosbrugger & Kelava, 2007; Schecker, 2014). Die Subskalen des Fachwissens mit den zum Teil sehr niedrigen Werten für Cronbachs α können dennoch nicht als sicher belastbar angenommen werden, weshalb auf eine tieferführende Analyse des Fachwissens auf Ebene der Themenbereiche verzichtet wird.

Auf Ebene der Praktikumsversuche ist aufgrund der geringen Anzahl von zum Teil nur zwei Items pro Praktikumsversuch keine reliable Messung zu erwarten. Dies bestätigen die Werte für Cronbachs α bei niedrigen mittleren Werten von $\bar{\alpha}_{Pre,VS} = .20$ im Prä-Test und $\bar{\alpha}_{Post,VS} = .12$ im Post-Test und sehr breiten Wertebereichen von $-.19 \leq \alpha_{Pre,VS} \leq .65$ und $-.26 \leq \alpha_{Post,VS} \leq .71$. Die Messung des Fachwissens auf Ebene der Praktikumsversuche kann damit nicht als belastbar angenommen werden. Um Fehlinterpretationen für die Versuche aufgrund der geringen und heterogenen Zuverlässigkeit zu vermeiden, wird das Fachwissen auf Ebene der Praktikumsversuche ebenfalls nicht weiterführend analysiert.

Als weiteres Kriterium zur Einschätzung der Güte des Fachwissenstests wurden die Itemtrennschärfen bestimmt. Die Trennschärfen liegen für die Aufgaben im Gesamttest zum Messzeitpunkt Prä im Bereich von $-.17 \leq r_{i(t-i)} \leq .57$ und zum Messzeitpunkt Post im Bereich von $-.18 \leq r_{i(t-i)} \leq .60$ (Tab. 19, Details siehe Anhang A.2.7). Für den Prä-Test haben dabei 29 von 48 Aufgaben eine Trennschärfe unter dem allgemein angenommenen Grenzwert von $r_{i(t-i)} < .3$ (Döring & Bortz, 2016) und sieben dieser Aufgaben haben eine negative Trennschärfe. Im Post-Test sind es 34 von 48 Aufgaben mit einer Trennschärfe unter dem Grenzwert und drei Aufgaben mit einer negativen Trennschärfe. Negative Trennschärfen von Fachwissensitems sind erstes ein Zeichen dafür, dass Studierende mit einem hohen Gesamttestwert diese Aufgaben schlechter lösen als Studierende mit einem niedrigen Gesamttestwert. In

diesem Fall sind die negativen Trennschärfen vom Betrag jedoch alle $|r_{i(t-i)}| < .30$ und damit nahe Null. Das bedeutet, dass die Aufgaben von Studierenden mit hohen und niedrigen Gesamttestwert etwa gleich oft korrekt beantwortet wurden, sodass anhand dieser Aufgabe die Studierenden mit hohem und niedrigen Fachwissen nicht unterschieden werden können. Die Fachwissensaufgaben mit geringer Trennschärfe sind dabei zu etwa gleichen Teilen neue und bewährte Aufgaben. Damit sind die neu entwickelten Aufgaben zum Beispiel aufgrund schlechter Formulierungen als Ursache für die geringen Trennschärfen unwahrscheinlich. Grund für die geringen Itemtrennschärfen ist bei diesem Instrument vor allem der breite Querschnitt der abgefragten fachlichen Inhaltsbereiche. Auch Kröger (2019) erhält aufgrund dessen für die Fachwissensaufgaben im Projekt KiL geringe Trennschärfen. Wie oben bereits für die Reliabilität diskutiert wurde, beziehen sich die zusammengestellten Aufgaben auf disjunkte Inhaltsbereiche innerhalb des physikalischen Fachwissens, sodass das Konstrukt nicht als homogen angenommen werden kann. Durch ein Entfernen der Aufgaben mit geringer Trennschärfe würde die gesamte Breite der fachlichen Inhalte der 24 Praktikumsversuche, die für die Untersuchung der angenommenen Wirkzusammenhänge in dieser Studie essentiell ist, verloren gehen. Dabei kann nicht ausgeschlossen werden, dass die in dieser Arbeit untersuchte unterschiedliche Wahrnehmung der Praktikumsversuche sogar zu geringeren Trennschärfen beiträgt. Döring und Bortz (2016) weisen darauf hin, dass insbesondere bei breit abgefragten Konstrukten die Bewertungskriterien für die Itemtrennschärfe nicht mechanisch anwendbar sind und man geringe Trennschärfe zugunsten der größeren inhaltlichen Breite der Items zulassen kann. Aus diesem Grund wurde entschieden, die Fachwissensaufgaben vollständig beizubehalten und die Güte des Testinstrumentes in der Bewertung der Ergebnisse in den abschießenden Diskussionen zu berücksichtigen.

Auch wenn die Trennschärfe zugunsten der Breite des Tests nicht als kritisches Itemmerkmal aufgefasst wird, soll dennoch geprüft werden, ob andere ungewollte Ursachen für die geringen Trennschärfen in Frage kommen. Auch der Schwierigkeitsgrad der Aufgaben spielt für die Güte eines Wissenstest eine Rolle. Mäßige Trennschärfen können ebenfalls aus sehr hohen oder sehr niedrigen Schwierigkeiten der Aufgaben resultieren (Döring & Bortz, 2016). Daher wird ergänzend die Lösungswahrscheinlichkeit betrachtet, die als Maß für die Schwierigkeit der Items interpretiert werden kann (Döring & Bortz, 2016). Die Lösungswahrscheinlichkeit ergibt sich als Anteil aus dem Verhältnis der Anzahl korrekt beantworteter Items zur Gesamtanzahl der Antworten. Der Grad der Bearbeitung der Items kann ebenso einen Ausdruck des Schwierigkeitsgrades der Aufgaben darstellen. Der Bearbeitungsgrad der Items wurde als Quotient aus der Anzahl beantworteter Items und der Gesamtanzahl der möglichen Antworten gebildet. Die große Bandbreite der Lösungswahrscheinlichkeit von $.13 \leq p_i < .87$ im Prä-Test (Tab. 19, Details siehe Anhang A.2.7) zeigt, dass Aufgaben unterschiedlicher Schwierigkeit im Test enthalten sind, wodurch Studierende mit hohem und niedrigem Fachwissen besser differenziert werden können. Die Extremwerte der Lösungswahrscheinlichkeit weisen allerdings

auch darauf hin, dass es Aufgaben gibt, die sehr schwer ($p_i \geq .80$) und sehr leicht ($p_i \leq .20$) für die Zielgruppe sind (Döring & Bortz, 2016). Der Wertebereich des Bearbeitungsgrades als indirektes Maß für die Schwierigkeit ist im Prä-Test ebenfalls breit mit Werten von $.24 \leq bg_i \leq .98$. Die mittlere Lösungswahrscheinlichkeit liegt im Prä-Test bei $\bar{p}_i = .52 \pm .03$ (Tab. 19). Die ideale mittlere Lösungswahrscheinlichkeit für Leistungstests beträgt $\bar{p}_i = .50$ (vgl. z. B. Kröger, 2019). Der hier genutzte Test ist folglich gut auf die Zielgruppe abgestimmt. Die Aufgabenschwierigkeit stellt somit kein Problem für die Testgüte dar.

Geringe Trennschärfen können auch durch mangelnde inhaltliche Passung hervorgerufen werden (Moosbrugger & Kelava, 2007). Um dies zu prüfen, wurden die Fachwissensaufgaben rein inhaltlich begründet an den Lerninhalten der Praktikumsversuche in „versuchsnahe“ und „versuchsferne“ Aufgaben kategorisiert (siehe Anhang A.2.7) und der „versuchsnahe Datensatz“ als reduzierter Teildatensatz eingehender betrachtet. Der versuchsnahe Datensatz ist auch in späteren Analysen der Wirkannahmen bzgl. des Fachwissens relevant. Versuchsfern bedeutet dabei nicht, dass die Aufgabe nichts mit dem Versuchsinhalt zu tun hat, sondern dass sich die Studierenden über den Lerninhalt der Praktikumsversuche hinaus beschäftigen müssen, um ein umfassenderes Verständnis für das Thema des Versuches zu gewinnen und damit die Aufgaben korrekt beantworten zu können. Die Aufgaben zum Praktikumsversuch „Wasserstromkreis“ im Themenbereich Elektrizitätslehre seien hier als Beispiel angeführt. In diesem Praktikumsversuch geht es um die Untersuchung der Möglichkeiten und Grenzen der Analogie des Wasserstromkreises anhand eines Experimentes (siehe auch Anhang A.1.23). In der ersten Aufgabe (Aufgabe FW_EL1_1 im digitalen Anhang B) zu diesem Versuch geht es um elektrischen Stromfluss aufgrund von Natrium- und Chloridionen in einem dünnen wassergefüllten Rohr, das an eine Batterie angeschlossen ist. Im Praktikumsversuch geht es zwar auch um elektrischen Strom, es wird aber kein Phänomen von elektrischem Stromfluss in Wasser durch Ionen untersucht oder erörtert. Da zum korrekten Lösen der Aufgabe folglich Wissen über den Versuch hinaus benötigt wird, wird diese Aufgabe als versuchsfern eingestuft. In der zweiten Aufgabe (Aufgabe FW_EL1_2 im digitalen Anhang B) geht es um die Parallel- und Reihenschaltung von elektrischen Widerständen sowie um die Einflussgrößen auf einen Drahtwiderstand. Im Praktikumsversuch werden diese Zusammenhänge experimentell für Strömungswiderstände untersucht und explizit mit den Zusammenhängen für elektrische Widerstände verglichen. Damit wurde diese Aufgabe als versuchsnah eingeordnet. Aus den insgesamt 48 Aufgaben wurden 18 als versuchsfern eingestuft. Um den Schwierigkeitsgrad wie oben diskutiert ebenfalls zu berücksichtigen, wurden vier weitere sehr leichte und sehr schwere Aufgaben mit einer Änderung der Lösungswahrscheinlichkeit vom Prä- zum Post-Test im Betrag $|p_i| < .05$ ebenfalls aus dem versuchsnahen Datensatz entfernt, sodass der reduzierte Datensatz aus 26 Aufgaben besteht. Durch die Reduktion werden einzelne Praktikumsversuche im Datensatz nicht berücksichtigt, weil diese nicht gut durch Aufgaben abbildbar sind. Da dies nur für sieben der 24 Praktikumsversuche zutrifft,

behält der Datensatz damit nach wie vor eine gewisse inhaltliche Breite, um als Referenz zum vollständigen Datensatz zu dienen. Eine Analyse zeigt für den versuchsnahen Datensatz an den beiden Messzeitpunkte Prä und Post Reliabilitäten von $\alpha_{Pre,VS-nah} = .63$ und $\alpha_{Post,VS-nah} = .61$ (Anhang A.2.7). Die Reduktion der Anzahl der Aufgaben des Fachwissenstests führt zu einer Verringerung von Cronbachs α im Bereich mäßiger Reliabilität. Die Trennschärfen liegen in ähnlichen Bereichen wie für den Gesamtdatensatz mit $-.23 \leq r_{i(t-i)} \leq .55$ für den Prä-Test und $-.17 \leq r_{i(t-i)} \leq .56$ für den Post-Test (Anhang A.2.7). Die Trennschärfen sind damit für diese Teilauswahl an Fachwissensaufgaben nach wie vor eher gering. Diese Analysen des Fachwissenstests zeigen, dass die Hauptursachen für die geringen Trennschärfen nicht naheliegende Faktoren wie extreme Schwierigkeitsgrade und geringe Passung der Aufgaben sind, sondern vor allem die gewünschte Heterogenität des Tests.

Somit liefert der Wissenstest auf Ebene des Gesamttests eine valide und belastbare Messung des Fachwissens Physik, die zudem vom Schwierigkeitsgrad gut auf die Zielgruppe abgestimmt ist. Der eingesetzte Fachwissenstest zeichnet sich besonders durch die breite Abdeckung physikalischer Inhalte aus.

Selbsteinschätzung Fachwissen

Wie die vorangegangene Analyse gezeigt hat, ist der Wissenstest für das Fachwissen Physik insgesamt reliabel, hingegen nicht für das Fachwissen in den Themenbereichen wie Mechanik und in den einzelnen Praktikumsversuchen aufgrund der Heterogenität der fachlichen Inhalte und der verhältnismäßig wenigen Aufgaben. Um dieser begrenzten Analysierbarkeit zu begegnen, wurde beim Design der Studie entschieden, zusätzlich eine kurze Abfrage des Fachwissenszuwachses auf einer übergeordneten Ebene vorzunehmen. Dazu wurde für die Untersuchung des Wirkmodells der Zuwachs an Fachwissen über das Praktikum zusätzlich zum Wissenstest in Form einer Selbsteinschätzung der Studierenden erhoben.

Die Skala zur Selbsteinschätzung des Zuwachses an Fachwissen ist wie die Selbsteinschätzung der Zunahme der experimentellen Kompetenz (Abschnitt 11.1.2) Teil des PraQ-Tests von Rehfeldt (2018a). Die Skala von Braun und Gusy (2006) bildet im PraQ-Test die Basis für die Messung des Fachwissenszuwachses. Da die Skala zum Fachwissen im PraQ-Test so konzipiert ist, den Zustand am Praktikumsende abzufragen, wurde das Instrument in dieser Studie in der Abschlussbefragung am letzten Termin des Praktikumssemesters eingesetzt. Die Studierenden wurden dabei im Fragebogen gebeten, rückblickend auf das Praktikum die Aussagen zur Praktikumsqualität auf einer 6-teiligen Likert-Skala bezüglich der Zustimmung zu bewerten.

Die eingesetzte Skala fragt den Zuwachs an Fachwissen auf zwei Subskalen mit je vier Items ab[6]. Eine Übersicht zu den Items ist in Anhang A.2.6 gegeben. Ein typisches Item der ersten Subskala ist dabei „Aufgrund dieses Praktikums kann ich wichtige Theorien zu dem behandelten Thema wiedergeben.“. Während dieses Item zu den behandelten theoretischen Inhalten einen deklarativen Charakter hat, wird in der zweiten Subskala das vernetzende Element zwischen Theorie und Praxis aufgenommen. Ein Beispielitem dazu lautet: „Aufgrund dieses Praktikums kann ich Bezüge zwischen Theorie und Experiment herstellen.“. Die Abfrage des Fachwissenszuwachs ist damit zeitlich sehr ökonomisch einzusetzen und ist nicht an die physikalischen Inhalte gebunden.

Da der PraQ-Test von Rehfeldt (2018a) umfangreich und dabei unter anderem auch im Anfängerpraktikum Physik der Christian-Albrechts-Universität zu Kiel (CAU) evaluiert wurde, wurde der Test nicht pilotiert. Tabelle 20 gibt einen Überblick über die Kennwerte des Instrumentes. Eine Übersicht über die Items und detailliertere Ergebnisse der Analysen sind im Anhang A.2.6 zu finden. Die Selbsteinschätzung zeigt gute Reliabilitäten zu beiden Messzeitpunkten mit Werten für Cronbachs α auf der Gesamtskala von $\alpha_{Post1} = .87$ und $\alpha_{Post2} = .86$ (Tab. 20). Die Werte für die Subskalen sind (auch aufgrund der niedrigeren Itemanzahl) etwas geringer und liegen im Bereich von $.76 \leq \alpha \leq .85$. Die Trennschärfen in dieser Studie sind mit $.50 \leq r_{i(t-i)} \leq .75$ hoch. Eine differenziertere Analyse der Items zeigt, dass ein Entfernen von Items die Reliabilität nicht wesentlich verbessern würde (Anhang A.2.6).

Messzeitpunkt	**Subskala**	**Deskription**				**Itemtrennschärfe**		
	des Fachwissens	M	SE	n	α	$\bar{r}_{i(t-i)}$	min	max
Post 1	Gesamtskala	4.90	0.06	87	.87	.63	.50	.74
	Theorie	4.70	0.07	88	.85	.69	.65	.75
	Praxis	5.09	0.07	87	.76	.56	.49	.59
Post 2	Gesamtskala	4.93	0.07	68	.86	.62	.50	.75
	Theorie	4.73	0.08	68	.83	.65	.61	.73
	Praxis	5.14	0.07	68	.77	.57	.49	.71

Tabelle 20: *Kennwerte für die Selbsteinschätzung des Fachwissenszuwachses auf der Gesamtskala und den Subskalen. Dabei wird unterschieden zwischen dem Messzeitpunkt nach dem ersten (Post 1) und nach dem zweiten (Post 2) Praktikumssemester. Für die Deskription des Datensatzes sind Mittelwert (M, 6-teilige Likert-Skala), Standardfehler (SE) und Stichprobengröße (n) angegeben. Als Kenngrößen für die Analyse des Erhebungsinstrumentes ist neben der Reliabilität (Cronbachs α) auch die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$), sowie deren Extremwerte (min, max) angegeben.*

[6] Version der eingesetzten Skala: PraQ4A vom 18.05.2015 (Rehfeldt, Persönliche Kommunikation, Januar 2016)

Insgesamt liefert die Selbsteinschätzung des Fachwissenszuwachses damit ein zeitlich ökonomisches sowie reliables und trennscharfes Messinstrument. Die Besonderheit ist, dass der Fachwissenszuwachs unabhängig von den physikalischen Inhalten abgefragt wird. Damit kann die Selbsteinschätzung als direkte Ergänzung zum zuvor beschriebenen Fachwissenstest aufgefasst werden.

Wissenstest und Selbsteinschätzung im Vergleich

Vorangehend wurden die Messinstrumente für das Fachwissen zum einen mittels eines Prä- und Post-Wissenstests und zum anderen mittels einer Selbsteinschätzung des Zuwachses vorgestellt und analysiert. Beide Instrumente haben unterschiedliche Qualitäten in Bezug auf die Art der Abfrage (objektiver Test versus subjektive Selbsteinschätzung) aber auch hinsichtlich der Spezifität des Bezuges des erfragten Fachwissens, sodass sich beide Tests in dieser Studie ergänzen. Während mittels des Wissenstests punktuell und begrenzt spezifisches Fachwissens über Lerninhalte der Versuche abgefragt wird, ermittelt die Selbsteinschätzung den Zuwachs des Fachwissens auf einer übergeordneten Ebene allgemein für alle Lerninhalte des Praktikums, die zum Beispiel auch die inhaltliche Breite der Versuchsvorbereitung umfasst.

Der tatsächliche Zuwachs an Fachwissen liegt irgendwo zwischen den beiden erhobenen Maßen. Dabei kommt der Fachwissenszuwachs, der sich aus der Differenz der Prä- und Post-Testwerte des Wissenstests für die Probanden berechnet, dem tatsächlichen Fachwissenszuwachs näher. Auch wenn mit einem Wissenstest nie der gesamte Umfang an fachlichen Inhalten im Rahmen des Praktikums abgefragt werden kann, wurde für jedes der insgesamt 24 verschiedenen physikalischen Themen (z. B. Newtonsche Axiome) zwei Aufgaben eingesetzt, um das Wissen spezifisch und auf großer Breite direkt zu messen. Trotz der diskutierten Heterogenität und der geringen Konsistenz des Tests auf Subskalen wie Fachwissen Mechanik, zeigt sich der Test als Gesamttest für das Fachwissen Physik reliabel.

Idealerweise müsste ein Wissenstest, der diese Breite an Themen abfragt, eine größere Anzahl an Aufgaben pro Thema haben, um auch auf den Subskalen reliabel zu sein. Da insbesondere Fachwissenstests eine hohe Testzeit haben, wäre eine Umsetzung eines deutlich umfangreicheren Fachwissenstest im Rahmen des Praktikums nicht möglich gewesen. Da die begrenzte Analysierbarkeit des Fachwissenstests bereits beim Design der Studie klar war, wurde zusätzlich eine ökonomische Abfrage des Fachwissenszuwachses auf einer übergeordneten Ebene mittels einer Selbsteinschätzung vorgenommen.

Allerdings birgt die Selbsteinschätzung von Wissen die Gefahr der Über- und Unterschätzung des tatsächlichen Wissens durch die Befragten. Diverse Metastudien (G. Brown et al., 2013; Dochy et al., 1999; Falchikov & Boud, 1989) untersuchen die Fähigkeit von Lernenden, den eigenen Wissenstand einzuschätzen. Die Studien zeigen, dass die Lernenden bedingt in der Lage sind, ihre Leistung akkurat einzuschätzen, und die Qualität der Selbsteinschätzung von einer Vielzahl an Faktoren beeinflusst wird. Wie auch für die Selbsteinschätzung des Zuwachses an experimen-

teller Kompetenz diskutiert (Abschnitt 11.1.2), spielen psychologische und situative Faktoren wie zum Beispiel Selbstkonzept, Lernerfahrung und die Formulierung der Items eine Rolle. Die von den Befragten herangezogene Referenz des maximal möglichen Wissens beispielsweise führt dazu, dass jeder der Probanden seinen Wissenstand auf einer eigenen Messskala bewertet (z. B. Moritz, 1996; Higgins et al., 1986). Die hohen Werte der Selbsteinschätzung (Tab. 20) sprechen jedoch dafür, dass die Einschätzung stark am maximal erreichbaren Wissen durch die Lerngelegenheit Praktikum orientiert war.

Auch Rehfeldt (2018a) diskutiert die Erfassung des Lernzuwachses mittels Selbsteinschätzungen in seinem PraQ-Test kritisch in Bezug darauf, wieweit die Einschätzung mit der tatsächlichen Kompetenz bzw. Kenntnisse zusammenhängt und weniger Bestandteile der Selbstwirksamkeitserwartung (z. B. Cramer, 2010) misst. Rehfeldt et al. (2015) stellen dabei auch den Vorteil der hier angesprochenen ökonomischen und zudem fachunabhängigen Anwendbarkeit des Tests gegenüber den Nachteil der geringeren Validität durch den Einsatz von Selbsteinschätzungen statt objektiver Tests. Die Autoren Braun et al. (2008), auf deren Instrument die Selbsteinschätzung des Fachwissenszuwachses des PraQ-Tests basiert, kommen in ihrer Diskussion zu dem Urteil, dass Selbsteinschätzungen einen validen Indikator für den Wissenserwerb darstellen. Um die Kritik auszuräumen und Verzerrungen zu vermeiden, strebt Rehfeldt (2018a) einen Vergleich der Selbsteinschätzung mit tatsächlichen Leistungen anhand etablierter objektiver Testinstrumente an. Allerdings legt er dabei in seiner Testevaluation den Fokus auf die Untersuchung des Lernzuwachses der experimentellen Kompetenz (siehe Abschnitt 11.1.2), sodass ein Vergleich mit einem Leistungstest für Fachwissen aussteht.

Das Streudiagramm in Abbildung 32 veranschaulicht den Zusammenhang der beiden Maße für die Änderung des Fachwissens anhand der Selbsteinschätzung und des Wissenstests, die in dieser Studie erhoben wurden. Die Werte weisen eine große Streuung um die ergänzte Regressionsgerade auf (siehe Abschnitt 8.3.5). Die Korrelation der Größen ist klein mit $\beta = .22, p = .037$. Wie vorangehend ausgeführt, findet Rehfeldt (2018a) für Korrelationen der Selbsteinschätzung mit Leistungstest bzw. Selbstwirksamkeitserwartungen für die experimentelle Kompetenz ebenfalls niedrige Korrelationen dieser Größenordnung. Im Sinne einer Validierung der Instrumente zum Fachwissenszuwachs würde man hingegen eine höhere Korrelation des Tests mit der Selbsteinschätzung erwarten. Da der Fachwissenstest größtenteils aus bewährten Aufgaben besteht, kann von der Validität des Tests ausgegangen werden. Daraus folgt, dass die Selbsteinschätzung etwas anderes (z. B. Selbstwirksamkeit) misst als der Wissenstests.

Der geringe Zusammenhang der Maße ist aber durchaus auch plausibel, wenn man sich die zuvor beschriebenen Unterschiedlichkeiten der beiden Messinstrumente in Erinnerung ruft. Die geringe Korrelation kann als Bestätigung gesehen werden, dass die Selbsteinschätzung über das Wissen auf einer übergeordneten und globaleren Ebene erfasst. Der Fachwissenszuwachs wurde dabei in Items wie „Aufgrund

dieses Praktikums kann ich wichtige Theorien zu dem behandelten Thema wiedergeben.“ auf einer Metaebene erfasst. Damit umfasst dieses Beispiel-Item nicht nur den Wissenszuwachs über eine einzelne Theorie des Versuches wie der Test, sondern auch zum Beispiel über Theorien, die in der Breite der Versuchsvorbereitung erlernt wurden. Damit schließt die Selbsteinschätzung zum einen die thematischen Lücken des Wissenstests. Anhand dessen würde sich die Selbsteinschätzung gut als Referenz zum Wissenstest eignen. Zum anderen fragt die Selbsteinschätzung im Vergleich zum Wissenstest, der über das Wiedergeben von Inhalten hinaus auch das Verständnis abfragt, einen niedrigeren Anforderungsbereich ab. Wie vorangehend beschrieben ist die Selbsteinschätzung des Fachwissens so angelegt, dass sie vor allem einen deklarativen Charakter in Bezug auf theoretische Inhalte und vernetzende Aspekte von Theorie und Praxis beinhaltet. Dabei zeigt sich anhand der Fragen zur Vernetzung von Theorie und Praxis auch, dass die Selbsteinschätzung der experimentellen Praxis näher ist. Auch aus Gesprächen mit den Studierenden in den Prüfungsgesprächen lässt sich schließen, dass für die Studierenden besonders im Laborpraktikum die Grenzen zwischen Fachwissen und dem Wissen über experimentelle Methoden, Messtechniken und dergleichen verschwimmen. Die hohe Korrelation ($\beta = .71, p < .001$) zwischen der Selbsteinschätzung des Fachwissenszuwachses und der Selbsteinschätzung des Zuwachses an experimenteller Kompetenz kann dabei als Indiz für diese Beobachtung gesehen werden. Andererseits können beide Größen zur Selbsteinschätzung des Lernzuwachses auch von gleichen personenbezogenen Faktoren wie zum Beispiel der Selbstwirksamkeitserwartung beeinflusst werden. Diese Ausführungen zeigen, dass der gemessene Fachwissenszuwachs anhand von Wissenstest und Selbsteinschätzung kein Maß für ein und dasselbe Merkmal sind. Geht man aber davon aus, dass die Selbsteinschätzung eine Mischung aus Fachwissenszuwachs, einem Anteil an Zuwachs an experimenteller Kompetenz und Selbstwirksamkeitserwartung bezüglich des Fachwissen ausmacht, dann ist diese Größe nach wie vor ein relevantes Maß zur Untersuchung des Erfolgs des Praktikums.

Zusammenfassend folgt aus diesen Betrachtungen, dass der Fachwissenstest und die Selbsteinschätzung sich aufgrund der Unterschiedlichkeit in Art und Spezifität ergänzen und der Mehrwert in der Betrachtung beider Maße für den Fachwissenszuwachs liegt. Zwar ist die Selbsteinschätzung aufgrund der fehlenden Kontrolle der Einflussgrößen weniger zuverlässig als der Fachwissenstest, aber sie erfasst dafür Bereiche des Fachwissens, die der Wissenstest nicht abdecken kann. Während der Fachwissenstest Faktenwissen und physikalisches Verständnis in einzelnen konkreten Themenbereichen erfasst, umfasst die Selbsteinschätzung ökonomisch den Fachwissenszuwachses in einer globalen Dimension. Die Ausführungen zu den Unterschieden zwischen den Maßen für den Fachwissenszuwachs und die geringe Korrelation machen aber auch deutlich, dass eine gemeinsame Analyse zum Beispiel als manifeste Indikatoren für ein latentes Maß an Fachwissenszuwachs nicht sinnvoll ist. Somit werden die Größen gesondert analysiert und die Ergebnisse zu einer gemeinsamen Schlussfolgerung herangezogen.

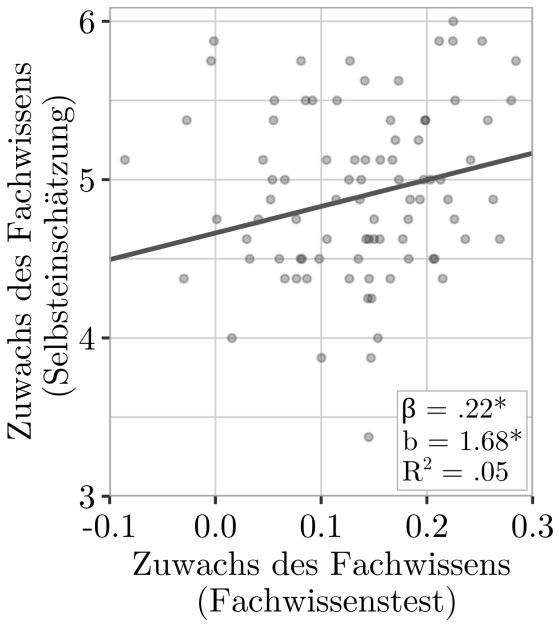


Abbildung 32: *Selbsteinschätzung des Zuwachses an Fachwissen (6-teilige Likert-Skala) abhängig von der Änderung des Fachwissens (anteilige Testwerte $\in [-1, 1]$ pro Person) auf der Datenebene „Zwischen Personen" für 88 Personen. Die Messwerte wurden semi-transparent eingezeichnet, um übereinanderliegende Datenpunkte deutlich zu machen. Ergänzt wurde die Regressionsgerade.*

11.2 Statistische Methodik

Nachdem in den vorangegangenen Abschnitten die Operationalisierung des Wirkmodells und die dabei eingesetzten Erhebungsinstrumente vorgestellt wurden, können auf dieser Basis im nächsten Schritt nun die statistischen Verfahren ausgewählt werden. Die Untersuchung eines Wirkmodells für einen komplexen Forschungsgegenstand wie einem Praktikum macht den Einsatz von mehreren abgestimmten statistischen Analyseverfahren notwendig. Zudem ist auch in dieser Analyse die in Teil III erläuterte Datenstruktur zu berücksichtigen (Abschnitt 8.3.1). In Teil III wurden zudem bereits Verfahren zur Prüfung von Stichprobenunterschieden und linearen Zusammenhängen eingeführt (siehe Abschnitte 8.3.4 und 8.3.5). Für die Analyse in diesem Teil der Arbeit kann in einigen Fällen auf diese Verfahren zurückgegriffen werden. Zur Analyse der angenommenen Wirkzusammenhänge müssen an dieser Stelle aber weitere Verfahren hinzugenommen werden, um diese komplexeren Zusammenhangsstrukturen prüfen zu können. Zu diesen Analysemethoden werden auch in diesem Teil nur die für diese Arbeit wichtigsten Informationen kurz dargestellt. Für mathematische Darstellungen und weitere Details sei auf die jeweils in den Abschnitten angegebenen Literaturquellen verwiesen.

11.2.1 Datenstruktur

In der Analyse in Teil III wurden versuchsbegleitend erhobene Daten betrachtet, die also für jeden einzelnen Versuch gemessen wurden. Insbesondere auf die Mehrebenenstruktur dieser Daten durch die Gruppierung in Personen und Versuche wurde dabei hingewiesen. In diesem Teil zur Analyse des Wirkmodells kommen Daten zur Erfassung des Lernerfolgs hinzu, die im Prä-Post-Test vor und nach dem Praktikumsteil erhoben wurden und sich damit nicht wie die versuchsbegleitenden Daten

auf die einzelnen Versuche, sondern auf die Personen und das Praktikum als Ganzes beziehen (siehe Abschnitte 8.1 und 11.1). Für die Prüfung des Wirkmodells bedeutet das, dass einige Zusammenhänge auf der Datenebene *Zwischen Personen* (Abb. 18B) untersucht werden. Um die versuchsbezogenen Daten zur wahrgenommenen Relevanz und zum situationalen Interesse auf die Datenebene *Zwischen Personen* zu beziehen, werden diese aggregiert. Dabei wird angenommen, dass ein Aufsummieren der Wahrnehmung aller einzelnen Teile des Praktikums, also aller durchgeführten Praktikumsversuche für eine Person, der Gesamtwahrnehmung des Praktikums dieser Person entspricht. Um diese Gesamtwahrnehmung trotz unterschiedlichen Anzahlen an durchgeführten Praktikumsversuchen oder fehlenden Daten vergleichbar zu machen, muss diese Summe für jede Person auf die Anzahl der Praktikumsversuche normiert werden. Dieser Personenmittelwert über alle Praktikumsversuche wird als Grundwahrnehmung der Relevanz des Praktikums bzw. des Interesses am Praktikums der jeweiligen Person interpretiert (siehe auch Abschnitt 8.3.2). Mit diesem Personenmittelwert für versuchsbegleitende Daten werden im Folgenden Analysen auf der Ebene *Zwischen Personen* betrachtet.

Zudem gilt die beschriebene Datenstruktur nur unter der Annahme, dass keine zeitlichen Effekte oder Kohorteneffekte die Daten beeinflussen und die Daten der einzelnen Kohorten damit zu einem Gesamtdatensatz zusammengefasst werden können. Daher müssen auch in dieser Analyse die Daten auf Kohorteneffekte durch die Laufzeit der Studie geprüft werden.

11.2.2 Vorgehen zur Prüfung des Wirkmodells

Nach der Prüfung der Annahmen zur Datenstruktur ist vor der eigentlichen Analyse des Wirkmodells in einem zweiten Schritt der Lernerfolg als wichtige Zielgröße des Praktikums selbst zu analysieren. Dazu werden die Methoden zur Prüfung von Stichprobenunterschieden in Abschnitt 8.3.4, die bereits für die Analyse der wahrgenommenen Relevanz in Teil III erläutert und eingesetzt wurden, genutzt. Wenn ein entsprechender Lernerfolg in den drei Bereichen Fachwissen, experimentelle Kompetenz und Interesse zu verzeichnen ist, dann kann die Analyse der Wirkzusammenhänge beginnen.

Bei dem aufgestellten Wirkmodell (Abb. 31) handelt es sich um ein Mediationsmodell. Als Mediator M wird eine Variable verstanden, die den Einfluss einer Variable X auf eine andere Variable Y vermittelt (z. B. Eid et al., 2010). Im angenommenen Wirkmodell wird der Effekt der wahrgenommenen Relevanz (X) auf den Lernerfolg (Y) über das Interesse (M) vermittelt. Dieser Effekt wird als indirekter Effekt bezeichnet. Der Effekt der wahrgenommenen Relevanz (X) auf den Lernerfolg (Y) ohne Vermittlung einer anderen Variable wird als direkter Effekt bezeichnet. Der totale Effekt oder Gesamteffekt setzt sich aus dem direkten und dem indirekten Effekt zusammen. Für eine Mediation sind unter anderem zwei Bedingungen zu erfüllen (siehe z. B. Eid et al., 2010). Für eine einfache Regression von X auf Y und auch von X auf M muss ein Zusammenhang vorliegen. In der Beschreibung des

statistischen Vorgehens im vorangegangenen Teil III wurden bereits Regressionsanalysen zur Untersuchung von Zusammenhängen von abhängigen und unabhängigen Variablen eingeführt (Abschnitt 8.3.5), die für diese Prüfungen der Zusammenhänge der einzelnen Variablen im dritten Schritt der Analyse genutzt werden. Wenn diese einzelnen Zusammenhänge zwischen den Variablen vorliegen, dann kann die Prüfung des gesamten Wirkmodells erfolgen.

Die finale Prüfung der komplexeren Zusammenhänge im aufgestellten Wirkmodell (Abb. 31) ist mit diesen Regressionsanalysen nicht möglich, da bei Regressionsanalysen abhängige und unabhängige Variablen strikt getrennt werden. Um solche vielschichtigeren Kausalhypothesen zur Erklärung von linearen Zusammenhängen zu prüfen, werden daher Pfadanalysen und Strukturgleichungsmodelle betrachtet, die nachfolgend in den Abschnitten 11.2.3 bis 11.2.5 kurz erklärt und auf die Anwendung in dieser Arbeit bezogen werden.

Aufgrund der Stichprobe lassen sich nicht beliebig komplexe Modelle belastbar schätzen, sodass das Wirkmodell jeweils einzeln für die drei Bereiche des Lernerfolgs analysiert wird, um die Modelle möglichst einfach zu halten. Die sechs Wirkmodelle mit den unterschiedlichen Zielvariablen bzgl. des Lernerfolgs werden im Folgenden auch mit den Buchstaben A bis F gekennzeichnet (A: situationales Interesse Hold, B: experimentelle Kompetenz, C und D: Selbstwirksamkeitserwartungen bzgl. Planung und Durchführung von Experimenten, E und F: Fachwissen aus Test und Selbsteinschätzung; siehe auch Zuordnung in Abb. 31).

Somit ergeben sich für die Prüfung des Wirkmodells für die Komponenten des Lernerfolgs jeweils folgende vier Analyseschritte:

1. Prüfung der Annahmen zur Datenstruktur
   (keine Kohorteneffekte und zeitl. Effekte)
2. Analyse des Lernerfolgs
3. Prüfung der Zusammenhänge der einzelnen Variablen
4. Prüfung des Wirkmodells

11.2.3 Pfadanalyse

Um Kausalhypothesen zur Erklärung von linearen Merkmalszusammenhängen wie im Wirkmodell zu prüfen, werden unter anderem Pfadanalysen betrachtet, wie zum Beispiel von Eid et al. (2010) und Bortz und Schuster (2010) beschrieben. Ein pfadanalytisches Modell kann als System von mehreren Regressionsgleichungen und somit als Erweiterung der multiplen Regression verstanden werden.

Die Zusammenhangshypothesen werden graphisch in einem sogenannten Pfaddiagramm dargestellt, aus dem das System dieser Modellgleichungen abgeleitet werden kann. In einem Pfaddiagramm werden die beobachteten (auch manifesten) Va-

riablen mithilfe von Rechtecken dargestellt. Gerichtete Pfeile repräsentieren den angenommenen Einfluss einer Variable auf die andere. Doppelpfeile stehen für symmetrische Effekte wie Korrelationen zweier unabhängiger Variablen. Variablen können im Pfadmodell sowohl unabhängige als auch abhängige Variable gleichzeitig sein. Jede Variable, auf die mindestens ein Pfeil zeigt, stellt eine abhängige Variable dar, für die eine Gleichung aufgestellt wird. Die Gleichung stellt die abhängige Variable als Summe der gewichteten unabhängigen Variablen und einem Achsenabschnitt dar, wie aus der multiplen Regression bekannt (siehe Abschnitt 8.3.5). Alle Gleichungen zusammengenommen stellen das finale Gleichungssystem der Pfadanalyse dar. Die Gewichtungen werden in dieser Analyse als Pfadkoeffizienten bezeichnet. In der finalen Darstellung des Pfaddiagrammes werden diese Pfadkoeffizienten an die Pfeile geschrieben. An die abhängigen Variablen wird zudem die Residualvarianz geschrieben, aus der hervorgeht, um welchen Anteil die Varianz dieser Variable nicht durch das Modell aufgeklärt wird. In dieser Arbeit werden zur besseren Vergleichbarkeit bei unterschiedlichen Skalen die standardisierten Pfadkoeffizienten (β) angegeben. Wie bei der multiplen Regression sind die Pfadkoeffizienten dabei nicht direkt mit den Korrelationen gleichsetzbar, können aber aus diesen abgeleitet werden (siehe z. B. Bortz & Schuster, 2010).

Die Schätzung der Modellparameter basiert dabei auf dem iterativen Maximum-Likelihood-Verfahren (siehe Abschnitt 8.3.5) anhand der Kovarianz-Matrix und der Mittelwertvektoren. In dieser Arbeit wurden fehlende Werte bei der Schätzung mit einem listenweisen Fallausschluss behandelt. Die Pfadanalysen wurden in dieser Arbeit mit dem R-Paket *lavaan* (Rosseel, 2012) berechnet.

Für die statistische Überprüfung eines indirekten Effektes, wie zum Beispiel die Annahme der Mediation von Relevanz und Lernerfolg durch das Interesse, ist zu berücksichtigen, dass dieser sich aus dem Produkt von Pfadkoeffizienten ergibt und dadurch nicht immer normalverteilt ist (vgl. z. B. Eid et al., 2010). Um die Standardfehler und damit die Konfidenzintervalle und p-Werte möglichst unverzerrt zu ermitteln, wird daher das Bootstrapping-Verfahren bei der Analyse von Mediatorhypothesen eingesetzt. Die Bootstrapping-Methode beruht darauf, dass die vorhandene Stichprobe quasi als Population betrachtet wird und wiederholt Stichproben daraus gezogen werden (z. B. Eid et al., 2010). Eine übliche Anzahl sind dabei 1000 Bootstrapping-Stichproben, die auch in diesen Analysen des Wirkmodells gezogen werden. Für jede dieser Bootstrapping-Stichproben wird nun das Pfadmodell geschätzt, sodass mithilfe der wiederholt bestimmten indirekten Effekte die Konfidenzintervalle ermittelt werden können.

Die Güte der Übereinstimmung der beobachteten und für das Modell geschätzten Kovarianzmatrix kann mit dem χ^2-Test geprüft werden (vgl. z. B. Bortz & Schuster, 2010; Urban & Mayerl, 2014). Die Nullhypothese ist entsprechend, dass beobachtete und geschätzte Kovarianzen gleich sind, also ein perfekter Modellfit vorliegt. Für die gewünschte gute Übereinstimmung sollte der Test folglich nicht signifikant sein. Der Test ist allerdings abhängig von der Stichprobengröße, sodass sich bei großen

Stichproben schon bei geringen Abweichungen die Wahrscheinlichkeit erhöht, die Nullhypothese zu verwerfen. Zur Beurteilung der Modellpassung gibt es eine Vielzahl weiterer Modelltests und Fit-Indizes, die verschiedene Vor- und Nachteile haben (vgl. z. B. Loehlin & Beaujean, 2017; Urban & Mayerl, 2014). In dieser Arbeit werden mehrere gängige Fit-Indizes zusätzlich zum χ^2-Test herangezogen, um die Modellgüte aus mehreren Perspektiven zu beurteilen. Urban und Mayerl (2014) empfehlen, sowohl absolute als auch relative Indizes zu nutzen. Absolute Indizes zeigen dabei wie der χ^2-Test auf, wie gut die beobachteten und durch das geschätzte Modell berechneten Kovarianzen übereinstimmen. Der **Standardized Root Mean Square of Residuals** (**SRMR**) beruht ebenfalls auf den Abweichungen der anhand des Modelles geschätzten und der empirisch beobachteten Kovarianzmatrix (vgl. z. B. Urban & Mayerl, 2014). Nach Hu und Bentler (1999) sollte der SRMR-Wert $< .08$ sein für einen guten Modellfit. Auch der **Root Mean Square Error of Approximation** (**RMSEA**) basiert wie der SRMR-Wert auf der Abweichung von geschätzter und beobachteter Kovarianzmatrix. Zusätzlich werden beim RMSEA die Freiheitsgrade des Modelles und die Fallzahl berücksichtigt (vgl. z. B. Urban & Mayerl, 2014). Ein RMSEA $< .06$ wird als guter Fit interpretiert (Hu & Bentler, 1999). Auch ein RMSEA $< .08$ wird noch als akzeptabel interpretiert (Browne & Cudeck, 1992). Zu den relativen Indizes zählt zum Beispiel der **Comperative Fit Index** (**CFI** $\in [0, 1]$). Die Grundidee bei einem relativen Index ist der Vergleich der Ergebnisse des geschätzten Modelles mit denen eines Vergleichsmodells, bei dem im Falle des CFI alle Kovarianzen in der Kovarianzmatrix auf Null gesetzt werden (vgl. z. B. Urban & Mayerl, 2014). Je näher der CFI am Maximalwert 1 liegt, desto besser gilt der Modellfit. Die ursprüngliche Daumenregel für einen akzeptablen Modellfit mit einem CFI $> .90$ (Bentler, 1992) wird inzwischen meist enger gesehen mit einem guten Modellfit bei einem CFI $> .95$ (Hu & Bentler, 1999). Für die Interpretation dieser Modellschätzungen ist neben der Modellgüte zudem zu bedenken, dass die Kausalhypothesen pfadanalytisch widerlegt, aber nicht bewiesen werden können, da immer äquivalente Alternativmodelle gefunden werden können (vgl. Bortz & Schuster, 2010).

11.2.4 Strukturgleichungsmodellierung

Wie bei Regressionsanalysen wird bei Pfadanalysen davon ausgegangen, dass alle Variablen im Modell messfehlerfrei erhoben wurden. Realistisch sind in der empirischen Bildungsforschung jedoch messfehlerbehaftete Variablen, die zu Verzerrungen der geschätzten Modellparameter und Standardfehler führen können. Strukturgleichungsmodelle ermöglichen eine Untersuchung messfehlerbereinigter Abhängigkeitsstrukturen, wie zum Beispiel von Eid et al. (2010), Urban und Mayerl (2014) und Döring und Bortz (2016) beschrieben. Aus diesem Grund werden Strukturgleichungsmodelle im Vergleich zu Pfadanalysen oft bevorzugt. Strukturgleichungsmodelle kombinieren dazu Pfadanalyse und Faktorenanalyse. Lineare Strukturgleichungsmodellierungen untersuchen entsprechend Zusammenhangsstrukturen auf Ebene latenter Variablen.

Strukturgleichungsmodelle setzen sich aus einem Messmodell und einem Strukturmodell zusammen. Im Messmodell wird auf Basis eines faktoranalytischen Modells definiert, wie sich die latenten Variablen aus den gemessenen Variablen ergeben. Unter latenten Variablen oder auch Konstrukten werden unbeobachtbare Variablen verstanden, die sich in anderen beobachtbaren Variablen manifestieren. Lernerfolg kann beispielsweise nicht direkt gemessen werden, sondern kann nur indirekt zum Beispiel aus gelösten Testaufgaben und praktischen Experimentiertests geschlussfolgert werden. Solche Testaufgaben enthalten zudem Messfehler wie zum Beispiel durch Raten von Antworten. Eine latente Variable kann als die Gemeinsamkeit der verschiedenen beobachteten Variablen ohne Messfehler aufgefasst werden. Zur Bestimmung der latenten Variable wird in der Faktorenanalyse ein lineares Gleichungssystem aufgestellt, bei der jede der i beobachteten Variablen x_i linear von dem Konstrukt η abhängt: $x_i = \lambda_i \cdot \eta + \delta_i$. Die Gewichtung λ_i wird als Faktorladung bezeichnet. δ_i ist die Messfehlervariable des Indikators x_i. Die Faktorladungen werden in dieser Arbeit nicht explizit berichtet, sofern die standardisierten Ladungen alle ähnlich nahe eins liegen und signifikant sind. Im Strukturmodell wird das Pfadmodell auf die latente Ebene übertragen und somit analog der Zusammenhang der latenten Variablen untereinander im Modell spezifiziert. Das Gesamtmodell ergibt sich damit als System linearer Gleichungen aus dem Messmodell und dem Strukturmodell. Um diese Beziehungen der Variablen im Strukturgleichungsmodell untereinander zu spezifizieren und die geschätzten Pfadkoeffizienten schließlich darzustellen, werden ebenfalls Pfaddiagramme genutzt (siehe Abschnitt 11.2.3). Im Pfaddiagramm werden die latenten Variablen mit Kreisen dargestellt und damit von den beobachteten Variablen (Rechtecke) abgegrenzt.

Die Schätzung der standardisierten Modellparameter erfolgt wie bei der Pfadanalyse kovarianzanalytisch über die Maximum-Likelihood-Methode. Dazu wurde in dieser Arbeit ebenfalls das R-Paket *lavaan* (Rosseel, 2012) eingesetzt. Das vorgestellte Bootstrapping-Verfahren findet ebenso Anwendung zur Prüfung der indirekten Effekte. Ebenso können die für die Pfadanalyse dargestellten Fit-Indizes zur Prüfung der Modellpassung herangezogen werden. Auch für das Strukturgleichungsmodell gilt jedoch dabei, dass es sich nicht um ein Kausalmodell handelt.

Für die Spezifikation des Messmodelles spielt in dieser Arbeit das Item-Parceling, das heißt das Bilden von Item-Paketen, eine Rolle. Für die Bestimmung des latenten Konstrukts im Modell werden dazu mehrere Items der Indikatorvariable in einem Mittelwert zusammengefasst, anstatt alle Items einzeln auf das latente Konstrukt laden zu lassen. Das Item-Parceling wird in dieser Arbeit insbesondere aufgrund des geringen Stichprobenumfangs eingesetzt, um die Anzahl der zu schätzenden Parameter verringern und so eine stabilere Parameterschätzung zu ermöglichen (z. B. Bandalos & Finney, 2001). Während die Item-Parcels in diesbezüglichen Diskussionen auch als reliablere Indikatoren gelten, wird gleichzeitig auch vor Missspezifikationen aufgrund der Multidimensionalität von Indikatoren gewarnt (vgl. z. B. Bandalos &

Finney, 2001; Little et al., 2002). In dieser Arbeit wurden daher bei Indikatorvariablen mit mehreren Dimensionen die einzelnen Subskalen als Item-Parcels genutzt.

11.2.5 Mehrebenenanalyse

Wie bei Regressionsanalysen wird bei den vorangehend beschriebenen Pfadanalysen und Strukturgleichungsmodellierungen vorausgesetzt, dass die gemessenen Daten unabhängig voneinander sind (z. B. Urban & Mayerl, 2014). Im Datensatz dieser Arbeit liegen Mehrebenenstrukturen durch Clusterung der versuchsbegleitenden Daten in Personen und Versuche vor. Werden derartige Mehrebenenstrukturen bei der Pfadanalyse oder Strukturgleichungsmodellierung von komplexeren Zusammenhangsstrukturen vernachlässigt, dann führt dies zu einer Verzerrung der Fit-Statistik und der Schätzung der Parameter und Standardfehler (Julian, 2001).

Die Mehrebenenanalyse als Erweiterung der Regressionsanalyse durch Betrachtung der Zusammenhänge innerhalb und zwischen Gruppen wurde bereits in Abschnitt 8.3.5 beschrieben. Bei der Pfadanalyse und der Strukturgleichungsmodellierung können ebenfalls Mehrebenenstrukturen berücksichtigt werden. Die Idee dahinter ist analog zur Mehrebenen-Regressionsanalyse, dass in einer Analyse gleichzeitig Modelle zu den Kovarianzstrukturen zwischen und innerhalb von Gruppen geschätzt werden (für mathematische Darstellungen siehe z. B. Muthén, 1994; Kaplan & Elliott, 1997; Kaplan, 2009; Hox, 2010; Urban & Mayerl, 2014).

11.2.6 Besonderheiten bei den Modellierungen des Lernerfolgs

Bevor anhand der vorgestellten Zusammenhangsanalysen auf das Vorgehen zur Prüfung des gesamten Wirkmodells in dieser Arbeit eingegangen wird, muss zum Verständnis zunächst noch eine Erläuterung erfolgen, in welcher Form die Komponente Lernerfolg in das Modell eingeht. Um den Lernerfolg zu erfassen, wurden das Fachwissen und die Selbstwirksamkeitserwartungen im Prä- und Post-Test erhoben, da in dieser Arbeit nicht das Erreichen eines bestimmten Endniveaus der Lernleistung (Post-Test) von Interesse ist, sondern der Lernzuwachs über das Praktikum. Das Vorwissen ist dabei einer der wichtigsten Prädiktoren für Lernerfolg (vgl. z. B. E. Wild & Möller, 2015). Dem Matthäus-Effekt zufolge (Begriff geprägt von Merton, 1968; im Kontext Lerntheorie siehe z. B. Renkl, 1996) führt ein höheres Vorwissen zu größerem Lernerfolg, da Vorwissen Vorteile in der Nutzung einer Lerngelegenheit bietet.

Um den Lernerfolg unter Berücksichtigung des Ausgangswertes im Wirkmodell zu analysieren, gibt es zwei klassische Vorgehensweisen. In einer der Vorgehensweisen wird der Post-Test (abhängige Variable) unter Kontrolle des Prä-Tests (unabhängige Variable) modelliert. In der anderen Vorgehensweise wird der absolute Lernzuwachs als Unterschied zwischen dem Post- und Prä-Testwert ($x_{abs,i} = x_{post,i} - x_{pre,i}$ für Person $i \in \{1, ..., n\}$) bestimmt und in dieser Form als abhängige Variable modelliert. Unter dem Begriff Lord's Paradoxon ist in den 1960er Jahren bekannt geworden,

dass diese vermeintlich äquivalenten Modelle unter bestimmten Bedingungen widersprüchliche Ergebnisse liefern können (Lord, 1967). Seitdem ist Lord's Paradoxon und die Frage nach der Modellwahl Thema zahlreicher Diskussionen und Veröffentlichungen (vgl. z. B. Allison, 1990; van Breukelen, 2013).

Für die Betrachtung der Wirkmodelle mit Fachwissen anhand des Wissenstests und Selbstwirksamkeitserwartungen (Wirkmodelle C bis E) gibt es sowohl Argumente für die Betrachtung des Post-Wertes unter Kontrolle des Prä-Wertes als auch für die Betrachtung des absoluten Lernzuwachses. Wenn der Prä-Wert mit einem anderen Prädiktor und/oder wie in dieser Arbeit mit dem Post-Wert in kausalem Zusammenhang steht, bevorzugt Allison (1990) beispielsweise die gängigere Modellierung des Post-Wertes unter Kontrolle des Prä-Wertes. In dieser Arbeit ist es aufgrund der geringen Stichprobe andererseits erstrebenswert, die Modellstruktur möglichst einfach zu halten, also folglich den absoluten Lernzuwachs zu modellieren. Zum anderen ist bei Betrachtung des absoluten Zuwachses die Modellstruktur übereinstimmend mit den Wirkmodellen B und F, in denen der im Post-Test von den Studierenden eingeschätzte Zuwachs an Fachwissen und experimenteller Kompetenz betrachtet wird (siehe Abschnitt 11.1). Anhand letzterer Aspekte wurde beschlossen, die Wirkmodelle C bis E mit dem absoluten Lernzuwachs für das Fachwissen anhand des Wissenstests bzw. der Selbstwirksamkeitserwartungen darzustellen. Zur Absicherung der Ergebnisse wurden die Modellierungen aber auch für den Post-Wert unter Kontrolle des Prä-Wertes berechnet. Abweichungen zwischen den Modellen werden im Folgenden berichtet und diskutiert.

Ein weiterer Aspekt, der in diesen Überlegungen zu bedenken ist, ist ein möglicher Deckeneffekt bei der Messung des Lernerfolgs. Die Grenze des Messinstrument nach oben kann dazu führen, dass ein höherer Ausgangswert mit einem geringeren absoluten Lernzuwachs verbunden ist. Im Falle eines Leistungstests spielt dabei auch die Aufgabenschwierigkeit eine Rolle, da ein Lernzuwachs für Studierende mit hohem Vorwissen erschwert wird, wenn nur wenige und damit auch die schwierigsten Aufgaben übrig bleiben. Damit kann das Vorwissen auch ein Prädiktor für den absoluten Lernzuwachs sein. Eine Aufnahme des Prä-Tests in die Modellierung des absoluten Lernzuwachses ist allerdings mit einer Verzerrung verbunden, da der Prä-Test damit zweimal in die Modellierung eingeht und damit die Annahme der Unabhängigkeit der Residuen für lineare Modelle verletzt wird (z. B. Eid et al., 2010, Abschnitt 8.3.5). Da die Vernachlässigung eines Prädiktors aber ebenfalls mit Verzerrungen verbunden sein kann (z. B. Eid et al., 2010), wurde in dieser Analyse trotzdem ergänzend ein Modell mit dem absoluten Lernzuwachs als abhängige Variable unter Kontrolle des Prä-Tests vorgenommen, um einen Hinweis auf die Stabilität der Ergebnisse zu erhalten. Wenn starke Abhängigkeiten von Lernzuwachs und Ausgangswert erwartet werden und dies inhaltlich der Fragestellung dienlich ist, sind allerdings idealerweise latente Veränderungsmodelle (latent change score models, siehe z. B. Kievit et al., 2018) heranzuziehen, um die Zusammenhänge genauer zu modellieren.

11.2.7 Besonderheiten bei den Modellierungen der Wirkannahmen

In Bezug auf die Auswahl der Modellierungsverfahren ist anhand der vorangegangenen Ausführungen zu den Zusammenhangsanalysen (Abschnitte 11.2.3 bis 11.2.5) zu schlussfolgern, dass das ideale statistische Vorgehen für die Prüfung des gesamten Wirkmodells die Mehrebenen-Strukturgleichungsmodellierung ist, um messfehlerbereinigte Zusammenhangsstrukturen bei Berücksichtigung der Mehrebenenstruktur der Daten zu untersuchen. Allerdings ist diese Analyse nicht für jede Komponente des Lernerfolgs ideal, da einige der Größen nicht versuchsbegleitend, sondern vor und nach dem Praktikumsteil gemessen wurden (siehe Abschnitt 8.1).

Nur beim Interesse als Lernerfolg (Wirkmodell A) kann tatsächlich eine Mehrebenenanalyse auf den Datenebenen *Zwischen* und *Innerhalb Personen* vorgenommen werden, da alle Variablen versuchsbegleitend erhoben wurden. Die versuchsbegleitenden Variablen werden dazu jeweils zu vier Item-Parcels mit je zwei Items zusammengefügt und als Indikatoren für die Konstrukte auf den beiden Datenebenen verwendet. Während die wahrgenommene Relevanz als eindimensional aufgefasst wird, teilen sich die Item-Parcels für das situationale Interesse in je zwei Parcels zu den jeweiligen Subskalen auf. Um Verzerrungen zu vermeiden, wurden die Faktorladungen für die beiden Datenebenen gleich gehalten.

Die experimentelle Kompetenz, die Selbstwirksamkeitserwartungen und das Fachwissen wurden hingegen im Prä- und/oder Post-Test erhoben, sodass in diesen Fällen (Wirkmodelle B bis F) eine einfache Strukturgleichungsmodellierung auf der Ebene *Zwischen Personen* vorgenommen wurde, die immer noch den Vorteil bietet, messfehlerbereinigte Zusammenhangsstrukturen zu untersuchen. Um die versuchsbegleitend erhobenen Daten zur wahrgenommenen Relevanz und zum situationalen Interesse auf der Ebene *Zwischen Personen* zu analysieren, werden Personenmittelwerte über alle Bewertungen der im Praktikum durchgeführten Versuche gebildet, wie in Abschnitt 11.2.1 erläutert.

In den Modellierungen wird für die wahrgenommene Relevanz und das situationale Interesse ein Item-Parceling wie in der Modellierung des Interesses als Lernerfolg vorgenommen. Für die Modellierung der Selbsteinschätzungen der Zunahme der experimentellen Kompetenz bzw. des Fachwissens (Wirkmodell B und F) werden dabei als Indikatoren Item-Parcels aus den jeweiligen neun bzw. zwei Subskalen gebildet, da die Variablen mehrere Subdimensionen haben. Für die Selbstwirksamkeitserwartungen werden zwei Modellierungen vorgenommen (Wirkmodelle C und D), da für die beiden Subskalen unterschiedliche Lerngelegenheiten im Praktikum vorhanden sind und daher unterschiedliche Effekte erwartet werden. In der jeweiligen Modellierung der Subskala werden die latenten Konstrukte zu den Selbstwirksamkeitserwartungen anhand vier Item-Parcels a zwei Items geschätzt.

Für die Zunahme des Fachwissens anhand des Wissenstests (Wirkmodell E) wird abweichend von der bisher angewendeten Strukturgleichungsmodellierung eine Pfadanalyse berechnet. Die Heterogenität der Fachwissensdaten führt dazu, dass eine

latente Modellierung des Fachwissenszuwachses in der Strukturgleichungsmodellierung auch trotz unterschiedlichem Parceling der Aufgaben keine konsistenten und belastbaren Ladungen der manifesten Indikatoren auf das latente Konstrukt liefert. Die Analyse des Testinstrumentes für das Fachwissen hat durch geringe Trennschärfen und fragwürdige Reliabilität auf Ebene der Subskalen bereits die Heterogenität des gemessenen Fachwissens gezeigt (siehe Abschnitt 11.1.3). Ursachen für die Heterogenität sind die breite Abfrage des Fachwissens zu verschiedenen fachlichen Themenbereichen sowie unterschiedliche Schwierigkeitsgrade. Unterschiedliche Operationalisierungen der manifesten Größen durch Teilskalen von beispielsweise Themenbereichen oder Aufgaben unterschiedlicher Themen und Versuche konnten nicht zu einer konsistenten Modellierung des latenten Fachwissenszuwachses beitragen. Gleiches gilt für die Betrachtung von Teildatensätzen bestimmter Personengruppen oder Fachwissensaufgaben, da dies insbesondere durch die Reduktion des Informationsgehalts zu größeren Messfehlern führt. Da in einer Pfadanalyse (siehe Abschnitt 11.2.3) nur Zusammenhänge zwischen beobachteten Größen betrachtet werden, entfällt die in diesem Fall schwierige Schätzung von latenten Konstrukten. Allerdings entfällt auch der Vorteil, dass Messfehler durch die Verwendung der latenten Variablen berücksichtigt werden.

12 Prüfung des Wirkmodells

Auf Basis der dargestellten Operationalisierung des Wirkmodells, der Erfassung des Lernerfolgs und der statistischen Methodik, kann nun die eigentliche Prüfung des aufgestellten Wirkmodells erfolgen. Im Rahmen der in Abschnitt 4.2.2 entwickelten Forschungsfrage 3 soll die Wirkung der wahrgenommenen Relevanz auf den Lernerfolg im Praktikum analysiert werden. Anhand der zuvor festgelegten Komponenten des Lernerfolgs gliedert sich diese Untersuchung der Wirkung des Lehramtspraktikums in drei Teiluntersuchungen bzgl. des Interesses (Abschnitt 12.1), der experimentellen Kompetenz (Abschnitt 12.2) und des Fachwissens (Abschnitt 12.3). Für die Komponenten des Lernerfolgs werden insgesamt sechs einzelne Modellierungen des aufgestellten Wirkmodells vorgenommen, die im Folgenden mit Wirkmodell A bis F bezeichnet werden. Die Ergebnisse dieser Teiluntersuchungen werden dann in Abschnitt 13 zu einem gemeinsamen Bild zusammengeführt, um das in Teil III gezeichnete Bild zur Wirkung des Praktikums in Bezug auf die wahrgenommene Relevanz zu erweitern.

12.1 Interesse

Das Interesse spielt im Wirkmodell des lehramtsspezifischen Praktikums zwei wichtige Rollen (siehe Abschnitt 3.2 und Abb. 31). Zum einen kommt dem situationalen Interesse als Proxy für die Lernmotivation die Rolle des Vermittlers zwischen wahrgenommener Relevanz und Lernerfolg zu. Dies ist der Grund, warum das Interesse bei der Prüfung des Wirkmodells zuerst untersucht werden muss. Zum anderen wird das individuelle Interesse und damit mittelbar auch die Motivation als Lernerfolg betrachtet. Dieses motivationale Ziel eines Praktikums ist bei Lehramtsstudierenden besonders bedeutsam, da Motivation und Enthusiasmus als Eigenschaften guter Lehrkräfte angesehen werden (vgl. z. B. E. Wild & Möller, 2015). Aus verschiedenen Studien geht hervor, dass sich günstigeres Lernverhalten und eine höhere Motivation bei Lernenden zeigt, wenn die Lehrkräfte mit Enthusiasmus und aus Interesse unterrichten (Rosenshine, 1970; T. C. Wild et al., 1992; Patrick et al., 2000; Feldman, 2007). Daher wird nun zunächst die Entwicklung des individuellen Interesses im Lehramtspraktikum im Hinblick auf den Lernerfolg analysiert. Basierend auf der Betrachtung der verschiedenen Phasen des Interesses kann dann die eigentliche Analyse des Wirkmodells zunächst in Bezug auf das Interesse (Wirkmodell A) und anschließend in Bezug auf die anderen Komponenten des Lernerfolgs folgen.

12.1.1 Individuelles und situationales Interesse

Individuelles Interesse an Physik und am Unterrichten von Physik

Die ersten zwei Aspekte des Interesses, die in dieser Untersuchung mit Bezug auf das Praktikum betrachtet werden, sind das individuelle Interesse an Physik und am Unterrichten von Physik. Diese Aspekte des Interesses bzgl. der Physik wurden in

den Prä-Post-Tests vor und nach den beiden Praktikumssemestern erfasst. Diese Messzeitpunkte werden im Folgenden mit Prä 1, Post 1, Prä 2 und Post 2 bezeichnet. Varianzanalysen haben im Vorfeld der Analyse gezeigt, dass die Daten zum Interesse zu den jeweiligen Messzeitpunkten (Prä 1 - Post 2) keine signifikanten Unterschiede bzgl. des Semesters der Erhebung (Sommersemester 2016 - 2019, Winter-/Sommersemester) aufweisen, womit gemäß der Annahmen zur Datenstruktur (siehe Abschnitt 8.3.1) keine Kohorteneffekte vorliegen. Für einen ersten Überblick ist die Deskription der Daten zum Interesse an den verschiedenen Messzeitpunkten in Tabelle 21 gelistet und in den Boxplots in Abbildung 33 veranschaulicht. Da an den Messzeitpunkten (Prä 1 - Post 2) die gleichen Studierenden befragt wurden, handelt es sich um abhängige Daten. Die Daten sind unbalanciert, da nicht alle Studierenden am zweiten Praktikumssemester teilgenommen haben.

Zunächst wird das Interesse der Lehramtsstudierenden an Physik betrachtet. Dies ist mit mittleren Werten $M \geq 3.75$ für alle Messzeitpunkte über dem mittleren Maß der 5-teiligen Likert-Skala erwartungsgemäß hoch (Tab. 21 oben). Der Boxplot in Abbildung 33a veranschaulicht die Verteilung des Interesses für die vier Messzeitpunkte. Der Messbereich ist insbesondere nach unten hin nicht voll ausgeschöpft, was für Lehramtsstudierende mit der Fächerwahl Physik zu erwarten ist. Trotzdem fallen einzelne Ausreißer mit Werten von < 2.5 auf. Der Median im Boxplot weist auf eine Schiefe der Verteilungen der Daten der Messzeitpunkte Prä 1 und Post 2 hin. Ein Shapiro-Test bestätigt, dass das Interesse an Physik zu diesen Messzeitpunkten nicht normalverteilt ist. Der Levene-Test zur Prüfung der Varianz-Homogenität zeigt, dass die Streuung an allen Messzeitpunkten ähnlich ist ($F(3, 309) = 0.021, p = .889$).

Im Folgenden sollen die Mittelwerte für die Messzeitpunkte mithilfe von Mittelwertvergleichen auf Unterschiede untersucht werden. Anhand der Deskription sind im Rahmen der Standardfehler keine Unterschiede des Interesses an Physik vor und nach den jeweiligen Praktikumssemestern (Prä 1 - Post 1, Prä 2 - Post 2) sowie vor und nach dem gesamten Praktikum (Prä 1 - Post 2) zu erwarten (Tab. 21 oben). Ein kleiner Unterschied im Rahmen der Standardfehler ist allerdings zwischen Ende des ersten und Anfang des zweiten Praktikumssemesters (Post 1 - Prä 2) beobachtbar. Um konkrete Aussagen treffen zu können, werden die Unterschiede der Stichproben auf Signifikanz geprüft. Aufgrund der Abhängigkeit der Daten in Bezug auf die Personen wird dazu eine einfaktorielle Varianzanalyse mit Messwiederholung genutzt und zur Absicherung aufgrund der nicht-parametrischen Daten (siehe oben) um einen Friedman-Test ergänzt (siehe Abschnitt 8.3.4). Unvollständige Datensatze werden in diesen Analysen abhängiger Daten nicht berücksichtigt, sodass nur die Daten der Studierenden einfließen, die an beiden Semestern teilgenommen haben ($n = 68$). Der Mauchly-Test weist darauf hin, dass die Annahme der Sphärizität verletzt ist ($W(3) = 0.83, p = .027$), weshalb für die Varianzanalyse der p-Wert zusätzlich mit Greenhouse-Geisser-Korrektur ($\epsilon_{GG} = .89$) berichtet wird (siehe Abschnitt 8.3.4). Die Varianzanalyse zeigt, dass im Rahmen des Signifikanzniveaus von .05 kein Unterschied zwischen den Messzeitpunkten feststellbar

ist ($F(3, 201) = 2.53, p = .059, p_{GG} = .066, \eta^2_G = .01$). Der Friedman-Test weist jedoch auf signifikante Unterschiede hin ($\chi^2(3) = 15.63, p = .001$). Um die Messzeitpunkte mit unterschiedlichem Interesse zu identifizieren, wurden paarweise abhängige t-Tests und Wilcoxon-Vorzeichen-Rang-Tests mit Bonferroni-Korrektur als Post-Hoc-Tests berechnet. Während die t-Tests keine signifikanten Unterschiede aufzeigen, zeigt der Wilcoxon-Test signifikante Unterschiede zwischen den Zeitpunkten Prä 1 und Prä 2 ($p = .037$) sowie Post 1 und Prä 2 ($p = .016$). Die Effektstärke ist in beiden Fällen allerdings klein mit einem Cohens $d \approx 0.20$. Damit bestätigen diese Tests die Vermutungen anhand der Deskription. Es liegen keine Unterschiede für das Interesse an Physik vor und nach den jeweiligen Praktikumssemestern (Prä 1 - Post 1, Prä 2 - Post 2) sowie vor und nach dem gesamten Praktikum (Prä 1 - Post 2) vor. Der Unterschied des Interesses vom ersten Praktikumssemester (Prä 1, Post 1) zum Beginn des zweiten Praktikumssemesters (Prä 2) kann nicht auf das Lehramtspraktikum zurückgeführt werden, da in diesem Zeitraum kein Lehramtspraktikum stattgefunden hat. Das Interesse an Physik wird folglich im Mittel als stabil über den Praktikumszeitraum angenommen werden.

Nun wird analog das Interesse der Lehramtsstudierenden am Unterrichten von Physik betrachtet (Tab. 21 unten, Abb. 33b). Das Interesse am Unterrichten von Physik liegt mit Mittelwerten von etwa $M \approx 4.3$ im oberen Viertel der 5-teiligen Likert-Skala (Tab. 21 unten). Die Verteilung des Interesses, veranschaulicht mithilfe von Boxplots in Abbildung 33b, zeigt deutlich, dass die Werte für die vier Messzeitpunkte hauptsächlich am oberen Rand der Skala liegen. Dies weist deutlich auf einen Deckeneffekt hin. Nur einzelne Ausreißer weisen Werte von < 2.5 auf. Die Daten weisen damit keine Normalverteilung auf, wie nicht nur die Boxplots zeigen, sondern auch Shapiro-Tests bestätigen. Auch wenn für die Daten Varianz-Homogenität anhand eines Levene-Tests besteht ($F(3, 309) = 2.12, p = .097$), sind die Daten somit nicht-parametrisch. Anhand der Deskription sind im Rahmen der Standardfehler keine Unterschiede des Interesses am Unterrichten von Physik zwischen den vier Messzeitpunkten zu erwarten. Wie zuvor für das Interesse an Physik wurden für das Interesse am Unterrichten von Physik eine Varianzanalyse mit Messwiederholung ($F(3, 201) = 2.53, p = .059, p_{GG} = .070$ mit $\epsilon_G G = .83, \eta^2_G = .01$, Mauchly-Test: $W(3) = 0.75, p = .002$) und ein Friedman-Test ($\chi^2(3) = 7.04, p = .071$) berechnet. Für keinen der Mittelwertvergleiche ergeben sich Unterschiede im Interesse am Unterrichten von Physik. Dies kann zum einen daran liegen, dass das Interesse am Unterrichten von Physik über das Praktikum hinweg stabil ist, aber zum anderen auch an dem Messinstrument, welches durch den Deckeneffekt nicht geeignet ist, um entsprechende Unterschiede zu detektieren.

Insgesamt zeigt diese Untersuchung, dass das Interesse an Physik und am Unterrichten von Physik generell sehr hoch von den Lehramtsstudierenden eingeschätzt wird. Das hat aber zur Folge, dass die Messungen durch Deckeneffekte geprägt sind. Zudem zeigt sich, dass für diese Interessen kein Einfluss durch das Praktikum nach-

gewiesen werden kann. Eine Ursache dafür kann sein, dass die Messinstrumente Unterschiede nicht erfassen können. Dazu kommt der Punkt, dass die Interessen gemäß dem Konstrukt des individuellen Interesses per se als relativ stabil aufgefasst werden können.

Subskala	Messzeitpunkt	n	M	SD	SE	MIN	MAX
Physik	Prä 1	89	3.76	0.50	0.05	2.67	4.83
	Post 1	88	3.75	0.51	0.05	2.50	4.83
	Prä 2	68	3.92	0.58	0.07	1.50	4.83
	Post 2	68	3.88	0.57	0.07	2.17	4.83
Unterrichten	Prä 1	89	4.42	0.51	0.05	2.17	5.00
	Post 1	88	4.33	0.53	0.06	3.00	5.00
	Prä 2	68	4.31	0.63	0.08	1.67	5.00
	Post 2	68	4.32	0.61	0.07	2.33	5.00

Tabelle 21: *Deskriptive Statistik für das Interesse an Physik und am Unterrichten von Physik (5-teilige Likert-Skala) vor und nach dem ersten (Prä 1, Post 1) und dem zweiten Praktikumssemester (Prä 2, Post 2).*

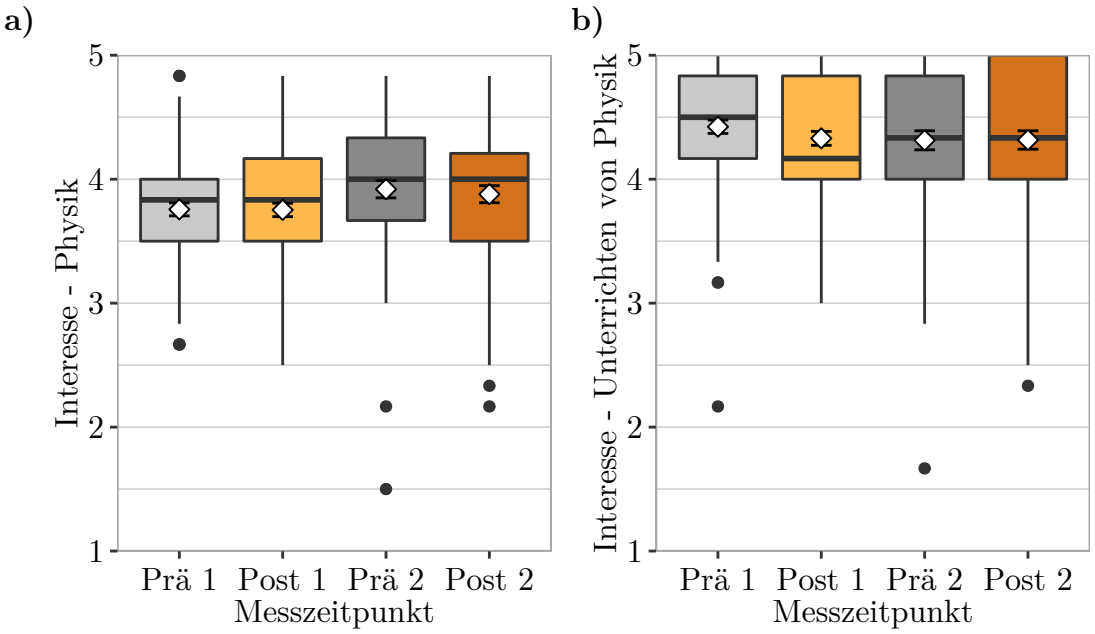


Abbildung 33: *Darstellung des Interesses a) an Physik und b) am Unterrichten von Physik (5-teilige Likert-Skala) vor (grau) und nach (orange) dem ersten (Prä 1, Post 1) und dem zweiten Praktikumssemester (Prä 2, Post 2). Der Boxplot zeigt neben Minimum, unterem Quartil, Median, oberem Quartil und Maximum zusätzlich die Mittelwerte (weiße Rauten) mit Standardfehlern. Die zugehörigen deskriptiven Werte sind in Tabelle 21 gelistet.*

Situationales Interesse

Im Gegensatz zum individuellen Interesse spiegelt das situationale Interesse den zeitlich eher variablen Anteil des Interesses dar. Dieses wurde in dieser Studie in Form der Catch- und Hold-Komponente des situationalen Interesses in Bezug auf die einzelnen Praktikumsversuche versuchsbegleitend erhoben. In diesem Abschnitt wird dazu zunächst untersucht, inwieweit Kohorteneffekte und zeitliche Effekte das situationale Interesse beeinflussen und entsprechend eine Vereinfachung der Datenstruktur zulässig ist. Dann werden die beiden Komponenten des situationalen Interesses bzgl. der einzelnen Praktikumsversuche untersucht und verglichen, um am Ende dieses Abschnittes Rückschlüsse für die Rollen des Interesses im Wirkmodell ziehen zu können.

Zu Beginn dieser Analyse wurde die Zulässigkeit der Vereinfachung der Datenstruktur des situationalen Interesses analog zur wahrgenommenen Relevanz hinsichtlich Kohorteneffekten und zeitlichen Effekten geprüft. Einfaktorielle Varianzanalysen und Kruskal-Wallis-Tests zeigen, dass sich das Personenmittel des situationalen Interesses Catch bzw. Hold nicht signifikant für die Winter- bzw. Sommersemester über den Befragungszeitraum der sieben Semester unterscheidet[7]. Daher werden für das situationale Interesse wie für die wahrgenommene Relevanz Kohorteneffekte im Folgenden vernachlässigt und die Daten der Semester zu einer Gesamtstichprobe zusammengefasst.

Wie für die wahrgenommene Relevanz können auch für das situationale Interesse zeitliche Effekte über den langen Befragungszeitraum der einzelnen Lehramtsstudierenden von bis zu einem Jahr relevant sein. Da für die wahrgenommene Relevanz kein solcher Effekt festgestellt werden konnte, wird für die Catch-Komponente des situationalen Interesses ebenso kein (großer) zeitlicher Effekt erwartet. Für die Hold-Komponente des situationalen Interesses, die die langfristigeren Aspekte des situationalen Interesses abdeckt, liegt hingegen die Hypothese eines zeitlichen Effektes näher. Die Catch-Komponente des Interesses könnte die Hold-Komponente regelmäßig positiv triggern, sodass die Hold-Komponente über den Praktikumsverlauf ansteigen könnte. Das hätte einen Einfluss auf angenommene Datenstruktur des situationalen Interesses und die weitere Analyse. Auch die Entwickler der eingesetzten Skala zum situationalen Interesse sehen das genaue Zusammenspiel der beiden Komponenten des situationalen Interesses und der zeitlichen Entwicklung grundlegend als noch ungeklärt (Lewalter & Knogler, 2014; Knogler et al., 2015). Wenn die Hypothese auf den zeitlichen Effekt auf das Interesse an den Praktikumsversuchen zutrifft, dann ist die zeitliche Entwicklung des Interesses ebenso ein wichtiger Aspekt, wenn man die Wirkmechanismen von adressatenspezifischen Praktika noch weiter im Detail verstehen möchte. Für eine optimale Gestaltung der Praktikumsversuche könnte der Zeiteffekt zum Beispiel auch bei der Festlegung der Reihenfolge der Versuche im Praktikum ausgenutzt werden.

[7] Catch SS: $F(3, 80) = 0.369, p = .775, \eta^2 = .01$, Catch WS: $F(2, 72) = 1.515, p = .227, \eta^2 = .04$; Hold SS: $F(3, 80) = 0.554, p = .647, \eta^2 = .02$, Hold WS: $F(2, 72) = 2.428, p = .095, \eta^2 = .06$

Um diese Hypothese bzgl. der zeitlichen Effekte zu prüfen, wurden analog zur wahrgenommenen Relevanz Wachstumskurvenmodelle der Komponenten des situationalen Interesses über die sechs Zeitabschnitte mit jeweils vier Praktikumsversuchen im Verlauf der zwei Praktikumsteile untersucht (siehe auch Abschnitt 8.3.6 für Modellspezifikation). Die Intraklassen-Korrelation (ICC) anhand der Nullmodelle (Modell 0P und 0V in Tab. 10 und 22) zeigt, dass für die Catch-Komponente des situationalen Interesses 22 % der Varianz auf Unterschiede zwischen den Personen ($n = 81$) und 4 % auf Unterschiede zwischen Versuchen ($n = 25$) zurückzuführen sind. Für die Hold-Komponente sind die Varianzanteile entsprechend 34 % zwischen den Personen und 1 % zwischen den Versuchen. Dies zeigt, dass der Varianzanteil für Versuche sehr klein ist und damit die Hauptvarianz in den Personen liegt. Die ICC liegt damit unterhalb der Grenze von 5 %, ab der eine Berücksichtigung der Gruppierung empfohlen wird (siehe Abschnitt 8.3.5). Für das Interesse Hold liegt der Designeffekt zudem unter dem Grenzwert von deff= 2, sodass demnach die Gruppierungen der Versuche in der Mehrebenenstruktur nicht berücksichtigt werden müssen (siehe Abschnitt 8.3.5). Zumal sich die Fragestellung auf die zeitliche Entwicklung der Personeneinstellung bezieht, wird daher nur die Personengruppierung in der nun folgenden Mehrebenenanalyse der Wachstumsverläufe berücksichtigt.

Die Modellparameter des Wachstumskurvenmodells (siehe Spezifikation Modell A in Tab. 10) für die Catch und Hold-Komponente des situationalen Interesses sind dargestellt in Tabelle 22. Für die Catch-Komponente des situationalen Interesses zeigt sich anhand der zufälligen Effekte des Modells, dass 23 % der Gesamtvarianz zurückgeht auf Unterschiede zwischen Personen am ersten Messzeitpunkt (Versuche 1-4), 77 % auf Unterschiede innerhalb der Personen und keine Varianz auf Unterschiede zwischen den Änderungsraten der Personen (Tab. 22). Die mittlere zeitliche Veränderungsrate (fester Effekt: $\gamma_{10} = .01, p > .05$) ist nicht signifikant von Null verschieden. Abbildung 34a veranschaulicht die zeitlichen Verläufe der nichtstandardisierten Catch-Komponente des Interesses für die einzelnen Personen. Im Mittel ist im Modell keine zeitliche Veränderung des Interesse Catch festzustellen ($g_{10} = .01, p > .05$, Abb. 34a). Für die einzelnen Studierenden liegen Verläufe mit unterschiedlichen Geradensteigungen um Null vor (Range: $-0.07 < b_{1j} < 0.07$, 95 %-Konfidenzintervall: $-0.03 < b_{1j} < 0.02$, Abb. 34a). Ein zeitlicher Effekt über den Praktikumsverlauf auf das situationale Interesse Catch ist damit nicht feststellbar.

Für die Hold-Komponente zeigen die zufälligen Effekte, dass 35 % der Gesamtvarianz des Interesses auf Unterschiede zwischen Personen am ersten Messzeitpunkt, 64 % auf Unterschiede innerhalb der Personen und 1 % auf Unterschied zwischen den Änderungsraten der Personen zurückzuführen sind (Tab. 22). Anhand der festen Effekte ist für das Interesse Hold feststellbar, dass die mittlere zeitliche Veränderungsrate ($\gamma_{10} = .04, p < .01$) signifikant von Null verschieden ist. Der gefundene zeitliche Effekt über den Praktikumsverlauf ist dabei allerdings sehr klein. Abbildung 34b veranschaulicht die zeitlichen Verläufe für die einzelnen Studierenden. Im Vergleich zur Catch-Komponente (Abb. 34a) ist dabei ein geringer Anstieg mit einer unstandardi-

AV: sit. Interesse, UV: Messzeitpunkt

	Sit. Interesse Catch			Sit. Interesse Hold		
	Modell 0P	**Modell 0V**	**Modell A**	**Modell 0P**	**Modell 0V**	**Modell A**
	Person	Versuch		Person	Versuch	
feste Effekte						
zeitliche Änderung (γ_{10})			$\gamma_{10} = -.01$ (.02) $g_{10} = -0.01$ (0.01)			$\gamma_{10} = .04^{**}$ (.01) $g_{10} = 0.03^{**}$ (0.01)
zufällige Effekte						
Varianz zwischen Gruppen (Var(u_{0j}))	0.22	0.04	0.23	0.34	0.01	0.35
Varianz innerhalb Gruppen (Var(e))	0.78	0.96	0.77	0.66	0.99	0.64
Varianz der Steigungen (Var(u_{1j}))			0.00			0.01
Korrelation Ausgangswert und Änderung			-0.21			-0.22
ICC	.22	.04	.23	.34	.01	.35
Designeffekt	5.5	3.4		8.0	1.7	
Anzahl Beobachtungen	1650	1650	1650	1650	1650	1650
Anzahl Gruppen	81	25	81	81	25	81
Devianz	4425	4652	4432	4195	4677	4179

Modell 0: Intercept-Only Modell; **Modell A:** Random-Coefficients-Modell

Signifikanzniveau für feste Effekte: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

Tabelle 22: *Ergebnisse der Mehrebenenanalysen (Intercept-Only-Modell 0 und Random-Coefficients-Modell A, Tab. 10) des zeitlichen Verlaufs des situationalen Interesses Catch und Hold. (γ: standardisiertes Regressionsgewicht, g: nicht standardisiertes Regressionsgewicht, in Klammern: Standardfehler; Intercepts nicht dargestellt, da bei standardisierten Daten nicht sinnvoll interpretierbar)*

sierten Veränderungsrate von $g_{10} = 0.03, p < .01$ erkennbar (Abb. 34b), der für sechs Praktikumsphasen einen Gesamtanstieg der Hold-Komponente von $g_{10} \cdot 6 = 0.18$ Einheiten auf der 5-teiligen Skala bzw. $\gamma_{10} \cdot 6 = .24$ Standardabweichungen bedeutet. Für die einzelnen Personen ergeben sich Entwicklungsverläufe mit Steigungen im 95 %-Konfidenzintervall von $0.01 < b_{1j} < 0.05$ (Range: $-0.05 < b_{1j} < 0.10$, Abb. 34b). Zwischen Ausgangswerten und Veränderungsraten liegt für beide Komponenten des Interesses eine Korrelation von etwa $r = -.2$ vor. Dies bedeutet, dass für Studierende, die ein hohes Interesse zu Beginn des Praktikums haben, das Interesse vermutlich aufgrund des Deckeneffekts über die Zeit tendenziell weniger ansteigt als für Studierende mit einem geringeren Interesse zu Praktikumsbeginn.

Die unterschiedlichen Zeitverläufe des Interesses auf Ebene der einzelnen Studierenden können durch die Konfundierung mit den Versuchen und andere Einflüsse erklärt werden. Die Versuchsreihenfolge wurde in dieser Studie nicht systematisch untersucht, sodass auf kürzeren Zeitskalen, also zwischen einzelnen Versuchen und für bestimmte Versuchsreihenfolgen keine Aussagen zu Übertragungseffekten getroffen werden können. Durch die immer unterschiedliche Abfolge der Versuche für die Studierenden kann geschlussfolgert werden, dass der festgestellte zeitliche Effekt auf das Interesse Hold auf dem gesamten Untersuchungszeitraum nicht durch die Versuchsreihenfolge bestimmt ist.

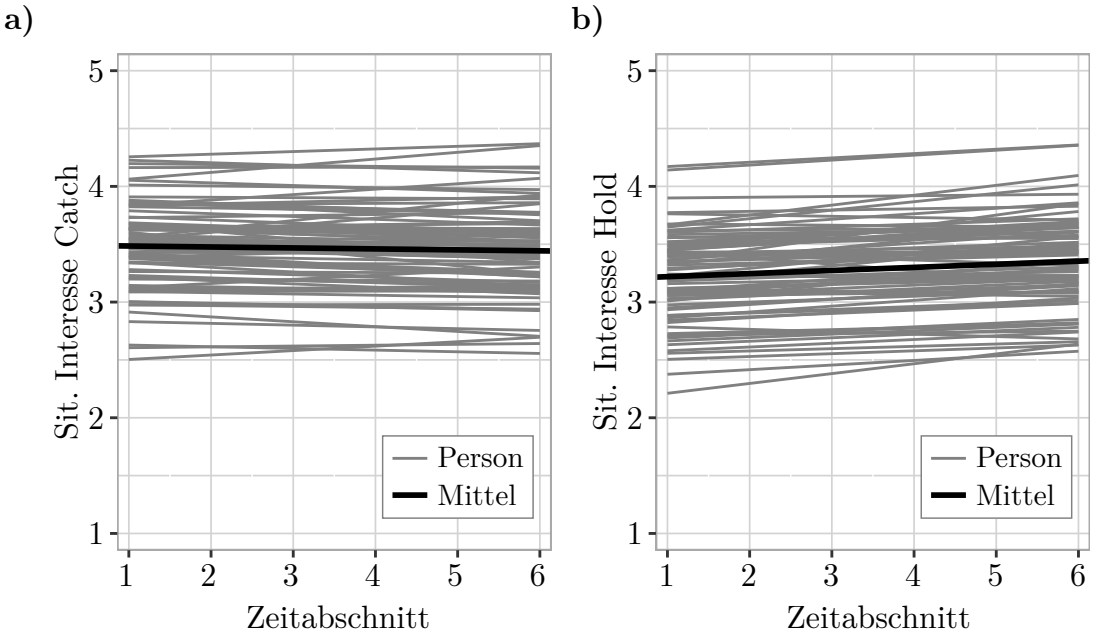


Abbildung 34: *Entwicklungsverläufe aus der Mehrebenenanalyse (Modell A, Tab. 10 und 22) für situationale Interesse Catch (a) und Hold (b) (5-teilige Likert-Skala, ohne Standardisierung) abhängig vom Zeitabschnitt der Beobachtung. Die grauen Linien zeigen die Verläufe für alle einzelnen Personen (Range Catch: $-0.07 < b_{1j} < 0.07$; Range Hold: $-0.05 < b_{1j} < 0.10$). Die dickere schwarze Linie ist der Verlauf über alle Personen (Catch: Ausgangswert im ersten Zeitabschnitt: $g_{oo} = 3.48$, Steigung: $g_{1o} = -0.01$; Hold: Ausgangswert im ersten Zeitabschnitt: $g_{oo} = 3.22$, Steigung: $g_{1o} = 0.03$).*

Anhand der Wachstumskurvenmodelle konnte somit festgestellt werden, dass der zeitliche Verlauf des Praktikums wie auch für die wahrgenommenen Relevanz im Mittel keinen Effekt auf die Catch-Komponente des situationalen Interesses hat. Damit ist das Interesse Catch kurzfristig geprägt durch die Praktikumssituation. Auf die Hold-Komponente konnte hingegen gemäß der Hypothese im Mittel ein signifikanter positiver Effekt gezeigt werden. Allerdings ist dieser Effekt sehr klein, sodass sich der Zuwachs des Interesse Hold über die Zeit wahrscheinlich nicht als Folge regelmäßiger Triggerung des Interesse Catch nachweisen lassen kann. Überlegungen zur Ausnutzung des zeitlichen Effektes für eine optimale Reihung der Versuche lassen damit nicht auf hohe Effekte hoffen. Aufgrund des kleinen Effektes werden zeitliche Effekte für das situationale Interesse in den Folgenden Analysen vernachlässigt und damit die Datenstruktur vereinfacht.

Nachdem diese Voruntersuchungen bzgl. zeitlicher Effekte abgeschlossen sind, wird zunächst ein Überblick über das situationale Interesse in seinen beiden Komponenten für den *Gesamtdatensatz* (Datenebene Abb. 18A), also für alle vorhandenen Kombinationen von 89 Personen und bis zu 24 Versuchen ($n = 1805$), gegeben. Für die Deskription sind die Kennwerte im oberen Abschnitt der Tabelle 23 gelistet und die Verteilungen in den Boxplots in Abbildung 35 dargestellt. Die Boxplots weisen auf eine Schiefe der Verteilungen hin. Ein Shapiro-Test bestätigt, dass keine Normalverteilung für die Daten vorliegt (Catch: $W = 0.98, p < .001$, Hold: $W = 0.99, p < .001$). Anhand der Boxplots wird deutlich, dass der Messbereich der beiden Komponenten zum Interesse voll ausgeschöpft ist. Damit sind Unterschiede des Interesses für die Bewertungen der einzelnen Versuche messbar. Die Mittelwerte für das situationale Interesse Catch und Hold liegen über alle Versuche und Personen hinweg über dem mittleren Maß der 5-teiligen Likert-Skala. Dabei ist der mittlere Wert für die Catch-Komponente mit $M_{CA} = 3.50 \pm 0.02$ größer als für die Hold-Komponente mit $M_{HO} = 3.30 \pm 0.02$. Es ist allerdings zu bedenken, dass die Daten in Bezug auf Personen und Versuche abhängig sind und deshalb die Standardabweichungen und Standardfehler in dieser ersten Beschreibung zu hoch geschätzt wurden (z. B. Field et al., 2012).

Um die Einschätzung des situationalen Interesses der Studierenden bzgl. der einzelnen Praktikumsversuche zu berücksichtigen und näher zu untersuchen, wird analog zur Untersuchung der wahrgenommenen Relevanz (siehe Abschnitt 9.2) die Datenebene gewechselt auf *Zwischen Versuchen* (Abb. 18C). Das situationale Interesse wird somit im Folgenden bezüglich der Versuche aufgeschlüsselt und über Personen gemittelt. Um beim Vergleich der Versuche die Abhängigkeit der Daten in Bezug auf die Studierenden zu berücksichtigen, werden die Personendaten, wie in Abschnitt 8.3.2 beschrieben, um den Personenmittelwert zentriert, sodass Daten um die Personenvarianz korrigiert werden. In den Abbildungen 36 und 37 ist das situationale Interesse Catch und Hold für alle Versuche in Form von Balkendiagrammen dargestellt. Tabelle 23 zeigt im unteren Abschnitt die deskriptive Statistik zur Übersicht für die Versuche, in denen die jeweilige Komponente des situatio-

nalen Interesses maximal und minimal bewertet wird (Deskription für alle Versuche im Anhang A.2.4). Für die Catch-Komponente sind das die Versuche Fallender Doppler-Effekt (Anhang A.1.21) und Feldeffekttransistor (FET) (Anhang A.1.24) mit Mittelwerten von $M_{Doppler} = 3.81 \pm 0.06$ und $M_{FET} = 3.07 \pm 0.11$. Für die Hold-Komponente sind es die Versuche Thermohaus (Anhang A.1.15) und Interferenz am Smartphone (Anhang A.1.11) mit Mittelwerten von $M_{Thermo} = 3.72 \pm 0.08$ und $M_{Interferenz} = 3.15 \pm 0.09$. Die Mittelwerte des Interesses liegen somit für alle Versuche um die Mitte der 5-teiligen Skala oder darüber. Die maximalen und minimalen Mittelwerte des Interesses unterscheiden sich um $\Delta_{CA} = 0.74 \pm 0.17$ für die Catch-Komponente und um $\Delta_{HO} = 0.57 \pm 0.17$ für die Hold-Komponente. Somit liegt der Hauptmessbereich etwas mehr als eine halbe Skaleneinheit oberhalb der Skalenmitte, sodass Decken- oder Bodeneffekte ausgeschlossen werden. Anhand eines Shapiro-Tests (siehe Abschnitt 8.3.3) sind die zentrierten Daten für etwas mehr als die Hälfte der Versuche normalverteilt. Ein Levene-Test (siehe Abschnitt 8.3.3) zeigt, dass die Varianzen der einzelnen Versuche nicht homogen sind (Catch: $F(24, 1780) = 2.46, p < .001$; Hold: $F(24, 1780) = 3.82, p < .001$). Die Standardabweichungen liegen für die Catch-Komponente im Bereich von $0.56 \leq SD_{CA} \leq 0.92$ und für die Hold-Komponente im Bereich von $0.41 \leq SD_{HO} \leq 0.80$, wobei der Versuch zum Roller Coaster in dem Wertebereich wegen der sehr geringen Stichprobe ($n = 9$) nicht berücksichtigt ist. Für beide Komponenten ist die Standardabweichung für den Versuch zum FET am größten. Das bedeutet, dass die Studierenden das Interesse an manchen Versuchen, wie besonders den FET, weniger einheitlich bewertet haben als das Interesse an anderen Versuchen.

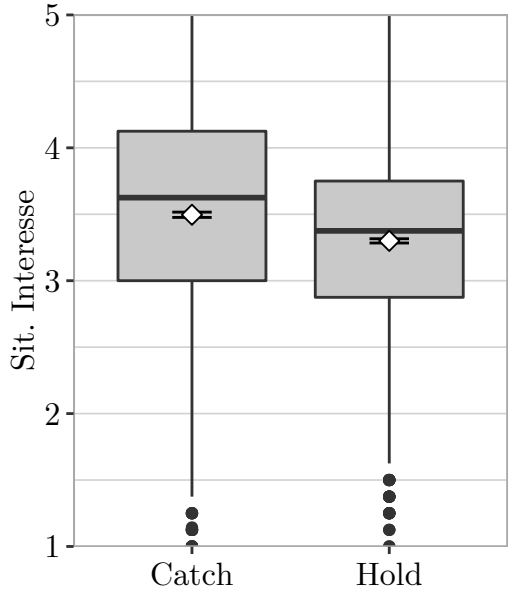


Abbildung 35: *Die Abbildung zeigt das situationale Interesse (5-teilige Likert-Skala) aufgeteilt in die Catch- (links) und Hold-Komponente (rechts) ($n = 1805$, Datenebene Gesamtdatensatz, Abb. 18A). Der Boxplot zeigt dabei neben Minimum, unterem Quartil, Median, oberem Quartil und Maximum zusätzlich die Mittelwerte ($M_{CA} = 3.50 \pm 0.02$, $M_{HO} = 3.30 \pm 0.02$, weiße Rauten) mit Standardfehlern. Im oberen Teil der Tabelle 23 ist die zugehörige deskriptive Statistik dargestellt.*

Skala	Versuch	n	M	SD	SE	MIN	MAX
Catch	alle	1805	3.50	0.85	0.02	1.00	5.00
Hold	alle	1805	3.30	0.68	0.02	1.00	5.00
Catch (Z)	Fallender Doppler-Effekt	78	3.81	0.57	0.06	2.48	5.08
Catch (Z)	Feldeffekttransistor	76	3.07	0.92	0.11	1.06	4.86
Hold (Z)	Thermohaus	65	3.72	0.67	0.08	2.00	5.65
Hold (Z)	Interferenz am Smartphone	73	3.15	0.78	0.09	1.23	5.21

Tabelle 23: *Deskriptive Statistik für das situationale Interesse Catch und Hold. Im oberen Teil sind die Kennwerte für die unzentrierten Daten über alle Kombinationen von Personen und Versuchen (Datenebene Gesamtdatensatz, Abb. 18A) tabelliert. Eine Visualisierung befindet sich in Abbildung 35. Im unteren Teil der Tabelle sind die deskriptiven Werte für die Versuche mit jeweils maximalem und minimalem zentriertem (Z) situationalen Interesse gelistet (Datenebene Zwischen Versuchen, Abb. 18C). Eine Darstellung für alle Versuche ist in den Abbildungen 36 und 37 gezeigt. Die Details für deskriptive Werte weiterer Versuche sind im Anhang A.2.4 zu finden.*

Nach der Deskription soll nun geprüft werden, ob sich die einzelnen Versuche hinsichtlich der Komponenten des situationalen Interesses signifikant unterscheiden. Aus den Balkendiagrammen in den Abbildungen 36 und 37 für das zentrierte situationale Interesse Catch und Hold aufgeschlüsselt auf die einzelnen Praktikumsversuche wird ersichtlich, dass besonders im mittleren Wertebereich des Interesses die Versuche sich rein visuell im Rahmen des Standardfehlers wenig unterscheiden. Um belastbare Aussagen treffen zu können, werden daher analog zur Analyse der wahrgenommenen Relevanz (Abschnitt 9.2) unabhängige einfaktorielle Varianzanalysen für das zentrierte situationale Interesse Catch und Hold bzgl. der Gruppierung in Praktikumsversuche berechnet (siehe auch Abschnitt 8.3.4). Die Varianzanalysen zeigen, dass es signifikante Unterschiede zwischen den Versuchen für das situationale Interesse Catch ($F(24, 1780) = 5.72, p < .001, \eta^2 = .07$) und Hold ($F(24, 1780) = 2.91, p < .001, \eta^2 = .04$) mit mittlerem bzw. kleinem Effekt gibt. Die Unterschiede werden durch die Kruskal-Wallis-Tests für die nichtparametrischen Daten (siehe oben) bestätigt (Catch: $H(24) = 127.6, p < .001$, Hold: $H(24) = 64.16, p < .001$).

Zur Identifikation der Praktikumsversuche mit signifikanten Unterschieden wurden paarweise unabhängige t-Tests und Wilcoxon-Rangsummentests mit Bonferroni-Korrektur als Post-Hoc-Tests mit den zentrierten Daten gerechnet. Beide Tests zeigen konsistente Ergebnisse. Die Abbildungen 38 und 39 zeigen in einer Matrix paarweise für die Versuche die Effektstärke (Cohens d) und das Signifikanzniveau des t-Tests für das situationale Interesse Catch und Hold. Die Versuche sind dabei nach absteigendem Interesse der jeweiligen Komponente sortiert. Für das Interesse Catch in Abbildung 38 sind für diese Stichprobe und die mehrfachen Vergleiche tendenziell Unterschiede ab einer mittleren Effektstärke von Cohens $d > 0.5$ signifikant. Signifikante Unterschiede liegen damit vor allem zwischen den Versuchen mit höchstem

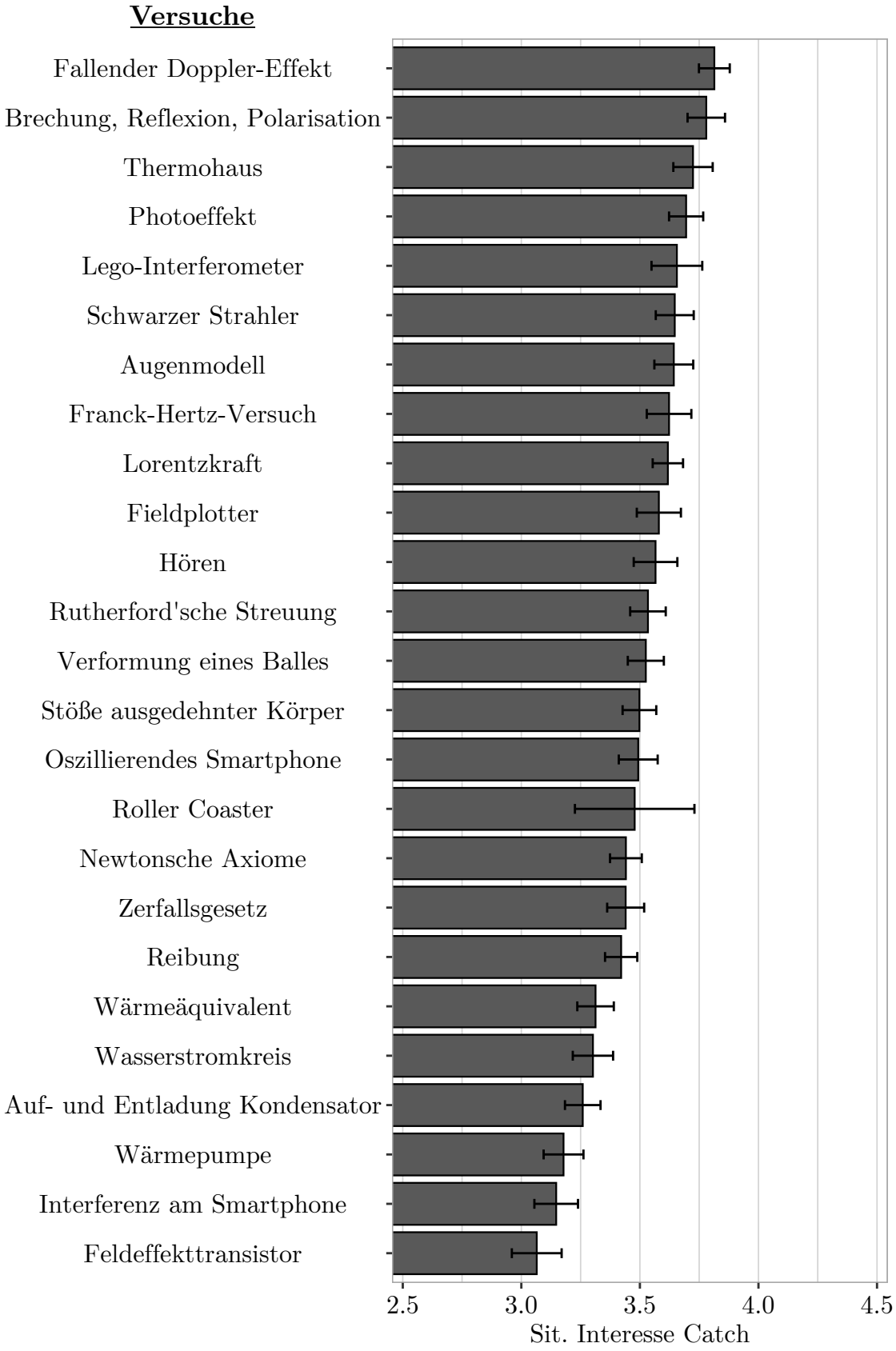


Abbildung 36: *Darstellung des* ***zentrierten situationalen Interesses Catch*** *(5-teilige Likert-Skala) für die einzelnen Praktikumsversuche (*$65 \leq n \leq 89$*,* $n_{RollerCoaster} = 9$*). Das Balkendiagramm zeigt Mittelwerte und Standardfehler.*

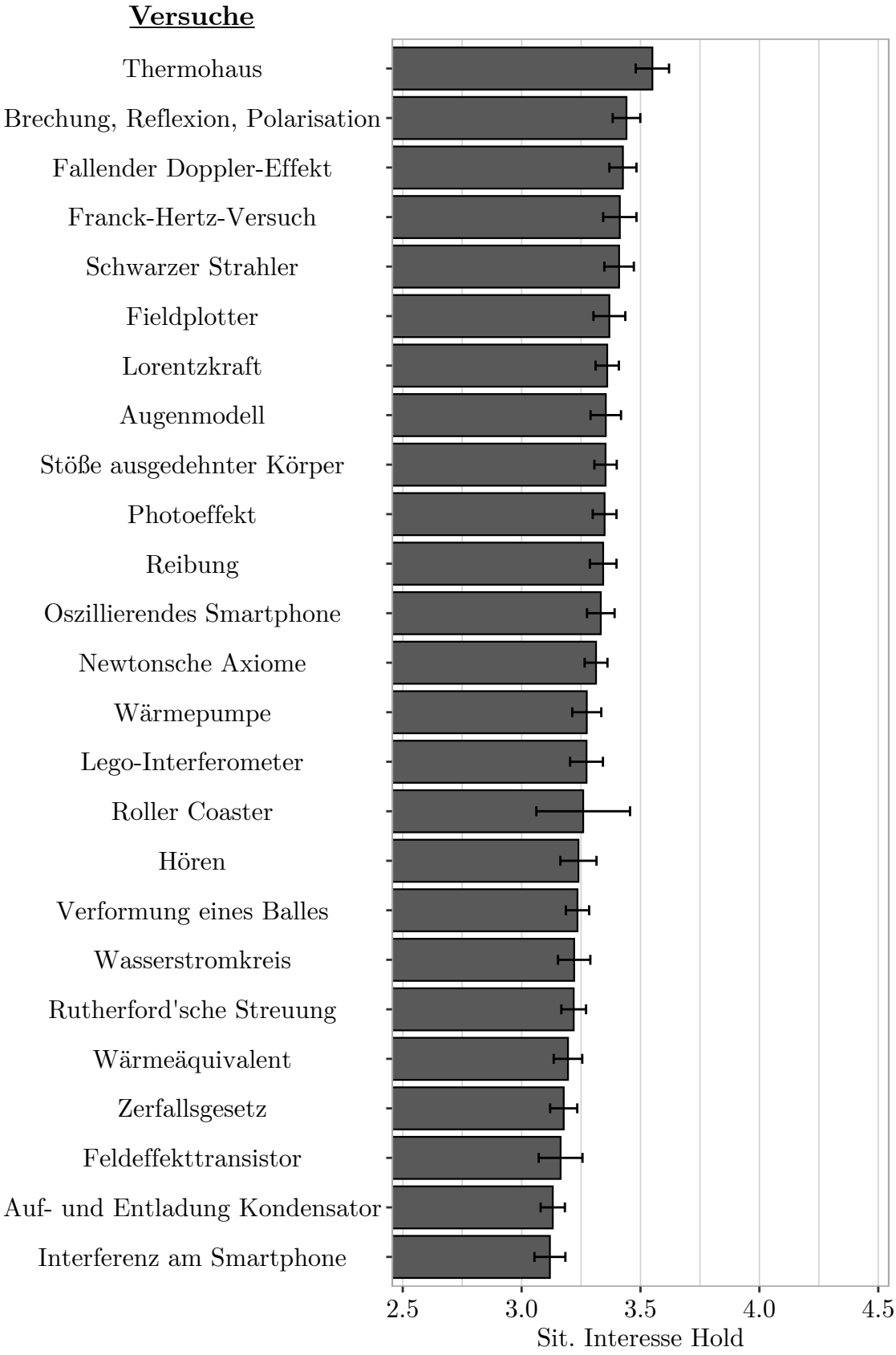


Abbildung 37: *Darstellung des* ***zentrierten situationalen Interesses Hold*** *(5-teilige Likert-Skala) für die einzelnen Praktikumsversuche (*$65 \leq n \leq 89$, $n_{RollerCoaster} = 9$*). Das Balkendiagramm zeigt Mittelwerte und Standardfehler.*

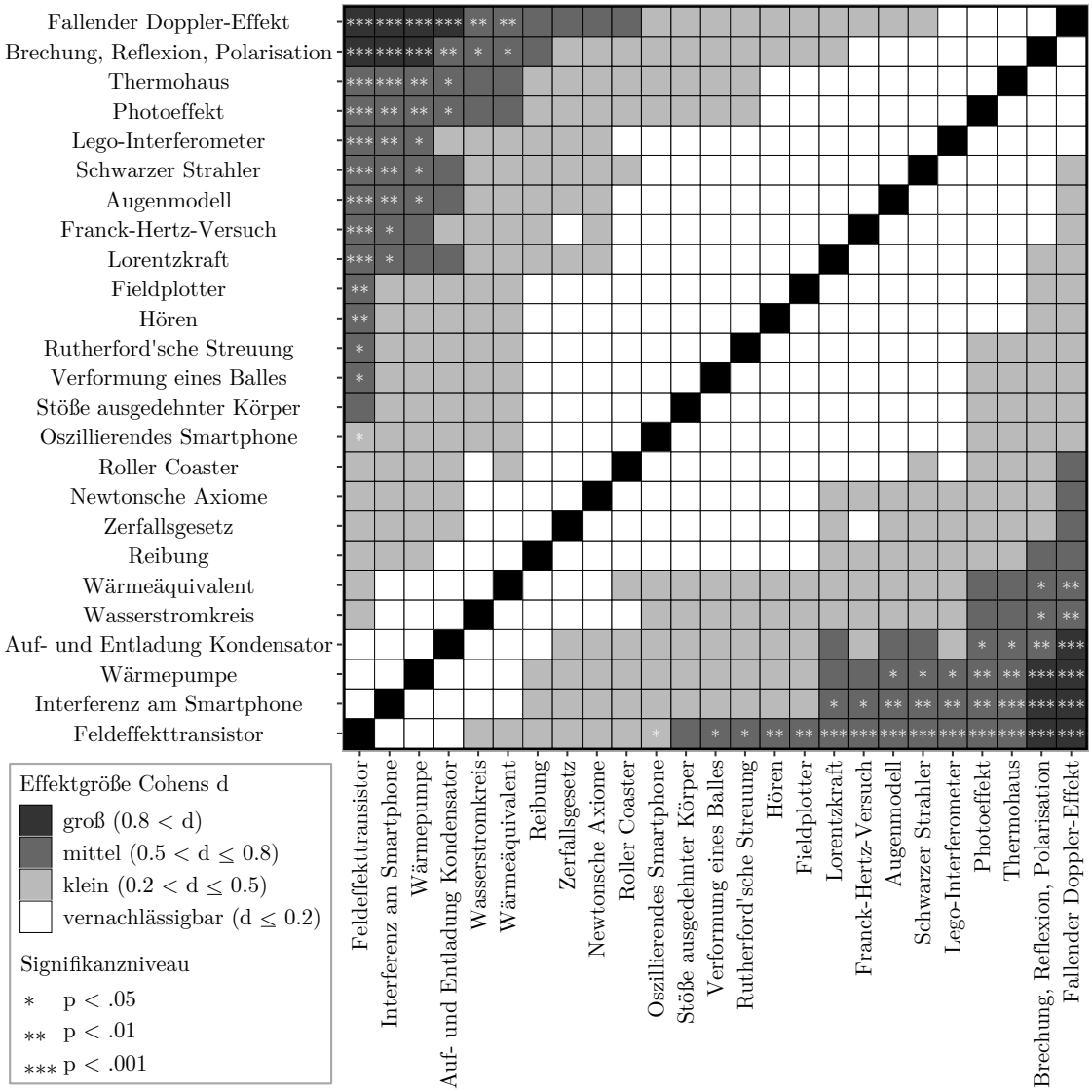


Abbildung 38: *Für die 25 Praktikumsversuche wurden paarweise Mittelwertvergleiche für das* ***zentrierte situationale Interesse Catch*** *mittels unabhängigen t-Test mit Bonferroni-Korrektur berechnet ($60 \leq n \leq 89$, $n_{RollerCoaster} = 9$). Angegeben sind für die Versuchskombinationen in Graustufen kodiert die Effektgröße Cohens d (Kategorien orientiert an Cohen (1988), siehe Abschnitt 8.3.4) und symbolisch kodiert das Signifikanzniveau p des t-Tests.*

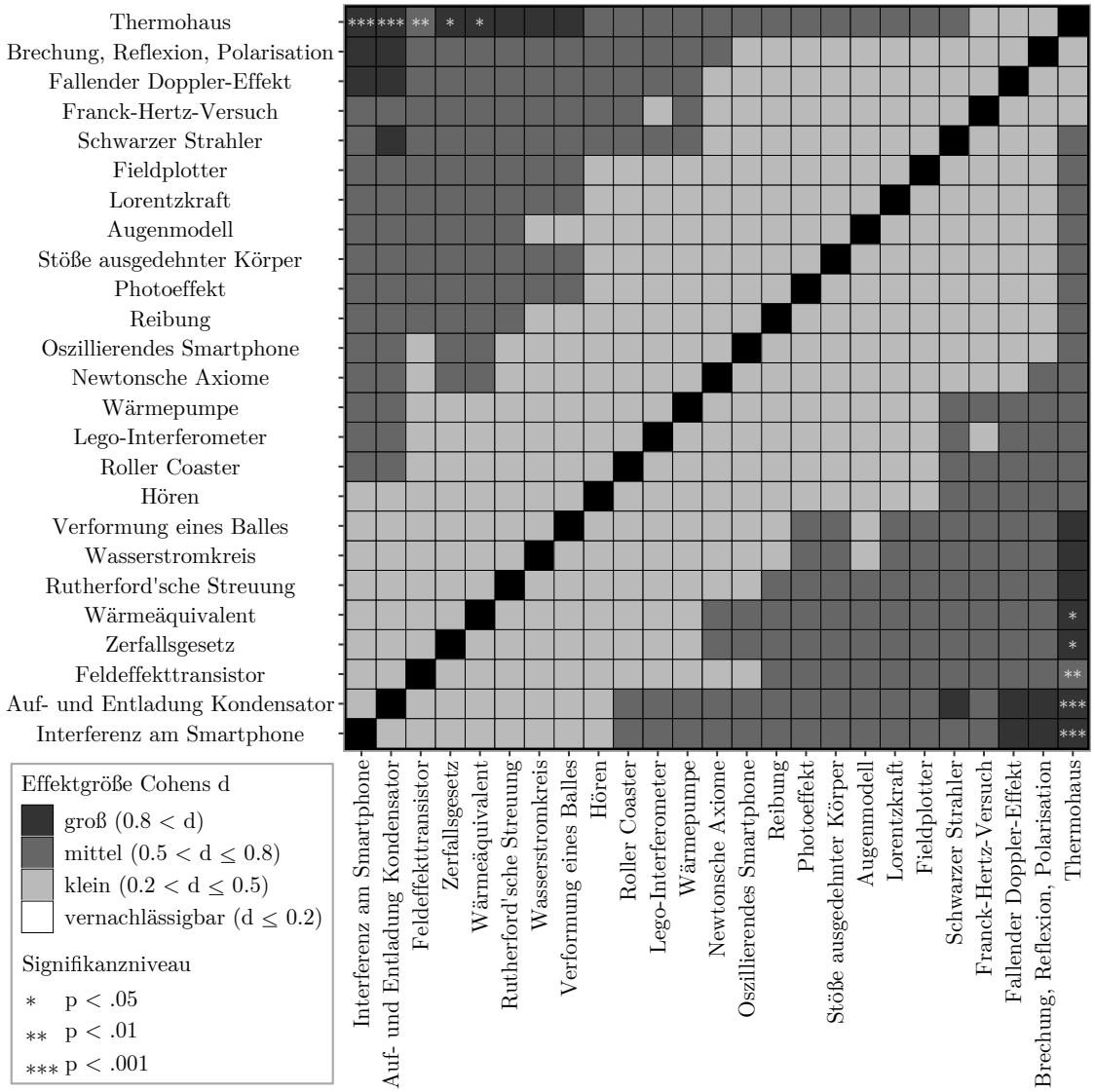


Abbildung 39: *Für die 25 Praktikumsversuche wurden paarweise Mittelwertvergleiche für das* ***zentrierte situationale Interesse Hold*** *mittels unabhängigen t-Test mit Bonferroni-Korrektur berechnet (*$60 \leq n \leq 89$, $n_{RollerCoaster} = 9$*). Angegeben sind für die Versuchskombinationen in Graustufen kodiert die Effektgröße Cohens d (Kategorien orientiert an Cohen (1988), siehe Abschnitt 8.3.4) und symbolisch kodiert das Signifikanzniveau p des t-Tests.*

und niedrigstem Interesse vor. Die Praktikumsversuche im mittleren Wertebereich unterscheiden sich, wie die Deskription in Abbildung 36 vermuten ließ, nicht signifikant hinsichtlich des mittleren Interesses Catch. Somit sind Unterschiede zwischen dem Interesse an den Praktikumsversuchen vorhanden, wenn auch nur für die Versuchen mit den stärksten minimalen und maximalen Ausprägungen des Interesses. Mit Rückblick auf die im Rahmen des Standardfehlers ähnlichen Mittelwerte für die Versuche im mittleren Wertebereich in Abbildung 36 sind diese Ergebnisse plausibel. In Hinblick auf das Interesse Hold in Abbildung 38 unterscheidet sich nur der Versuch zum Thermohaus mit dem größten Interesse signifikant von den fünf Versuchen mit dem niedrigsten Interesse Hold. Für die Hold-Komponente wurden damit deutlich weniger signifikante unterschiedliche Versuche identifiziert als für die höhere Catch-Komponente. In Bezug auf das Konstrukt ist dies ein plausibles Ergebnis, da die Catch-Phase des Interesses kurzfristigere und damit noch versuchsspezifischere Aspekte umfasst als die langfristigere Hold-Phase (siehe Abschnitt 3.2).

Insgesamt zeigt diese Analyse des situationalen Interesses, dass die Praktikumsversuche über dem Mittelmaß als interessant wahrgenommen werden, was in Bezug auf die Neuentwicklung der lehramtsspezifischen Praktikumsversuche sehr positiv zu bewerten ist. Die Praktikumsversuche werden in Bezug auf das situationale Interesse zu einem Teil signifikant unterschiedlich von den Studierenden bewertet. Dabei sind für die geringere Hold-Komponente gemäß der Theorie zum Konstrukt deutlich weniger Unterschiede als für die höhere Catch-Komponente feststellbar. Der Varianzanteil, der auf Unterschiede im Interesse zwischen Versuchen zurückzuführen ist, liegt nur bei 4 % für die Catch-Komponente und bei 1 % für die Hold-Komponente (Tab. 22). Im Gegensatz zum geringen Varianzanteil durch Unterschiede der Versuche resultieren für das Interesse Catch 22 % und für Hold 34 % der Gesamtvarianz aus Unterschieden zwischen Personen. Damit sind auf Ebene der Personeneinschätzungen weiterführende Analysen sehr vielversprechend.

Schlussfolgerungen für das Interesse im Wirkmodell

Im Wirkmodell ist das individuelle Interesse ein wichtiger Aspekt des Lernerfolgs. Das individuelle Interesse an Physik kann dabei zu Beginn eines adressatenspezifischen Physikpraktikums je nach Studierendengruppe unterschiedlich ausgeprägt sein. Allein aufgrund der Studienwahl ist es naheliegend, dass Studierende von physik-nahen Studiengängen wie beispielsweise Lehramt Physik oder Meteorologie (T. Haller, 2011) ein höheres und stabiles Grundinteresse an Physik haben als von physik-fernen Studienrichtungen wie zum Beispiel Medizin. Es wurde vorangehend gezeigt, dass die Studierenden im Lehramtspraktikum erwartungsgemäß ein sehr ausgeprägtes individuelles Interesse am Fach Physik und am Unterrichten von Physik haben. Damit ist nicht verwunderlich, dass sich dieses hohe, recht allgemein gehaltene und aufgrund der Studienwahl vermutlich bereits über einen längeren Zeitraum bestehende Grundinteresse an Physik durch das Lehramtspraktikum, als

eine von mehreren physikalischen Lehrveranstaltungen, im Mittel nicht nachweisbar verändert. Für physik-fernere Studienrichtungen wie beispielsweise Medizin könnte sich das individuelle Interesse an Physik auch anders entwickeln. Zudem bringt das stark ausgeprägte individuelle Interesse im Lehramtspraktikum messtechnische Herausforderungen wie Deckeneffekte mit sich, sodass geringe Änderungen des individuellen Interesses auch schwer mit den hier eingesetzten Instrumenten nachweisbar sein dürften. Aus diesem Grund ist die Operationalisierung des Wirkmodells für das Lehramtspraktikum in Bezug auf das Interesse noch einmal vor dem Hintergrund der vorangegangenen Analyse der verschiedenen Aspekte des Interesses zu bewerten.

Gemäß der theoretischen Grundidee des Konstruktes des situationalen Interesses wird das Interesse in der ersten Phase der Interessensentwicklung kurzfristig angeregt (Catch), um anschließend ggf. langfristiger aufrecht erhalten zu werden (Hold) und sich ggf. schließlich in ein stabileres individuelles Interesse zu entwickeln (z. B. Hidi und Renninger, 2006; Knogler et al., 2015, siehe auch Abschnitt 3.2). Vorangehend wurde gezeigt, dass das situationale Interesse Hold für die einzelnen Versuche einheitlicher bewertet wurde als das Interesse Catch. Zudem konnte der Trend beobachtet werden, dass sich das Interesse Hold über den Verlauf des Praktikums erhöht. Für das situationale Interesse Hold stimmen diese Ergebnisse mit den theoretischen Überlegungen überein, dass die Hold-Komponente einen langfristigeren Charakter als die Catch-Komponente hat. Da das situationale Interesse Hold gemäß dem Vier-Phasen-Modell als Vorstufe für das individuelle Interesse gilt (Hidi & Renninger, 2006), kann das situationale Interesse Hold als Proxy für das individuelle Interesse herangezogen werden, wie auch in Abschnitt 3.2 erläutert. Hinter dem Konstrukt des Interesses Hold verbirgt sich der persönliche Wert und der Wunsch, über das Praktikum hinaus mehr über die Inhalte der Versuche erfahren zu wollen. Somit umfasst das situationale Interesse Hold das aufrecht erhaltene Interesse aufgrund des Praktikumsversuches. Das situationale Interesse Hold kann zudem in gewissem Maß als Proxy für aufrecht erhaltene Lernmotivation aufgefasst werden, da Interesse ein Bedingungsfaktor für Motivation ist (z. B. Rheinberg et al., 2001; Schiefele, 2008). Insbesondere der epistemische Charakter dieser Interessen-Komponente kann als Proxy für langfristigere Lernmotivation angehender Lehrkräfte im Praktikum aufgefasst werden. Aufgrund der messtechnischen Herausforderungen des individuellen Interesses und vorangegangener Analysen und Überlegungen bzgl. des situationales Interesses Hold wird das situationale Interesse Hold als Proxy für die Entwicklung von individuellem Interesse durch die Praktikumsversuche aufgefasst und in den folgenden Analysen des Wirkmodells als motivationaler Lernerfolg betrachtet.

Das Interesse spielt im Wirkmodell auch eine zweite Rolle. Dabei ist das situationale Interesse Vermittler zwischen wahrgenommener Relevanz und Lernerfolg. Die vorangegangenen Analysen des situationalen Interesses Catch geben Hinweise darauf, dass das situationale Interesse Catch gemäß der Theorie als kurzfristiger und stärker situationsbezogen, also stärker versuchsbezogen als die Hold-Komponente aufgefasst werden kann. Damit fügt sich die kurzfristigere Catch-Komponente des

Interesses gut als Prädiktor für die im Praktikumsversuch angeregte Lernmotivation und damit als vermittelnde Größe in das Wirkmodell ein, wie in Abschnitt 3.2 bereits ausgeführt.

Somit sind die Rollen des Interesses im Wirkmodell des Lehramtspraktikums anhand der vorangegangenen Überlegungen und Untersuchungen wie folgt zu besetzen: Das kurzfristige situationale Interesse Catch ist der vermittelnde Aspekt des Interesses und das langfristigere situationale Interesse Hold wird als Proxy für das individuelle Interesse und somit als Lernerfolg im Praktikum gewertet.

12.1.2 Wirkmodell mit Interesse als Lernerfolg

Vorangehend wurden die verschiedenen Aspekte des erhobenen Interesses analysiert und damit die ersten Schritte in Richtung Prüfung des Wirkmodells für das Interesse als Komponente des Lernerfolgs gemacht. Damit können nun in den nächsten Schritten die paarweisen Zusammenhänge zwischen der wahrgenommenen Relevanz und dem situationalen Interesse analysiert sowie das angenommene Wirkmodell mit Bezug auf der Interesse insgesamt (Wirkmodell A) geprüft werden.

Voraussetzung für die Modellierung der Wirkannahmen in Bezug auf das Interesse ist, dass zwischen Paaren der einzelnen Merkmale wahrgenommene Relevanz und situationales Interesse Catch und Hold Zusammenhänge bestehen (siehe Abschnitt 11.2.2). Aus den vorangegangenen Analysen kann bereits auf Parallelen der Merkmale, insbesondere für Versuche mit niedrigen Merkmalsausprägungen, geschlossen werden (Abb. 22, 36, 37). Somit ist eine weiterführende Analyse der Zusammenhänge der Merkmale zur Prüfung des Wirkmodells vielversprechend. Für die Prüfung auf einzelne Zusammenhänge werden unter Berücksichtigung der Datenstruktur Mehrebenen-Regressionsanalysen durchgeführt. Um zu entscheiden, welche Level der Mehrebenenstruktur dabei berücksichtigt werden müssen, wird Rückbezug auf die Betrachtungen der Nullmodelle in den Abschnitten 9.1 und 12.1 genommen (Tab. 11 und Tab. 22). Wie schon in den vorherigen Mehrebenenanalysen wird die Gruppierung in Versuche zur Vereinfachung des Modelles auch im Folgenden vernachlässigt, da der Varianzanteil für Versuche für alle drei Merkmale sehr klein ist ($ICC < .10$) und die Fragestellung sich auf Personeneinschätzungen bezieht.

Die einzelnen Zusammenhänge der wahrgenommenen Relevanz und des situationalen Interesses werden anhand der Mehrebenenanalysen auf den zwei Datenebenen *Zwischen Personen* und Beobachtungen *Innerhalb Personen* untersucht (siehe Abb. 18). Das Modell spezifiziert sich wie in der Analyse der wahrgenommenen Relevanz in Abschnitt 9.4 (siehe Modell 1 in Tab. 9) als Random-Intercept-Modell mit dem zentrierten Merkmal als unabhängige Variable auf Level 1 und dem Personenmittelwert des Merkmals als unabhängige Variable auf Level 2. Die geschätzten Modellparameter der Mehrebenenanalyse sind in Tabelle 24 dargestellt. Die zufälligen Effekte der Modelle zeigen, dass für die Merkmalskombinationen rund 20 % bis 30 % der Gesamtvarianz auf Unterschiede zwischen Personen und der restliche Anteil auf

Unterschiede innerhalb der Personen zurückgeht. Auf der Betrachtungsebene *Zwischen Personen* hat die wahrgenommene Relevanz einen positiven mittleren Effekt von $\gamma_{01} = .38, p < .001$ auf das Interesse Catch und einen positiven mittleren Effekt von $\gamma_{01} = .48, p < .001$ auf das Interesse Hold. *Innerhalb Personen* hat die wahrgenommene Relevanz einen positiven mittleren Effekt von $\gamma_{10} = .44, p < .001$ sowohl auf die Catch-, als auch auf die Hold-Komponente des Interesses. Für den Zusammenhang der beiden Komponenten des situationales Interesses untereinander ergibt sich auf der Ebene *Zwischen Personen* ein positiver mittlerer Effekt von $\gamma_{01} = .48, p < .001$ und auf der Ebene *Innerhalb Personen* ein positiver hoher Effekt von $\gamma_{10} = .60, p < .001$. Die betrachteten Zusammenhänge sind damit signifikant. Im Vergleich zum Nullmodell werden durch die Aufnahme der Prädiktoren in Modell 1 jeweils mehr als 20 % der Varianz auf Level 1 und 50 % der Varianz auf Level 2 zusätzlich aufgeklärt.

Damit sind anhand der festgestellten signifikanten Zusammenhänge zwischen den Merkmalen die Voraussetzungen erfüllt, um im nächsten Schritt die einzelnen Wirkzusammenhänge bezüglich des Interesses in einem Gesamtmodell unter Berücksichtigung aller Merkmale zusammenzuführen.

Das Mehrebenen-Strukturgleichungsmodell (siehe Abschnitte 11.2.5 und 11.2.7) zur Prüfung des Mediationsmodells mit Interesse als Lernerfolg (Wirkmodell A) ist in Abbildung 40 dargestellt. Entsprechend der Untersuchung der paarweisen Zusammenhänge der Merkmale ist auch in diesem Modell die Gruppierung der Personen berücksichtigt und wird auf den Ebenen *Zwischen Personen* und *Innerhalb Personen* betrachtet. Zur Beurteilung der Modellgüte werden die in Abschnitt 11.2.3 vorgestellten Fit-Kennwerte herangezogen. Der χ^2-Test ($\chi^2(111) = 1220, p < .001$) ist signifikant und deutet damit auf eine geringere Passung des Modells hin. Die SRMR-Werte für beide Datenebenen weisen hingehen auf eine gute Modellpassung hin ($\text{SRMR}_{zwischen} = .08$, $\text{SRMR}_{innerhalb} = .06$), während die Fit-Indizes RMSEA= .08 und CFI= .91 eher am jeweiligen Grenzwert für gute Modellpassungen liegen. Insgesamt wird anhand dieser Werte geschlossen, dass das Modell die Daten damit zumindest akzeptabel widerspiegelt und die Ergebnisse damit interpretierbar sind.

Die Datenebene *Zwischen Personen* wird als zeitlich konstanter Anteil der Wahrnehmung der einzelnen Personen, also als eine Art Grundeinstellung interpretiert. Alle Pfadkoeffizienten sind auf dieser Datenebene *Zwischen Personen* hoch signifikant (Abb. 40 oben). Dabei hat die wahrgenommene Relevanz einen signifikanten Gesamteffekt von $\gamma_{tot,zw} = .79$ auf das situationale Interesse Hold. Dieser Effekt teilt sich zu 59 % auf den direkten Effekt ($\gamma_{di,zw} = .46, p < .001$) und 41 % auf den indirekten Effekt ($\gamma_{ind,zw} = .32, p < .001$) auf. Der indirekte, über das situationale Interesse Catch mediierte Effekt ergibt sich dabei aus einem signifikanten Effekt von $\gamma_{zw} = .77$ zwischen Relevanz und Interesse Catch und einem signifikanten Effekt von $\gamma_{zw} = .42$ zwischen Interesse Catch und Interesse Hold. Die Ergebnisse auf der Datenebene *Zwischen Personen* bedeuten, dass Studierende, die Praktikumsversuche

generell als relevanter wahrnehmen als andere, auch grundsätzlich kurzfristig und langfristiger interessierter an den Inhalten sind.

Die Ebene *Innerhalb Personen* entspricht dem zeitlich variablen Anteil des Konstruktes, der von der Situation, also dem Praktikumsversuch, abhängt. Für die Datenebene *Innerhalb Personen* sind ebenfalls alle Pfadkoeffizienten signifikant (Abb. 40 unten). Der direkte Effekt der wahrgenommenen Relevanz auf das Interesse Hold ist dabei $\gamma_{di,in} = .28$ und der indirekte Effekt $\gamma_{ind,in} = .43$. Letzteren Effekt erhält man aus einem signifikanten Effekt von $\gamma_{in} = .60$ zwischen Relevanz und Interesse Catch und einem signifikanten Effekt von $\gamma_{in} = .71$ zwischen Interesse Catch und Hold. Der Gesamteffekt teilt sich von der Größe damit zu 39 % auf den direkten Effekt und 61 % auf den indirekten Effekt auf. Dabei liegt ein Gesamteffekt von wahrgenommener Relevanz auf das Interesse Hold von $\gamma_{tot,in} = .71$ vor. Auf der Datenebene *Innerhalb Personen* wird mit 83 % ein noch höherer Anteil der Varianz des motivationalen Lernerfolgs aufgeklärt als für die Datenebene *Zwischen Personen*. Diese Ergebnisse auf der Datenebene *Innerhalb Personen* bedeuten, dass eine Person, die einen Versuch relevanter wahrnimmt als einen anderen, diesen Versuch und dessen Inhalte kurzfristig und langfristiger auch als interessanter einstuft.

Für beide Datenebenen zeigt sich ein signifikanter Gesamteffekt der wahrgenommenen Relevanz auf das situationale Interesse Hold mit einer hohen Varianzaufklärung des Lernerfolgs Interesse Hold. Während der direkte Effekt auf der Datenebene *Zwischen Personen* größer ist, ist auf der Datenebene *Innerhalb Personen* der indirekte Effekt größer. Diese Ergebnisse zeigen insgesamt, dass die wahrgenommene Relevanz einen hohen positiven Effekt auf das situationale Interesse Catch und Hold auf beiden Datenebene hat. Aus den Unterschieden der indirekten Teileffekte auf den Datenebenen kann man zusätzlich schließen, dass das Interesse Catch dem kurzfristigeren Charakter dieses Konstrukts entsprechend bei der Wahrnehmung der einzelnen Praktikumsversuche (*Innerhalb Personen*) eine größere Rolle spielt als bei der grundsätzlichen Wahrnehmung des Praktikums (*Zwischen Personen*).

Übergeordnet ist aus dieser Analyse (Wirkmodell A) zu schließen, dass das theoriebasierte Wirkmodell in Bezug auf das Interesse als motivationale Komponente des Lernerfolgs im Lehramtspraktikum durch belastbare empirische Ergebnisse eindeutig gestützt wird.

Unabhängige Variable **Abhängige Variable**	**Relevanz** **Interesse Catch**	**Relevanz** **Interesse Hold**	**Interesse Catch** **Interesse Hold**
feste Effekte			
Zwischen Personen (γ_{01})	$\gamma_{01} = .38^{***}(0.04)$	$\gamma_{01} = .48^{***}(0.04)$	$\gamma_{01} = .48^{***}(0.04)$
	$g_{01} = 0.89^{***}(0.09)$	$g_{01} = 0.90^{***}(0.08)$	$g_{01} = 0.74^{***}(0.06)$
Innerhalb Personen (γ_{10})	$\gamma_{10} = .44^{***}(0.02)$	$\gamma_{10} = .44^{***}(0.02)$	$\gamma_{10} = .60^{***}(0.01)$
	$g_{10} = 0.67^{***}(0.03)$	$g_{10} = 0.54^{***}(0.02)$	$g_{10} = 0.56^{***}(0.01)$
zufällige Effekte			
Varianz zwischen Personen ($\text{Var}(u_{0j})$)	0.10	0.12	0.12
Varianz innerhalb Personen ($\text{Var}(e)$)	0.57	0.46	0.29
ICC	.15	.21	.29
Anzahl Beobachtungen	1791	1791	1791
Anzahl Personen	89	89	89
Pseudo-R^2 Level 1	0.23	0.38	0.63
Pseudo-R^2 Level 2	0.59	0.51	0.50

Modell 1: Random-Intercept-Modell

Signifikanzniveau für feste Effekte: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

Tabelle 24: *Ergebnisse der Mehrebenenanalyse (Random-Intercept-Modell, analog zu Modell 1 in Tab. 9) zur Untersuchung der paarweisen Zusammenhänge zwischen den Merkmalen wahrgenommene Relevanz und situationales Interesse Catch und Hold. (γ: standardisierte Regressionsgewichte, g: nicht standardisierte Regressionsgewichte, in Klammern: Standardfehler; Intercepts wegen Standardisierung nicht dargestellt)*

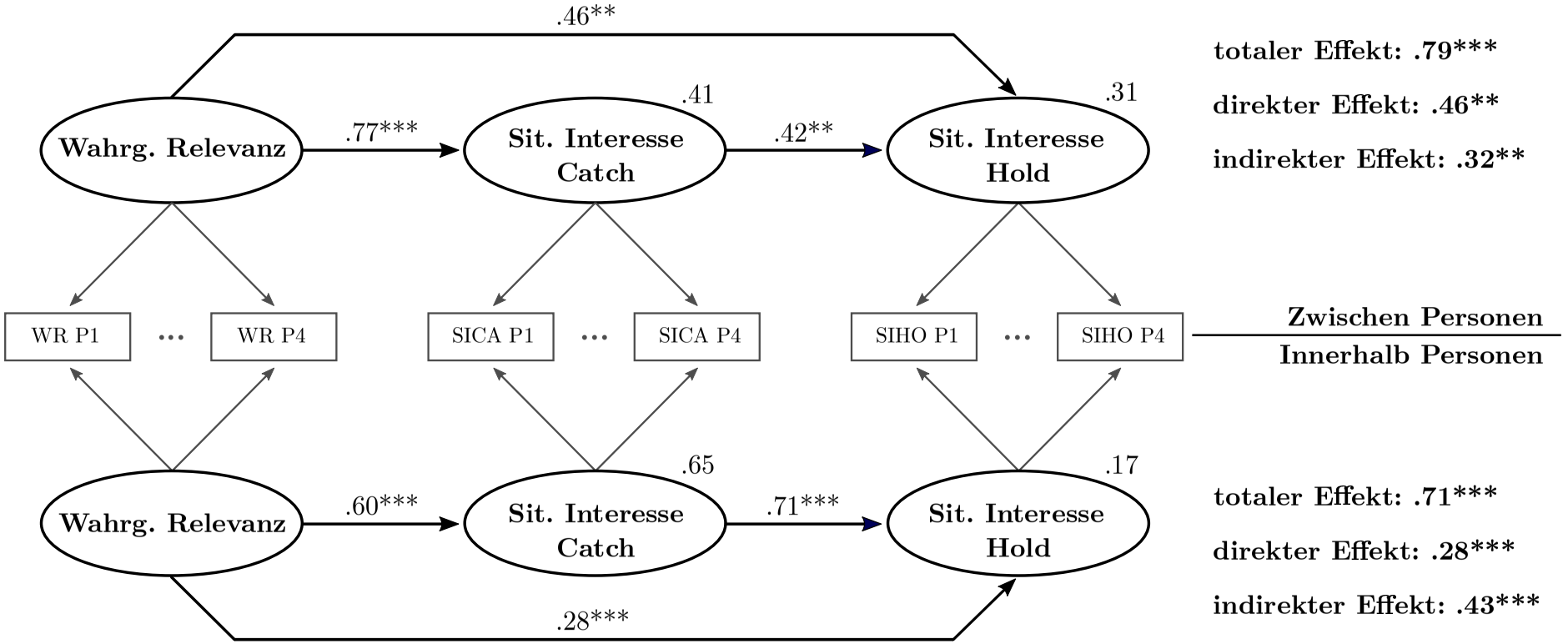


Abbildung 40: *Ergebnis des Mehrebenen-Strukturgleichungsmodells der Mediation von situationalem Interesse Catch zwischen wahrgenommener Relevanz und situationalem Interesse Hold (Wirkmodell A). In der oberen Hälfte der Abbildung ist die Ebene Zwischen Personen und in der unteren Hälfte die Ebene Innerhalb Personen dargestellt. Abgebildet sind die standardisierten Koeffizienten. Die Werte an den abhängigen Variablen stellen die nicht aufgeklärte Varianz dar. Das Signifikanzniveau ist an folgender Kodierung zu erkennen:* $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$. *Für das Modell liegen folgende Fit-Kennwerte vor:* χ^2*-Test:* $\chi^2(111) = 1220, p < .001$, $SRMR_{zwischen} = .08$, $SRMR_{innerhalb} = .06$, *RMSEA*= .08, *CFI*= .91.

12.2 Experimentelle Kompetenz

Nach der Prüfung des Wirkmodells für das Interesse wird nun die experimentelle Kompetenz als zentrale Komponente des Lernerfolgs im Praktikum unter die Lupe genommen. Dazu werden sowohl eine Selbsteinschätzung der Zunahme der experimentellen Kompetenzen als auch die Selbstwirksamkeitserwartungen in Bezug auf das Experimentieren im Unterricht herangezogen, wie in Abschnitt 11.1.2 erläutert. Die Analysen orientieren sich jeweils an den Analyseschritten, die in Abschnitt 11.2.2 aufgestellt wurden. Zunächst ist jeweils der Lernerfolg als Zielgröße selbst zu analysieren, bevor dann die einzelnen Wirkzusammenhänge untersucht werden und abschließend das gesamte Wirkmodell in Bezug auf die experimentelle Kompetenz (Wirkmodell B) und Selbstwirksamkeitserwartungen (Wirkmodelle C und D) geprüft wird.

12.2.1 Zunahme experimenteller Kompetenz

Zur Deskription ist die Selbsteinschätzung des Zuwachses an experimenteller Kompetenz in Abbildung 41 als Boxplot für die zwei Messzeitpunkte Post 1 und Post 2 nach dem ersten bzw. zweiten Semester dargestellt. Ergänzend finden sich deskriptive Werte in Tabelle 25. Die Zusammenfassung der Stichprobe ist dabei zulässig, da Varianzanalysen im Vorfeld gezeigt haben, dass die Daten keine signifikanten Unterschiede bzgl. der Semesterart (Winter-/Sommersemester)[8] und der Kohorte (Sommersemester 2016 - 2019)[9] aufweisen. Der Zuwachs der experimentellen Kompetenz ist im Mittel für die beiden Messzeitpunkte mit $M_{Post1} = 4.68 \pm 0.07$ und $M_{Post2} = 4.86 \pm 0.07$ auf einer 6-teiligen Likert-Skala hoch (Tab. 25). Der Boxplot veranschaulicht deutlich, dass die Messwerte des Zuwachses an experimenteller Kompetenz hauptsächlich am oberen Rand der Skala liegen. Dies deutet auf eine Tendenz zum Deckeneffekt hin, auch wenn dies im engeren Sinne kein Leistungstest ist. Anhand des Shapiro-Tests ist die Größe zum Zeitpunkt Post 1 nicht normalverteilt ($W = 0.95, p = .002$) und bei Post 2 normalverteilt ($W = 0.98, p = .410$). Ersteres ist vermutlich durch Ausreißer und die dadurch schiefe Verteilung begründet (Abb. 41). Die Streuung zu beiden Zeitpunkten ist ähnlich, wie ein Levene-Test zur Prüfung der Varianz-Homogenität ($F(1, 154) = 0.018, p = .893$) zeigt.

Um die Unterschiede der experimentellen Kompetenz zwischen den beiden Messzeitpunkten auf Signifikanz zu prüfen, wird ein abhängiger t-Test durchgeführt. Da die Differenz zwischen den Daten der beiden Messzeitpunkte normalverteilt ist (Shapiro-Test: $W = 0.99, p = .892$), sind die Testvoraussetzungen für den t-Test erfüllt. Der t-Test bestätigt den Eindruck anhand der Deskription, dass keine signifikanten Unterschiede zwischen den Messzeitpunkten vorliegen ($t(67) = 1.86, p = .067, d = 0.19$). Auch wenn die experimentellen Kompetenzen im ersten Semester des Anfängerpraktikums erstmals gefördert und im zweiten Semester vertieft werden,

[8] t-Tests bzgl. Semesterart: Post 1: $t(27.0) = 0.59, p = .562$; Post 2: $t(7.2) = 1.84, p = .101$

[9] Varianzanalysen bzgl. Kohorte: Post 1: $F(6, 81) = 1.07, p = .386$; Post 2: $F(5, 62) = 1.43, p = .225$

waren aufgrund der übergeordneten Formulierung der Items der Selbsteinschätzung ohne direkten Bezug auf die einzelnen spezifischen Methoden (z. B. „Durch das Praktikum kann ich Rohdaten besser graphisch darstellen.“, siehe auch Abschnitt 11.1.2) auch keine großen Unterschiede für die Selbsteinschätzung zwischen den Semestern mit den unterschiedlichen Versuchen und Methoden zu erwarten.

Zusammengenommen wird der Zuwachs an experimenteller Kompetenz durch das Praktikum im Mittel hoch eingeschätzt. Rehfeldt (2018b) hat das Instrument zur Selbsteinschätzung der experimentellen Kompetenz zur Evaluation in mehreren Praktika Deutschlands eingesetzt und dabei leicht niedrigere Mittelwerte der Subskalen erhalten ($3.94 \pm 0.07 \leq M \leq 4.60 \pm 0.06, n = 237$). Das weist darauf hin, dass der selbst eingeschätzte Zuwachs an experimenteller Kompetenz ($M_{Post1} = 4.68 \pm 0.07$, $M_{Post2} = 4.86 \pm 0.07$) auch im Vergleich zu anderen Praktika im höheren Bereich liegt. In Bezug auf das neu entwickelte Lehramtspraktikum ist dieser hoch eingeschätzte Lernerfolg sehr positiv zu bewerten.

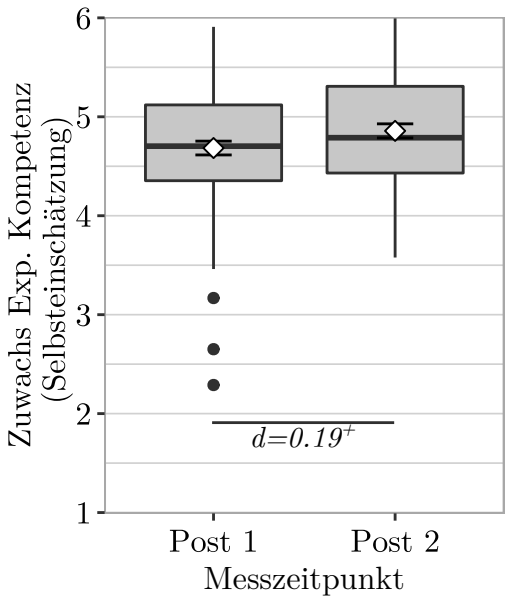


Abbildung 41: *Darstellung der Selbsteinschätzung des Zuwachses der experimentellen Kompetenz (6-teilige Likert-Skala) am Ende des ersten (Post 1) und zweiten Praktikumssemesters (Post 2). Der Boxplot zeigt neben Minimum, unterem Quartil, Median, oberem Quartil und Maximum zusätzlich die Mittelwerte mit Standardfehlern. Ergänzt wurde zudem die Effektgröße Cohen's d und das Signifikanzniveau aus dem Vergleich der Mittelwerte anhand eines t-Tests. Die zugehörigen deskriptiven Werte sind in Tabelle 25 gelistet.*

Messzeitpunkt	n	M	SD	SE	MD	MIN	MAX
Post 1	88	4.68	0.65	0.07	4.70	2.29	5.91
Post 2	68	4.86	0.59	0.07	4.79	3.58	6.00

Tabelle 25: *Deskriptive Statistik für die Selbsteinschätzung des Zuwachses der experimentellen Kompetenz (6-teilige Likert-Skala) am Ende des ersten (Post 1) und zweiten Praktikumssemesters (Post 2).*

12.2.2 Wirkmodell mit experimenteller Kompetenz als Lernerfolg

Nachdem in den ersten vorangehenden Schritten der Zuwachs an experimenteller Kompetenz untersucht wurde, kann darauf aufbauend nun der Zusammenhang dieser Größe mit der wahrgenommenen Relevanz und dem situationalen Interesse Catch betrachtet werden. Wenn entsprechende Zusammenhänge gefunden wurden, kann dann im finalen Schritt das Wirkmodell als Ganzes geprüft werden (Wirkmodell B).

Da die experimentellen Kompetenzen im Fach Physik im ersten Semester dieses Anfängerpraktikums erstmals gefördert und gefordert werden, ist davon auszugehen, dass Effekte auf die experimentelle Kompetenz im ersten Praktikumssemester größer sind als im zweiten Praktikumssemester, in dem Detailkenntnisse über spezifische Methoden vertieft werden. Aus diesem Grund wird in den weiteren Analysen der Zusammenhänge hauptsächlich die Selbsteinschätzung für den Messzeitpunkt Post 1 betrachtet. Da die experimentelle Kompetenz nicht versuchsbezogen erhoben wurde, können die Wirkzusammenhänge nur auf der Datenebene *Zwischen Personen* (siehe Abb. 18B) für die $n = 88$ Lehramtsstudierenden analysiert werden. Um die versuchsbegleitend erhobenen Daten zur wahrgenommene Relevanz und zum situationalen Interesse Catch auf der Ebene *Zwischen Personen* zu analysieren, werden Personenmittelwerte über alle im ersten Praktikumssemester durchgeführten Versuche gebildet wie in Abschnitt 11.2.1 erläutert.

Die Ergebnisse der Untersuchung der einzelnen paarweisen Zusammenhänge mittels linearen Regressionen (siehe Abschnitt 8.3.5) sind in Tabelle 26 gelistet und in den Streudiagrammen in Abbildung 42 und 43 veranschaulicht. Wie bereits in Abschnitt 12.1.2 gezeigt, ist der Zusammenhang zwischen wahrgenommener Relevanz und situationalem Interesse Catch hoch signifikant ($\beta = .70, p > .001$, Abb. 42). Die Effekte der beiden Größen auf den Zuwachs der experimentellen Kompetenz sind ebenfalls signifikant mit mittlerem bis hohem Effekt. Für die wahrgenommene Relevanz ist der Effekt $\beta = .42, p > .001$ und für das situationale Interesse $\beta = .49, p > .001$ (Tab. 26, Abb. 43a und b). Damit werden 17 % bzw. 24 % der Varianz der experimentellen Kompetenz durch die Prädiktoren aufgeklärt.

Damit liefern die gefundenen Zusammenhänge Hinweise auf die Plausibilität der Hypothese, dass sich eine größere wahrgenommene Relevanz bzw. ein höheres situationales Interesse positiv auf den Zuwachs an experimenteller Kompetenz auswirkt. Damit ist eine wichtige Voraussetzungen erfüllt, um die Analyse im nächsten Schritt zu vertiefen und die einzelnen Wirkzusammenhänge im Mediationsmodell zu analysieren.

Anhand der nachfolgenden Strukturgleichungsmodellierung des Wirkmodells werden die Effekte unter Berücksichtigung aller betrachteten Einflussgrößen geschätzt und der Anteil der direkten und indirekten Effekte differenziert. Grundlage bilden die Daten, die auch in der vorangegangenen Regressionsanalyse auf Ebene *Zwischen*

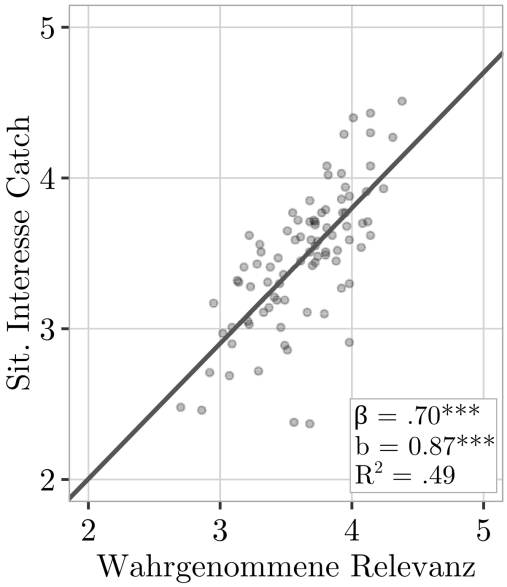


Abbildung 42: *Das situationale Interesses Catch abhängig von der wahrgenommenen Relevanz (jeweils Personenmittelwerte, 5-teilige Likert-Skala) auf der Datenebene Zwischen Personen für $n = 88$ Probanden. Die Messwerte wurden semi-transparent eingezeichnet, um übereinanderliegende Datenpunkte deutlich zu machen. Die Regressionsparameter entsprechen denen in Tabelle 26.*

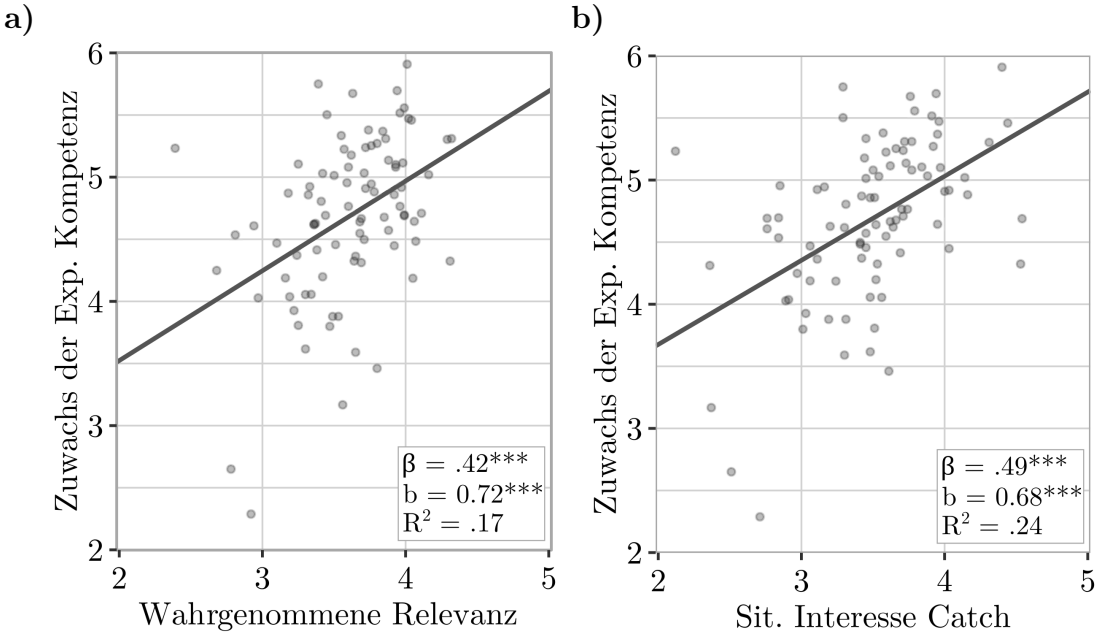


Abbildung 43: *Selbsteinschätzung des Zuwachses an experimenteller Kompetenz (6-teilige Likert-Skala) abhängig von a) der wahrgenommenen Relevanz und b) vom situationalen Interesse Catch (jeweils Personenmittelwerte, 5-teilige Likert-Skala) auf der Datenebene Zwischen Personen für $n = 88$ Personen im ersten Praktikumssemester. Die Messwerte wurden semi-transparent eingezeichnet, um übereinanderliegende Datenpunkte deutlich zu machen. Die Regressionsparameter entsprechen denen in Tabelle 26.*

Abh. Variable	**Unabh. Variable**	β	b	p	R^2
Sit. Interesse Catch	Relevanz	$.70^{***}$	0.87^{***}	$< .001$	.49
Zuwachs exp. Kompetenz	Relevanz	$.42^{***}$	0.72^{***}	$< .001$	.17
	Sit. Interesse Catch	$.49^{***}$	0.68^{***}	$< .001$	.24

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $n = 88$

Tabelle 26: *Ergebnisse der Regressionsanalysen einzelner Zusammenhänge von wahrgenommener Relevanz, situationalem Interesse Catch (jeweils Personenmittelwerte, 5-teilige Likert-Skala) und der Selbsteinschätzung des Zuwachses an experimenteller Kompetenz (6-teilige Likert-Skala) auf der Datenebene Zwischen Personen für $n = 88$ Personen im ersten Praktikumssemester. (Erläuterungen der Regressionsparameter sind in Abschnitt Methoden 8.3.5 zu finden.)*

Personen analysiert wurden. In Abbildung 44 ist das Strukturgleichungsmodell der Mediation von situationalem Interesse zwischen wahrgenommener Relevanz und dem Zuwachs an experimenteller Kompetenz dargestellt. Die Passung des Modelles wird anhand der in Abschnitt 11.2.3 vorgestellten Fit-Kennwerte betrachtet. Während der χ^2-Test ($\chi^2(116) = 284, p < .001$), der RMSEA-Wert (RMSEA= .13) und der CF-Index (CFI= .89) auf eine weniger gute Passung des Modelles hinweisen, kann die Modellpassung anhand des SRMR-Wertes (SRMR= .07) als gut beurteilt werden. Insgesamt wird anhand dieser Werte geschlossen, dass die Modellpassung damit noch akzeptabel ist. Daher werden die geschätzten Parameter des Modelles im nächsten Schritt betrachtet und anschließend unter Einbeziehung der Modellgüte diskutiert.

Wie aus der Regressionsanalyse zu erwarten ist, ist der Gesamteffekt von wahrgenommener Relevanz auf den Zuwachs an experimenteller Kompetenz mit $\beta_{tot} = .46, p = .005$ signifikant und von mittlerer Größe. Der Gesamteffekt teilt sich von der Größe zu 41 % auf den direkten Effekt ($\beta_{di} = .19, p = .262$) und 59 % auf den indirekten, über das situationale Interesse mediierten Effekt ($\beta_{ind} = .27, p = .017$) auf. Der direkte Effekt von Relevanz auf Zuwachs an experimenteller Kompetenz ist dabei allerdings nicht signifikant. Von der Größe ist dieser Effekt etwa halb so groß wie der Effekt von wahrgenommener Relevanz auf den Zuwachs an experimenteller Kompetenz in der Regression ($\beta = .42, p < .001$, Tab. 26). Der indirekte Effekt ist hingegen signifikant. Dieser Effekt ergibt sich aus einem signifikanten Effekt von $\beta = .74, p < .001$ zwischen Relevanz und Interesse und einem signifikanten Effekt $\beta = .36, p = .016$ zwischen Interesse und Zuwachs an experimenteller Kompetenz. Durch dieses Modell werden 27 % der Varianz des selbst eingeschätzten Zuwachses an experimenteller Kompetenz aufgeklärt. Damit zeigt diese Strukturgleichungsmodellierung einen signifikanten, mittleren Gesamteffekt der Mediation von situationalem Interesse zwischen wahrgenommener Relevanz und dem Zuwachs an experimenteller Kompetenz. Die bestimmten Teileffekte zeigen, dass der Zusammenhang zwischen wahrgenommener Relevanz und dem Zuwachs des Fachwissens zum Großteil über

das situationale Interesse mediiert wird. Dabei kann keine genaue Aufteilung in den direkten und indirekten Teileffekt bestimmt werden, da der direkte Effekt auf den Zuwachs zu gering ist, um diesen mittels dieser Stichprobe ($n = 88$) signifikant nachzuweisen.

Für die Strukturgleichungsmodellierung wurde sich auf Daten aus dem ersten Praktikumssemester beschränkt, da größere Effekte vermutet wurden. Modellierungen der experimentellen Kompetenz über beide Semester zeigen ähnliche Modellschätzungen mit einem geringeren und auch aufgrund der geringeren Stichprobe nicht signifikanten Gesamteffekt, was die Vermutungen bezüglich des Datensatzes zum zweiten Messzeitpunkt stützt.

Aus den Kennwerten zur Güte des hier betrachteten Modells, die zum Teil unter den empfohlenen Grenzwerten für gute Modellpassungen liegen (z. B. CFI= .89), kann der Rückschluss gezogen werden, dass das geschätzte Modell zwar nicht verworfen werden muss, aber die Daten auch nicht optimal widerspiegelt. Eine mögliche Ursache kann die Messung der experimentellen Kompetenz sein, die, wie in Abschnitt 11.1.2 diskutiert, als Selbsteinschätzung in der Kritik einer geringen Validität steht, da auch andere Faktoren wie das Selbstkonzept die Selbsteinschätzung bedingen. Zudem wurden keine Referenzwerte wie eine Einschätzung des Ausgangswertes und der maximal möglichen experimentellen Kompetenz in die Analyse einbezogen, sodass durch die resultierenden individuellen Messskalen der Studierenden Verzerrungen nicht ausgeschlossen werden können. Gleichermaßen wurde in Abschnitt 11.1.2 auch diskutiert, dass die Selbsteinschätzung trotzdem gute Anhaltspunkte für die experimentelle Kompetenz liefert, sodass die Ergebnisse der Strukturgleichungsmodellierung durchaus als wertvolle Hinweise bei der Überprüfung der Wirkannahmen aufgefasst werden können.

Somit bleibt festzuhalten, dass die wahrgenommene Relevanz einen signifikant positiven, mittleren Effekt auf den selbst eingeschätzten Zuwachs an experimenteller Kompetenz hat. Die Prüfung des Wirkmodells in Bezug auf die experimentelle Kompetenz (Wirkmodell B) stellt damit ein wichtiges Indiz für die Plausibilität des Wirkmodells dar.

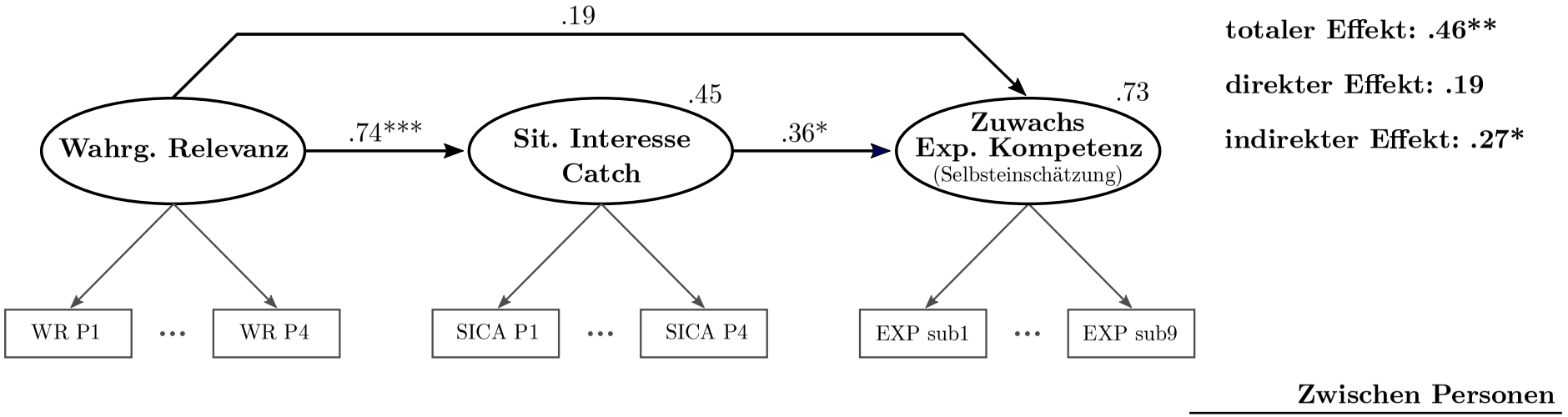


Abbildung 44: *Ergebnis des Strukturgleichungsmodells der Mediation von situationalem Interesse Catch zwischen wahrgenommener Relevanz und der Selbsteinschätzung des Zuwachses der experimentellen Kompetenz für $n = 88$ Studierende für das erste Praktikumssemester (Wirkmodell B). Abgebildet sind die standardisierten Koeffizienten. Die Werte an den abhängigen Variablen stellen die Residualvarianz dar. Das Signifikanzniveau ist an folgender Kodierung zu erkennen: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$. Für das Modell liegen folgende Fit-Kennwerte vor: χ^2-Test: $\chi^2(116) = 284, p < .001$, SRMR = .07, RMSEA = .13, CFI = .89.*

12.2.3 Entwicklung Selbstwirksamkeitserwartungen

Die Lehramtsstudierenden sollen sich durch die Aneignung experimenteller Kompetenzen im Praktikum mehr und sicherer zutrauen, Experimente im Schulunterricht zu planen und durchzuführen (Abschnitt 3.1). Somit zählt die Selbstwirksamkeitserwartung der Lehramtsstudierenden, im Folgenden auch Selbstwirksamkeit, zu den Lernerfolgen im Lehramtspraktikum. Gleichzeitig sind die Selbstwirksamkeitserwartungen auch ein wichtiger Indikator für die experimentelle Kompetenz. Daher stehen die Selbstwirksamkeitserwartungen im Fokus der nachfolgenden Betrachtungen zum Wirkmodell.

Bevor die Selbstwirksamkeit in Relation zu anderen Größen gebracht und das Wirkmodell diesbezüglich geprüft werden kann, ist es erforderlich, die Selbstwirksamkeit selbst zu analysieren. Für die Einschätzung der Wirksamkeit des Praktikums stellt sich in diesem Abschnitt vor allem die Kernfrage, ob und in welchem Maß sich die Selbstwirksamkeitserwartungen der Lehramtsstudierenden nach dem Praktikum gemäß Theorie (Abschnitt 3.1) gesteigert haben. Zur vollständigen Einschätzung des Lernerfolgs bezüglich der Selbstwirksamkeit tragen weitere Teilfragen bei. Vor Betrachtung des zeitlichen Verlaufs sind die Selbstwirksamkeitserwartungen, mit denen die Studierenden in das Praktikum starten, von Interesse, um die Ausgangssituation der Studierenden zu bewerten. Bei der anschließenden Untersuchung der zeitlichen Änderungen der Selbstwirksamkeit wird nicht nur betrachtet, in welchem Maß die Selbstwirksamkeit im Mittel über das gesamte Praktikum und die einzelnen Semester ansteigt. Untersucht wird auch, wie unterschiedlich das Praktikum auf die Selbstwirksamkeitserwartungen der einzelnen Personen wirkt und ob darin eine Systematik feststellbar ist. Zudem stellt sich die Frage, ob nicht nur die Selbstwirksamkeitserwartungen in Bezug auf die Durchführung von Experimenten, sondern auch die Selbstwirksamkeitserwartungen in Bezug auf die Planung durch dieses neue Anfängerpraktikum gefördert werden können, obwohl der Fokus im Anfängerpraktikum deutlich auf der Durchführung von Experimenten liegt.

Für den ersten Einblick ist die deskriptive Statistik zur Selbstwirksamkeitserwartung an den verschiedenen Messzeitpunkten in Tabelle 27 bereitgestellt. Die Boxplots in Abbildung 45 veranschaulichen zudem die Verteilung der Daten. Die erfasste Selbstwirksamkeitserwartung spaltet sich in die zwei Komponenten Planung und Durchführung von Experimenten im Unterricht auf. Die Selbstwirksamkeit wurde in den Prä-Post-Tests vor und nach den beiden Praktikumssemestern (Prä 1, Post 1, Prä 2 und Post 2) erfasst. Die Daten sind damit abhängig und außerdem unbalanciert, da nicht alle Studierenden am zweiten Praktikumssemester und den zugehörigen Befragungen teilgenommen haben. Die Stichproben der einzelnen Semester wurden dabei zu einer Gesamtstichprobe zusammengefasst, weil Varianzanalysen im Vorfeld gezeigt haben, dass die Daten keine signifikanten Unterschiede bzgl. des Semesters der Erhebung (Sommersemester 2016 - 2019) aufweisen.

Zunächst soll die Selbstwirksamkeitserwartung in Bezug auf die Planung von Experimenten im Unterricht deskriptiv betrachtet werden. Die mittlere Selbstwirksamkeit Planung liegt für die vier Messzeitpunkte im Wertebereich von $2.92 \pm 0.07 \leq M_P \leq 3.60 \pm 0.06$ (Tab. 27 oben). Zu Beginn des Praktikums liegt die mittlere Selbstwirksamkeitserwartung Planung damit in der Mitte des Messbereichs der 5-teiligen Likert-Skala und hat die größte Streuung. Dies kann zum einen auf die Bandbreite der Einschätzungen zu Beginn des Praktikums hindeuten. Zum anderen kann hier auch die Schwierigkeit der Lehramtsstudierenden hineinspielen, die bisher gegebenenfalls nicht geförderten experimentellen Kompetenzen einordnen zu können und sich ohne bisherige Erfahrung zur Beantwortung der Items schwer in die Lage des Unterrichtens zu versetzen zu können. Die Mittelwerte der Selbstwirksamkeitserwartungen nehmen über den Zeitraum des Praktikums hypothesenkonform zu und die Streuung nimmt ab. Der Unterschied der Mittelwerte vor und nach dem Praktikums (Prä 1 und Post 2) beträgt rund 0.7 Skaleneinheiten. Der Boxplot in Abbildung 45a veranschaulicht die Verteilung der Selbstwirksamkeit Planung für die vier Messzeitpunkte. Der Messbereich ist nach unten nicht ganz ausgenutzt. Shapiro-Tests zeigen, dass die Selbstwirksamkeit für alle Messzeitpunkte normalverteilt ist. Anhand der Boxplots und mittels eines Levene-Tests ($F(3, 308) = 3.25, p = .022$) zeigt sich, dass die Streuung für die Messzeitpunkte unterschiedlich ist.

Subskala	**Messzeitpunkt**	**n**	**M**	**SD**	**SE**	**MIN**	**MAX**
Planung	Prä 1	89	2.92	0.68	0.07	1.50	4.62
	Post 1	88	3.47	0.51	0.05	1.75	4.75
	Prä 2	68	3.42	0.53	0.06	2.00	4.50
	Post 2	67	3.60	0.50	0.06	2.50	4.88
Durchführung	Prä 1	89	3.20	0.60	0.06	1.00	4.50
	Post 1	88	3.52	0.55	0.06	1.75	4.88
	Prä 2	67	3.51	0.58	0.07	1.75	5.00
	Post 2	66	3.68	0.48	0.06	2.38	4.75

Tabelle 27: *Deskriptive Statistik für Selbstwirksamkeitserwartungen in Bezug auf die Planung und Durchführung von Experimenten im Unterricht (5-teilige Likert-Skala) vor und nach dem ersten (Prä 1, Post 1) und dem zweiten Praktikumssemester (Prä 2, Post 2).*

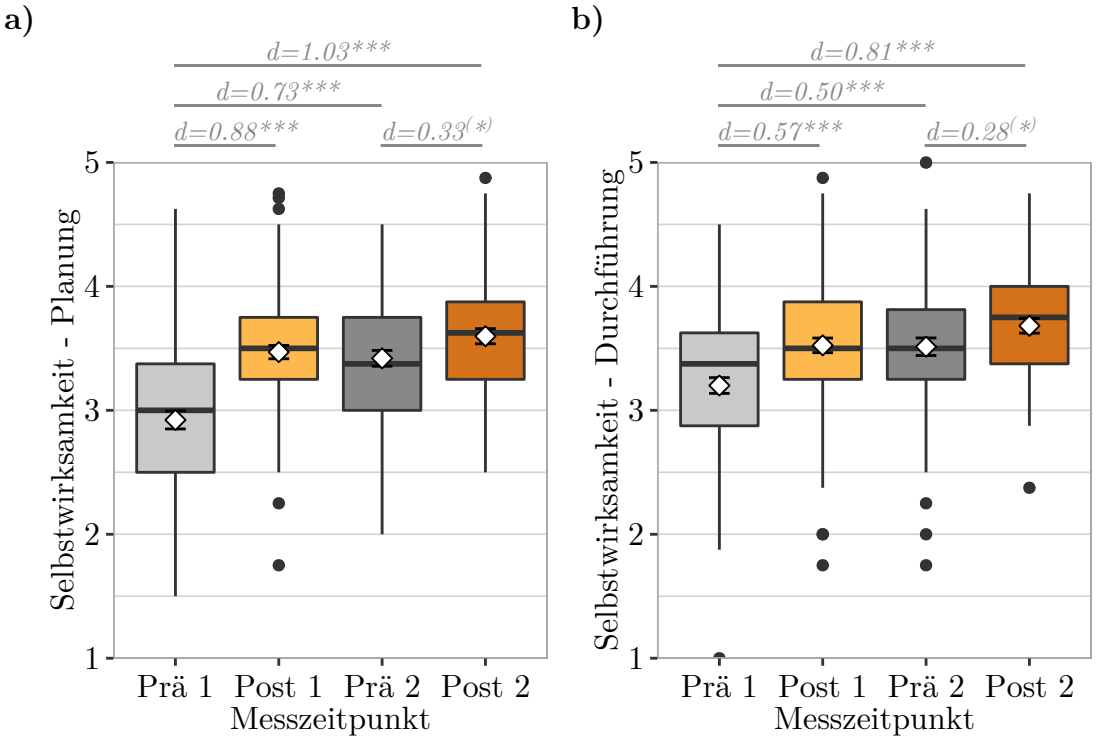


Abbildung 45: *Darstellung der Selbstwirksamkeitserwartung in Bezug auf a) Planung und b) Durchführung von Experimenten im Unterricht (5-teilige Likert-Skala) vor (grau) und nach (orange) dem ersten (Prä 1, Post 1) und dem zweiten Praktikumssemester (Prä 2, Post 2). Der Boxplot zeigt neben Minimum, unterem Quartil, Median, oberem Quartil und Maximum zusätzlich die Mittelwerte (weiße Rauten) mit Standardfehlern. Die zugehörigen deskriptiven Werte sind in Tabelle 27 gelistet. Zudem sind auch die Signifikanzniveaus der t-Tests (*** $p < .001$, (*) t-Test: $p < .05$ und Wilcoxon-Test: $p > .05$) und Cohens d als Effektgröße dargestellt. Nicht signifikante Unterschiede wurden zur Übersichtlichkeit nicht dargestellt.*

Für die Selbstwirksamkeitserwartung in Bezug auf die Durchführung von Experimenten im Unterricht zeigt sich ein ähnliches deskriptives Bild. Für die vier Messzeitpunkte liegt die mittlere Selbstwirksamkeit Durchführung im Wertebereich von $3.20 \pm 0.06 \leq M_D \leq 3.68 \pm 0.06$ und damit über dem Mittelmaß der 5-stufigen Skala (Tab. 27 unten). Der Mittelwert der Selbstwirksamkeit Durchführung liegt nach dem Praktikum (Post 2) um rund eine halbe Skaleneinheit höher als vor dem Praktikum (Prä 1). Die Verteilungen der Daten zur Durchführung in den Boxplots in Abbildung 45b liefern in Bezug auf die Ausnutzung des Messbereiches ein ähnliches Bild wie für die Planung. Die Boxplots weisen auf eine Schiefe der einzelnen Verteilungen hin. Shapiro-Tests bestätigen, dass die Daten für alle Messzeitpunkte außer Post 2 nicht normalverteilt sind. Für die Messzeitpunkte besteht Varianzhomogenität, wie ein Levene-Test zeigt ($F(3, 306) = 0.821, p = .483$).

Im Vergleich zeigen sich ähnliche deskriptive Werte für die Selbstwirksamkeitserwartungen in Bezug auf Planung und Durchführung. Auffällig ist allerdings der Ausgangsmittelwert (Prä 1). Dieser ist für die Durchführung mit kleinem bis mittle-

rem Effekt signifikant höher als für die Planung ($M_P = 2.92 \pm 0.07$, $M_D = 3.20 \pm 0.06$, abh. t-Test: $t(88) = 5.38, p < .001$, Wilcoxon-Vorzeichen-Rang-Test: $W = 683.5, p < .001$, Cohens $d = 0.43$, $n = 89$). Auch Rabe et al. (2012) stellen fest, dass die Selbstwirksamkeitserwartungen für die Durchführung von Studierenden höher eingeschätzt wird als die Planung. Rabe et al. (2012) gehen ebenfalls von der Ursache aus, dass den Studierenden die Durchführung von Experimenten im Unterricht als passive Erfahrung bekannt ist, sie aber wenig Einblicke in die Planung haben. Am Ende des Praktikums (Post 2) lässt sich anhand eines Wilcoxon-Vorzeichen-Rang-Tests nur ein sehr kleiner signifikanter Unterschied für die Selbstwirksamkeit in Bezug auf Planung und Durchführung feststellen ($M_P = 3.60 \pm 0.06$, $M_D = 3.68 \pm 0.06$, abh. t-Test: $t(64) = 1.67, p = .100$, Wilcoxon-Vorzeichen-Rang-Test: $W = 559, p = .033$ Cohens $d = 0.15$, $n = 68$). Dies zeigt, dass sich die Lehramtsstudierenden am Ende des Lehramtspraktikums sowohl die Planung als auch die Durchführung von Experimenten im Unterricht fast gleichermaßen besser zutrauen, obwohl die Planung im Gegensatz zur Durchführung von Experimenten nicht Schwerpunkt des Praktikums ist.

Die Boxplots in Abbildung 45a und 45b verdeutlichen, dass die mittleren Werte der Selbstwirksamkeitserwartungen gemäß der Theorie über die Messzeitpunkte im Praktikumsverlauf zunehmen. Die Standardfehler lassen zudem vermuten, dass es Unterschiede für die mittlere Selbstwirksamkeit an den Messzeitpunkten gibt. Wie für das individuelle Interesse (siehe Abschnitt 12.1.1) werden zur Absicherung dieser Beobachtung einfaktorielle Varianzanalysen mit Messwiederholung und aufgrund der nicht-parametrischen Daten Friedman-Tests auf die vollständigen Datensätze der Selbstwirksamkeitserwartungen bzgl. der Messzeitpunkte angewendet. Da anhand eines Mauchly-Tests hervorgeht, dass die Annahme der Sphärizität verletzt ist (Planung: $W(3) = 0.67, p < .001$, Durchführung: $W(3) = 0.74, p = .002$), wird zusätzlich für die Varianzanalyse der Greenhouse-Geisser-korrigierte p-Wert (Planung: $\epsilon_{GG} = .79$, Durchführung: $\epsilon_{GG} = .83$) berichtet. Die Varianzanalyse für die Selbstwirksamkeitserwartungen in Bezug auf Planung zeigt, dass signifikante Unterschiede zwischen den Messzeitpunkten vorliegen ($F(3, 198) = 35.8, p < .001, p_{GG} < .001, \eta^2_G = .17$). Dies wird bestätigt durch den nicht-parametrischen Friedman-Test ($\chi^2(3) = 55.49, p < .001$). Auch für die Komponente Durchführung zeigt die Varianzanalyse signifikante Unterschiede auf ($F(3, 192) = 19.41, p < .001, p_{GG} < .001, \eta^2_G = .09$), die durch den Friedman-Test bestätigt werden ($\chi^2(3) = 32.33, p < .001$).

Um die signifikanten Unterschiede der Selbstwirksamkeit Planung zwischen den Messzeitpunkten zu identifizieren, wurden paarweise abhängige t-Tests mit Bonferroni-Korrektur als Post-Hoc-Tests berechnet. Da die Differenzen der Messwertpaare nicht in jedem Fall normalverteilt sind und die Daten damit nicht alle die Testvoraussetzungen erfüllen, wurden zur Absicherung der Ergebnisse zusätzlich Wilcoxon-Vorzeichen-Rang-Tests mit Bonferroni-Korrektur angewendet. Über den Boxplots in Abbildung 45 wurden die Signifikanzniveaus dieser t-Tests ($p < .05$) und Cohens d als Effektgröße ergänzt. Die Ergebnisse der t-Tests für die Selbstwirksamkeit in

Bezug auf Planung in Abbildung 45a zeigen, dass signifikante Unterschiede zwischen den Zeitpunkten vor und nach dem jeweiligen Praktikumssemester Prä 1 und Post 1 ($d = 0.88$) sowie Prä 2 und Post 2 ($d = 0.33$) bestehen. Letzterer Unterschied für das zweite Praktikumssemester (Prä 2 und Post 2) ist dabei deutlich kleiner als für das erste Praktikumssemester und ist mit Vorbehalt zu sehen, da der Wilcoxon-Test diesen Unterschied anhand des t-Tests nicht bestätigt. Weitere signifikante Unterschiede liegen für die Messzeitpunkte Prä 1 und Prä 2 ($d = 0.73$) sowie Prä 1 und Post 2 ($d = 1.03$) vor. Für die Selbstwirksamkeitserwartungen in Bezug auf Durchführung von Experimenten zeigen sich für die gleichen Messzeitpunkte signifikante Unterschiede mit ähnlicher Relation in Bezug auf die Effektgrößen (Abb. 45b). Dies spricht für die Konsistenz der erhobenen Daten. Die Effektgrößen sind für die Durchführung allerdings geringer. So ist zum Beispiel die Effektgröße über das gesamte Praktikum mit $d = 0.81$ noch im hohen Bereich, für den ersten Praktikumsteil mit $d = 0.57$ im mittleren Bereich.

Diese Ergebnisse zeigen, dass die Selbstwirksamkeitserwartungen in Bezug auf Planung und auch Durchführung von Experimenten über das Praktikum hinweg im Mittel ansteigen und die Unterschiede dabei sogar von großer Effektstärke sind. Der Unterschied der Selbstwirksamkeit ist dabei gemäß Hypothese für beide Komponenten der Selbstwirksamkeit im ersten Praktikumssemester deutlich größer als im zweiten Praktikumssemester. Für die praktikumsfreie Zeit (Post 1 - Prä 2) konnte keine Änderung festgestellt werden. Dies ist ein starkes Indiz dafür, dass die festgestellte Änderung der Selbstwirksamkeitserwartungen durch das neue Praktikum bewirkt wurde.

Im nächsten Schritt wird eine Mehrebenenanalyse im Sinne eines Wachstumskurvenmodells (siehe Abschnitt 8.3.6) zur detaillierteren Untersuchung der Zunahme der Selbstwirksamkeitserwartungen aller Studierenden über das Praktikum herangezogen. Die Analyse erfolgt analog zur zeitlichen Analyse der wahrgenommenen Relevanz und des situationalen Interesses (Abschnitte 9.1 und 12.1.1). Dazu wird zunächst die Mehrebenenstruktur der Selbstwirksamkeitserwartungen in einem Nullmodell mit Level 1 als Beobachtungen und den Level 2 Gruppierungen in Personen betrachtet (Modell 0P in Tabellen 10, 28 und 29). Für die Selbstwirksamkeit sind anhand der Intraklassen-Korrelation (ICC) für die Komponente Planung 37 % und für die Komponente Durchführung 48 % der Varianz auf Unterschiede zwischen den Personen ($n = 89$) zurückzuführen. Damit ist es notwendig, die Personenvarianz bei den zeitlichen Betrachtungen zu berücksichtigen.

Für die Wachstumskurvenmodelle dient der Messzeitpunkt als kontinuierlich angenommener Prädiktor. Die Modelle $A12$, $A11$ und $A22$ in den Tabellen 28 und 29 basieren dabei auf dem Grundmodell A aus Tabelle 10 und unterscheiden sich hinsichtlich der Messzeitpunkte, die in die Betrachtung einbezogen werden. Für Modell $A12$ werden alle vier Beobachtungszeitpunkte Prä 1 bis Post 2 und damit das gesamte Praktikum berücksichtigt. Für die Modelle $A11$ und $A22$ werden jeweils nur das

erste (Prä 1 - Post 1) und zweite (Prä 2 - Post 2) Praktikumssemester einbezogen. Da in der Zeit zwischen Ende des ersten Praktikumssemester und Anfang des zweiten Praktikumssemesters (Post 1 - Prä 2) kein Praktikum stattfindet und dementsprechend auch anhand der Mittelwertvergleiche keine signifikanten Unterschiede für die Zeitpunkte festgestellt wurden, wurden die Beobachtungen an den Messzeitpunkten Post 1 und Prä 2 auf einen gemeinsamen Zeitpunkt kodiert.

Die zufälligen Effekte des Modells $A12$ für die Selbstwirksamkeit Planung zeigen, dass 61 % der Gesamtvarianz zurückgeht auf Unterschiede zwischen Personen am ersten Messzeitpunkt, 10 % auf Unterschiede zwischen den Änderungsraten der Personen und 29 % auf Unterschiede innerhalb der Personen (Tab. 28). Dies ist für die Selbstwirksamkeit Durchführung sehr ähnlich mit 60 % der Gesamtvarianz durch Unterschiede zwischen Personen am ersten Messzeitpunkt, 8 % durch Unterschiede zwischen den Veränderungsraten der Personen und 32 % durch Unterschiede innerhalb der Personen (Tab. 29). Anhand der festen Effekte ergeben sich im Mittel für den gesamten Praktikumszeitraum signifikante Veränderungsraten von $\gamma_{10} = .56$ ($g_{1o} = 0.35, p < .001$) für Planung und $\gamma_{10} = .40$ ($g_{1o} = 0.23, p < .001$) für Durchführung. Die Abbildungen 46a und b zeigen die Entwicklungsverläufe ohne Standardisierung für die einzelnen Personen und den mittleren Verlauf, sodass die Varianz der Änderungsraten für die Personen deutlich wird. Für die einzelnen Personen sind Verläufe mit unterschiedlichen Anstiegen zu erkennen (95 %-Konfidenzintervall Planung: $0.43 < b_{1j} < 0.68$, Durchführung: $0.28 < b_{1j} < 0.51$).

Da sich in den vorangegangenen Mittelwertvergleichen gezeigt hat, dass die Prä-Post-Unterschiede für das erste Praktikumssemester größer sind als für das zweite, wurden in Modell $A11$ und $A22$ (Tab. 28 und 29) jeweils nur das erste bzw. zweite Semester berücksichtigt, um diese Veränderungsraten näher zu untersuchen. Diese Untersuchungen bestätigen, dass der Anstieg für das erste Praktikumssemester für beide Komponenten der Selbstwirksamkeit im Mittel größer ist (Planung: $\gamma_{10} = .81$, $g_{10} = 0.51, p < .001$; Durchführung: $\gamma_{10} = .51$, $g_{10} = 0.30, p < .001$) als für das zweite Semester (Planung: $\gamma_{10} = .28$, $g_{10} = 0.14, p = .011$; Durchführung: $\gamma_{10} = .25$, $g_{10} = 0.14, , p = .006$). Für die einzelnen Studierenden bedeutet das Anstiege der Selbstwirksamkeitserwartungen bis zu eineinhalb Skaleneinheiten für das erste Semester. Die experimentellen Kompetenzen werden im ersten Praktikumssemester (z. T. neu) erlernt und im zweiten Semester darauf aufbauend vertieft. Damit kann die Hypothese, dass sich entsprechend auch die Selbstwirksamkeitserwartungen bezüglich des Experimentierens im ersten Semester deutlicher verändern, als bestätigt angesehen werden.

Bei dieser Betrachtung stellt sich außerdem die Frage, ob einer der beiden Praktikumsteile WS und SS, also die unterschiedlichen Praktikumsthemen im Winter- und Sommersemester (siehe Abschnitt 8.1), die Selbstwirksamkeit der Studierenden mehr erhöht. Eine Hinzunahme der Praktikumsteile als Prädiktor in das Modell $A12$ zeigt, dass die Praktikumsteile keinen signifikanten Effekt haben und damit vernachlässigbar sind. Im ersten Praktikumssemester werden offenbar die experimentellen

Selbstwirksamkeit Planung

	Modell 0P	**Modell** $A12$	**Modell** $A11$	**Modell** $A22$
Messzeitpunkte (MZP)		Prä 1 - Post 2	Prä 1 - Post 1	Prä 2 - Post 2
feste Effekte				
Ausgangswert (g_{00})		$g_{00} = 3.00^{***}$	$g_{00} = 2.92^{***}$	$g_{00} = 3.44^{***}$
zeitliche Änderung (γ_{10})		$\gamma_{10} = .56^{***}$	$\gamma_{10} = .81^{***}$	$\gamma_{10} = .28^{*}$
		(.06)	(.10)	(.11)
		$g_{10} = 0.35^{***}$	$g_{10} = 0.51^{***}$	$g_{10} = 0.14^{*}$
		(0.04)	(0.06)	(0.05)
zufällige Effekte				
Varianz zwischen Personen ($\text{Var}(u_{0j})$)	0.37	0.75	0.93	0.66
Varianz innerhalb Personen ($\text{Var}(e)$)	0.63	0.35	0.22	0.34
Varianz der Steigungen ($\text{Var}(u_{1j})$)		0.12	0.45	0.27
Korrelation Ausgangswert und Änderung		-0.77	-0.73	-0.40
ICC	.37	.58	.74	.66
Designeffekt	2.0			
Anzahl Beobachtungen	312	312	245	223
Anzahl Personen	89	89	89	68
Devianz	838	738	568	562

Modell 0: Intercept-Only-Modell; **Modell A:** Random-Coefficients-Modell

Signifikanzniveaus für feste Effekte: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

Tabelle 28: *Ergebnisse der Mehrebenenanalysen (Intercept-Only-Modell 0 und Random-Coefficients-Modell A, Tab. 10) des zeitlichen Verlaufs der Selbstwirksamkeitserwartungen in Bezug auf* ***Planung*** *von Experimenten. (*$A12$*: Zeitraum Prä 1 - Post 2,* $A11$*: Zeitraum Prä 1 - Post 1,* $A22$*: Zeitraum Prä 2 - Post 2,* g_{00}*: gruppenunspezifischer Achsenabschnitt,* $g_1 0$*: nicht standardisiertes gruppenunspezifisches Regressionsgewicht,* $\gamma_1 0$*: standardisiertes gruppenunspezifisches Regressionsgewicht)*

Selbstwirksamkeit Durchführung

	Modell 0P	**Modell** $A12$	**Modell** $A11$	**Modell** $A22$
Messzeitpunkte (MZP)		Prä 1 - Post 2	Prä 1 - Post 1	Prä 2 - Post 2
feste Effekte				
Ausgangswert (g_{00})		$g_{00} = 3.24^{***}$	$g_{00} = 3.20^{***}$	$g_{00} = 3.50^{***}$
zeitliche Änderung (γ_{10})		$\gamma_{10} = .40^{***}$	$\gamma_{10} = .51^{***}$	$\gamma_{10} = .25^{**}$
		(.06)	(.10)	(.09)
		$g_{10} = 0.23^{***}$	$g_{10} = 0.30^{***}$	$g_{10} = 0.14^{**}$
		(0.03)	(0.06)	(0.05)
zufällige Effekte				
Varianz zwischen Personen ($\text{Var}(u_{0j})$)	0.47	0.67	0.76	0.75
Varianz innerhalb Personen ($\text{Var}(e)$)	0.51	0.36	0.24	0.29
Varianz der Steigungen ($\text{Var}(u_{1j})$)		0.09	0.46	0.11
Korrelation Ausgangswert und Änderung		-0.57	-0.50	-0.70
ICC	.48	.60	.74	.70
Designeffekt	2.3			
Anzahl Beobachtungen	310	310	244	221
Anzahl Personen	89	89	89	68
Devianz	798	742	593	526

Modell 0: Intercept-Only-Modell; **Modell A:** Random-Coefficients-Modell

Signifikanzniveaus für feste Effekte: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

Tabelle 29: *Ergebnisse der Mehrebenenanalysen (Intercept-Only-Modell 0 und Random-Coefficients-Modell A, Tab. 10) des zeitlichen Verlaufs der Selbstwirksamkeitserwartungen in Bezug auf* ***Durchführung*** *von Experimenten. (A12: Zeitraum Prä 1 - Post 2, A11: Zeitraum Prä 1 - Post 1, A22: Zeitraum Prä 2 - Post 2, g_{00}: gruppenunspezifischer Achsenabschnitt, $g_1 0$: nicht standardisiertes gruppenunspezifisches Regressionsgewicht, $\gamma_1 0$: standardisiertes gruppenunspezifisches Regressionsgewicht)*

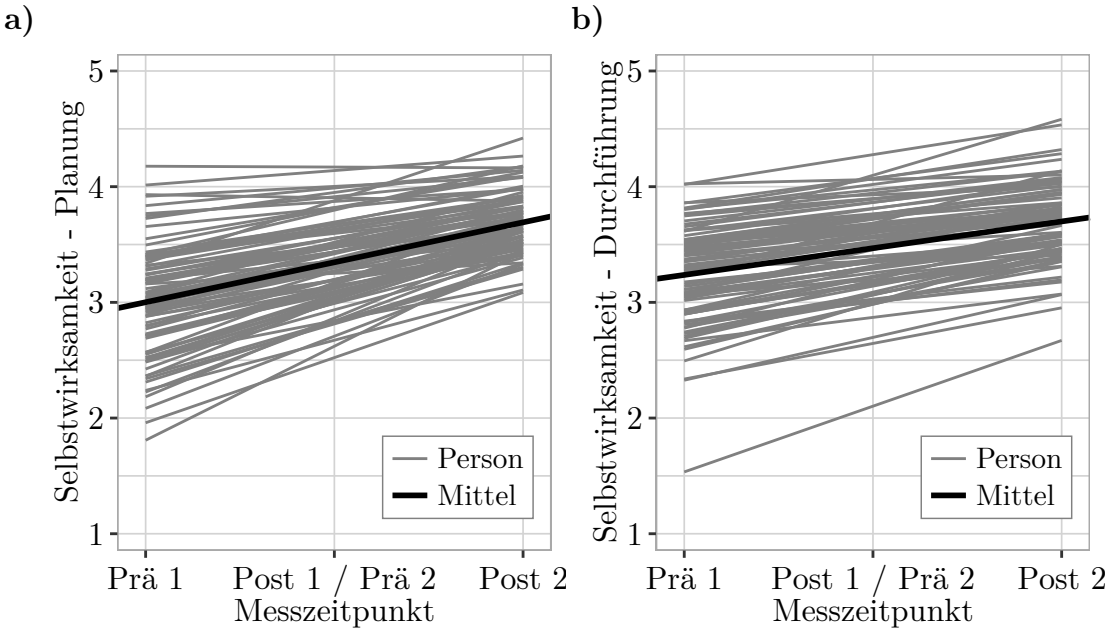


Abbildung 46: *Entwicklungsverläufe aus der Mehrebenenanalyse (Modell A, Tab. 10, 28 und 29) für die Selbstwirksamkeitserwartungen Bezug auf a) Planung und b) Durchführung von Experimenten im Unterricht (5-teilige Likert-Skala, ohne Standardisierung) abhängig vom Messzeitpunkt (Prä- und Post-Test im ersten (Prä 1, Post 1) und zweiten (Prä 2, Post 2) Praktikumssemester). Die grauen Linien zeigen die Verläufe für alle einzelnen Personen (Range Planung:* $-0.03 < b_{1j} < 0.81$, 95 %*-Konfidenzintervall Planung:* $0.43 < b_{1j} < 0.68$, *Range Durchführung:* $0.04 < b_{1j} < 0.59$, 95 %*-Konfidenzintervall Durchführung:* $0.28 < b_{1j} < 0.51$*). Die dickere schwarze Linie ist der Verlauf über alle Personen (Ausgangswert bei Prä 1* g_{oo} *und Steigung* g_{1o}*: Planung:* $g_{oo} = 3.00$, $g_{1o} = 0.35$, *Durchführung:* $g_{oo} = 3.24$, $g_{1o} = 0.23$*).*

Kompetenzen unabhängig von den Versuchsthemen erworben und im zweiten Semester vorwiegend vertieft.

Anhand der Entwicklungsverläufe der Selbstwirksamkeit Planung über den gesamten Praktikumszeitraum (Abb. 46a) wird deutlich, dass für Studierende, die mit niedriger Selbstwirksamkeit in das Praktikum einsteigen, die Selbstwirksamkeit über das Praktikum hinweg stärker ansteigt als für Studierende mit einem hohen Ausgangswert. Dies bestätigt auch die Korrelation zwischen Ausgangswert und Veränderungsrate, die für Planung negativ und hoch ist mit $-.77$. Für die Selbstwirksamkeit Durchführung ist die Korrelation im Vergleich etwas geringer mit $r = -.57$. Dies zeigt, dass das Praktikum besonders wirksam ist für Studierende mit niedrigen Selbstwirksamkeitserwartungen zu Beginn des Praktikum. Eine mögliche Ursache dafür kann ein Deckeneffekt des Messinstrumentes sein. Es schließt sich die Frage an, ob auch für Studierende mit hohen Selbstwirksamkeitserwartungen zu Beginn des Praktikum trotzdem eine Wirksamkeit festgestellt werden kann. Gruppiert man die Studierenden anhand eines Mediansplits in zwei Gruppen mit hoher und niedriger Selbstwirksamkeit zu Beginn des Praktikums und nimmt man diese Gruppierung

in das Modell $A12$ auf, lässt sich diese Frage mit Ja beantworten. Die resultierenden Veränderungsraten sind zwar für die Studierenden mit höherem Ausgangswert für die Selbstwirksamkeit (Planung: $\gamma_{10} = .22$, $g_{10} = 0.14, p = 002$; Durchführung: $\gamma_{10} = .16$, $g_{10} = 0.10, p = .029$) deutlich niedriger als für Studierenden mit niedrigem Ausgangswert (Planung: $\gamma_{10} = .93$, $g_{10} = 0.58, p =< .001$; Durchführung: $\gamma_{10} = .64$, $g_{10} = 0.38, p =< .001$), aber dennoch signifikant und positiv. Da die Selbstwirksamkeit Planung im Vergleich zur Durchführung niedrigere Ausgangswerte vor Beginn des Praktikums (Prä 1) und ähnliche Werte am Ende des Praktikums aufweist (Post 2) (siehe Vergleich oben und Tab. 27), kann der höhere Anstieg für die Planung zum Teil auch durch niedrigere Ausgangswerte begründet werden. Die Abhängigkeit der Änderung der Selbstwirksamkeitserwartungen vom Ausgangswert bedeutet zudem für die weitere Untersuchung des Wirkmodells, dass der Ausgangswert berücksichtigt werden sollte.

Damit untermauern diese Ergebnisse, dass die Selbstwirksamkeitserwartungen der Studierenden über das Praktikum hinweg deutlich zunehmen. Dieser Effekt ist sogar für die erste Praktikumshälfte noch größer, da die experimentelle Kompetenz zu Beginn des Praktikums von den Studierenden zum Teil neu erlernt wird und im weiteren Praktikumsverlauf vertieft wird. Die Lehramtsstudierenden trauen sich somit deutlich mehr zu, Experimente für den Physikunterricht zu planen und dort durchzuführen. Besonders stark profitieren die Lehramtsstudierenden, die mit niedrigeren Selbstwirksamkeitserwartungen in Bezug auf das Experimentieren in das Praktikum starten. Doch auch die Studierenden mit hohen Ausgangswerten können an wahrgenommener Selbstwirksamkeit gewinnen. Besonders hervorzuheben ist an dieser Stelle, dass, obwohl der Fokus in diesem Anfängerpraktikum auf der Durchführung von Experimenten liegt und die Studierenden im Regelfall noch keine Erfahrung in der Planung von Experimenten gesammelt haben, nicht nur Selbstwirksamkeitserwartungen im Bereich Durchführung, sondern auch die Selbstwirksamkeitserwartungen im Bereich Planung durch das Praktikum gefördert werden. Diese Arbeit steuert mit diesem Nachweis, dass die Selbstwirksamkeitserwartungen während des Lehramtspraktikums mit hohem Effekt zunehmen, neue Erkenntnisse zum eher jüngeren Forschungsfeld bzgl. der Entwicklung der Selbstwirksamkeitserwartungen von angehenden Lehrkräften bei (z. B. Weß et al., 2020).

Über das Maß des Zusammenhangs der Zunahme der Selbstwirksamkeitserwartungen der Studierenden und der Zunahme tatsächlicher experimenteller Fähigkeiten im Unterricht, kann an dieser Stelle keine belastbare Aussage getroffen werden (siehe auch Abschnitt 11.1). Anhand der Daten in dieser Arbeit korrelieren die beiden Komponenten der Selbstwirksamkeitserwartungen Planung und Durchführung mit der Selbsteinschätzung der experimentellen Kompetenz signifikant in mittlerer Höhe mit $r \approx .3, p < .010$. Diese Korrelation kann nicht als Beweis für den Zusammenhang der Selbstwirksamkeitserwartungen mit der tatsächlichen experimentellen Kompetenz dienen, da beide Konstrukte auch durch andere Faktoren wie das

Selbstkonzept bedingt sind (siehe Abschnitt 11.1.2), kann aber als Hinweis dienen. Die Pilotierung zur Erhebung der Selbstwirksamkeit (Abschnitt 11.1.2) hat zudem gezeigt, dass es manchen Lehramtsstudierenden schwer fällt, sich zur Beantwortung der Items in die Lage des Experimentierens im Unterrichten zu versetzen. Der Anstieg der Selbstwirksamkeitserwartungen über das Praktikum kann daher zum Teil auch bedeuten, dass die Studierenden sich besser in die Situation des Experimentierens hineinversetzen können. Eine Über- oder Unterschätzung der tatsächlichen experimentellen Fähigkeiten ist somit wie für die Selbsteinschätzung der Experimentierfähigkeit auch für die Selbstwirksamkeitserwartungen nicht ausgeschlossen. Dies schmälert allerdings nicht das positive Ergebnis dieser Teiluntersuchung der Wirkung des Lehramtspraktikums, denn eine entsprechend hohe Selbstwirksamkeitserwartungen ist ein wichtiger motivationaler Faktor und damit Voraussetzung dafür, dass eine Lehrkraft überhaupt das Experimentieren im Physikunterricht einsetzt.

Zusammenfassend kann anhand dieser Teiluntersuchung festgestellt werden, dass die Selbstwirksamkeitserwartungen der Lehramtsstudierenden im Verlauf des Lehramtspraktikums signifikant mit großem Effekt zunehmen. Durch das Lehramtspraktikum wird damit ein wesentliches Lernziel in Bezug auf die Selbstwirksamkeitserwartungen der Lehramtsstudierenden erreicht.

12.2.4 Wirkmodell mit Selbstwirksamkeit als Lernerfolg

Nachdem im vorangegangenen Abschnitt gezeigt wurde, dass durch das Lehramtspraktikum die Selbstwirksamkeitserwartungen der Lehramtsstudierenden in Bezug auf Planung und Durchführung von Experimenten im Unterricht mit hohen Effekten gesteigert werden, können als Nächstes die Zusammenhänge zwischen Selbstwirksamkeitserwartungen und den Größen wahrgenommene Relevanz und situationales Interesse untersucht werden. Darauf aufbauend kann schließlich im letzten Schritt geprüft werden, inwiefern eine hohe wahrgenommene Relevanz bzw. ein größeres situationales Interesse sich gemäß Wirkmodell positiv auf die Zunahme der Selbstwirksamkeit auswirkt (Wirkmodelle C und D).

Die Selbstwirksamkeitserwartungen der Studierenden in Bezug auf das Experimentieren wurde vor und nach jedem Praktikumssemester erhoben. Daher wird analog zur experimentellen Kompetenz in Abschnitt 12.2.2 auch für diese Untersuchung der Zusammenhänge die Datenebene *Zwischen Personen* betrachtet. Für den Zuwachs an Selbstwirksamkeit wird die Differenz aus Post- und Prä-Testwert betrachtet, wie in Abschnitt 11.2.6 ausführlicher erläutert. Da für die Selbstwirksamkeit die größte Zunahme im ersten Praktikumssemester festgestellt wurde, wird im Folgenden die Änderung der Selbstwirksamkeit wie in der Analyse der experimentiellen Kompetenz für das erste Praktikumssemester betrachtet. In den vorangegangenen Betrachtungen wurde gezeigt, dass es keinen signifikanten Effekt hat, welcher der beiden Praktikumsteile WS und SS, also welche der unterschiedlichen

Praktikumsthemen, im ersten Praktikumssemester durchgeführt wurden, sodass die Praktikumsteile in dieser Analyse keine Berücksichtigung finden. Die erhobenen Selbstwirksamkeitserwartungen gliedern sich in die zwei Komponenten Planung und Durchführung, die im Folgenden parallel betrachtet werden.

Zunächst werden paarweise Zusammenhänge zwischen den Größen wahrgenommene Relevanz, situationales Interesse Catch (jeweils Personenmittelwerte) und der Änderung der Selbstwirksamkeitserwartungen in Bezug auf Planung und Durchführung von Experimenten im Unterricht anhand von linearen Regressionen betrachtet. Die Regressionsparameter sind in Tabelle 30 dargestellt. Die Abbildung 47 zeigt zudem Streudiagramme, die die Zusammenhänge der Änderung der Selbstwirksamkeitserwartungen mit der wahrgenommenen Relevanz und dem situationalen Interesse veranschaulichen. Der Zusammenhang zwischen situationalem Interesse Catch und wahrgenommener Relevanz ist mit $\beta = .70, p < .001$ hoch (Tab. 30). Die wahrgenommene Relevanz hat einen signifikanten, positiven Effekt von $\beta = .24, p = .021$ auf die Änderung der Selbstwirksamkeit in Bezug auf Planung von Experimenten für den Unterricht. Das Streudiagramm der Personenwerte in Abbildung 47a veranschaulicht die große Streuung der Werte. Der Effekt vom situationalen Interesse auf die Änderung der Selbstwirksamkeit bzgl. Planung ist signifikant mit $\beta = .21, p = .044$ (Abb. 47b). Die Effekte von wahrgenommener Relevanz und situationalem Interesse auf den Zuwachs der Selbstwirksamkeit in Bezug auf die Planung von Experimenten im Unterricht sind dabei allerdings jeweils klein bis mittelstark und klären etwa $R^2 \approx 5\,\%$ der Varianz der Selbstwirksamkeitserwartungen auf. Für die Selbstwirksamkeit in Bezug auf die Durchführung von Experimenten im Unterricht sind die Effekte von ähnlicher Größenordnung. Für die wahrgenommene Relevanz ist der Effekt auf die Änderung der Selbstwirksamkeit bzgl. Durchführung signifikant mit $\beta = .22, p = .041$ und für das situationale Interesse signifikant mit $\beta = .21, p = .045$ (Tab. 30, Abb. 47c und d).

Zusammenfassend zeigen die einzelnen Regressionen signifikante, positive Zusammenhänge zwischen den untersuchten Größen. Die gezeigten Effekte sind ein Indikator für Plausibilität der Wirkannahme, dass eine erhöhte wahrgenommene Relevanz bzw. eine größeres situationales Interesse sich positiv auf den Zuwachs an Selbstwirksamkeit auswirkt. Die Effekte zeigen damit auch, dass die Voraussetzungen für ein Mediationsmodell gegeben sind.

Damit kann nun im nächsten Schritt das gesamte Wirkmodell betrachtet werden, um die Effekte unter Berücksichtigung aller Einflussgrößen zu schätzen. Analog zur Untersuchung des Wirkmodells in Bezug auf die experimentelle Kompetenz wird dafür nachfolgend eine Strukturgleichungsmodellierung auf der Datenebene *Zwischen Personen* ($n = 88$) für die Daten der Selbstwirksamkeitserwartungen im ersten Praktikumssemester durchgeführt. Obwohl die Selbstwirksamkeitserwartungen bzgl. der Planung und Durchführung einen hohen positiven Zusammenhang von $\beta = .59, p < .001$ aufweisen (Tab. 30), werden die beiden Komponenten nicht in ei-

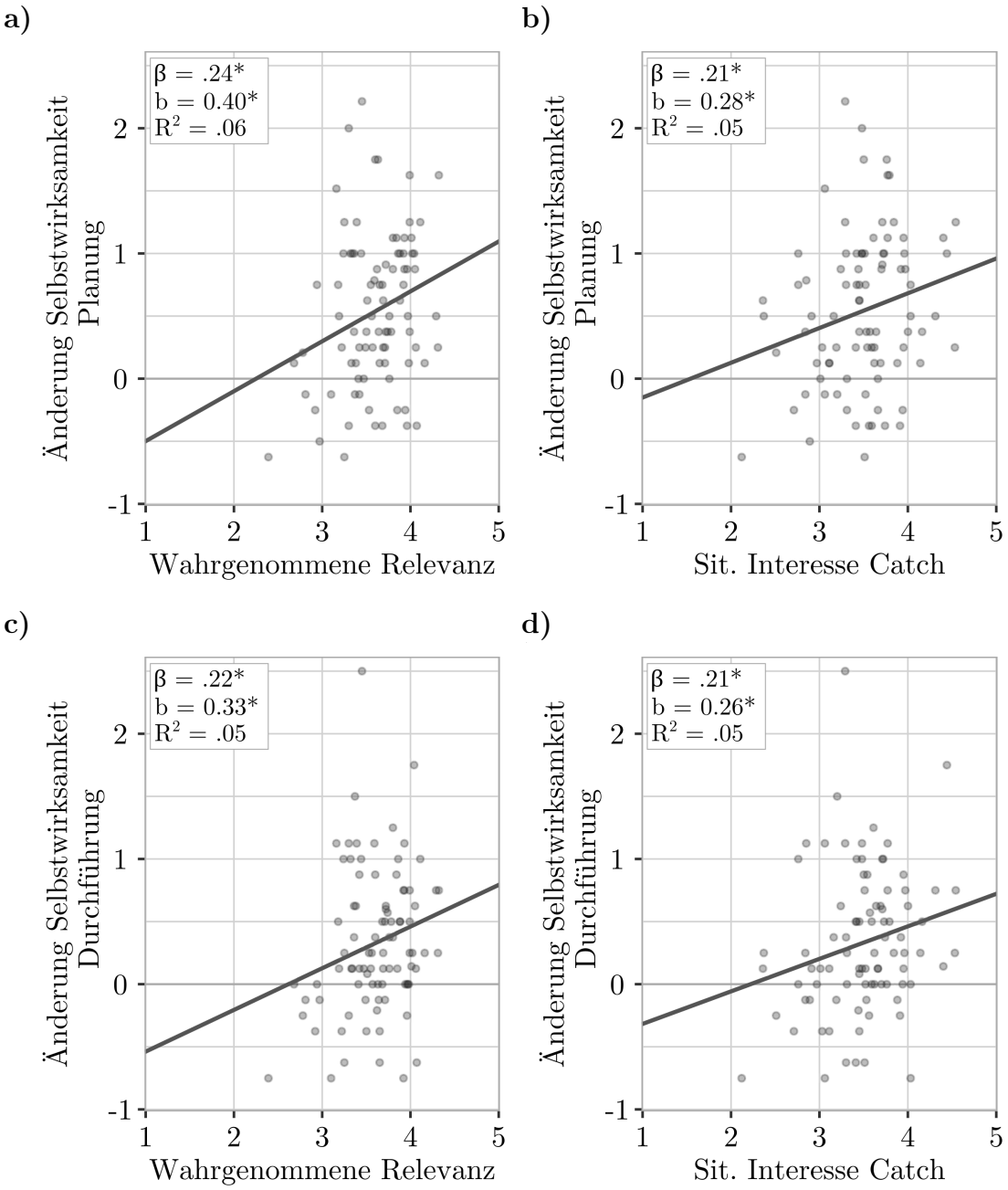


Abbildung 47: *Änderung der Selbstwirksamkeit in Bezug auf Planung (obere Reihe a) und b)) und Durchführung (untere Reihe c) und d)) von Experimenten (5-teilige Likert-Skala) abhängig von der wahrgenommenen Relevanz (linke Spalte a) und c)) und dem situationalen Interesse Catch (rechte Spalte b) und d)) (jeweils Personenmittelwerte, 5-teilige Likert-Skala) auf der Datenebene Zwischen Personen für $n = 88$ Personen. Die Messwerte wurden semi-transparent eingezeichnet, um übereinanderliegende Datenpunkte deutlich zu machen. Die Regressionsparameter entsprechen denen in Tabelle 30.*

Abh. Variable	**Unabh. Variable**	β	b	p	R^2
Sit. Interesse Catch	Relevanz	$.70^{***}$	0.87^{***}	$< .001$	.49
SWE Planung	Relevanz	$.24^{*}$	0.40^{*}	.021	.06
	Sit. Interesse Catch	$.21^{*}$	0.28^{*}	.044	.05
SWE Durchführung	Relevanz	$.22^{*}$	0.33^{*}	.041	.05
	Sit. Interesse Catch	$.21^{*}$	0.26^{*}	.045	.05
	SWE Planung	$.59^{***}$	0.56^{***}	$< .001$	.35

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$, $n = 88$

Tabelle 30: *Ergebnisse der Regressionsanalysen einzelner Zusammenhänge von wahrgenommener Relevanz, situationalem Interesse Catch (jeweils Personenmittelwerte, 5-teilige Likert-Skala) und Änderung der Selbstwirksamkeitserwartungen (SWE, Post-Prä) in Bezug auf Planung und Durchführung von Experimenten im Unterricht (5-teilige Likert-Skala) auf der Datenebene Zwischen Personen für $n = 88$ Personen im ersten Praktikumssemester.*

nem Indikator für das Messmodell des Strukturgleichungsmodells zusammengefasst, sondern nacheinander in einzelnen Modellen betrachtet (Wirkmodelle C und D). Dadurch soll in der Modellierung berücksichtigt werden, dass aufgrund der Gestaltung der Praktikumsversuche mehr Lerngelegenheiten für die Durchführung als für die Planung von Experimenten vorhanden sind und deshalb beide Komponenten der Selbstwirksamkeitserwartungen im Praktikum, wie aus Abschnitt 12.2.3 hervorgeht, unterschiedlich beeinflusst werden.

Abbildung 48 zeigt das Strukturgleichungsmodell der Mediation von situationalem Interesse zwischen wahrgenommener Relevanz und der Änderung der Selbstwirksamkeit bzgl. der Planung (Wirkmodell C). Zur Beurteilung der Modellgüte werden Fit-Kennwerte der Modellierung bestimmt (siehe Abschnitt 11.2.3). Alle Fit-Kennwerte weisen auf eine gute Modellpassung hin (χ^2-Test: $\chi^2(51) = 49.7, p = .525$, SRMR= .04, RMSEA= .00, CFI= 1.00). Das Modell spiegelt die Daten damit gut wider, sodass die geschätzten Parameter des Modells sinnvoll interpretiert werden können. Der Gesamteffekt $\beta_{tot} = .29, p = .046$ zwischen wahrgenommener Relevanz und Änderung der Selbstwirksamkeit als Summe des direkten und indirekten Effektes ist positiv und von mittlerer Größe. Nach Aufteilung in direkte und indirekte Anteile sind die Teileffekte hingegen nicht signifikant von Null verschieden. Der Effekt teilt sich dabei etwa zur Hälfte in einen direkten ($\beta_{di} = .16, p = .437$) und einen indirekten Teileffekt ($\beta_{ind} = .13, p = .401$) auf, die beide nicht signifikant sind. Der indirekte Effekt ist bestimmt durch einen höheren signifikanten Effekt zwischen Relevanz und Interesse ($\beta = .74, p < .001$) und einem kleineren nicht signifikanten Effekt zwischen Interesse und Änderung der Selbstwirksamkeit ($\beta = .18, p = .399$). Dieses Modell klärt damit 10 % der Varianz der Änderung der Selbstwirksamkeitserwartungen bzgl. der Planung von Experimenten auf.

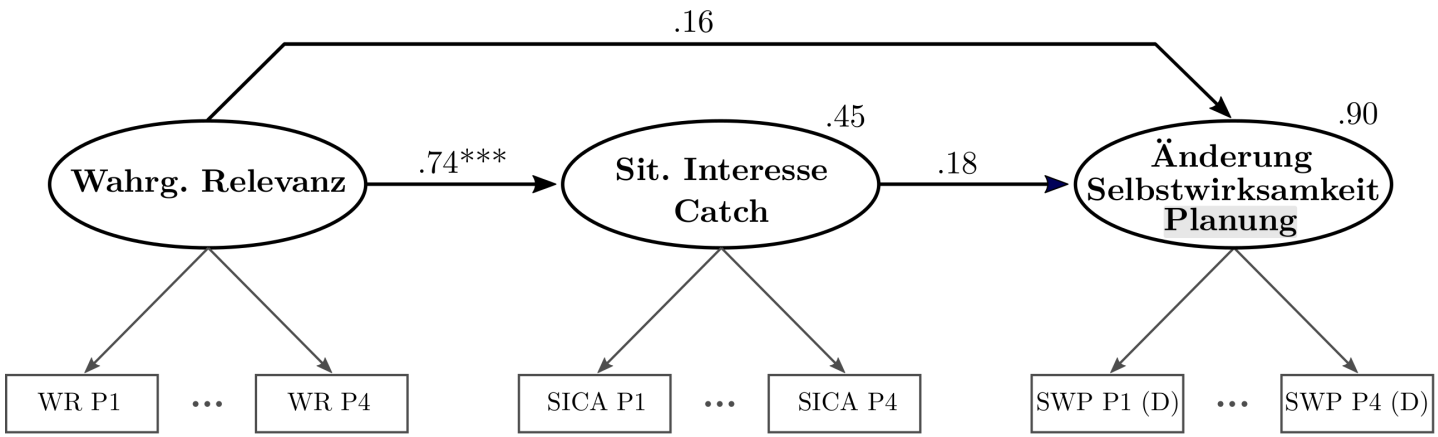


Abbildung 48: *Ergebnis des Strukturgleichungsmodells der Mediation von situationalem Interesse Catch zwischen wahrgenommener Relevanz und Änderung der Selbstwirksamkeit in Hinsicht auf die* ***Planung*** *von Experimenten im Physikunterricht für $n = 88$ Studierende im ersten Praktikumssemester auf der Datenebene Zwischen Personen (Wirkmodell C). Abgebildet sind die standardisierten Pfadkoeffizienten. Die Werte an den abhängigen Variablen stellen die Residualvarianz dar. Das Signifikanzniveau ist an folgender Kodierung zu erkennen: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$. Für das Modell liegen folgende Fit-Kennwerte vor: χ^2-Test: $\chi^2(51) = 49.7, p = .525$, SRMR= .04, RMSEA= .00, CFI= 1.00.*

Abbildung 49 zeigt entsprechend das Strukturgleichungsmodell der Änderung der Selbstwirksamkeitserwartungen in Bezug auf die Durchführung von Experimenten (Wirkmodell D). Auch für diese Modellierung wird aus den Fit-Kennwerten eine gute Modellpassung geschlussfolgert (χ^2-Test: $\chi^2(51) = 58.5, p = .220$, SRMR= .05, RMSEA= .04, CFI= .99). Die wahrgenommene Relevanz hat einen mittleren positiven, signifikanten Gesamteffekt $\beta_{tot} = .25, p = .033$ auf die Änderung der Selbstwirksamkeit. Nach Aufteilung in direkte und indirekte Anteile sind die Teileffekte hingegen nicht signifikant von Null verschieden. Betrachtet man nur die Größen der Teileffekte, dann teilt sich der Gesamteffekt etwa zu zwei Drittel auf den direkten ($\beta_{di} = .16, p = .308$) und zu einem Drittel auf den indirekten Effekt ($\beta_{ind} = .09, p = .508$) auf. Der indirekte Effekt ergibt sich dabei aus einem höheren signifikanten Effekt ($\beta = .74, p < .001$) zwischen Relevanz und Interesse und einem kleineren nicht signifikanten Effekt ($\beta = .12, p = .498$) zwischen Interesse und Änderung der Selbstwirksamkeit. Insgesamt werden 7 % der Varianz der Änderung der Selbstwirksamkeitserwartungen bzgl. der Durchführung von Experimenten in diesem Modell aufgeklärt.

Zur Absicherung der hier geschätzten Modellparameter wurden alternative Modellierungen durchgeführt, wie in Abschnitt 11.2.6 ausgeführt. Modellierungen der Selbstwirksamkeitserwartungen über beide Semester oder mit beiden Komponenten als Indikatoren für ein latentes Konstrukt in einem gemeinsamen Modell zeigen bei einer guten Modellpassung ähnliche Modellschätzungen. Werden die Selbstwirksamkeitserwartungen im Wirkmodell statt über die absolute Änderung über den Wert des Post-Tests unter Kontrolle des Prä-Tests modelliert, ist der Gesamteffekt der wahrgenommenen Relevanz auf die Änderung der Selbstwirksamkeitserwartungen gleichwertig mit dem im hier dargestellten Modell. Im vorigen Abschnitt 12.2.3 hat sich gezeigt, dass die Änderung der Selbstwirksamkeit für beide Komponenten negativ von der Ausgangsausprägung abhängt. Eine Aufnahme der Ausgangsausprägung in das hier betrachtete Modell zeigt ebenfalls einen äquivalenten Gesamteffekt. Der Effekt der wahrgenommenen Relevanz auf die Änderung der Selbstwirksamkeit ist in allen alternativen Modellierungen ebenfalls von mittlerer Größenordnung, sodass auch in diesen ein Nachweis der direkten und indirekten Teileffekte für diese Stichprobe ($n = 88$) nicht gelingt. Diese zusätzlich betrachteten Modellierungen stützen die hier gefundenen Effekte.

Insgesamt kann auf Basis dieser Analyse der Zusammenhänge (Wirkmodelle C und D) festgehalten werden, dass für Studierende, die die Praktikumsversuche im Mittel als relevanter wahrnehmen, die Zunahme der Selbstwirksamkeitserwartungen über das Praktikum signifikant höher ist als für die anderen Studierenden. In Bezug auf die Analyse des Wirkmodells ist dies ein Hinweis auf die Plausibilität der angenommenen Wirkzusammenhänge.

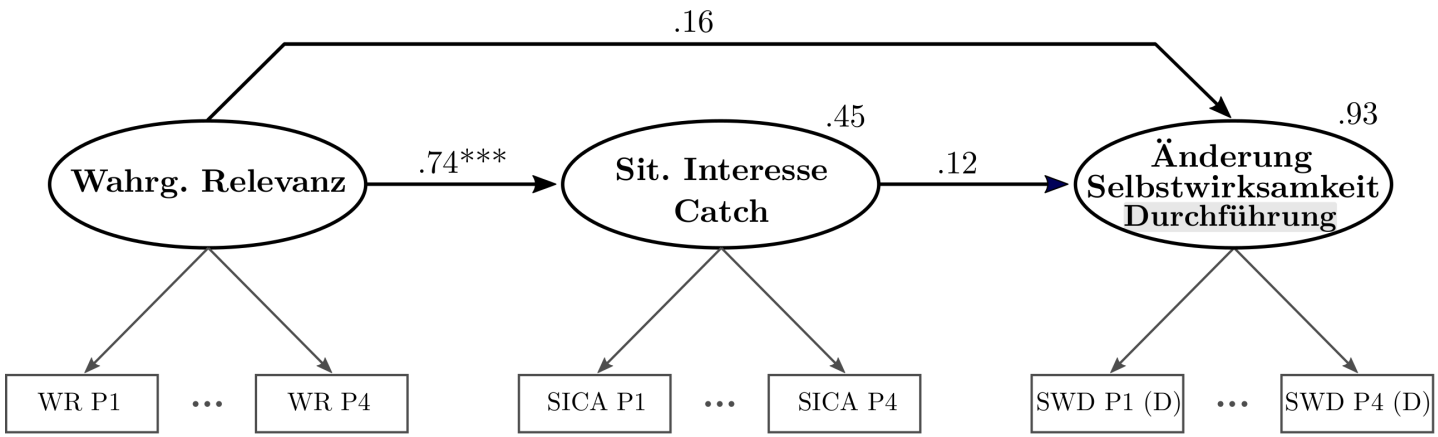


Abbildung 49: *Ergebnis des Strukturgleichungsmodells der Mediation von situationalem Interesse Catch zwischen wahrgenommener Relevanz und Änderung der Selbstwirksamkeit in Hinsicht auf die* ***Durchführung*** *von Experimenten im Physikunterricht für $n = 88$ Studierende im ersten Praktikumssemester (Wirkmodell D). Abgebildet sind die standardisierten Effekte. Die Werte an den abhängigen Variablen stellen die Residualvarianz dar. Das Signifikanzniveau ist an folgender Kodierung zu erkennen: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$. Für das Modell liegen folgende Fit-Kennwerte vor: χ^2-Test: $\chi^2(51) = 58.5, p = .220$, SRMR= .05, RMSEA= .04, CFI= .99.*

12.3 Fachwissen

Das Laborpraktikum hat unter anderem die Funktion, das fachphysikalische Wissen, das sich die Studierenden in den Grundvorlesungen mit zugehörigen Übungen angeeignet haben, aufzufrischen und zu vertiefen (siehe z. B. Abschnitt 3.1). Nicht zuletzt soll daher nach den Analysen der Wirkmodelle mit Interesse und experimenteller Kompetenz nun auch das Fachwissen als dritte Hauptkomponente des Lernerfolgs im Wirkmodell des Lehramtspraktikums untersucht werden. Um die Prüfung des Wirkmodells als Hauptziel dieser Analyse angehen zu können, sind wie in den vorangegangenen Analysen die Analyseschritte aus Abschnitt 11.2.2 auch für die Betrachtung des Fachwissens erforderlich. Anhand dessen gliedert sich dieser Abschnitt in die Analyse des Lernerfolgs in Bezug auf das Fachwissen (Abschnitt 12.3.1) und die darauf aufbauende Prüfung der diesbezüglich angenommenen Wirkzusammenhänge (Abschnitt 12.3.2).

12.3.1 Fachwissenszuwachs

Um die Wirksamkeit des Praktikums beurteilen zu können, wird zunächst der Frage nachgegangen, inwieweit das Fachwissen der Lehramtsstudierenden über das Praktikum hinweg zunimmt. Dabei wird unterschieden zwischen Fachwissenszuwachs, der mittels Prä-Post-Fachwissenstests erhoben wurde, und jenem, der über eine Selbsteinschätzung erfasst wurde (siehe Abschnitt 11.1.3).

Fachwissenszuwachs anhand des Wissenstests

Mithilfe der Fachwissenstests vor und nach dem Praktikumssemester wird objektiv physikalisches Faktenwissen und Verständnis zu den einzelnen Themenbereichen erfasst, wie in Abschnitt 11.1.3 erläutert. Um den Lernerfolg anhand des Wissenstests beurteilen zu können, sind verschiedene Teilaspekte zu betrachten. Eine erste Frage ist die nach der Zulässigkeit der Vereinfachung der Datenstruktur durch Annahme der Vernachlässigbarkeit von Kohorteneffekten (siehe Abschnitt 11.2.1). Für den Fachwissenszuwachs ist dabei auch besonders der mögliche Einfluss der Einführung der Benotung des Praktikums ab dem Wintersemester 2017/2018 relevant. Nach Prüfung dieser Bedingungsfaktoren kann dann eine Aussage zum Lernerfolg bzgl. des Fachwissens getroffen werden.

Im ersten Schritt wird das Fachwissen mithilfe von Stichprobenvergleichen auf Kohorteneffekte geprüft. Das Fachwissen wird dazu als der Anteil der erreichten Punktzahl von der maximal möglichen Punktzahl betrachtet. Der Testwert liegt damit innerhalb des Intervalls [0,1] und gibt den Anteil des korrekt ausgefüllten Tests an (siehe Abschnitt 11.1.3). Anhand unabhängiger einfaktorieller Varianzanalysen und Kruskal-Wallis-Tests wird geschlossen, dass das Fachwissen innerhalb der Semesterart mit den unterschiedlichen Themenbereichen (Winter-/Sommersemester, WS/SS) für die unterschiedlichen Messzeitpunkte (Prä, Post) und dem absoluten

Zuwachs (Delta=Post-Prä) keine Unterschiede bzgl. der Kohorten aufweist[10]. Zudem hat ein abhängiger t-Test gezeigt, dass für den Zuwachs an Fachwissen kein signifikanter Unterschied zwischen der Semesterart nachweisbar ist[11]. In Bezug auf die Kohorteneffekte ist dabei insbesondere auch von Interesse, ob die Einführung der Benotung des Praktikums mit der neuen Fachprüfungsordnung (CAU, 2017b) ab dem Wintersemester 2017/2018 (siehe Abschnitt 5.3), also dem vierten von sieben Semestern der Laufzeit dieser Studie, einen Einfluss auf den Fachwissenszuwachs hat. In die Modulnote fließt die Bewertung von Vorgesprächen und Protokollen ein, die auch gleichermaßen vor Einführung der Benotung wesentliche Elemente im Lehramtspraktikum dargestellt haben. Die Benotung dieser Elemente könnte jedoch eine extrinsische Motivation darstellen, durch die die Studierenden die fachlichen Inhalte besser vorbereiten, im Protokoll sorgfältiger nacharbeiten und dadurch im Praktikum insgesamt mehr lernen. Um diesen Aspekt zu untersuchen, wurden Stichprobenvergleiche durchgeführt. Unabhängige t-Tests und Wilcoxon-Rangsummen-Tests zeigen, dass die Benotung des Vorgespräches keinen signifikanten Unterschied bzgl. des Fachwissenszuwachses hervorruft[12]. Die Ergebnisse sprechen somit gegen die aufgestellte These, dass eine Benotung des Praktikums zu einer signifikanten Erhöhung des Lernerfolges bzgl. des Fachwissens führt und damit einen wesentlichen Einflussfaktor darstellt. Insgesamt zeigt diese Prüfung des Einflusses von Kohorteneffekt und Benotung somit, dass die Zusammenfassung der Kohorten (Sommersemester 2016 - 2019) ohne Berücksichtigung der Benotung für das Fachwissen möglich ist. Aufgrund des ähnlichen Zuwachses des Fachwissens in Winter- und Sommersemester kann für folgende Untersuchungen auch die Vereinfachung vorgenommen werden, Personenwerte des Fachwissenszuwachses als Anteil der maximal möglichen Punktzahl über die durchgeführten Praktikumsteile zu bilden und zu vergleichen, auch wenn Themenbereiche im Praktikum nicht bearbeitet wurden.

Im nächsten Schritt kann nun die eigentliche Analyse des Fachwissens erfolgen. Tabelle 31 fasst die wichtigsten deskriptiven Werte für die Personenwerte des Fachwissens im Prä- und Post-Test zusammen. Die Mittelwerte für die beiden Messzeitpunkte zeigen mit $M_{Pre} = .46 \pm .01$ und $M_{Post} = .59 \pm .01$, dass die Studierenden im Mittel etwa die Hälfte des Fachwissenstests korrekt gelöst haben (Tab. 31). Diese rund 50 % im Mittel korrekten Antworten entsprechen der idealen mittleren Lösungswahrscheinlichkeit für Leistungstests, um das Wissen möglichst differenziert zu messen und Decken- oder Bodeneffekte möglichst gering zu halten, wie in Abschnitt 11.1.3 beschrieben. Der Boxplot in Abbildung 50 veranschaulicht die Verteilung

[10] WS: Prä: $F(2, 71) = 0.10, p = .905$; Post: $F(2, 72) = 1.38, p = .259$, $H(2) = 2.35, p = .308$; Delta: $F(2, 72) = 1.32, p = .273$; SS: Prä: $F(3, 80) = 2.06, p = .112$; Post: $F(3, 79) = 2.80, p = .045$, $H(3) = 7.21, p = .066$; Delta: $F(3, 79) = 0.12, p = .947$

[11] abhängiger t-Test: $M_{WS} = .13 \pm .01, n_{WS} = 75, M_{SS} = .14 \pm .01, n_{SS} = 83, t(69) = 1.31, p = .193, d = 0.23$

[12] WS: $M_{WS,Pruef} = .14 \pm .01, n_{WS,Pruef} = 57, M_{WS,ohne} = .10 \pm .02, n_{WS,ohne} = 18, t(31.0) = 1.39, p = .174, d = 0.36$, SS: $M_{SS,Pruef} = .14 \pm .01, n_{SS,Pruef} = 53, M_{SS,ohne} = .15 \pm .02, n_{SS,ohne} = 30, t(44.1) = 0.45, p = .657, d = 0.11$

der Daten für die beiden Messzeitpunkte. Der Shapiro-Test bestätigt den visuellen Eindruck anhand des Boxplots, dass die Testwerte des Fachwissens zu beiden Zeitpunkten normalverteilt sind (Prä: $W = 0.99, p = .552$; Post: $W = 0.98, p = .350$). Der Levene-Test zur Prüfung der Varianzhomogenität zeigt, dass die Streuung zu beiden Zeitpunkten ähnlich ist ($F(1, 175) = 1.00, p = .320$). Das mittlere Fachwissen im Post-Test ist im Vergleich zum Prä-Test um $M_{Delta} = .13 \pm .01$ zu höheren Werten verschoben. Ein Vergleich dieser Mittelwerte mithilfe eines t-Tests für abhängige Stichproben bei gegebener Varianzhomogenität (s. o.) und Normalverteilung der Differenzen (Shapiro-Test: $W = 98, p = .272$) zeigt, dass sich das Fachwissen an den beiden Zeitpunkten vor und nach dem Praktikum signifikant unterscheidet ($t(87) = 16.45, p < .001$). Die Berechnung der Effektstärke zeigt entsprechend einen großen Effekt (Cohens $d = 1.05$).

Messzeitpunkt	n	M	SD	SE	MD	MIN	MAX
Prä	89	.46	.13	.01	.46	.14	.73
Post	88	.59	.12	.01	.60	.29	.84

Tabelle 31: *Deskriptive Statistik für das Fachwissen im Prä- und Post-Test als personenbezogener Anteil des maximal möglichen Testwertes.*

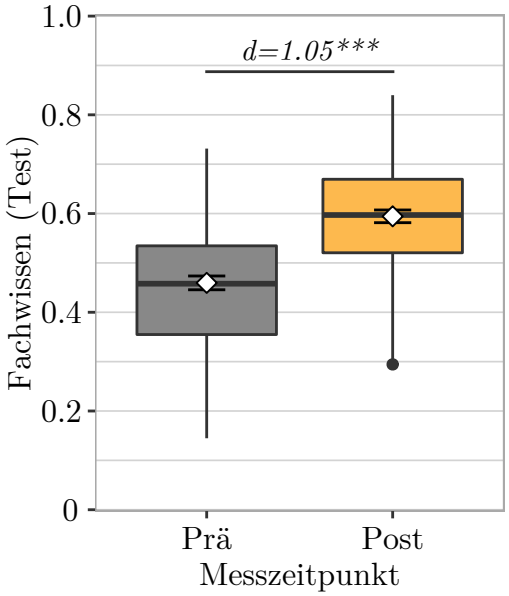


Abbildung 50: *Darstellung des Fachwissens Physik als Anteil des maximal möglichen Testwertes vor (Prä) und nach (Post) dem Praktikum. Der Boxplot zeigt neben Minimum, unterem Quartil, Median, oberem Quartil und Maximum zusätzlich die Mittelwerte (weiße Rauten) mit Standardfehlern. Ergänzt wurde zudem die Effektgröße Cohens d und das Signifikanzniveau aus dem Vergleich der Mittelwerte anhand eines t-Tests. Die zugehörigen deskriptiven Werte sind in Tabelle 31 gelistet.*

In Abbildung 51 sind die anteiligen Testwerte im Prä- und Post-Test für den gesamten Fachwissenstest und für die einzelnen Themenbereiche des Tests gegenübergestellt. In der Betrachtung des Testinstrumentes in Abschnitt 11.1.3 wurde deutlich, dass die Reliabilitätswerte auf diesen Subskalen aufgrund der Heterogenität des Instrumentes fragwürdig sind. Dieser Einflussfaktor muss bei der Betrachtung berücksichtigt werden, weshalb ein verstärktes Augenmerk auf die Plausibilität der Ergebnisse gelegt wird. Die Testmittelwerte für die sechs Themenbereiche liegen im Prä-Test im Bereich von $.33 \leq M_{Pre,Thema} \leq .57$ und im Post-Test im Bereich von $.43 \leq M_{Post,Thema} \leq .72$ (Abb. 51, Anhang A.2.7). Für Atomphysik, Mechanik und Messtechnik sind die anteiligen Testwerte im Prä- und Post-Test höher und für Wärmelehre, Optik und Elektrizitätslehre (E-Lehre) niedriger als die Testmittelwerte des Gesamttests. Die hohen Testwerte für Mechanik sind dabei plausibel, da Mechanik ein zentrales Thema in der schulischen Physikausbildung ist und die hochschulische Physikausbildung ebenfalls viele Lerngelegenheiten bietet. Die Testwerte für den Bereich Wärmelehre sind dagegen erwartungsgemäß niedrig, da dieser Themenbereich in der Schule deutlich weniger behandelt wird als die anderen Themenbereiche. In der Hochschulausbildung wird Wärmelehre am Ende des ersten Semesters behandelt und nimmt dadurch oft einen geringeren Umfang in den Lehrveranstaltungen und Prüfungen im Vergleich zur Mechanik ein, die am Anfang des Semesters gelehrt wird. Die Fachwissensaufgaben im Themenbereich Messtechnik erheben keine messtechnischen Aspekte, sondern setzen sich zu drei Viertel aus der Mechanik und zu einem Viertel aus der Elektrizitätslehre zusammen, da anhand dieser Themen in den Versuchen messtechnischer Inhalt vermittelt wird. Die Testwerte für Messtechnik sind als eine Zusammensetzung der beiden Wissensbereiche Mechanik mit den höchsten Testwerten und Elektrizitätslehre mit den niedrigsten Testwerten nachvollziehbar und aufgrund des hohen Anteils an Mechanik auch entsprechend hoch. Damit sind die Testwerte für die einzelnen Themenbereiche relativ zueinander plausibel, sodass die Betrachtung des Fachwissens auf Themenebene trotz der fragwürdigen Werte für Cronbachs α (siehe Abschnitt 11.1.3) vertretbare Hinweise auf den Fachwissenszuwachs liefert.

Für die einzelnen Themenbereiche ist der absolute Lernzuwachs in Abbildung 51 als Änderung der Testwerte vom Prä- zum Post-Test bestimmbar und liegt im Bereich von ($.08 \leq M_{Delta,Thema} \leq .19$). Mithilfe von abhängigen t-Tests wurde überprüft, ob diese Änderung jeweils statistisch signifikant ist. Das Signifikanzniveau und die Effektgröße Cohens d sind dafür in der Abbildung 51 ergänzt. Da nicht in allen Fällen die Voraussetzung von Varianzhomogenität und Normalverteilung für den abhängigen t-Test gegeben waren, wurden die Ergebnisse mit dem Wilcoxen-Vorzeichen-Rang-Test abgesichert. Die Ergebnisse aus dem Mittelwertvergleich zeigen, dass die Unterschiede auch für alle einzelnen Themenbereiche signifikant sind. Die Effektstärken sind dabei im mittleren bis großen Bereich von $0.47 \leq d \leq 0.96$. Besonders auffällig ist, dass der Fachwissenszuwachs für den Themenbereich Messtechnik minimal ist ($M_{Delta,MT} = .08 \pm .02$), obwohl der Großteil der Fachwissensauf-

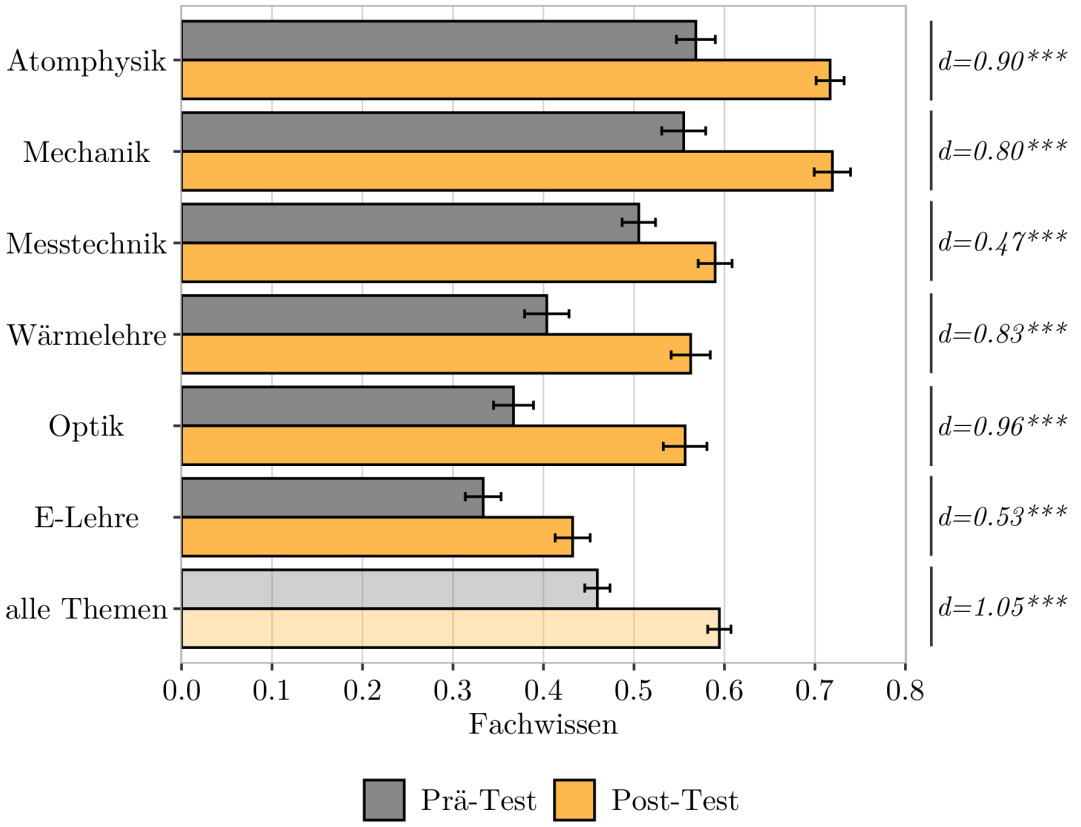


Abbildung 51: *Das mittlere Fachwissen Physik mit Standardfehler im Prä- (grau) und Post-Test (orange) für die einzelnen Themenbereiche ($73 \leq n \leq 88$, weitere deskriptive Werte im Anhang A.2.7) und für den Gesamttest ($n = 88$, helleres grau und orange). Das Fachwissen ist als Anteil des maximal möglichen Testwertes angegeben. Ergänzt wurde zudem die Effektgröße Cohens d und das Signifikanzniveau aus dem Vergleich der Mittelwerte anhand eines t-Tests.*

gaben zu den Messtechnik-Versuchen aus dem Themenbereich Mechanik stammen, für den der Fachwissenszuwachs zu den größten zählt ($M_{Delta,ME} = .17 \pm .02$). Dies kann daran liegen, dass der Schwerpunkt der Versuche stärker die Messtechnik und weniger die fachlichen Inhalte betrifft und damit der Lernzuwachs nicht vollständig mit einem Fachwissenstest erfasst werden kann. Da die Themenbereiche mit den einzelnen Praktikumsversuchen konfundiert sind, wird diese Auffälligkeit auch als Hinweis auf Varianz im Fachwissenszuwachs gesehen, die durch die einzelnen, unterschiedlich gestalteten Praktikumsversuche hervorgerufen wird. Festzuhalten ist, dass das Fachwissen über alle Themenbereiche hinweg zunimmt und sich nicht auf einzelne Bereiche beschränkt.

Einen zusätzlichen Hinweis auf die Zunahme des Fachwissens liefert der Bearbeitungsgrad des Fachwissenstests, der im Rahmen der Itemanalyse betrachtet wurde (Tab. 19; Anhang A.2.7). Geht man davon aus, dass die Studierenden Fachwissensaufgaben überspringen, für die sie die Antwort nicht kennen, dann ist der Bearbeitungsgrad implizit ein Maß für das Fachwissen. Erhöht sich der Bearbeitungs-

grad, erhöht sich für die Studierenden die Einschätzung der eigenen Kompetenz, die Fachwissensaufgabe korrekt zu beantworten. Sehr wahrscheinlich beeinflussen den Bearbeitungsgrad allerdings zusätzlich auch andere Faktoren wie die Motivation, sich Zeit zum Ausfüllen des Tests zu nehmen. Die mittleren Bearbeitungsgrade des Fachwissenstests im Prä- und Post-Test ($M_{Pre} = .71 \pm .03$, $M_{Post} = .77 \pm .02$) unterscheiden sich bei gegebener Normalverteilung und Varianzhomogenität anhand eines abhängigen t-Tests signifikant (Shapiro-Test: Post-Prä: $W = 0.98, p = .649$; Levene-Test: $F(1, 94) = 2.82, p = .072$; t-Test: $t(47) = 4.25, p < .001$). Die Effektstärke zeigt einen kleinen Effekt mit Cohens $d = 0.32$. Die Zunahme des mittleren Bearbeitungsgrades weist somit ebenfalls auf eine signifikante Erhöhung des Fachwissens hin.

Die Untersuchungen der Fachwissensdaten zeigen außerdem, dass das Vorwissen gemäß dem Matthäus-Effekt (z. B. Renkl, 1996) einen hohen positiven Effekt auf das Fachwissen im Post-Test hat ($\beta = .81, p < .001, R^2 = .65$). Zudem zeigt sich ein mittlerer negativer Effekt des Vorwissens auf den absoluten Zuwachs an Fachwissen ($\beta = -.44, p < .001, R^2 = .19$). Studierende mit einem hohen Vorwissen haben damit nach dem Praktikum ebenfalls ein hohes Fachwissen. Allerdings ist der Lernzuwachs durch die Begrenzung der Fachwissensaufgaben erschwert, sodass der Lernzuwachs für Studierende mit hohem Vorwissen geringer ist als für Studierende mit geringem Vorwissen (siehe Abschnitt 11.2.6). Das Vorwissen ist somit eine nicht vernachlässigbare Einflussgröße bei der weiteren Analyse des Fachwissens.

Andere Einflussgrößen, die sich verzerrend auf die Messergebnisse des Fachwissenstests auswirken können, wurden bei der Erhebung möglichst gering gehalten. Um beispielsweise Übungseffekte durch das wiederholte Bearbeiten der gleichen Fachwissensaufgaben im Post- wie im Prä-Test, zu verringern, wurde die Reihenfolge der Aufgaben in den Testheften variiert. Ebenfalls könnte es zu einer Verzerrung der Messergebnisse des Fachwissenszuwachses im Praktikum führen, wenn die Studierenden in der Zeit des Praktikums andere Lerngelegenheiten oder Prüfungen mit den Inhaltsbereichen des Praktikums wahrgenommen haben. Bei einem Regelstudium werden die Themenbereiche Mechanik, Wärmelehre, Optik und Elektrizitätslehre in den ersten beiden Semestern vor dem Anfängerpraktikum in den Lehrveranstaltungen *Physik I und II* bearbeitet (siehe Abb. 4). Das Bestehen von *Physik I* ist dabei zudem Zulassungsvoraussetzung für das Praktikum. Laut Studienplan liegen parallel zum Anfängerpraktikum Lehramt die Lehrveranstaltungen zur *Physik der Materie I und II*. In diesen Lehrveranstaltungen ist Atomphysik Schwerpunkt des dritten Semesters, während Atomphysik im Anfängerpraktikum im Regelfall erst im vierten Semester Thema ist. Sofern die Studierenden somit nicht deutlich vom Regelstudienplan abweichen, kann davon ausgegangen werden, dass keine weiteren Lerngelegenheiten parallel zum Praktikum bestehen, in denen ebenfalls die Praktikumsinhalte vermittelt werden. Ohne Frage kann trotzdem nicht ausgeschlossen werden, dass die Studierenden vom Studienplan abgewichen sind. Es wird allerdings davon ausgegangen, dass es sich in dem Fall um Einzelfälle handelt, die in der Summe

der rund 90 Studierenden einen geringen Störeinfluss haben. Auch wenn Störeffekte somit nicht vollständig ausgeschlossen werden können, weisen die gefundenen hohen Effektstärken auf belastbare Effekte in Bezug auf den Zuwachs an Fachwissen hin.

Insgesamt wird für diese Teilanalyse des Lernerfolges bzgl. des Fachwissens abschließend festgehalten, dass die Lehramtsstudierenden mit einem mittleren Maß an Vorwissen ins Praktikum starten und dieses Wissen im Laufe des Praktikums ausbauen. Die Zunahme des Fachwissens ist dabei nachweisbar und sogar mit großem Effekt. Dies zeigt, dass dieses wesentliche Lernziel bzgl. der Wiederholung und Vertiefung von Fachwissen im neuen Lehramtspraktikum erreicht wurde.

Fachwissenszuwachs anhand der Selbsteinschätzung

Der Fachwissenstest beruht auf einer sehr breiten Abfrage von fachlichen Inhalten mit verhältnismäßig wenig Aufgaben und ist dadurch sehr heterogen. Aufgrund dieser Heterogenität ist die Analysierbarkeit des Fachwissenstests begrenzt, wie in Abschnitt 11.1.3 erörtert. Aus diesem Grund wurde zusätzlich eine Abfrage des Fachwissenszuwachses auf einer übergeordneten, globaleren Ebene mittels Selbsteinschätzung vorgenommen. Aus diesem Grund eignet sich die Selbsteinschätzung sehr gut um eine Referenz zum Wissenstest zu bilden. Diese Selbsteinschätzung des Fachwissenszuwachses soll nun im nächsten Schritt hinsichtlich des Lernerfolges und damit der Wirkung des Praktikums analysiert werden.

Für diese Analyse wird analog zur Analyse des Lernerfolges in Bezug auf die Selbsteinschätzung des Zuwachses an experimenteller Kompetenz vorgegangen (siehe Abschnitt 12.2.1). Der Boxplot in Abbildung 52 veranschaulicht für die Deskription die Selbsteinschätzung des Fachwissenszuwachses an den beiden Messzeitpunkten nach dem ersten bzw. zweiten Semester (Post 1 und Post 2). Die Stichproben werden dabei zusammengefasst, da mittels entsprechender Tests auf Stichprobenunterschiede im Vorfeld für die Daten keine signifikanten Unterschiede bzgl. der Semesterart (Winter-/Sommersemester)[13] und keine Kohorteneffekte bzgl. des Semesters der Erhebung (Sommersemester 2016 - 2019)[14] festgestellt wurden. Die deskriptiven Werte, die zusätzlich in Tabelle 32 gelistet wurden, zeigen anhand der Mittelwerte $M_{Post1} = 4.89 \pm 0.06$ und $M_{Post2} = 4.93 \pm 0.07$ auf einer 6-teiligen Likert-Skala für die beiden Messzeitpunkte einen hohen Zuwachs des Fachwissens. Der Boxplot (Abb.52) veranschaulicht, dass die Messwerte den Fachwissenszuwachs hauptsächlich am oberen Rand der Skala liegen. Dies deutet auf eine Tendenz zum Deckeneffekt, auch wenn dies im engeren Sinne kein Leistungstest ist. Außerdem veranschaulicht der Boxplot die Verteilung des Fachwissenszuwachses für die beiden Messzeitpunkte. Dabei fällt die leichte Schiefe der Verteilung und der Ausreißer zum Zeitpunkt Post 1 im Vergleich zur symmetrischen Verteilung zum Zeitpunkt Post 2 auf. Der Shapiro-

[13] t-Tests bzgl. Semesterart: Post 1: $t(26.6) = 3.35, p = .189$; Post 2: $t(7.7) = 1.73, p = .123$

[14] Varianzanalysen bzgl. Kohorte: Post 1: $F(6, 81) = 1.51, p = .185$; Post 2: $F(5, 62) = 0.91, p = .480$

Test bestätigt den visuellen Eindruck, dass der Fachwissenszuwachs bei Zeitpunkt Post 1 nicht normalverteilt ($W = 0.95, p = .001$) und bei Post 2 normalverteilt ist ($W = 0.98, p = .267$). Der Levene-Test zur Prüfung der Varianz-Homogenität zeigt, dass die Streuung zu beiden Zeitpunkten ähnlich ist ($F(1, 154) = 0.008, p = .931$).

Ein Vergleich der Mittelwerte mithilfe eines t-Tests für abhängige Stichproben zeigt für die vollständigen Datensätze, wie anhand der Mittelwerte erwartet, keine Unterschiede zwischen dem Fachwissenszuwachs an den beiden Messzeitpunkten ($t(67) = 0.76, p = .450$). Die Berechnung der Effektstärke zeigt entsprechend einen vernachlässigbar kleinen Effekt (Cohens $d = 0.08$). Da die Differenz der Daten der Messzeitpunkte nicht normalverteilt ist (Shapiro-Test: $W = 0.93, p = .001$), ist diese Testvoraussetzung für einen abhängigen t-Test nicht erfüllt. Daher wurde ergänzend ein Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben und nichtparametrische Daten gerechnet ($W = 896, p = .426$), der das Ergebnis des t-Tests bestätigt. Trotz der unterschiedlichen inhaltlichen Themenbereiche in den Zeiträumen ist diese ähnliche Einschätzung der Zunahme des Fachwissens in beiden Semestern erwartungsgemäß, da die Items global und ohne Bezug auf den spezifischen Lerninhalt formuliert sind (z. B. „Aufgrund dieses Praktikums kann ich wichtige Theorien zu dem behandelten Thema wiedergeben.“, siehe auch Abschnitt 11.1.3).

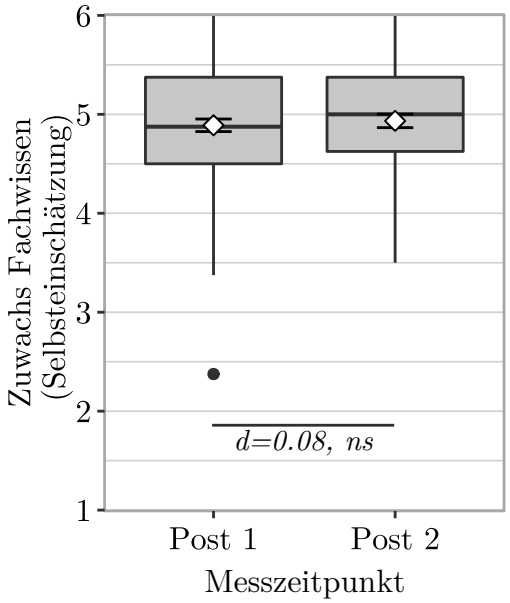


Abbildung 52: *Darstellung der Selbsteinschätzung des Zuwachses des Fachwissens (6-teilige Likert-Skala) am Ende des ersten (Post 1) und zweiten Praktikumssemesters (Post 2). Der Boxplot zeigt neben Minimum, unterem Quartil, Median, oberem Quartil und Maximum zusätzlich die Mittelwerte mit Standardfehlern. Ergänzt wurde zudem die Effektgröße Cohens d und das Signifikanzniveau aus dem Vergleich der Mittelwerte anhand eines t-Tests. Die zugehörigen deskriptiven Werte sind in Tabelle 32 gelistet.*

Messzeitpunkt	n	M	SD	SE	MD	MIN	MAX
Post 1	88	4.89	0.60	0.06	4.88	2.38	6.00
Post 2	68	4.93	0.56	0.07	5.00	3.50	6.00

Tabelle 32: *Deskriptive Statistik für die Selbsteinschätzung des Zuwachses des Fachwissens (6-teilige Likert-Skala) am Ende des ersten (Post 1) und zweiten Praktikumssemesters (Post 2).*

Damit ist das Ergebnis der Analyse in diesem Teilabschnitt, dass der Zuwachs an Fachwissen durch das Praktikum im Mittel sehr hoch von den Studierenden eingeschätzt wird. Rehfeldt (2018b) gibt basierend auf der Testevaluation in mehreren Praktika Deutschlands leicht niedrigere Werte für die Subskalen an ($M_{theo} = 4.30 \pm 0.06$ und $M_{prax} = 4.78 \pm 0.06$, $n = 237$). Dies kann als Indiz gewertet werden, dass die Selbsteinschätzung des Fachwissenszuwachses dieses neu entwickelten Praktikums ($M_{Post1} = 4.89 \pm 0.06, M_{Post2} = 4.93 \pm 0.07$) über dem Mittelwert der Praktika liegt.

Die hohen Werte der Selbsteinschätzung des Fachwissenszuwachses bestätigen das Ergebnis der Analyse des Fachwissenstests im vorangegangenen Abschnitt. Zudem stützt die hohe Einschätzung des Fachwissenszuwachses die These, dass die Selbsteinschätzung den Fachwissenszuwachs in einer größeren Breite als der Wissenstest erfasst und dadurch thematische Lücken schließt. Da davon ausgegangen werden kann, dass die Selbsteinschätzung des Fachwissenszuwachses auch eine gewisse Schnittmenge mit der experimentellen Kompetenz aufweist (siehe Abschnitt 11.1.3), sind die hohen Werte der Selbsteinschätzung des Fachwissenszuwachses somit auch als ein ergänzender Hinweis für das Erreichen des Lernziels zur experimentellen Kompetenz zu sehen.

Insgesamt zeigen die Ergebnisse des Abschnittes damit, dass im neu entwickelten Laborpraktikum für Lehramtsstudierende ein wesentliches Lernziel erreicht wurde: Die Studierenden wiederholen und vertiefen im Verlauf des Laborpraktikums nachweisbar ihr physikalisches Fachwissen. Dies wird als Erfolg für die Neukonzeption des Lehramtspraktikums verbucht.

12.3.2 Wirkmodell mit Fachwissen als Lernerfolg

Mit dem nachgewiesenen Zuwachs an Fachwissen kann nun der angenommene Wirkzusammenhang geprüft werden, ob eine erhöhte wahrgenommene Relevanz bzw. ein größeres situationales Interesse sich positiv auf das Fachwissen auswirkt (Wirkmodelle E und F).

Wirkmodell mit Fachwissen anhand des Wissenstests

Das Fachwissen wurde vor und nach dem Praktikumssemester mit einem Wissenstest gemessen, sodass auch in dieser Untersuchung des Wirkmodells die Untersuchungsebene *Zwischen Personen* eingenommen wird. Für den Zuwachs an Fachwissen wird die Differenz aus Post- und Prä-Testwert als Anteil an der maximal möglichen Punktzahl ($\in [-1, 1]$) betrachtet, wie in Abschnitt 11.2.6 ausführlicher erläutert.

Im Folgenden werden analog zu den vorangegangenen Wirkanalysen zunächst einzelne Zusammenhänge mittels Regressionsanalysen zwischen den Größen wahrgenommene Relevanz, situationales Interesse Catch und Fachwissenszuwachs anhand der Wissenstests betrachtet. Der obere Teil der Tabelle 33 (schwarz) zeigt die be-

stimmten Regressionsparameter zwischen den Variablen für den Gesamtdatensatz. Im unteren Tabellenteil (blau) befinden sich die Regressionsparameter für den auf versuchsnahes Fachwissen reduzierten Datensatz (siehe Abschnitt 11.1.3), der im Anschluss betrachtet wird. Abbildung 53 veranschaulicht ergänzend die Zusammenhänge der Änderung des Fachwissens mit der wahrgenommenen Relevanz bzw. dem situationalen Interesse in Streudiagrammen.

Die wahrgenommene Relevanz hat einen signifikanten, positiven Effekt von $\beta = .29, p = .005$ auf den Zuwachs des Fachwissens. Die Streuung der Personenwerte im Streudiagramm der Größen in Abbildung 53a veranschaulicht, dass es sich dabei um einen mittleren Effekt handelt. Das situationale Interesse Catch hat einen ähnlichen signifikanten Effekt von $\beta = .24, p = .025$ auf den Zuwachs des Fachwissens (Abb. 53b). Die nachgewiesenen Effekte zwischen den Größen zeigen, dass die Wirkannahmen plausibel sind und die Voraussetzungen für die Analyse des Mediationsmodells (siehe Abschnitt 11.2.2) erfüllt sind.

Auf dieser Basis können im nächsten Schritt die einzelnen Wirkzusammenhänge in einem Modell zusammengeführt werden, um Effekte unter Berücksichtigung aller hier betrachteten Einflussgrößen zu bestimmen. Dazu wird in diesem Fall aufgrund der Heterogenität des Fachwissenstests abweichend von der bisher angewendeten Strukturgleichungsmodellierung eine Pfadanalyse berechnet (Wirkmodell E), wie in Abschnitt 11.2.7 ausführlicher erläutert. Die Parameter des Pfadmodelles werden mit den auch zuvor betrachteten Personenmittelwerten für die wahrgenommene Relevanz und das situationale Interesse Catch und den personenbezogenen Testwerten für den Fachwissenszuwachs als manifeste Variablen geschätzt. Das Ergebnis ist im Pfadmodell in Abbildung 54 dargestellt. In schwarz sind die standardisierten Pfadkoeffizienten des hier zunächst betrachteten Gesamtdatensatzes dargestellt. Zur Bewertung der Modellpassung werden die in Abschnitt 11.2.3 vorgestellten Fit-Kennwerte betrachtet. Bis auf den χ^2-Test ($\chi^2(3) = 70, p < .001$) zeigen die Fit-Kennwerte, dass das Pfadmodell die Daten gut widerspiegelt (SRMR= .00, RMSEA= .00, CFI= 1.00) und die Ergebnisse im nächsten Schritt analysiert werden können.

Der signifikante Gesamteffekt $\beta_{tot} = .30, p = .003$ zwischen wahrgenommener Relevanz und Zuwachs an Fachwissen ist positiv und von mittlerer Größe, wie anhand der Betrachtungen der Einzelzusammenhänge (Tab. 33) erwartet. Die direkten und indirekten Teileffekte sind im 5 %-Signifikanzniveau hingegen nicht signifikant. Betrachtet man nur die Größen der Teileffekte, dann teilt sich der Gesamteffekt zu etwa 85 % auf den direkten Effekt ($\beta_{di} = .25, p = .059$) und zu 15 % auf den über das situationale Interesse mediierten indirekten Effekt ($\beta_{ind} = .04, p = .708$) auf. Der Anteil des direkten Effekts, dessen p-Wert im 10 %-Signifikanzniveau liegt, scheint in diesem Fall tendenziell größer zu sein, als der direkte Effekt. Der indirekte Effekt ergibt sich aus einem höheren, signifikanten Effekt zwischen Relevanz und Interesse ($\beta = .71, p < .001$) und einem kleineren, nicht signifikanten Effekt zwischen Interesse und Fachwissenszuwachs ($\beta = .06, p = .707$). Insgesamt werden 9 % des Fachwissenszuwachses durch dieses Modell aufgeklärt. Die Analyse gibt damit zwar

Abh. Variable	Unabh. Variable	β	b	p	R^2
Gesamtdatensatz					
Sit. Interesse Catch	Relevanz	$.71^{***}$	0.89^{***}	$< .001$	.50
Fachwissenszuwachs	Relevanz	$.29^{**}$	0.06^{**}	.005	.09
	Sit. Interesse Catch	$.24^{*}$	0.04^{*}	.025	.06
Datensatz versuchsnah					
Sit. Interesse Catch	Relevanz	$.66^{***}$	0.86^{***}	$< .001$	.44
Fachwissenszuwachs	Relevanz	$.32^{**}$	0.10^{**}	.002	.11
	Sit. Interesse Catch	$.32^{**}$	0.08^{**}	.002	.10

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$, $n = 88$

Tabelle 33: *Ergebnisse der Regressionsanalysen einzelner Zusammenhänge von wahrgenommener Relevanz, situationalem Interesse Catch (jeweils Personenmittelwerte, 5-teilige Likert-Skala) und Fachwissenszuwachs (anteilige Testwerte $\in [-1, 1]$ pro Person, Post-Prä als Zuwachs über das Praktikum) auf der Datenebene Zwischen Personen für $n = 88$ Studierende. Im oberen Tabellenteil (schwarz) sind die Zusammenhänge für den Gesamtdatensatz dargestellt und im unteren (blau) für den auf versuchsnahes Fachwissen reduzierten Datensatz.*

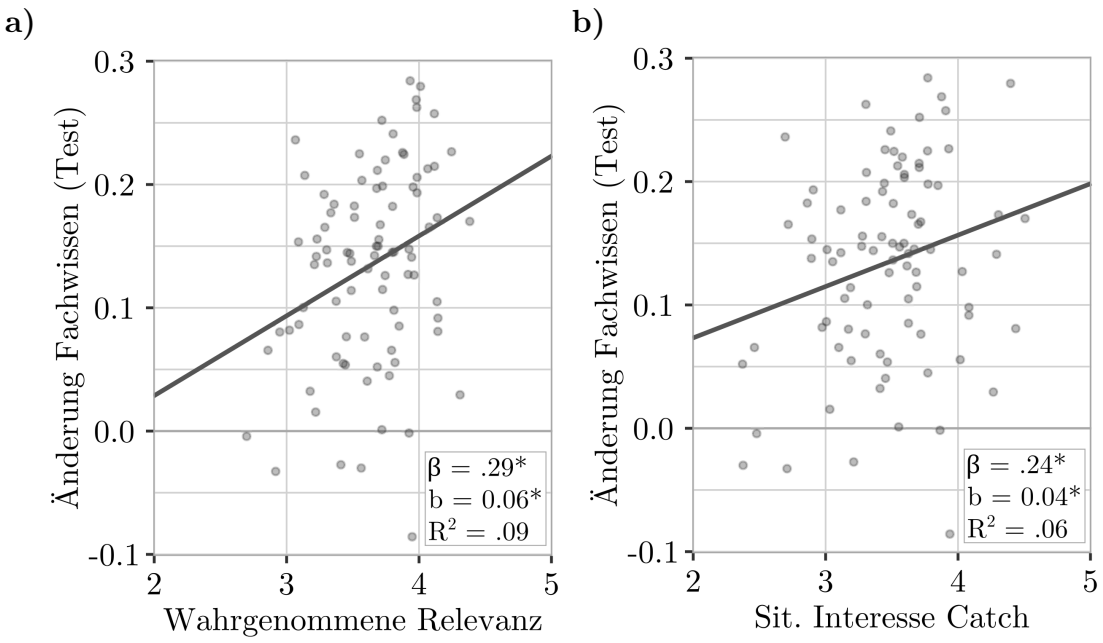
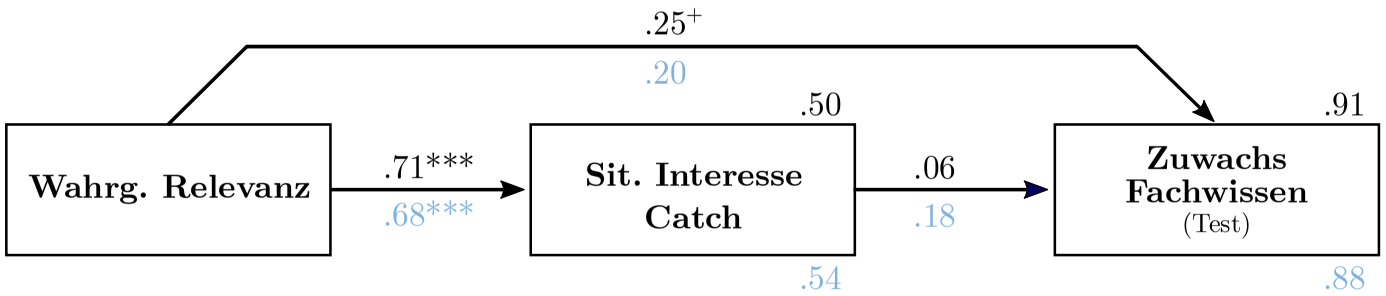


Abbildung 53: *Die Änderung des Fachwissens (anteilige Testwerte $\in [-1, 1]$ pro Person) abhängig von a) der wahrgenommenen Relevanz und b) dem situationalen Interesse Catch (jeweils Personenmittelwerte, 5-teilige Likert-Skala) auf der Datenebene Zwischen Personen für $n = 88$ Personen. Die Messwerte wurden semi-transparent eingezeichnet, um übereinanderliegende Datenpunkte deutlich zu machen. Die Regressionsparameter entsprechen denen in Tabelle 33.*



Gesamtdatensatz:
totaler Effekt: .30**
direkter Effekt: .25+
indirekter Effekt: .04

Datensatz versuchsnah:
totaler Effekt: .32**
direkter Effekt: .20
indirekter Effekt: .12

Zwischen Personen

Abbildung 54: *Ergebnis der Pfadanalyse der Mediation von situationalem Interesse Catch zwischen wahrgenommener Relevanz und Änderung des Fachwissens (Test) für $n = 88$ Studierende auf der Datenebene Zwischen Personen (Wirkmodell E). Abgebildet sind die standardisierten Pfadkoeffizienten. Die Werte an den abhängigen Variablen stellen die Residualvarianz dar. In diesem Pfadmodell sind oberhalb der Pfade die Werte für den Gesamtdatensatz dargestellt (schwarz) und unterhalb für den auf versuchsnahes Fachwissen reduzierten Datensatz (blau). Das Signifikanzniveau ist an folgender Kodierung zu erkennen: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$. Für das Modell des Gesamtdatensatzes liegen folgende Fit-Kennwerte vor: χ^2-Test: $\chi^2(3) = 70, p < .001$, SRMR= .00, RMSEA= .00, CFI= 1.00. Für das Modell für den versuchsnahen Datensatz sind die Fit-Kennwerte: χ^2-Test: $\chi^2(3) = 66, p < .001$, SRMR= .00, RMSEA= .00, CFI= 1.00.*

Hinweise darauf, dass der nachgewiesene Gesamteffekt vor allem auf dem direkten Teileffekt beruht. Die Teileffekte auf den Fachwissenszuwachs sind aber mit Vorsicht zu behandeln, da diese zu gering sind, um sie mit dieser Stichprobengröße ($n = 88$) signifikant nachweisen zu können.

Es wurden alternative Modellierungen zur Absicherung der hier geschätzten Modellparameter durchgeführt. Um Lord's Paradoxon auszuschließen wurde das Fachwissen im Wirkmodell statt über die absolute Änderung über den Wert des Post-Test unter Kontrolle des Prä-Tests modelliert, wie in Abschnitt 11.2.6 erörtert. Diese Modellschätzung weist mit einer guten Modellpassung auf einen kleineren signifikanten Gesamteffekt der wahrgenommenen Relevanz auf das Fachwissen hin ($\beta_{tot} = .13, p = .042$). Um zusätzlich einen ersten Eindruck zu erhalten, inwieweit die in Abschnitt 12.3.1 festgestellte negative Abhängigkeit des Fachwissenszuwachses vom Vorwissen einen Einfluss auf die Wirkzusammenhänge hat, wurde der Prä-Testwert in einer weiteren Modellierung in das hier betrachtete Modell aufgenommen. Auch dieses Modell zeigt bei einer guten Modellpassung einen kleineren signifikanten Gesamteffekt ($\beta_{tot} = .20, p = .044$). Der Effekt der wahrgenommenen Relevanz auf die Änderung des Fachwissens ist in allen alternativen Modellierungen von kleiner Größenordnung, sodass auch in diesen ein Nachweis der direkten und indirekten Teileffekte für diese Stichprobe ($n = 88$) nicht gelingt. Diese alternativen Modellierungen weisen darauf hin, dass die Zusammenhänge zwischen Fachwissen im Prä-Test und im Post-Test bzw. der Fachwissensänderung sowie ggf. den anderen Prädiktoren komplexer sind und sich daher unterschiedlich in den Modellparametern niederschlagen. Da die Messfehler in den alternativen Modellen unterschiedlich Eingang finden, kann sich auch dies in den unterschiedlichen Modellschätzungen zeigen. In Bezug auf die Messfehler ist bei der Bewertung der Modelle grundsätzlich auch einzubeziehen, dass bei einer Pfadanalyse im Gegensatz zur Strukturgleichungsmodellierung der Vorteil entfällt, dass Messfehler durch die Verwendung der latenten Variablen berücksichtigt werden. Anhand der alternativen Modellierungen ist damit davon auszugehen, dass der Gesamteffekt der wahrgenommenen Relevanz auf das Fachwissen zwar signifikant ist, aber etwas überschätzt worden sein kann und damit eher im breiteren Bereich von kleiner bis mittlerer Größenordnung verortet werden sollte.

Um einen weiteren Anhaltspunkt für die Stabilität und Gültigkeit der gefundenen Zusammenhänge im Gesamtdatensatz zu erhalten, soll die Analyse der Wirkannahmen im Folgenden auch analog für den versuchsnahen Datensatz durchgeführt werden. Dieser Teildatensatz wurde aus Aufgaben gebildet, die, rein inhaltlich begründet, nah an den Lerninhalten der Praktikumsversuche sind (siehe Abschnitt 11.1.3). Mit dieser Modellierung soll somit auch der Frage nachgegangen werden, ob der Zusammenhang für das direkt in den Praktikumsversuchen erworbene Fachwissen höher ist als für das Fachwissen, das auf umfassenderem Verständnis beruht.

Im unteren Teil der Tabelle 33 (blau) sind die Regressionsparameter für den versuchsnahen Datensatz dargestellt. Der Effekt der wahrgenommenen Relevanz auf

den Zuwachs des Fachwissens ist mit $\beta = .32, p = .002$ von mittlerer Größe. Dieser Effekt unterscheidet sich nur gering von dem des Gesamtdatensatzes (Tab. 33 schwarz oben). Das situationale Interesse zeigt einen ebenfalls mittleren Effekt von $\beta = .32, p = .002$ auf den Zuwachs des Fachwissens für den versuchsnahen Datensatz und ist damit etwas größer als für den Gesamtdatensatz. Insgesamt sind im Vergleich zu den Effekten im Gesamtdatensatz die Effekte im versuchsnahen Datensatz tendenziell etwas größer, aber von der gleichen Größenordnung. Die Voraussetzungen für die Analyse des Mediationsmodells mit dem versuchsnahen Datensatz sind ebenfalls gegeben.

Abbildung 54 zeigt im Pfadmodell der Wirkannahmen unter den Pfadkoeffizienten für den Gesamtdatensatz (schwarz) die Koeffizienten für den versuchsnahen Datensatz (blau). Die Pfadanalyse weist auch für den versuchsnahen Datensatz eine gute Modellpassung auf (χ^2-Test: $\chi^2(3) = 66, p < .001$, SRMR= .00, RMSEA= .00, CFI= 1.00). Der signifikante Gesamteffekt der Mediation ist mit $\beta_{tot} = .32, p = .001$ für beide Datensätze von gleicher Größenordnung. Auch in den anderen Pfadkoeffizienten ähneln sich die Pfadanalysen der beiden Datensätze, wie auch anhand der Betrachtungen der Einzelzusammenhänge (Tab. 33) zu erwarten ist. Direkter und indirekter Teileffekt ($\beta_{di} = .20, p = .108$, $\beta_{ind} = .12, p = .293$) sind für die versuchsnahen Daten ebenfalls nicht signifikant von Null verschieden. Für die beiden Datensätze liegen damit keine wesentlichen Unterschiede in der pfadanalytischen Modellierung vor. Diese Konsistenz spricht für die Stabilität der Pfadanalyse. Zudem bedeutet es, dass sich der Effekt der wahrgenommenen Relevanz auf den Fachwissenszuwachs nicht hauptsächlich im direkt in den Praktikumsversuchen erworbenen Fachwissen begründet, sondern auch gleichermaßen auf das umfassendere Verständnis, das über die direkten Lerninhalte der Praktikumsversuche hinaus geht.

Nach der vorangehenden Analyse der Wirkhypothesen unter Berücksichtigung der Versuchsnähe des Fachwissens kann als weiterführende Frage aufkommen, ob die Studierenden nicht nur für die Praktikumsversuche und damit für deren physikalische Inhalte eine gewisse berufliche Relevanz erkennen oder nicht, sondern unabhängig davon auch weitergehend für diese fachlichen Inhalte zwischen relevantem und nicht relevantem Fachwissen differenzieren. Gute Anhaltspunkte für diese Vermutungen liefern die lehramtsbezogenen Studien von Lorentzen et al. (2019) und Massolt und Borowski (2018), die erfolgreich in der Lehramtsausbildung das universitäre mit dem schulischen Fachwissen durch eine dritte Wissenskomponente verknüpfen, um die Relevanzwahrnehmung der Studierenden zu erhöhen. Für die Analyse der Wirkzusammenhänge würde das bedeuten, dass der Fachwissenszuwachs durch eine weitere Relevanzwahrnehmung der Inhalte mit eher schulischem oder universitärem Fachwissen beeinflusst sein kann. Da man die versuchsnahen Fachwissensaufgaben zu einem Großteil auch als schulisch und die versuchsfernen eher universitär einordnen kann, kann mit ähnlichen Effekten wie in den vorangegangenen Analysen gerechnet werden. Daher wird mit diesem Datensatz nicht davon ausgegangen, diese möglichen Effekte nachweisen und differenzieren zu können.

Eine Vereinfachung bei den Analysen des Wirkmodells in Bezug auf den Fachwissenszuwachs aufgrund der vorhandenen Datenlage ist die Untersuchung auf der Datenebene *Zwischen Personen* anhand von Personenwerten über alle Praktikumsversuche hinweg. Allerdings könnten Betrachtungen auf den anderen Datenebenen (*Innerhalb* und *Zwischen Versuche*, Abb. 18) die Wirkannahmen für die unterschiedlichen Datenebenen differenzieren und Besonderheiten der Wirkung einzelner Praktikumsversuche identifizieren. Dabei könnten beispielsweise Versuche aufgedeckt werden, die sehr gut konzipiert sind und den Studierenden dadurch den Lerninhalt unabhängig von der wahrgenommenen Relevanz sehr gut vermitteln. Andersherum kann es Versuche geben, die den Studierenden aus der Schulerfahrung unbekannt sind (z. B. Feldeffekttransistor, siehe Abschnitt 9.4), wodurch die Relevanz niedriger wahrgenommen wird und der Lernzuwachs aber eben aufgrund des fehlenden Vorwissens groß sein kann. Somit wären Betrachtungen auf den anderen Datenebenen wünschenswert. Der Fachwissenstest wurde jedoch in Hinblick auf die inhaltliche Breite und die Umsetzbarkeit so konzipiert, dass fachliche Inhalte zu einzelnen Praktikumsversuchen mithilfe von je zwei Aufgaben abgefragt wurden. Betrachtungen auf Versuchsebene waren daher nur theoretisch möglich. Eine Operationalisierung mit nur zwei Aufgaben pro Versuch bildet das Fachwissen nicht zuverlässig ab und stellt somit keine belastbare Basis dar (Abschnitt 11.1.3). Sowohl auf der Ebene zwischen Versuchen, als auch auf der Ebene innerhalb Versuchen kann daher anhand der vorliegenden Daten keine Modellierung erfolgen.

Als Gesamtergebnis dieser Analyse (Wirkmodell E) ist festzuhalten, dass die wahrgenommene Relevanz nachweislich einen positiven Effekt auf den Fachwissenszuwachs hat. Dies ist ein Indiz für die Plausibilität des Wirkmodells. Der Analyse sind durch die Stichprobengröße aber Grenzen gesetzt.

Wirkmodell mit Fachwissen anhand der Selbsteinschätzung

Da der Fachwissenszuwachs aus dem Wissenstest wie beschrieben auf spezifisches Fachwissen ausgerichtet ist, soll die nachfolgende Analyse der globaleren Selbsteinschätzung des Fachwissenszuwachses ergänzende Informationen für die Prüfung des angenommenen Wirkmodells liefern (Wirkmodell F). In diesem Abschnitt bezeichnet Fachwissenszuwachs, sofern nicht anders gekennzeichnet, daher den Fachwissenszuwachs aus der Selbsteinschätzung anhand des PraQ-Fragebogens.

Zunächst werden in diesem Abschnitt einzelne Zusammenhänge zwischen den Größen wahrgenommene Relevanz, situationales Interesse Catch und der Selbsteinschätzung des Fachwissenszuwachses anhand von paarweisen Regressionsanalysen betrachtet. Für die Untersuchung der Zusammenhänge mit Bezug auf die Selbsteinschätzung wird die Datenebene *Zwischen Personen* mit $n = 88$ Studierenden betrachtet, da die Selbsteinschätzung nicht versuchsbezogen, sondern am Ende jedes Semesters vorgenommen wurde. Aus Konsistenzgründen zur Selbsteinschätzung der

Zunahme der experimentellen Kompetenz wird in den folgenden Analysen ebenfalls nur die Selbsteinschätzung des Fachwissens zum ersten Messzeitpunkt betrachtet.

Die Regressionsparameter sind in Tabelle 34 dargestellt. Abbildung 55 veranschaulicht die Zusammenhänge und die Streuung der Wertepaare. Die wahrgenommene Relevanz hat einen signifikanten, positiven Effekt von $\beta = .37, p < .001$ auf den Zuwachs des Fachwissens. Der Effekt $\beta = .45, p < .001$ vom Interesse Catch auf den Zuwachs des Fachwissens liegt ebenso im mittleren Bereich. 13 % bzw. 20 % der Varianz der experimentellen Kompetenz werden durch diese Prädiktoren aufgeklärt.

Abh. Variable	**Unabh. Variable**	β	b	p	R^2
Sit. Interesse Catch	Relevanz	$.70^{***}$	0.87^{***}	$< .001$	.49
Zuwachs Fachwissen (SE)	Relevanz	$.37^{***}$	0.58^{***}	$< .001$	.13
Zuwachs Fachwissen (SE)	Sit. Interesse Catch	$.45^{***}$	0.56^{***}	$< .001$	.20

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $n = 88$

Tabelle 34: *Ergebnisse der Regressionsanalysen einzelner Zusammenhänge von wahrgenommener Relevanz, situationalem Interesse Catch (jeweils Personenmittelwerte, 5-teilige Likert-Skala) und der Selbsteinschätzung des Zuwachses an Fachwissen (6-teilige Likert-Skala) auf der Datenebene Zwischen Personen für $n = 88$ Personen im ersten Praktikumssemester. (Erläuterungen der Regressionsparameter sind in Abschnitt Methoden 8.3.5 zu finden.)*

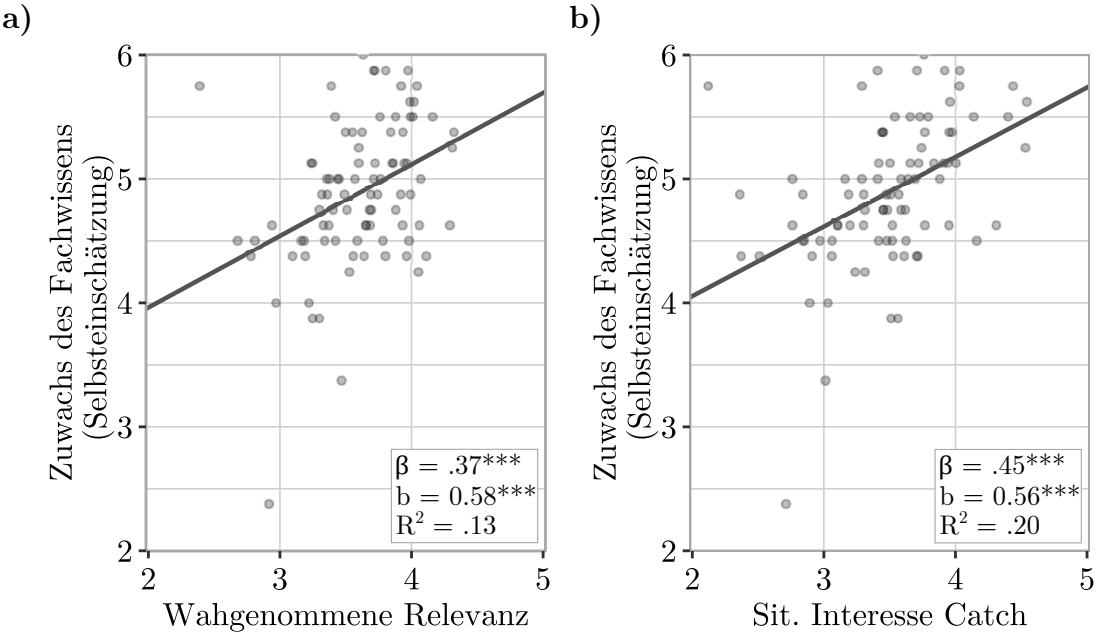


Abbildung 55: *Selbsteinschätzung des Zuwachses an Fachwissen (6-teilige Likert-Skala) abhängig von a) der wahrgenommene Relevanz und b) vom situationalen Interesse Catch (jeweils Personenmittelwerte, 5-teilige Likert-Skala) auf der Datenebene Zwischen Personen für $n = 88$ Personen im ersten Praktikumssemester. Die Messwerte wurden semitransparent eingezeichnet, um übereinanderliegende Datenpunkte deutlich zu machen. Die Regressionsparameter entsprechen denen in Tabelle 34.*

Die gefundenen positiven Zusammenhänge zwischen den Größen Relevanz, Interesse und Fachwissenszuwachs weisen auf die Plausibilität der Wirkannahmen hin. Damit sind die Voraussetzungen für die weiterführende Betrachtung des Mediationsmodell im nächsten Schritt gegeben.

Mithilfe einer Strukturgleichungsmodellierung können damit nun die Effekte aller betrachteten Einflussgrößen im Wirkmodell gemeinsamen untersucht werden. Für die Analyse werden die gleichen Daten wie für Betrachtung der Einzelzusammenhänge im oberen Teil dieses Abschnittes auf der Datenebene *Zwischen Personen* betrachtet. Das Strukturgleichungsmodell mit den geschätzten Parametern der Mediation von situationalem Interesse zwischen wahrgenommener Relevanz und Fachwissenszuwachs ist in Abbildung 56 dargestellt (Wirkmodell F). Zur Beurteilung der Modellgüte werden Fit-Kennwerte herangezogen (siehe Abschnitt 11.2.3). Während der χ^2-Test ($\chi^2(32) = 50.3, p = .021$) und der RMSEA-Wert (RMSEA = .08) weniger für eine gute Modellpassung sprechen, zeigen der SRMR-Wert (SRMR = .04) und der CF-Index (CFI = .98), dass das Modell die Daten gut widerspiegelt. Die Modellpassung ist damit ausreichend belastbar und die geschätzten Parameter des Modelles können sinnvoll diskutiert werden.

Die wahrgenommene Relevanz hat einen mittleren Gesamteffekt von $\beta_{tot} = .43, p < .001$ auf den Zuwachs an Fachwissen, der etwas höher ist als die Regressionsanalyse vermuten ließ (Tab. 34). Dieser Effekt teilt sich vom Wert her zu 21 % auf einen direkten Effekt von $\beta_{di} = .09, p = .093$ und zu 79 % auf einen indirekten, über das situationale Interesse mediierten Effekt von $\beta_{ind} = .34, p = .006$ auf. Während der direkte Effekt dabei nicht signifikant ist, ist der höhere indirekte Effekt signifikant. Letzterer Teileffekt ergibt sich aus einem signifikanten Effekt zwischen Relevanz und Interesse von $\beta = .74, p < .001$ und einem signifikanten Effekt zwischen Interesse und Fachwissenszuwachs von $\beta = .46, p = .007$. Damit werden 29 % der Varianz des selbst eingeschätzten Zuwachses an Fachwissens aufgeklärt. Diese Strukturgleichungsmodellierung zeigt somit einen signifikanten, mittleren Gesamteffekt zwischen wahrgenommener Relevanz und dem selbst eingeschätzten Zuwachs an Fachwissen auf, der hauptsächlich durch das situationale Interesse mediiert wird.

Eine Strukturgleichungsmodellierung für den Zuwachs an Fachwissen über beide Praktikumssemester stützt diese gefundenen Effekte für das erste Praktikumssemester. Zu beachten ist hierbei jedoch, dass wie bei der Selbsteinschätzung der experimentellen Kompetenz (Abschnitt 12.2.2) auch für die Selbsteinschätzung des Fachwissens keine Referenz- und Ausgangswerte berücksichtigt werden konnten, sodass eine Verzerrung möglich ist. Zudem muss die geringere Validität der Selbsteinschätzung im Vergleich zum Wissenstest bedacht werden (siehe Abschnitt 11.1.3).

Alles in allem zeigt die Strukturgleichungsmodellierung (Wirkmodell F), dass für Studierende, die die Praktikumsversuche im Mittel relevanter und interessanter wahrnehmen, die Zunahme des Fachwissens anhand der Selbsteinschätzung signifikant höher ist. Dies ist ein wichtiger Nachweis für die Plausibilität des Wirkmodells.

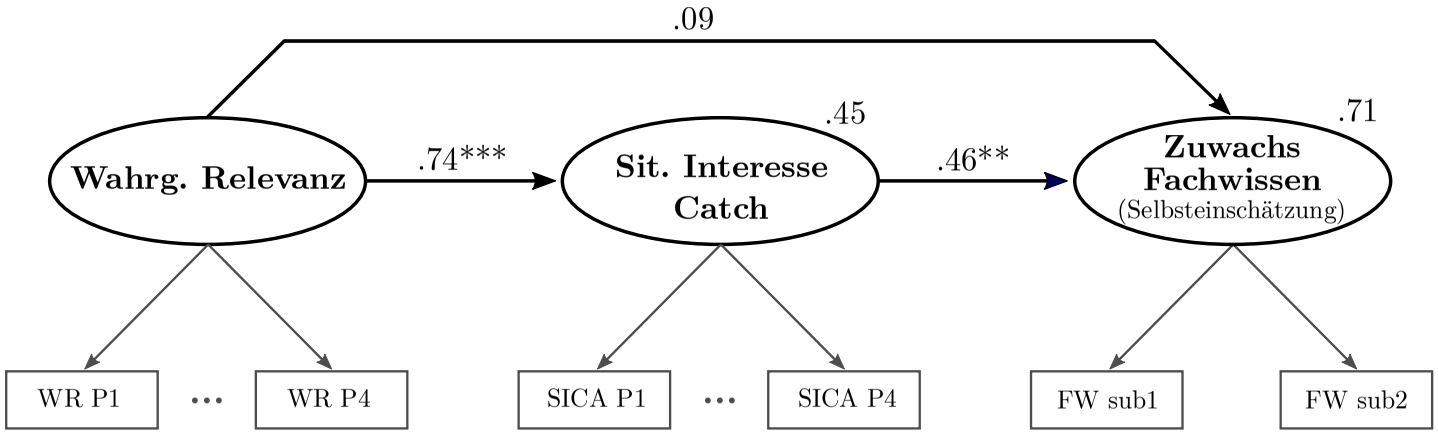


Abbildung 56: *Ergebnis des Strukturgleichungsmodells der Mediation von situationalem Interesse Catch zwischen wahrgenommener Relevanz und der Selbsteinschätzung des Zuwachses des Fachwissens für $n = 88$ Studierende für das erste Praktikumssemester auf der Datenebene Zwischen Personen (Wirkmodell F). Abgebildet sind die standardisierten Koeffizienten. Die Werte an den abhängigen Variablen stellen die Residualvarianz dar. Das Signifikanzniveau ist an folgender Kodierung zu erkennen: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$. Für das Modell liegen folgende Fit-Kennwerte vor: χ^2-Test: $\chi^2(32) = 50.3, p = .021$, SRMR = .04, RMSEA = .08, CFI = .98.*

Synthese der Wirkmodelle zum Fachwissen

In Form eines Wissenstests und einer Selbsteinschätzung wurden in dieser Arbeit zwei Instrumente ökonomisch zur Erfassung des Lernerfolgs bzgl. des Fachwissens eingesetzt. Bedingt durch die Art und die Spezifität haben beide Instrumente unterschiedliche Qualitäten, sodass folglich der Mehrwert in der gemeinsamen Betrachtung beider Maße für den Fachwissenszuwachs liegt. Daher werden an dieser Stelle die Ergebnisse der Modellierungen der Wirkmodelle mit Lernerfolg Fachwissen anhand des Tests und der Selbsteinschätzung (Wirkmodelle E und F) zusammengeführt.

In Bezug auf die Prüfung des angenommenen Wirkmodells zeigen sich leicht unterschiedliche Effekte der wahrgenommenen Relevanz auf den Fachwissenszuwachs anhand des Wissenstests und der Selbsteinschätzung. Für den Wissenstest ist der Effekt signifikant, von mittlerer Größe und lässt in der Pfadanalyse für diese Stichprobe keine belastbaren Rückschlüsse auf den direkten und indirekten Teileffekt zu. Für die Selbsteinschätzung zeigt die Modellierung, dass der Effekt von Relevanz auf Fachwissenszunahme von mittlerer bis großer Stärke ist und hauptsächlich über das situationale Interesse mediiert wird. Unter Einbezug der bereits diskutierten Vor- und Nachteile muss dabei bedacht werden, dass die Selbsteinschätzung als weniger valide im Vergleich zum Wissenstest gilt (siehe Abschnitt 11.1.3) und auch aufgrund der fehlenden Berücksichtigung von Referenzwerten verzerrt sein kann. Die Unterschiede der Effekte der beiden Untersuchungen lassen sich aber auch anhand der Spezifität der Instrumente erklären: Während der Fachwissenstest auf konkrete fachliche Inhalte der Praktikumsversuche ausgerichtet ist und den Zuwachs an Fachwissen zwar objektiv aber nur punktuell misst, umfasst die Selbsteinschätzung den Fachwissenszuwachs übergeordneter und schließt damit zum Beispiel auch die Breite der Vorbereitung der Versuche und die Vernetzung von Theorie und experimenteller Praxis ein. Anhand dessen wäre denkbar, dass der Wissenstest den Fachwissenszuwachs nicht auf allen angesprochenen Inhaltsebenen umfasst und der Effekt damit etwas kleiner ist. Hinzukommen kann auch, dass die Studierenden das in besonders relevant wahrgenommenen Versuchen erlernte Fachwissen selbst auch als besonders relevant einstufen. Aufgrund der wahrgenommenen Bedeutsamkeit könnte dieser Fachwissenszuwachs auch als besonders hoch in der Selbsteinschätzung gewichtet werden, sodass der Effekt der wahrgenommene Relevanz auf den Fachwissenszuwachs für diesen Test deutlicher hervortritt. Die tatsächliche Größe des Effekts von Relevanz auf Fachwissenszuwachs liegt damit sehr wahrscheinlich zwischen den Ergebnissen der beiden Analysen bei einem mittleren Effekt.

Während die Analysen der Wirkannahmen bzgl. des Fachwissenszuwachses mit den Daten des Wissenstests und der Selbsteinschätzung für sich genommen keine direkte generische Schlussfolgerung zulassen, liefern beide Analysen (Wirkmodelle E und F) zusammengenommen ein stärkeres Indiz für die Plausibilität der Wirkannahme, dass die wahrgenommene Relevanz der Praktikumsversuche einen positiven Effekt auf den Fachwissenszuwachs hat.

13 Synopse zum Wirkmodell

Nachdem im vorangegangenen Teil III dieser Arbeit eindeutige Erkenntnisse über die Eigenschaften relevant wahrgenommener Praktikumsversuche im Lehramtspraktikum gewonnen werden konnten, wurde in dieser Analyse ein Schritt darüber hinaus gewagt und die Wirkung der wahrgenommenen Relevanz auf den festgestellten Lernerfolg im Praktikum erforscht. Die gewonnenen Erkenntnisse zu den zuvor einzeln betrachteten Komponenten des Lernerfolgs bzgl. Interesse, experimentelle Kompetenz und Fachwissen werden in dieser Synopse zusammengeführt und gedeutet, um Forschungsfrage 3 übergeordnet bzgl. des Lernerfolgs insgesamt zu beantworten.

Die drei Hauptkomponenten des Lernerfolgs konnten in dieser Arbeit ökonomisch und zuverlässig im Lehramtspraktikum erfasst werden, obwohl der Lernprozess im Praktikum hochkomplex ist und von vielen Faktoren beeinflusst wird. Analysen dieser Daten haben einen hohen Lernerfolg für alle Komponenten nachgewiesen. Damit ist die Entwicklung des Praktikums überaus erfolgreich gewesen.

Der Entwicklung des Lehramtspraktikums liegt ein in dieser Arbeit theoretisch abgeleitetes Wirkmodell zugrunde. Demnach wirkt sich eine Förderung der wahrgenommenen Relevanz positiv auf das situationale Interesse sowie mittelbar auf die Lernmotivation und damit wiederum förderlich auf den Lernerfolg aus. Aufbauend auf den Nachweis des Lernerfolgs wurden Modellierungen zur Prüfung dieser angenommenen Wirkzusammenhänge für die einzelnen Komponenten des Lernerfolgs berechnet. Hierbei muss beachtet werden, dass trotz der Vollerhebung über mehrere Semester die Stichprobe der Lehramtsstudierenden klein ist. Damit sind dem Modellumfang und der Schätzung von Modellparametern Grenzen gesetzt. In der Vielzahl von Bedingungsfaktoren im Praktikum und deren Überlagerung begründet sich zudem ein starkes Rauschen der Daten. Dies bedeutet, dass ein Nachweis von kleinen bis mittleren Effekten schwierig ist. Trotz dieser statistischen Herausforderungen konnten in dieser Arbeit signifikante Effekte für paarweise Zusammenhänge der Größen des Wirkmodells und für das Wirkmodell selbst festgestellt werden.

Für die paarweisen Zusammenhänge der Größen können Befunde aus der Literatur zur Prüfung der Konsistenz der Ergebnisse herangezogen werden. Die gefundenen hohen Zusammenhänge zwischen wahrgenommener Relevanz und situationalem Interesse sind in guter Übereinstimmung mit der Literatur bzgl. Interesse. Beispielsweise haben Lewalter und Geyer (2009) einen hohen Effekt der wahrgenommenen inhaltlichen Relevanz auf das situationale Interesse für Schulklassen in naturwissenschaftlich-technischen Museen festgestellt. Für Studierende unterschiedlicher Studienrichtungen konnte ein positiver, mittlerer Zusammenhang zwischen Relevanz und Studieninteresse gezeigt werden (Schiefele et al., 2003). Auch für die mittleren bis hohen Zusammenhänge zwischen situationalem Interesse und Lernerfolg in diesen Untersuchungen zeigt die Literatur ähnliche Größenordnungen der Effekte. Schiefele, Krapp und Schreyer (1993) fanden in ihrer Metaanalyse zum Beispiel einen mittleren Zusammenhang zwischen dem Interesse am Schulfach und der Leis-

tung darin. Im akademischen Kontext konnte in einer neueren Studie ein ähnlicher positiver Zusammenhang gezeigt werden (Rotgans & Schmidt, 2011). Damit sind die in dieser Studie festgestellten einzelnen Zusammenhänge konsistent mit bisherigen Erkenntnissen der Bildungsforschung. In dieser Arbeit wurden diese paarweisen Zusammenhänge zudem in einem Wirkmodell für das Lehramtspraktikum gemeinsam analysiert. Eine empirische Prüfung der hier aufgestellten Wirkannahmen als Gesamtmodell in der Literatur ist hingegen nicht bekannt.

In den einzelnen Modellierungen des Wirkmodells (Wirkmodelle A-F), die in Tabelle 35 zum direkten Vergleich nochmal gegenüber gestellt wurden, zeigen sich dabei signifikante Gesamteffekte der wahrgenommenen Relevanz auf den Lernerfolg bzgl. Interesse, experimentelle Kompetenz, die Selbstwirksamkeitserwartungen und das Fachwissen. Die statistischen Grenzen der Modellierung zeigen sich an dieser Stelle darin, dass direkte und über das situationale Interesse Catch mediierte indirekte Teileffekte nur bei höheren Gesamteffekten belastbar differenzierbar sind. Die Gesamteffekte der wahrgenommenen Relevanz auf die einzelnen Komponenten des Lernerfolgs sind von unterschiedlicher Größenordnung. Dabei hebt sich der motivationale Lernerfolg in Bezug auf das situationale Interesse Hold mit sehr hohen Gesamteffekten und signifikanten Teileffekten ab (Wirkmodell A). Die wahrgenommene Relevanz beeinflusst das situationale Interesse an den Praktikumsinhalten hypothesenkonform positiv. Für den fachlichen Lernerfolg in Bezug auf das Experimentieren und Fachwissen zeigen die Modellierungen mittlere bis große Gesamteffekte (Wirkmodelle B-F). Die Gesamteffekte der wahrgenommene Relevanz sind aber für alle Komponenten des Lernerfolgs im Lehramtspraktikum positiv gerichtet und spiegeln damit konsistent die gleiche Aussage wider.

Wirk-modell	**Zielvariable Lernerfolg**	**Daten-ebene**	**totaler Effekt**	**direkter Effekt**	**indirekter Effekt**
A	Sit. Interesse Hold	Zwischen	$.79^{***}$	$.46^{**}$	$.32^{**}$
		Innerhalb	$.71^{***}$	$.28^{***}$	$.43^{***}$
B	exp. Kompetenz (SE)	Zwischen	$.46^{**}$	.19	$.27^{*}$
C	SWE Planung	Zwischen	$.29^{*}$	.16	.13
D	SWE Durchführung	Zwischen	$.25^{*}$	.16	.09
E	Fachwissen (Test)	Zwischen	$.30^{**}$	.25	.04
F	Fachwissen (SE)	Zwischen	$.44^{**}$	.09	$.34^{**}$

Tabelle 35: *Gegenüberstellung der standardisierten Effekte der Modellierungen der Wirkannahmen bzgl. der unterschiedlichen Komponenten des Lernerfolgs (SWE: Selbstwirksamkeitserwartungen; Signifikanzniveau: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$; Werte aus Abb. 40, 44, 48, 49, 54, 56). Die Datenebene bezieht sich auf Personen.*

In dieser Arbeit konnten damit einige konsistente empirische Indizien für die Plausibilität der theoretisch begründeten Wirkannahmen zusammen getragen werden. Jede Modellierung der Wirkannahme (Wirkmodelle A-F) für sich allein hat keine Beweiskraft für eine Generalisierbarkeit der gefundenen Zusammenhänge, da keine Kausalmodelle vorliegen und Alternativmodelle nicht ausgeschlossen werden können. In ihrer Gesamtheit über alle Komponenten des Lernerfolgs hinweg lässt sich aber aus den in dieser Arbeit gefundenen konsistenten empirischen Indizien mit großer Wahrscheinlichkeit auf das Zutreffen des aufgestellten Wirkmodells schließen.

Auf den Punkt gebracht bedeutet diese Schlussfolgerung, dass eine höhere Relevanzwahrnehmung der Praktikumsversuche im Lehramtspraktikum zu einer Verbesserung des Lernerfolgs im Praktikum führt. Diese Erkenntnis wird nochmal besonders interessant bei gemeinsamer Diskussion und Bewertung mit den Möglichkeiten, die in Teil III zu den Eigenschaften von relevant wahrgenommenen Praktikumsversuchen aufgezeigt wurden.

Teil V
Abschließende Gesamtbetrachtung

Vorangehend wurden nun die drei großen Schritte zur Entwicklung und Analyse des Lehramtspraktikums im Rahmen dieser Arbeit dargestellt. In dieser Arbeit sollte durch die Neukonzeption eines lehramtsspezifischen Anfängerpraktikums die Trennung von Fach und Fachdidaktik im Bachelorstudium Physik Lehramt durch Vernetzungselemente aufgebrochen und die spezifischen Bedürfnisse der Lehramtsstudierenden berücksichtigt werden, um das fachliche und fachdidaktische Professionswissen zu vermitteln, das für die zukünftige Tätigkeit als Lehrperson erforderlich ist, und um die wahrgenommene Relevanz zu fördern (Teil II). Durch die Beforschung des Lehramtspraktikums konnten zum einen Erkenntnisse darüber gewonnen werden, welche Eigenschaften relevant wahrgenommene Praktikumsversuche haben (Teil III). Zum anderen wurde aus einer Mehrzahl an empirischen Indizien geschlossen, dass sich diese wahrgenommene Relevanz förderlich auf das situationale Interesse und den Lernerfolg der Studierenden auswirkt (Teil IV).

In diesem Teil V ist es zur abschließenden Gesamtbetrachtung der Arbeit zunächst erforderlich, diese Erkenntnisse aus der Entwicklung und Beforschung des Lehramtspraktikums zusammenzuführen, in den Forschungskontext einzuordnen und übergeordnet zu bewerten (Abschnitt 14). Darauf aufbauend werden dann die praktischen Implikationen dieser Befunde auf Praktika diskutiert, die in den einzelnen Synopsen bereits angedeutet wurden (Abschnitt 15). Auf dieser Basis wird dann abschließend das Potential der Ergebnisse für zukünftige Forschung und Entwicklungen erörtert (Abschnitt 16).

14 Diskussion der Befunde im Forschungskontext

Im Rahmen dieser Arbeit wurde ein spezifisches Laborpraktikum für Lehramtsstudierende des Faches Physik neu konzipiert und umgesetzt (Teil II). Der Entwicklung liegt ein dafür theoretisch hergeleitetes Wirkmodell zugrunde, wonach die wahrgenommene Relevanz die Lernmotivation und damit den Lernerfolg positiv beeinflusst. Die 25 neu entwickelten Praktikumsversuche wurden in einem ersten Schritt dieser Forschung in Hinblick auf eine optimale Gestaltung bzgl. der wahrgenommenen Relevanz quantitativ analysiert. Dabei wurde die Rolle der Kontextualisierung der Versuchsbeschreibungen (Forschungsfragen 1) und weitere Eigenschaften relevant wahrgenommener Versuche (Forschungsfragen 2) untersucht (Teil III). Im Einzelnen hat sich dabei gezeigt, dass schon geringe Modifikationen der Versuchsbeschreibung in Form von schulbezogener Kontextualisierung der Einleitungen ausreichend sind, damit der Schulkontext der Versuche besser erkannt wird und die Versuche für die spätere Lehrtätigkeit als relevanter wahrgenommen werden (Forschungsfrage 1). Durch die Analyse der Forschungsfrage 2 konnte die Schlüsselrolle des Schulkontextes für die wahrgenommene Relevanz, die in adressatenspezifischen Praktika als zentrales Gestaltungselement genutzt wird (z. B. Theyßen, 2000; Nagel & Wolny, 2013), auch in Relation zu anderen Versuchseigenschaften eindeutig bestätigt werden. Zusätzlich wurden weitere Eigenschaften relevanter Versuche identifiziert, die die Gestaltungsprinzipien des neuen Praktikums bestätigen. Übergeordnet lassen sich diese Eigenschaften, die die größten Effekte auf die wahrgenommene Relevanz haben, wie zum Beispiel ein schulnahes Versuchsthema und die Anschaulichkeit, als schulnah kategorisieren. Aus den Antworten auf die Forschungsfragen 1 und 2 folgt somit, dass durch eine entsprechend spezifische Konzeption von Praktikumsversuchen die wahrgenommene Relevanz im Lehramtspraktikum gezielt erhöht werden kann. Im nächsten Schritt wurde in dieser Arbeit im Rahmen von Forschungsfrage 3 die Gültigkeit der aufgestellten Wirkannahmen quantitativ untersucht (Teil IV). Zusätzlich konnten dadurch erste Indizien zur Plausibilität des Wirkmodells gefunden werden. Durch die Prüfung des Wirkmodells konnte gezeigt werden, dass die wahrgenommene Relevanz einen positiven Einfluss auf das situationale Interesse an den Inhalten der Praktikumsversuche hat. Zudem haben sich eine Reihe konsistenter Indizien aufgetan, aus deren Gesamtheit geschlossen wird, dass sich die wahrgenommene Relevanz gemäß des Wirkmodells förderlich auf den Lernerfolg in Bezug auf das Fachwissen und die experimentelle Kompetenz der Studierenden auswirkt.

Die Möglichkeit der gezielten Beeinflussung der wahrgenommenen Relevanz durch Nutzung der Erkenntnisse zu den Versuchseigenschaften lässt zusammen mit dem Befund zur Plausibilität des Wirkmodells die These zu, dass folglich durch die adressatenspezifische Konzeption eines Praktikums das situationale Interesse am Lerninhalt und der Lernerfolg der Studierenden gefördert werden kann. Die Prüfung und Quantifizierung letzterer These durch Modellierung eines Gesamtwirkmodells wäre in dieser Arbeit allerdings ein Schritt über den schon gewagten Schritt in Forschungs-

frage 3 hinaus. Für die Identifikation der Eigenschaften relevanter Versuche mit Ausnahme der Kontextualisierung wurde in dieser Arbeit explorativ vorgegangen. Dies ist auch ein Grund dafür, dass die nachgewiesenen einzelnen Effekte von kleiner bis mittlerer Größenordnung sind. Dies begründet sich aber auch in der Vielzahl der Bedingungsfaktoren der wahrgenommenen Relevanz und des Lernerfolgs im Praktikum, die in einem Gesamtmodell Berücksichtigung finden sollten. Auf Basis des Studiendesigns, der Stichprobe und der Daten dieser Arbeit ist damit das Gesamtmodell nicht belastbar testbar. Für die Analyse des Effekts der Versuchseigenschaften auf den Lernerfolg müsste vor allem eine systematische Variation der Versuchseigenschaften basierend auf den Befunden dieser Arbeit erfolgen, wie auch nochmal im Ausblick in Abschnitt 16 aufgegriffen wird. Dennoch geben die hier gefundenen Effekte allen Grund zur Schlussfolgerung, dass die These zum Gesamtwirkmodell plausibel ist und die spezifische Konzeption eines Praktikums das situationale Interesse sowie in dessen Folge die Lernmotivation und damit den Lernerfolg der Studierenden fördert. Dabei lässt sich insbesondere vermuten, dass je mehr der identifizierten relevanzschaffenden Eigenschaften ein Praktikumsversuch hat und umso deutlicher und umfassender die Kontextualisierung der Versuchsbeschreibung ist, desto größer ist der Gesamteffekt auf die wahrgenommene Relevanz der Versuche, sodass damit die Lernmotivation und der Lernerfolg bestmöglich gefördert werden kann.

Nach der Zusammenführung der Haupterkenntnisse der Beforschung des Lehramtspraktikums in Kiel stellt sich die Frage danach, inwieweit diese Erkenntnisse auch auf Praktika an anderen Standorten und für andere Studierendengruppen übertragbar sind. Die Diskussion der Übertragbarkeit ist nicht nur essenziell, um die Befunde im Forschungskontext einzuordnen und zu bewerten. Die Diskussion ist auch im Anschluss für die Beurteilung wichtig, welcher praktische Nutzen folglich aus den Erkenntnissen dieser Arbeit für die Verbesserung anderer (adressatenspezifischer) Praktika gezogen werden kann.

In einem ersten Schritt wird dazu die Übertragbarkeit auf physikalische Lehramtspraktika an anderen Standorten betrachtet. Die gewonnenen Erkenntnisse sind keineswegs als standortbezogen für das Lehramtspraktikum in Kiel zu betrachten. Die Lehramtsstudierenden des Faches Physik durchlaufen mit ähnlichen schulischen Vorerfahrungen an verschiedenen Standorten in Deutschland eine ähnliche Ausbildung mit ähnlichen Lernzielen, die auf den Lehrberuf vorbereiten (z. B. Großmann & Urban, 2006). Die Studierenden werden an diesen Standorten daher auch sehr ähnliche Vorstellungen von relevanten Praktikumsversuchen haben. Folglich sollten die Forschungserkenntnisse übertragbar für die Studierenden des Lehramts Physik allgemein sein. Mit den Befunden dieser Arbeit kann damit die bisher in der Forschung noch offene Frage, ob auch Lehramtsstudierende des Faches Physik durch ein spezifisches Lehramtspraktikum im Hinblick auf wahrgenommene Relevanz und motivationale Faktoren profitieren können, eindeutig positiv beantwortet werden.

Im zweiten Schritt wird die Übertragbarkeit der Ergebnisse dieser Arbeit und deren Bedeutung im Forschungskontext für Lehramtsstudierende anderer Naturwissenschaften und anderer Studierendengruppen diskutiert. Dazu wird zunächst die Frage nach der Übertragbarkeit und der Bedeutung der Befunde zu den Eigenschaften relevant wahrgenommener Versuche gestellt. Laborpraktika von Lehramtsstudierenden anderer naturwissenschaftlicher Fächer (z. B. Chemie, Biologie) kann bei dieser Frage eine besondere Rolle zugesprochen werden. Da ebenfalls als zukünftiges Tätigkeitsfeld die Schule angestrebt wird und Experimente in den naturwissenschaftlichen Fächern der Schule einen ähnlichen Stellenwert haben. ist es wahrscheinlich, dass die festgestellten Eigenschaften (wie z. B. Anschaulichkeit, schulnahes Versuchsthema) der Versuche, die von Studierende des Lehramts Physik als relevant eingestuft werden, für Lehramtsstudierende anderer naturwissenschaftlicher Fächer in einem großen Maß übertragbar sind. Für die Chemie wäre damit beispielsweise die Anschaulichkeit eines Experimentes ebenfalls ein wichtiges Kriterium für die Eignung im Schulunterricht und damit auch ein Kriterium für die wahrgenommene Relevanz in einem Laborpraktikum. Damit sind diese Forschungsergebnisse für das Lehramtspraktikum Physik in einem gewissen Rahmen auch exemplarisch für andere naturwissenschaftliche Laborpraktika der Lehramtsausbildung zu verstehen. Die für das Fach Physik gefundenen Versuchseigenschaften können daher als vielversprechender Ausgangspunkt für die Entwicklung und ggf. weiterführende Beforschung eines lehramtsspezifischen Laborpraktikums einer anderen naturwissenschaftlichen Fachrichtung dienen.

So stellt sich nun die Frage, ob auch andere Studierendengruppen von den Erkenntnissen über die Eigenschaften relevanter Praktikumsversuche profitieren können. Auch wenn zum Beispiel Nebenfachstudierende im Fach Medizin, Biologie oder Geophysik andere Ausbildungsziele verfolgen und andere Tätigkeitsfelder anvisieren und daher die identifizierten relevanzstiftenden Versuchseigenschaften nicht direkt eins-zu-eins übertragen werden können, dann bietet jedoch die Kontextualisierung interessante Optionen für Praktikumsversuche spezifisch für diese Studierendengruppen. Die Kontextualisierung hat eine Sonderrolle, da sie auf die Studierendengruppe angepasst werden kann und daher auch in anderen Praktika als zentrales Gestaltungskriterium adressatenspezifischer Versuche genutzt wird (z. B. Theyßen, 2000). Für Nebenfachstudierende wie zum Beispiel der Medizin kann dabei nicht wie bei Lehramtsstudierenden mit der Fächerwahl Physik davon ausgegangen werden, dass sie ebenso ein grundsätzlich hohes Interesse an Physik mitbringen. Daher wäre es nicht verwunderlich, wenn der Effekt der Kontextualisierung auf die wahrgenommene Relevanz für physik-ferne Nebenfachstudierende sogar höher ausfällt als für die Lehramtsstudierenden des Faches Physik oder als für Studierende anderer physiknaher Studienrichtungen (z. B. Geophysik).

Mit den Befunden dieser Studie konnten damit exemplarisch Antworten auf die in der Beforschung adressatenspezifischer Praktika offene Frage nach der optimalen Versuchsgestaltung insbesondere auf Ebene der Versuche gefunden werden. Aus den

Ergebnissen dieser Arbeit können damit konkrete Empfehlungen für die Gestaltung (lehramtsspezifischer) Praktika an anderen Standorten abgeleitet werden können, die im Abschnitt 15 weiterführend diskutiert werden. Dabei konnte gezeigt werden, dass eine Beforschung der wahrgenommenen Relevanz auf Ebene der Versuche trotz der Herausforderung, die das Praktikum als vielschichtiger Untersuchungsgegenstand birgt, möglich ist. Diese Arbeit steuert dazu eine Vorgehensweise und ein neu entwickeltes Instrument zur zeitlich ökonomischen Erfassung der Versuchseigenschaften bei, mit denen entsprechende Erkenntnisse über relevant wahrgenommene Praktikumsversuche für spezifische Studierendengruppen erlangt werden können.

Nun soll ein Blick auf die Übertragbarkeit der Erkenntnisse dieser Arbeit zum Wirkmodell und deren Bedeutung geworfen werden. Das in dieser Arbeit aufgestellte Wirkmodell basiert auf grundlegenden Theorien und empirischen Befunden zu psychologischen Verhaltensmustern, die unabhängig von der Personengruppe sind, und enthält keine spezifische Anpassung auf die Studierendengruppe des Lehramts Physik (siehe Abschnitt 3). Aus diesem Grund kann geschlussfolgert werden, dass die grundlegenden Erkenntnisse zum Wirkmodell ebenfalls für andere Studierendengruppen in spezifischen Praktika plausibel sind. Damit können das Wirkmodell und die Befunde aus dem Lehramtspraktikum Physik der CAU in Bezug auf das Wirkmodell genauso wie die Kontextualisierung als prototypisch für adressatenspezifische Praktika allgemein betrachtet werden. Nachdem die in Abschnitt 3.6 genannten Studien (z. B. Theyßen, 2000; Nagel & Wolny, 2013) erste Einsichten zu den Wirkannahmen adressatenspezifischer Praktika gegeben haben, konnte auf dieser Basis die breit angelegte Analyse dieser Arbeit vorgenommen werden und die Größen wahrgenommene Relevanz, situationales Interesse und Lernerfolg anhand des hier hergeleiteten Wirkmodells erstmals in direkten Zusammenhang gebracht und quantitativ analysiert werden. Dabei ist es gelungen, die Plausibilität des angenommenen Wirkmodells aus verschiedenen Modellierungen der Zusammenhänge zu schlussfolgern. Diese Erkenntnisse leisten damit einen bedeutsamen Beitrag zum Verständnis der Rolle der wahrgenommenen Relevanz und des Wirkmechanismus in adressatenspezifischen Praktika. Außerdem liefert diese Arbeit damit eine Reihe von Anregungen für weiterführende Forschung, die in Abschnitt 16 diskutiert werden.

Diese Arbeit leistet damit wichtige Beiträge in zwei Bereichen aktueller Forschung. Zum einen trägt diese Arbeit zur Verbesserung der zukünftige Lehramtsausbildung bei, die aktuell ein großes Ziel in Deutschland ist, wie das vom Bundesministerium für Bildung und Forschung geförderte Projekt *Qualitätsoffensive Lehrerbildung* zeigt. Das in dieser Arbeit konzipierte Lehramtspraktikum bietet durch die Vernetzung von Fach und Fachdidaktik und durch die lehramtsspezifischen Gestaltungskriterien eine wirksame Möglichkeit, um den Lehramtsstudierenden die Relevanz der Lerninhalte für ihre zukünftige Lehrtätigkeit zu verdeutlichen und auch das situationale Interesse sowie mittelbar die Lernmotivation und den Lernerfolg zu fördern. Das lehramtsspezifisches Laborpraktikum stellt damit auch einen Weg

dar, um das Lehramtsstudium Physik in die Richtung der von der DPG geforderten Lehramtsausbildung „sui generis“ (Großmann & Hertel, 2014) zu orientieren, um die kumulative Entwicklung der fachlichen und fachdidaktischen Komponenten des Professionswissens zu stärken und die Studierenden ideal auf die Anforderungen ihrer späteren Lehrtätigkeit vorzubereiten. Die Erhöhung der wahrgenommenen Relevanz in der frühen Phase des Studiums könnte auch ein erstes Weichenstellen zur Verringerung von Studienabbruchentscheidungen im naturwissenschaftlichen Lehramtsstudium bedeuten. Zum anderen bringen diese Erkenntnisse die Forschung bzgl. der Gestaltung und Wirkung adressatenspezifischer Praktika im Allgemeinen voran. Das bedeutet, dass nicht nur das Lehramtsstudium von den Erkenntnissen dieser Studie profitieren kann, sondern viele Studienrichtungen, in denen Laborpraktika durchgeführt werden.

15 Implikationen für die Praxis

Die Erkenntnisse dieser Arbeit leisten nicht nur einen Forschungsbeitrag bzgl. Gestaltung und der Wirkungsweise adressatenspezifischer Praktika. Aus den Ergebnissen folgen auch konkrete praktische Empfehlungen für die Verbesserung der universitären Lehramtsausbildung in den naturwissenschaftlichen Fächern im Besonderen und die Gestaltung von Laborpraktika in der Hochschullehre im Allgemeinen.

Bei der Neukonzeption des lehramtsspezifischen Anfängerpraktikums in dieser Arbeit wurden ausgewählte Gestaltungsprinzipien umgesetzt, um die gängige Trennung von Fach und Fachdidaktik sowie den hohen fachlichen Anteil in der frühen Phase des Bachelorstudiums aufzubrechen und um dadurch unter anderem die wahrgenommene Relevanz der Lehramtsstudierenden zu fördern. Die empirischen Ergebnisse zeigen, dass durch die lehramtsspezifische Konzeption die wahrgenommene Relevanz der Praktikumsversuche erhöht wird, die sich gemäß der angenommen Wirkmechanismen wiederum förderlich auf das situationale Interesse am Lerninhalt und den Lernerfolg auswirkt. Damit stellt das lehramtsspezifische Anfängerpraktikum nicht nur ein wichtiges Element einer Verbesserung der Lehramtsausbildung an der CAU in Kiel dar, sondern wird auch als exemplarische Innovation zu einer wirksamen Option, um Fach und Fachdidaktik an anderen Standorten im Lehramtsstudium Physik enger zu vernetzen und damit die wahrgenommene Relevanz und den Lernerfolg zu fördern. Wie im Abschnitt zuvor diskutiert, ist auch ein Transfer auf Laborpraktika anderer naturwissenschaftlicher Fächer im Lehramtsstudium wie zum Beispiel Chemie und Biologie durchaus stimmig, sodass auch dort diese lohnenswerte Möglichkeit der Verbesserung der Lehramtsausbildung durch eine lehramtsspezifische Ausrichtung der Praktika stärker ins Interessenfeld rücken sollte.

Die in dieser Arbeit eingesetzten Gestaltungsprinzipien wurden anhand adressatenspezifischer Praktika aus bisheriger Forschung abgeleitet und für die lehramtsspezifische Konzeption des Praktikums spezifiziert. Es wurde empirisch gezeigt, dass einige dieser Designkriterien nachweisbar förderlich für die wahrgenommene Relevanz des Praktikums sind und damit gemäß der gezeigten Wirkzusammenhänge auch förderlich für den Lernerfolg der angehenden Lehrkräfte sind. Zudem konnten dabei weitere Erkenntnisse über Eigenschaften relevant wahrgenommener Versuche gewonnen werden, sodass die Gestaltungsprinzipien der Versuche noch weiter ausdifferenziert werden konnten. Auf Basis dieser aktuellen Erkenntnisse wird daher zur Erhöhung der wahrgenommenen Relevanz des Praktikums empfohlen, für die Entwicklung von Praktikumsversuchen für ein lehramtsspezifisches Anfängerpraktikum folgende Gestaltungsprinzipien zu nutzen: die Ausrichtung der fachlichen Lernziele auf den Erwerb des fachlichen Professionswissens für die spätere Lehrtätigkeit, die Berücksichtigung des Vorwissens, die Auswahl schulrelevanter Versuchsthemen, die Anschaulichkeit der Versuche zum Beispiel durch Analogie-Experimente, der Einsatz schulrelevanter Messtechniken und Software, die Kontextualisierung der Versuchs-

beschreibung sowie die Verknüpfung fachlicher und fachdidaktischer Fragestellungen (siehe Details in den Abschnitten 7 und 10).

Da die Betreuung eines bestehenden Laborpraktikums alleine bereits mit nicht vernachlässigbaren Kosten und einem hohem Zeitaufwand verbunden ist, kann eine komplette Neuentwicklung eines lehramtsspezifischen Praktikums an anderen Standorten schwierig bis schlichtweg nicht praktikabel sein. Daher sollen hier insbesondere lohnende Möglichkeiten betrachtet werden, wie auch mit geringerem Aufwand und innerhalb bestehender Rahmenbedingungen an anderen Standorten lehramtsspezifische Praktika individuell gestaltet und damit die wahrgenommene Relevanz und der Lernerfolg gefördert werden können.

Als eine erste Möglichkeit kann im Fall eines physikalischen Praktikums auf die hier entwickelten lehramtsspezifischen Praktikumsversuche zurückgegriffen werden, sodass eine vollständige Neuentwicklung gar nicht notwendig ist. Im Rahmen dieser Arbeit ist ein Pool von lehramtsspezifischen Praktikumsversuchen mit hoher wahrgenommener Relevanz entstanden, der inzwischen über die 25 hier dargestellten und untersuchten Versuche (Anhang A.1) hinaus geht, stetig anhand der empirischen Ergebnisse dieser Studie und Rückmeldungen der Betreuenden und Studierenden weiterentwickelt (z. B. Marten, 2018; Poeppel, 2019) und ausgebaut wird (z. B. Wolff-Riechert, 2017; Wyrwich, 2017; Brüning, 2018; Heer, 2018; Jacke, 2018; Culemeyer, 2019; Mandelkow, 2020). Ein weiterer interessanter Punkt ist, dass die Neukonzeptionen des Lehramtspraktikums an der CAU im Rahmen enger Randbedingungen u. a. bzgl. Lernzielen, Schwierigkeitsgrad und Studienordnung umgesetzt wurden, die nicht nur am Standort Kiel eine Rolle spielen. Das hier vorgestellte lehramtsspezifische Anfängerpraktikum konnte so ein klassisches Praktikum ohne Veränderung der Studienordnung vollständig ersetzen. Das bedeutet aus Praxissicht, dass diese erprobten lehramtsspezifischen Praktikumsversuche genutzt werden können, um sie sukzessiv in ein bestehendes klassisches Praktikum ohne eine Veränderung der bestehenden Rahmenbedingungen zu integrieren. So kann im Rahmen der Kapazitäten des jeweiligen Standortes ein klassisches Anfängerpraktikum auf Basis der hier neu entwickelten Versuche schrittweise lehramtsspezifisch ausgerichtet werden.

Statt lehramtsspezifische Versuche anhand der Gestaltungsempfehlungen neu zu entwickeln oder erprobte Versuche zu übernehmen, kann bei engeren Rahmenbedingungen als weitere Möglichkeit eine lehramtsspezifische Ausrichtung bereits bestehender Versuche eines Praktikums vorgenommen werden. Diese Arbeit hat gezeigt, dass auch bekannte Praktikumsversuche, die anhand der Designprinzipien eine neue lehramtsspezifische Ausrichtung erfahren haben, als sehr relevant wahrgenommen werden können (z. B. Newtonsche Axiome, Anhang A.1.19). In dieser Arbeit konnten viele kleine versuchsbedingte Faktoren als Stellschrauben in Bezug auf die wahrgenommene Relevanz identifiziert werden. Dies lässt zwar vermuten, dass ein deutlich größerer Effekt auf Relevanz und motivationale Faktoren erreicht werden kann, wenn mehrere der Designprinzipien genutzt werden. Doch können natürlich auch nur ein-

zelne Designprinzipien genutzt werden, um bestehende Versuche lehramtsspezifisch auszurichten. In diesem Fall wird insbesondere eine Kontextualisierung der Versuchsbeschreibungen empfohlen, da diese Arbeit gezeigt hat, dass eine Kontextualisierung der Einleitung allein bereits schon ausreicht, damit der Schulkontext des Versuches besser erkannt wird und die wahrgenommene Relevanz des Versuches erhöht wird. Anhand dessen ist zu erwarten, dass ein deutlich größerer Effekt erreicht werden kann, wenn der Gestaltungsspielraum der Versuchsbeschreibung über die Einleitung hinaus genutzt wird, um den Schulkontext stärker in der Versuchsanleitung zu verankern. Die Identifikation passender schulischer Kontexte und ggf. sogar für eine Verknüpfung der fachlichen Fragestellungen mit fachdidaktischen Aspekten, kann auch als Anlass genommen werden, Unterstützung in der Fachdidaktik zu erfragen und damit die Kooperation von Fach und Fachdidaktik am Standort für die Lehramtsausbildung gewinnbringend auszubauen. Die Kontextualisierung der Versuchsbeschreibungen ist dabei sogar nicht nur eine attraktive Möglichkeit für Laborpraktika für Lehramtsstudierende, sondern auch auf adressatenspezifische Praktika im Allgemeinen. Die Kontextualisierung bietet damit eine weniger aufwendige, aber trotzdem nachweislich wirksame Maßnahme zur (lehramts-) spezifischen Ausrichtung von Praktikumsversuchen und Verbesserung der wahrgenommenen Relevanz.

Somit sind lehramtsspezifische Praktikumsversuche nicht nur nachweislich wirksam in Bezug auf die wahrgenommene Relevanz, das situationale Interesse sowie den Lernerfolg und damit auch für andere Standorte und Fächer empfehlenswert, sondern vor allem auch individuell im Rahmen der Möglichkeiten und Grenzen des jeweiligen Standortes zielführend umsetzbar. Damit leistet diese Arbeit auch sehr konkrete und praktikable Vorschläge zur Umsetzung der gewonnen Forschungsergebnisse in der Hochschullehre.

16 Desiderat

An dieser Stelle soll nun abschließend das Potential der Befunde für zukünftige Forschung und Entwicklungen beleuchtet werden. Dabei ergeben sich sowohl Desiderata für das sehr aktuelle Feld der Lehramtsausbildung als auch für das sehr forschungsaufwendige Feld der adressatenspezifischen Praktika.

In dieser Arbeit ist es gelungen, explorativ eine Reihe von Eigenschaften relevant wahrgenommener Praktikumsversuche zu identifizieren und damit Prinzipien für die Gestaltung von lehramtsspezifischen Praktikumsversuchen zu formulieren. Daraus folgt nicht nur die Frage nach den optimalen Versuchseigenschaften für andere Studierendengruppen, sondern weiterführend vor allem auch, wie zum Beispiel durch eine ideale Kombination der empfohlenen Versuchseigenschaften andersherum nun systematisch Versuche mit hoher wahrgenommener Relevanz konzipiert werden können. Dazu bedarf es weiterführend zur Analyse der wahrgenommenen Relevanz einer systematischen Variation der für das Lehramt Physik explorativ identifizierten Versuchseigenschaften. Die Befunde zur Wirkung adressatenspezifischer Praktikumsversuche stellen zudem einen Ausgangspunkt dar, um die Gültigkeit der Wirkannahmen abzusichern und tiefer gehendes Verständnis der Wirkungsweise in Bezug auf den Lernerfolg auf Ebene der einzelnen Versuche allgemein für adressatenspezifische Praktika zu erlangen. Zudem ist auch der empirische Nachweis des Gesamtwirkmodells von Interesse, wonach die spezifische Praktikumsgestaltung schlussendlich den Lernerfolg begünstigt, wie in Abschnitt 14 andiskutiert. Auch dazu ist eine gezielte Manipulation der Versuchseigenschaften sowie eine Anpassung der Erhebungsinstrumente z. B. durch Ausweitung des Fachwissenstest für bestimmte Versuchsthemen in Breite und Tiefe notwendig, um auf Versuchsebene das Wirkmodell differenzierter zu analysieren. Auch sind die langfristigen Auswirkungen eines solchen Praktikums auf die Lernmotivation und den Lernerfolg in der Studieneingangsphase und in der weiterführenden Ausbildung sowie im Fall des Lehramtspraktikums auf den Einsatz von Experimenten in der späteren Schulpraxis von Interesse. Als eine mögliche Folgestudie würde sich anbieten, mit diesem Projekt im Sinne des Design-Based-Research in eine zweite Phase zu starten. So können die bestehenden Versuche, die aktuell anhand der Erkenntnisse dieser Arbeit weiterentwickelt werden, und weitere neue Lehramtsversuche mit gezielter Manipulation der Versuchseigenschaften genutzt werden, um das Verständnis über die Eigenschaften und Wirkungsweise der spezifischen Praktikumsversuche zu vertiefen.

Auch für eine Untersuchung der langfristigeren Wirkung einer lehramtsspezifischen Lehrveranstaltung bietet das hier konzipierte Praktikum eine Basis. Eine interessante Frage diesbezüglich ist, ob durch das lehramtsspezifische Praktikum die angestrebte Verbesserung der kumulativen Entwicklung der fachlichen und fachdidaktischen Komponenten des Professionswissens erreicht werden und weiterführend damit neue Argumente in Bezug auf das von der Deutschen Physikalischen Gesellschaft (DPG) geforderte eigenständige Lehramtsstudium „sui generis“ (Großmann

& Hertel, 2014) gewonnen werden können. Außerdem ist interessant, ob sich langfristig durch die lehramtsspezifische Konzeption der Laborpraktika der Einsatz von Experimenten im Physikunterricht der Schule verbessert. In Hinblick auf aktuelle Diskussionen ist eine weitere äußerst spannende Fragestellung, ob eine Erhöhung der wahrgenommene Relevanz in der frühen Phase des Studiums sogar die hohen Studienabbruchzahlen im naturwissenschaftlichen Lehramtsstudium verringern kann.

Diese Arbeit bereitet mit den gewonnen Erkenntnissen zur wahrgenommenen Relevanz und den Wirkzusammenhängen damit nicht nur den Boden für weiterführende Beforschung von lehramts- und adressatenspezifischen Praktika. Aus praktischer Perspektive der Hochschullehre sollen die Ergebnisse dieser Arbeit gleichzeitig Impulse, Anregungen und konkrete Empfehlungen zur Verbesserung der Lehramtsausbildung im Speziellen und für die Umsetzung und Weiterentwicklung von adressatenspezifischen Praktika im Allgemeineren bieten.

Literaturverzeichnis

AAPT. (1998). Goals of the introductory physics laboratory. *American Journal of Physics*, *66*(6), 483–485.

AG-Studienqualität. (2011). *Allgemeiner Bericht Onlinebefragung Professionsorientierung/Berufsqualifizierung im Lehramtsstudium an der Universität Potsdam.* [letzter Zugriff am 18.04.2020]. https://pep.uni-potsdam.de/media/reports/up_zfl_umfrage-professionsorientierung-lehramt_2011.pdf

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological methodology*, 93–114.

Andersen, J., Block, D., Neumann, K. & Wendlandt, H. (2018). Lehramtsausbildung Physik 2.0 - Vernetzung von Fach, Fachdidaktik und schulpraktischen Aspekten. In B. Brouër, A. Burda-Zoyke, J. Kilian & I. Petersen (Hrsg.), *Vernetzung in der Lehrerinnen-und Lehrerbildung: Ansätze, Methoden und erste Befunde aus dem LeaP-Projekt an der Christian-Albrechts-Universität zu Kiel* (S. 23–35). Waxmann Verlag.

Audacity. (o.D.). *Audacity* [letzter Zugriff am 03.02.2020]. https://www.audacity%5C-team.org/

Aufschnaiter, S. v. & Welzel, M. (1996). Beschreibung von Lernprozessen. *Lernen in den Naturwissenschaften. Kiel: IPN*, 301–327.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2016). *Multivariate Analysemethoden.* Springer.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, *37*(3), 379–384.

Bandalos, D. L. & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Hrsg.), *New developments and techniques in structural equation modeling* (S. 289–316). Psychology Press.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, *52*(1), 1–26.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bauer, A., Reinhold, P. & Sacher, M. (2018). Operationalisierung der experimentellen Kompotenz (Physik-)Studierender. *GDCP-Jahrestagung Regensburg*, 919–922.

Bauer, A. & Sacher, M. (2018). Kompetenzorientierte, universitäre Laborpraktika - Das Paderborner Physik Praktikum (3P). *PhyDid B-Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung.*

Bauer, J., Diercks, U., Rösler, L., Möller, J. & Prenzel, M. (2012). Lehramtsstudium in Deutschland: Wie groß ist die strukturelle Vielfalt. *Unterrichtswissenschaft*, *40*(2), 101–120.

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaf*, *9*(4), 469–520.

Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Waxmann Verlag.

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American journal of Physics*, *62*(8), 750–762.

Bennett, J., Lubben, F. & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science education*, *91*(3), 347–370.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin.. *Psychological bulletin*, *112*(3), 400.

Berger, R. (2002). Einfluss kontextorientierten Physikunterrichts auf Interesse und Leistung in der Sekundarstufe II. *Zeitschrift für Didaktik der Naturwissenschaften*, *8*(1), 119–132.

Bergin, D. A. (1999). Influences on classroom interest. *Educational psychologist*, *34*(2), 87–98.

Blossfeld, H.-P. (2019). *Education As a Lifelong Process: The German National Educational Panel Study (NEPS).* (Bd. 3). Springer.

Bortz, J. & Schuster, C. (2010). *Statistik für Human-und Sozialwissenschaftler* [7. Auflage]. Springer-Verlag.

Boud, D., Dunn, J., Kennedy, T. & Thorley, R. (1980). The aims of science laboratory courses: A survey of students, graduates and practising scientists. *European Journal of Science Education*, *2*(4), 415–428.

Brauer, A.-K. (2017). *Analyse von Streuexperimenten an Magneten.* Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Braun, E. & Gusy, B. (2006). Perspektiven der Lehrevaluation. *Didaktik und Evaluation in der Psychologie*, 152.

Braun, E., Gusy, B., Leidner, B. & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, *54*(1), 30–42.

Brown, D. (2019). *Tracker – Video Analysis and Modeling Tool* [letzter Zugriff am 03.02.2020]. https://physlets.org/tracker/

Brown, D. & Cox, A. J. (2009). Innovative uses of video analysis. *The Physics Teacher*, *47*(3), 145–150.

Brown, G., Harris, L. & McMillan, J. (2013). Sage handbook of research on classroom assessment. *Chap. Student self-assessment, pp. 367–393). Sage Thousand Oaks, CA. doi*, *10*(4135), 9781452218649.

Browne, M. W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, *21*(2), 230–258.

Brüggemann, V. (2015). Validierung einer Skala zur Erfassung der Skriptqualität in Physik-Experimentierpraktika, Abschlussarbeit zum Erwerb des akademischen Grades Master of Education. *Freie Universität Berlin.*

Brüning, C. (2018). *Entwicklung eines Praktikumsversuches mit dem Themenschwerpunkt Fehlerbetrachtung.* Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Cannam, C., Landone, C. & Sandler, M. (2010). Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. *Proceedings of the 18th ACM international conference on Multimedia*, 1467–1468.

CAU. (2007a). *Fachprüfungsordnung der Christian-Albrechts-Universität zu Kiel für 2-Fach-Bachelor Physik (Version 2007)* [letzter Zugriff am 24.04.2020]. http://www.studservice.uni-kiel.de/sta/fachpruefungsordnung-physik-bachelor-master-2-faecher-2007.pdf

CAU. (2007b). *Modulhandbuch der Christian-Albrechts-Universität zu Kiel für 1-Fach-Bachelor Physik (Version 2007)* [letzter Zugriff am 24.04.2020]. https://www.physik.uni-kiel.de/de/studium/bama/modulhandbuch-1-fach-bachelor

CAU. (2007c). *Modulhandbuch der Christian-Albrechts-Universität zu Kiel für 2-Fach-Bachelor Physik (Version 2007)* [letzter Zugriff am 24.04.2020]. https://www.physik.uni-kiel.de/de/studium/bama/modulhandbuch-2-faecher-bachelor

CAU. (2017a). *Fachprüfungsordnung der Christian-Albrechts-Universität zu Kiel für 1-Fach-Bachelor Physik (Version 2017)* [letzter Zugriff am 24.04.2020]. http://www.studservice.uni-kiel.de/sta/fachpruefungsordnung-physik-bachelor-master-1-fach.pdf

CAU. (2017b). *Fachprüfungsordnung der Christian-Albrechts-Universität zu Kiel für 2-Fach-Bachelor Physik (Version 2017)* [letzter Zugriff am 24.04.2020]. http://www.studservice.uni-kiel.de/sta/fachpruefungsordnung-physik-bachelor-master-2-faecher.pdf

CAU. (2017c). *Modulhandbuch der Christian-Albrechts-Universität zu Kiel für 1-Fach-Bachelor Physik (Version 2017)* [letzter Zugriff am 24.04.2020]. https://www.physik.uni-kiel.de/de/studium/studiengang-physik/Modulhandbuch_Physik_1Fach_BSc_27052019.pdf

CAU. (2017d). *Modulhandbuch der Christian-Albrechts-Universität zu Kiel für 2-Fach-Bachelor Physik (Version 2017)* [letzter Zugriff am 24.04.2020]. https://www.physik.uni-kiel.de/de/studium/studiengang-physik/Modulhandbuch_Physik_2Faecher_BA_27052019.pdf

CAU. (2018). *Zwei-Fächer-Fachprüfungsordnung der Christian-Albrechts-Universität zu Kiel (Version 2018)* [letzter Zugriff am 18.05.2020]. http://www.studservice.uni-kiel.de/sta/zwei-faecher-pruefungsordnung-bachelor-master.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2. Aufl.). Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Cramer, C. (2010). Kompetenzerwartungen Lehramtsstudierender. Grenzen und Perspektiven selbsteingeschätzter Kompetenzen in der Lehrerbildungsforschung. *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht*, 85–97.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297–334.

Culemeyer, H. C. (2019). *Entwicklung und Test einer Lernstation zur Untersuchung von Wasserwellen*. Unveröffentlichte Masterarbeit an der Christian-Albrechts-Universität zu Kiel.

Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, *39*(2), 223–238.

Demtröder, W. (2016a). *Experimentalphysik 3 - Atome Moleküle und Festkörper* (5. Aufl.). Springer-Verlag.

Demuth, R., Gräsel, C. & Parchmann, I. (2008). *Chemie im Kontext: von der Innovation zur nachhaltigen Verbreitung eines Unterrichtskonzepts*. Waxmann.

Ding, L., Chabay, R., Sherwood, B. & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical review special Topics-Physics education research*, *2*(1).

Dochy, F., Segers, M. & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, *24*(3), 331–350.

Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer.

Dounas-Frazer, D. R., Van De Bogart, K. L., Stetzer, M. R. & Lewandowski, H. (2016). Investigating the role of model-based reasoning while troubleshooting an electric circuit. *Physical Review Physics Education Research*, *12*(1), 010137.

Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden* [1. Auflage]. Weinheim: Beltz.

Emden, M. (2011). *Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens: eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I* (Bd. 118). Logos Verlag Berlin GmbH.

English, M. C. & Kitsantas, A. (2013). Supporting student self-regulated learning in problem-and project-based learning. *Interdisciplinary journal of problem-based learning*, *7*(2), 6.

Falchikov, N. & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*(4), 395–430.

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. *The scholarship of teaching and learning in higher education: An evidence-based perspective* (S. 93–143). Springer.

Feynman, R., Leighton, R. & Sands, M. (1963). The Feynman Lectures on Physics, Vol. 1, Chap.1, Addison Wesley. *Reading, MA*.

Field, A., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. Sage publications.

Finkelstein, N. (2005). Learning physics in context: A study of student learning about electricity and magnetism. *International Journal of Science Education*, *27*(10), 1187–1209.

Fox, J. & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third). Sage.

Frühwein, O. & Heinke, H. (2005). Entwicklung eines Anfängerpraktikums für Studierende des Lehramtes Physik. *Didaktik der Physik–Beiträge zur DPG-Frühjahrstagung. Berlin.*

Frühwein, O. & Heinke, H. (2007). Umsetzung eines Anfängerpraktikums für Lehramtsstudierende mit Fach Physik. Berufsfeldorientierte Gestaltung einer Fachveranstaltung. *Didaktik der Physik–Beiträge zur DPG-Frühjahrstagung. Regensburg.*

Geertz, M. (2016). *Streuversuche mit Permanentmagneten auf einem Luftkissentisch.* Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Gersner, J. (2016). *Versuche zur Interferenz von Licht an einem Handy-Display.* Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. *Re-examining pedagogical content knowledge in science education* (S. 38–52). Routledge.

Gramzow, Y. (2015). Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. *Modellierung und Testkonstruktion. Berlin: Logos.*

Grossman, P., Schoenfeld, A. & Lee, C. (2005). Teaching subject matter. In L. Darling-Hammond, J. Bransford, P. LePage, K. Hammerness & H. Duffy (Hrsg.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (S. 201–231). San Francisco, CA: Jossey Bass.

Großmann, S. & Urban, K. (2006). Thesen für ein modernes Lehramtsstudium im Fach Physik. Deutsche Physikalische Gesellschaft.

Großmann, S. & Hertel, I. (2014). Zur fachlichen und fachdidaktischen Ausbildung für das Lehramt Physik.

Häder, M. (2015). *Empirische Sozialforschung - Eine Einführung* (3. Aufl.). Springer.

Haller, K. (1999). *Über den Zusammenhang von Handlungen und Zielen: eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*. Logos-Verlag.

Haller, T. (2011). *Didaktische Rekonstruktion eines Physikpraktikums für Meteorologie*. Diplomarbeit an der Uni Wien.

Hammann, M. (2004). Kompetenzentwicklungsmodelle Merkmale und ihre Bedeutung - dargestellt anhand von Kompetenzen beim Experimentieren. *Mathematische und Naturwissenschaftliche Unterricht*, 196–203.

Harlen, W. (1999). *Effective Teaching of Science. A Review of Research. Using Research Series, 21.* ERIC.

Hartig, J. & Rakoczy, K. (2010). Mehrebenenanalyse. In H. Holling & B. Schmitz (Hrsg.), *Handbuch Statistik, Methoden und Evaluation*. Hogrefe Verlag.

Häussler, P. & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of research in science teaching*, *39*(9), 870–888.

Hedderich, J. & Sachs, L. (2016). *Angewandte Statistik* (15. Aufl.). Springer.

Heer, A. (2018). *Entwicklung einer Experimentierstation zur akustischen Unschärfe*. Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Heidrich, J., Neumann, K. & Petersen, S. (2015). Strukturelle- und externe Validierung eines Experimentiertests. *Naturwissenschaftliche Bildung zwischen Science-und Fachunterricht Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Bremen*, 262–264.

Heidrich, J. (2016). *Erfassung von Experimentierkompetenz im universitären Kontext-Entwicklung und Validierung eines Experimentiertests zum Themenbereich Optik* (Diss.). Christian-Albrechts Universität Kiel.

Heptin, V. & Fromme, B. (2016). Mechanisches Modell zur Rutherfordstreuung. *Praxis der Naturwissenschaften-Physik in der Schule*, *65*(5).

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, *30*(3), 141–158.

Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., König, R., Richter, J. & Woisch, A. (2017). Zwischen Studienerwartungen und Studienwirklichkeit. *Forum Hochschule*, *1*(2017), 134–136.

Hidi, S. & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist*, *41*(2), 111–127.

Higgins, E. T., Strauman, T. & Klein, R. (1986). Standards and the process of self-evaluation: Multiple affects from multiple stages.

Hodson, D. (1996). Laboratory work as scientific method: Three decades of confusion and distortion. *Journal of Curriculum studies*, *28*(2), 115–135.

Hofstein, A. & Lunetta, V. N. (1982). The role of the laboratory in science teaching: Neglected aspects of research. *Review of educational research*, *52*(2), 201–217.

Hofstein, A. & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science education*, *88*(1), 28–54.

Holmes, N. G. & Wieman, C. E. (2018). Introductory physics labs: We can do better. *Physics Today*, *71*(1), 38–45.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2. Aufl.). Routledge.

Hu, L.-t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1–55.

Huber, S. R. & Jessen, K. (2013). Inhalte für das Physikpraktikum der Biologen. *PhyDid B-Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung*.

Hucke, L. (2000). *Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums* (Diss.). Universität Dortmund.

Jacke, F. (2018). *Videobasierte, quantitative Analyse des Millikanversuches*. Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Joachim Herz Stiftung. (o.D.). *Photo: Polarisation von Licht - Einführung - Analogie zu mechanischen Wellen* [letzter Zugriff am 14.07.2020]. https://www.leifiphysik.de/optik/polarisation/grundwissen/polarisation-von-licht-einfuehrung

Joußen, N., Thiel, J. & Heinke, H. (2019). Scaffolding im Anfängerpraktikum für Lehramtsstudierende. *PhyDid B-Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung*, *1*.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, *8*(3), 325–352.

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2. Aufl.). Sage Publications.

Kaplan, D. & Elliott, P. R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(1), 1–24.

Kattmann, U., Duit, R., Gropengießer, H. & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, *3*(3), 3–18.

Keller, M. (2011). *Teacher enthusiasm in physics instruction* (Diss.). University of Duisburg-Essen, Germany.

Keller, M., Neumann, K. & Fischer, H. (2014). Enthusiastic teaching and its impact on students' interest and self-concept: an investigation of German physics classrooms. In H. E. Fischer, P. Labudde, K. Neumann & J. Viiri (Hrsg.), *Quality of Instruction in Physics – Comparing Finland, Germany and Switzerland* (S. 129–144). Waxmann.

Kievit, R. A., Brandmaier, A. M., Ziegler, G., Van Harmelen, A.-L., de Mooij, S. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P. et al. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental cognitive neuroscience*, *33*, 99–117.

Kircher, E., Girwidz, R. & Häußler, P. (2015). *Physikdidaktik* (3. Aufl.). Springer.

Kish, L. (1965). *Survey sampling*. New York, NY: Wiley.

Kissmann, F., Müller, R., Schumacher, D. & Theyßen, H. (2010). Brauchen Studierende der Biologie ein Physikpraktikum? *PhyDid B-Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung*.

Klahr, D., Dunbar, K., Fay, A., Penner, D. & Schunn, C. (2000). Exploring science: the cognition and development of discovery processes (Cambridge, Massachusetts, Massachusetts Institute of Technology Press).

Klassen, R. M., Tze, V. M., Betts, S. M. & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational psychology review*, *23*(1), 21–43.

Knogler, M. (2014). *Investigating student interest in the context of problem-based learning* (Diss.). Technische Universität München.

Knogler, M., Harackiewicz, J. M., Gegenfurtner, A. & Lewalter, D. (2015). How situational is situational interest? Investigating the longitudinal structure of situational interest. *Contemporary Educational Psychology*, *43*, 39–50.

Köller, O., Trautwein, U., Lüdtke, O. & Baumert, J. (2006). Zum Zusammenspiel von schulischer Leistung, Selbstkonzept und Interesse in der gymnasialen Oberstufe. *Zeitschrift für pädagogische Psychologie*, *20*(1/2), 27–39.

Krapp, A. (1992). Interesse, Lernen und Leistung. Neue Forschungsansätze in der Pädagogischen Psychologie. *Zeitschrift für Pädagogik*, *38*(5), 747–770.

Kratschmar, J. (2017). *Erhebung zur Experimentellen Kompetenz in physikalischen Praktika und Validierung eines Evaluationsinstruments zur Praktikumsqualität (PraQ)*. Masterarbeit an der Freien Universität Berlin in der Didaktik der Physik.

Krey, O. & Rabe, T. (2012). Physikdidaktischer Kompetenzerwerb im Praktikum „Physikalische Schulexperimente “. *Inquiry-based Learning–Forschendes Lernen*, 164–166.

Kröger, J. (2019). *Struktur und Entwicklung des Professionswissens angehender Physiklehrkräfte* (Diss.). Christian-Albrechts-Universtität zu Kiel.

Kultusministerkonferenz. (1994). Richtlinie zur Sicherheit im Unterricht. *Stand: 26.02.16*.

Kunter, M., Frenzel, A., Nagy, G., Baumert, J. & Pekrun, R. (2011). Teacher enthusiasm: Dimensionality and context specificity. *Contemporary Educational Psychology*, *36*(4), 289–301.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.

Lai, M. H. & Kwok, O.-m. (2015). Examining the rule of thumb of not using multilevel modeling: The "design effect smaller than two" rule. *The Journal of Experimental Education*, *83*(3), 423–438.

Lawrence, M. A. (2016). *ez: Easy Analysis and Visualization of Factorial Experiments* [R package version 4.4-0]. https://CRAN.R-project.org/package=ez

LD-Didactic. (2014). *Handbuch CASSY Lab* [Version 26.05.2014]. https://www.ld-didactic.de/software/524221de.pdf

LD-Didactic. (o.D. a). *CASSY – Datalogging und Messtechnik für alle Ausbildungsniveaus* [letzter Zugriff am 03.02.2020]. https://www.ld-didactic.de/produkteloesungen/cassy-datalogging-und-messtechnik.html

LD-Didactic. (o.D. b). *Gebrauchsanweisung 388 00-06: Gerätegruppe zum mechanischen und elektrischen Wärmeäquivalent* [letzter Zugriff am 11.04.2020]. https://www.ld-didactic.de/documents/de-DE/GA/GA/3/388/38800de.pdf

LD-Didactic. (o.D. c). *Versuchsanleitung P2.6.3.1: Determining the efficiency of the heat pump as a function ofthe temperature differential* [letzter Zugriff am 10.04.2020]. https://www.ld-didactic.de/documents/de-DE/EXP/P/P2/P2631_e.pdf

LeBreton, J. M. & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational research methods*, *11*(4), 815–852.

Lersch, R. (2006). Lehrerbildung im Urteil der Auszubildenden. Eine empirische Studie zu beiden Phasen der Lehrerausbildung. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung vonLehrerinnen und Lehrern.* (S. 164–181). Weinheim: Beltz.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow & H. B. Mann (Hrsg.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (S. 278–292). Stanford University Press: Palo Alto, CA, USA.

Lewalter, D. & Geyer, C. (2009). Motivationale Aspekte von schulischen Besuchen in naturwissenschaftlich-technischen Museen. *Zeitschrift für Erziehungswissenschaft*, *12*(1), 28–44.

Lewalter, D. & Knogler, M. (2014). *A questionnaire to assess situational interest – theoretical considerations and findings*. Poster presented at Annual Meeting of the American Educational Research Association (Philadelphia, 2014).

Lewalter, D. & Willems, A. S. (2009). Die Bedeutung des motivationsrelevanten Erlebens und des individuellen Fachinteresses für das situationale Interesse im Mathematikunterricht. *Psychologie in Erziehung und Unterricht*.

Linke, R. (2017). *Ein Michelson-Interferometer aus LEGO®-Bausteinen: Untersuchung technischer und affektiver Aspekte*. Springer-Verlag.

Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural equation modeling*, *9*(2), 151–173.

Loehlin, J. C. & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural equation analysis* (5. Aufl.). Taylor & Francis.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological bulletin*, *68*(5), 304.

Lorentzen, J. (2020). *Entwicklung und Evaluation eines Lernangebots im Lehramtsstudium Chemie zur Förderung von Vernetzungen innerhalb des fachbezogenen Professionswissens* (Diss.). Christian-Albrechts Universität Kiel.

Lorentzen, J., Friedrichs, G., Ropohl, M. & Steffensky, M. (2019). Förderung der wahrgenommenen Relevanz von fachlichen Studieninhalten: Evaluation einer Intervention im Lehramtsstudium Chemie. *Unterrichtswissenschaft*, *47*(1), 29–49.

Lubben, F., Bennett, J., Hogarth, S. & Robinson, A. (2005). A systematic review of the effects of context-based and Science-Technology-Society (STS) approaches in the teaching of secondary science on boys and girls, and on lower ability pupils.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior research methods*, *49*(4), 1494–1502.

Maiseyenka, V., Nawrath, D. & Schecker, H. (2010). Modellbasierte Diagnose und Förderung experimenteller Kompetenz. *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Potsdam*.

Maiseyenka, V., Schecker, H. & Nawrath, D. (2013). Kompetenzorientierung des naturwissenschaftlichen Unterrichts-Symbiotische Kooperation bei der Ent-

wicklung eines Modells experimenteller Kompetenz. *PhyDid A-Physik und Didaktik in Schule und Hochschule*, *1*(12), 1–17.

Makrinus, L. (2012). *Der Wunsch nach mehr Praxis: Zur Bedeutung von Praxisphasen im Lehramtsstudium*. Springer-Verlag.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J. & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69*(S1), S12–S23.

Mandelkow, T. (2020). *Digitale Messwerterfassung von U-Rohr-Schwingungen und deren didaktische Analyse*. Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Mangiafico, S. (2020). *rcompanion: Functions to Support Extension Education Program Evaluation* [R package version 2.3.25]. https://CRAN.R-project.org/package=rcompanion

Marten, J. N. (2018). *Quantitative Analyse von Stoßprozessen mit attraktivem Wechselwirkungspotential*. Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Massolt, J. & Borowski, A. (2018). Increasing the Perceived Relevance of University Physics Problems by Focusing on School-Related Content Knowledge. *heiEDUCATION Journal. Transdisziplinäre Studien zur Lehrerbildung*, (1-2), 99–125.

Massolt, J. & Borowski, A. (2020). Perceived relevance of university physics problems by pre-service physics teachers: personal constructs. *International Journal of Science Education*, 1–23.

Meier, K. (2011). *Video zu Polarisation* [letzter Zugriff am 19.11.2016]. https://www.youtube.com/watch?v=eYguAcvjN2o

Meinhardt, C. (2018). *Entwicklung und Validierung eines Testinstruments zu Selbstwirksamkeitserwartungen von (angehenden) Physiklehrkräften in physikdidaktischen Handlungsfeldern*. Logos Verlag.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.

Mešić, V., Hajder, E., Neumann, K. & Erceg, N. (2016). Comparing different approaches to visualizing light waves: An experimental study on teaching wave optics. *Physical Review Physics Education Research*, *12*(1), 010135.

Ministerium für Schule und Berufsbildung Schleswig-Holstein. (2019). Fachanforderungen Physik [Stand: Februar 2019].

MINT-Forum, N. (2013). Zehn Thesen und Forderungen zur MINT-Lehramtsausbildung – Empfehlungen des Nationalen MINT Forums. *München: Herbert Utz Verlag*, *1*.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of educational psychology*, *85*(3), 424.

Moosbrugger, H. & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion*. Springer.

Moritz, C. E. (1996). Student Self-Assessment of Language Proficiency: Perceptions of Self and Others.

Müller, S., Vogt, P. & Kuhn, J. (2010). Das Handy im Physikunterricht: Anwendungsmöglichkeiten eines bisher wenig beachteten Mediums. *PhyDid B - Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung*.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological methods & research*, *22*(3), 376–398.

myphotonics. (o.D.). *Optik und Photonik mit LEGO-Bausteinen: Michelson-Interferometer*. [Webseite Imlau Research Group Ultrafast Physics Universität Osnabrück; letzter Zugriff am 02.02.2020]. https://www.ufp.uni-osnabrueck.de/en/education/myphotonics

Nagel, C. (2009). *eLearning im physikalischen Anfängerpraktikum* (Bd. 96). Logos Verlag Berlin GmbH.

Nagel, C., Scholz, R. & Weber, K.-A. (2018). Umfrage zu den Lehr/Lernzielen in physikalischen Praktika. *PhyDid B-Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung*.

Nagel, C. & Wolny, B. (2013). Ein adressatenspezifisches Physikpraktikum für Ernährungswissenschaften - Didaktische Rekonstruktion und Evaluation. *PhyDid A-Physik und Didaktik in Schule und Hochschule*, *1*(12), 48–61.

Neumann, K. (2004). *Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker*. Logos-Verlag.

Neumann, K. & Borowski, A. (2011). Monoedukativer Physikunterricht. *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie*, *31*, 123.

Neumann, K., Härtig, H., Harms, U. & Parchmann, I. (2017). Science teacher preparation in Germany. In J. Pedersen, T. Isozaki & T. Hirano (Hrsg.), *Model science teacher preparation programs: An international comparison of what works* (S. 29–52). IAP.

Neumann, K. & Welzel, M. (2007). A new labwork course for physics students: Devices, Methods and Research Projects. *European Journal of Physics*, *28*(3), S61.

Okada, K. (2013). Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika*, *40*(2), 129–147.

Olejnik, S. & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary educational psychology*, *25*(3), 241–286.

Olejnik, S. & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological methods*, *8*(4), 434.

Parchmann, I. & Kuhn, J. (2018). Lernen im Kontext. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 193–207). Springer.

Patrick, B. C., Hisley, J. & Kempler, T. (2000). "What's everybody so excited about?": The effects of teacher enthusiasm on student intrinsic motivation and vitality. *The Journal of Experimental Education*, *68*(3), 217–236.

Phywe. (o.D.). *Experiment Instructions: Mechanical Equivalent of Heat* [letzter Zugriff am 14.07.2020]. http://www.phywe-es.com/index.php/fuseaction/download/lrn_file/versuchsanleitungen/p2330200/p2330200e.pdf

Plomer, M. (2011). *Physik physiologisch passend praktiziert: Eine Studie zur Lernwirksamkeit von traditionellen und adressatenspezifischen Physikpraktika für die Physiologie* (Bd. 112). Logos Verlag Berlin GmbH.

Plomer, M., Jessen, K., Rangelov, G. & Meyer, M. (2010). Lernwirksamkeit adressatenspezifischer Praktikumsversuche aus Sicht der Physiologie. *PhyDid B-Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung*.

Poeppel, J. (2019). *Entwicklung eines Praktikumsversuchs - Einführung in phyphox.* Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Prenzel, M. (1995). Zum Lernen bewegen. Unterstützung von Lernmotivation durch Lehre. *Blick in die Wissenschaft*, *4*(7), 58–66.

Prenzel, M. & Drechsel, B. (1996). Ein Jahr kaufmännische Erstausbildung: Veränderungen in Lernmotivation und Interesse. *Unterrichtswissenschaft*, *24*(3), 217–234.

Psillos, D. & Niedderer, H. (2002). *Teaching and learning in the science laboratory.* Springer.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing - R Version 3.6.1.* R Foundation for Statistical Computing. Vienna, Austria.

Rabe, T., Meinhardt, C. & Krey, O. (2012). Entwicklung eines Instruments zur Erhebung von Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern. *Zeitschrift für Didaktik der Naturwissenschaften*, *18*, 293–315.

Rammert, M. & Webler-Pijahn, K. (2001). Dokumentation der Skalen-und Item-Auswahl für den Kinderfragebogen zur Lernmotivation und zum emotionalen Erleben. *Universität Bielefeld, Abteilung für Pädagogische Psychologie.*

Rapior, G., Sengstock, K. & Baev, V. (2006). New features of the Franck-Hertz experiment. *American journal of physics*, *74*(5), 423–428.

Rehfeldt, D. (2018a). *Erfassung der Lehrqualität naturwissenschaftlicher Experimentalpraktika* (Diss.). Freie Universität Berlin.

Rehfeldt, D. (2018b). Skalenhandbuch PraQ-A: Lernzuwachs der Studierenden Evaluation von naturwissenschaftlichen Experimentalpraktika.

Rehfeldt, D., Mühlenbruch, T. & Nordmeier, V. (2015). Fragebogen zu Praktikumskompetenzen (PraKo): Erforschung naturwissenschaftlicher Praktika.

Renkl, A. (1996). Vorwissen und Schulleistung. *Emotionen, Kognitionen und Schulleistung*, 175–190.

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research* [R package version 1.9.12]. Northwestern University. Evanston, Illinois.

Rheinberg, F. (2002). Motivation. *Kohlhammer, Stuttgart.*

Rheinberg, F., Vollmeyer, R. & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern-und Leistungssituationen (Langversion, 2001). *Diagnostica*, *2*, 57–66.

Rosenshine, B. (1970). Enthusiastic teaching: A research review. *The School Review*, *78*(4), 499–514.

Rosenthal, L., Robert & Jacobson. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological reports*, *19*(1), 115–118.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Rotgans, J. I. & Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction*, *21*(1), 58–67.

Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *33*(10), 1045–1063.

Rutherford, E. (1911). LXXIX. The scattering of α and β particles by matter and the structure of the atom. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *21*(125), 669–688.

Samsung. (o.D.). *Technisches Datenblatt Samsung Galaxy S3 NEO* [letzter Zugriff 23.09.2016]. http://%20www.samsung.com/%20de/consumer/mobiledevices/smartphones/galaxy-s/%20GT-I9301OKZDBT

Schaadt, T. (2014). Roller Coaster Bahn - eine physikalische und didaktische Analyse [Unveröffentlichte Abschlussarbeit an der Christian-Albrechts-Universität zu Kiel].

Schecker, H. (2014). Überprüfung der Konsistenz von Itemgruppen mit Cronbachs α. *Methoden in der naturwissenschaftsdidaktischen Forschung*. Springer-Verlag, Springer Spektrum.

Schecker, H. & Ralle, B. (2009). Naturwissenschaftsdidaktik und Lehrerbildung-Chancen und Risiken aktueller Entwicklungen. *PhyDid A-Physik und Didaktik in Schule und Hochschule*, *2*(9), 73–83.

Scheirer, C. J., Ray, W. S. & Hare, N. (1976). The analysis of ranked data derived from completely randomized factorial designs. *Biometrics*, 429–434.

Schiefele, U. (2004). Förderung von Interessen. In M. Grünke & J. Brunstein (Hrsg.), *Interventionen bei Lernstörungen - Förderung, Training und Therapie in der Praxis.* (S. 134–144). Göttingen: Hogrefe.

Schiefele, U. (2008). Lernmotivation und Interesse. *Handbuch der pädagogischen Psychologie*, *10*, 38–49.

Schiefele, U. (2009). Situational and individual interest. *Handbook of motivation at school*, 197–222.

Schiefele, U., Krapp, A. & Schreyer, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, *10*(2), 120–148.

Schiefele, U., Krapp, A., Wild, K. P. & Winteler, A. (1993). Der Fragebogen zum Studieninteresse (FSI). *Diagnostica*, *39*(4), 335–351.

Schiefele, U., Moschner, B. & Husstegge, R. (2002). Skalenhandbuch SMILE-Projekt. *Bielefeld, Universität Bielefeld, Abteilung für Psychologie, Unveröffentlichtes Manuskript*.

Schiefele, U. & Schaffner, E. (2015). Motivation. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 154–175). Springer-Verlag.

Schiefele, U., Streblow, L., Ermgassen, U. & Moschner, B. (2003). Lernmotivation und Lernstrategien als Bedingungen der Studienleistung. Ergebnisse einer Längsschnittstudie. *Zeitschrift für pädagogische Psychologie*, *17*(3/4), 185–198.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, *8*(4), 350.

Schmitz, G. S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology*.

Schneider, M. & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological bulletin*, *143*(6), 565.

Schreiber, N. (2012). *Diagnostik experimenteller Kompetenz: Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*. Logos Verlag.

Schreiber, N. & Theyßen, H. (2019). Selbstbeurteilungen zur formativen Diagnostik experimenteller Performanzen–Was zeichnet genau urteilende Schülerinnen und Schüler aus? *Zeitschrift für Didaktik der Naturwissenschaften*, *25*(1), 43–61.

Schreiber, N., Theyßen, H. & Dickmann, M. (2016). Wie genau beurteilen Schülerinnen und Schüler ihre eigenen experimentellen Fähigkeiten?–Ein Ansatz zur praktikablen Diagnostik experimenteller Fähigkeiten im Unterrichtsalltag–. *PhyDid A-Physik und Didaktik in Schule und Hochschule*, *1*(15), 49–63.

Schreiber, N., Theyßen, H. & Schecker, H. (2009). Experimentelle Kompetenz messen?! *PhyDid A-Physik und Didaktik in Schule und Hochschule*, *3*(8), 092–101.

Schreiber, N., Theyßen, H. & Schecker, H. (2014). Diagnostik experimenteller Kompetenz: Kann man Realexperimente durch Simulationen ersetzen? *Zeitschrift für Didaktik der Naturwissenschaften*, *20*(1), 161–173.

Schroedter, S. & Körner, H. (2011). Entwicklung eines Fragebogens zur Selbstwirksamkeitserwartung beim Experimentieren (SWE_EX). *Konzepte fachdidaktischer Strukturierung für den Unterricht: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg*, 164–166.

Schuster, K. (2018). *Entwicklung eines Praktikumsversuchs zur Videoanalyse von beschleunigten Bewegungen.* Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Schwarz, I., Effertz, C. & Heinke, H. (2013a). Entwicklung eines Physikpraktikums für Biologiestudierende-der Umgang mit Messunsicherheiten. *PhyDid B-Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung.*

Schwarz, I., Effertz, C. & Heinke, H. (2013b). Untersuchungen zur Nachbereitungsform „Protokollïm Physikpraktikum für Biologiestudierende. *GDCP-Jahrestagung Hannover*, 686–688.

Schwarzer, R. & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. *Zeitschrift für Pädagogik, Beiheft*, *44*, 28–53.

Shavelson, R. J., Baxter, G. P. & Gao, X. (1993). Sampling variability of performance assessments. *Journal of educational Measurement*, *30*(3), 215–232.

Shavelson, R. J., Ruiz-Primo, M. A. & Wiley, E. W. (1999). Note on sources of sampling variability in science performance assessments. *Journal of Educational Measurement*, *36*(1), 61–71.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, *15*(2), 4–14.

Siebert, M. (2016). *Entwicklung einer Experimentierstation zum Thema Thermographie*. Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Simon, K. (2015). *Fehleranalyse und Optimierung von Versuchen zur Bestimmung des Planckschen Wirkungsquantums*. Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Snijders, T. A. & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

SonicVisualiser. (o.D.). *Sonic Visualiser – Visualisation, analysis, and annotation of music audio recordings* [letzter Zugriff am 03.02.2020]. https://www.sonicvisualiser.org/

Sorge, S., Keller, M., Neumann, K. & Möller, J. (2019). Investigating the relationship between pre-service physics teachers' professional knowledge, self-concept, and interest. *Journal of Research in Science Teaching*.

Sorge, S., Kröger, J., Petersen, S. & Neumann, K. (2019). Structure and development of pre-service physics teachers' professional knowledge. *International Journal of Science Education*, *41*(7), 862–889.

Sorge, S., Petersen, S. & Neumann, K. (2016). Die Bedeutung der Studierfähigkeit für den Studienerfolg im 1. Semester in Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, *22*(1), 165–180.

Staacks, S. (o.D.). *phyphox - physical phone experiments* [Webseite der Entwickler*innen der RWTH Aachen, letzter Zugriff am 03.02.2020]]. https://phyphox.org/de/home-de/

Staacks, S., Hütz, S., Heinke, H. & Stampfer, C. (2018). Advanced tools for smartphone-based experiments: phyphox. *Physics Education*, *53*(4), 045009.

Stefani, L. A. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, *19*(1), 69–75.

Straube, P. (2016). *Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung bei (Lehramts-) Studierenden im Fach Physik* (Bd. 209). Logos Verlag Berlin GmbH.

Stuckey, M., Hofstein, A., Mamlok-Naaman, R. & Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. *Studies in Science Education*, *49*(1), 1–34.

Supermagnete. (o.D. a). *Riesenmagnet* [letzter Zugriff am 21.05.2016]. https://www.supermagnete.de/Magnetanwendungen/Riesenmagnet-Smash-das-Making-Of

Tesch, M. & Duit, R. (2004). Experimentieren im Physikunterricht–Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, *10*(10), 51–69.

Theyßen, H. (2000). *Ein Physikpraktikum für Studierende der Medizin: Darstellung der Entwicklung und Evaluation eines adressatenspezifischen Praktikums nach dem Modell der didaktischen Rekonstruktion.* Logos-Verlag.

Theyßen, H. (2005). Didaktische Rekonstruktion eines Physikpraktikums für Medizinstudierende. *Zeitschrift für Didaktik der Naturwissenschaften*, *11*, 57–27.

Theyßen, H. (2009). Mythos Bierschaumzerfall-Ein Analogon für den radioaktiven Zerfall?-. *PhyDid A-Physik und Didaktik in Schule und Hochschule*, *2*(9), 49–57.

Toothacker, W. (1983). A critical look at introductory laboratory instruction. *American Journal of Physics*, *51*(6), 516–520.

Torchiano, M. (2020). *effsize: Efficient Effect Size Computation* [R package version 0.7.8]. https://CRAN.R-project.org/package=effsize

Tschannen-Moran, M., Hoy, A. W. & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of educational research*, *68*(2), 202–248.

Tukey, J. W. (1959). A quick compact two sample test to Duckworth's specifications. *Technometrics*, *1*(1), 31–48.

Tukey, J. W. (1977). *Exploratory data analysis* (Bd. 2). Reading, Mass.

Urban, D. & Mayerl, J. (2014). *Strukturgleichungsmodellierung: ein Ratgeber für die Praxis.* Springer-Verlag.

Van Oers, B. (1998). From context to contextualizing. *Learning and instruction*, *8*(6), 473–488.

van Breukelen, G. J. (2013). ANCOVA versus CHANGE from baseline in nonrandomized studies: The difference. *Multivariate Behavioral Research*, *48*(6), 895–922.

van Vorst, H., Dorschu, A., Fechner, S., Kauertz, A., Krabbe, H. & Sumfleth, E. (2015). Charakterisierung und Strukturierung von Kontexten im naturwis-

senschaftlichen Unterricht–Vorschlag einer theoretischen Modellierung. *Zeitschrift für Didaktik der Naturwissenschaften*, *21*(1), 29–39.

Vieyra, C. & Vieyra, R. (o.D.). *Physics Toolbox Apps* [letzter Zugriff am 03.02.2020]. https://www.vieyrasoftware.net/

Vieyra, R., Vieyra, C., Jeanjacquot, P., Marti, A. & Monteiro, M. (2015). Turn your smartphone into a science laboratory. *The science teacher*, *82*(9), 32.

Vogt, P. & Kuhn, J. (2014). Acceleration Sensors of Smartphones - Possibilities and Examples of Experiments for Application in Physics Lessons. *Frontiers in Sensors*, *2*, 1–9.

Weber, A. (2004). Problem-based learning. *Im Spannungsfeld von Schule und Gesellschaft*, 200435.

Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen-eine umstrittene Selbstverständlichkeit. *Leistungsmessungen in schulen* (S. 17–32). Beltz.

Weißbach, J. (2013). *Entwicklung eines Lehrmoduls zur geometrischen Optik am menschlichen Auge*. Unveröffentlichte Examensarbeit an der Christian-Albrechts-Universität zu Kiel.

Wellnitz, N., Hecht, M., Heitmann, P., Kauertz, A., Mayer, J., Sumfleth, E. & Walpuski, M. (2017). Modellierung des Kompetenzteilbereichs naturwissenschaftliche Untersuchungen. *Zeitschrift für Erziehungswissenschaft*, *20*(4), 556–584.

Welzel, M., Haller, K., Bandiera, M., Hammelev, D., Koumaras, P., Niedderer, H., Paulsen, A., Robinault, K. & von Aufschnaiter, S. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden–Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, *4*(1), 29–44.

Weß, R., Priemer, B., Weusmann, B., Ludwig, T., Sorge, S. & Neumann, I. (2020). Der Verlauf von lehrbezogenen Selbstwirksamkeitserwartungen angehender MINT-Lehrkräfte im Studium. *Zeitschrift für Pädagogische Psychologie*.

Westphal, W. (1937). Die physikalischen Übungen an der Technischen Hochschule Berlin. *Z. Für Den Phys. Chem. Unterr., Bd*, *4*, 147.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wild, E. & Möller, J. (2015). *Pädagogische Psychologie*. Springer-Verlag.

Wild, T. C., Enzle, M. E. & Hawkins, W. L. (1992). Effects of perceived extrinsic versus intrinsic teacher motivation on student reactions to skill acquisition. *Personality and social psychology bulletin*, *18*(2), 245–251.

Woitkowski, D. (2017). Studieneingangsprofile in Fach-und Lehramts-Studiengängen Physik Eine kontrastierende Analyse auf Basis eines Kompetenzstrukturmodells für Fach-Physiker. *PhyDid A-Physik und Didaktik in Schule und Hochschule*, *1*(16), 43–56.

Wolff-Riechert, T. (2017). *Quantitative Videoanalyse der Bewegung von Permanentmagneten auf einem Luftkissentisch.* Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Wyrwich, T. (2017). *Entwicklung eines Praktikumsversuchs zur Absorptionsspektroskopie.* Unveröffentlichte Bachelorarbeit an der Christian-Albrechts-Universität zu Kiel.

Zell, E. & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125.

A Anhang

A.1 Kurzbeschreibungen der Praktikumsversuche

Dieser Anhang gibt eine Übersicht über alle entwickelten Praktikumsversuche in Form von Kurzbeschreibungen. Die vollständigen Versuchsbeschreibungen in der Version aus dem Beforschungszeitraum (Sommersemester 2016-2019) sind im digitalen Anhang zu finden. In den Kurzbeschreibungen werden Aufbau und Aufgaben der Praktikumsversuche skizziert, um die wichtigsten Punkte in Bezug auf Versuchsziele, Durchführung und Auswertung deutlich zu machen. Anschließend werden die Herausforderungen und der Schwierigkeitsgrad dargestellt, die ein wichtiger Eckpunkt zur inhaltlichen Schwerpunktlegung in der Entwicklung der Praktikumsversuche waren (siehe auch Anhang A.1.1). Die Stichworte zur Relevanz des Versuchsthemas für die Schule und die Einordnung des Themas in die schleswig-holsteinischen Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019), sind insbesondere zum Verständnis der lehramtsspezifischen Ausrichtung (Abschnitt 6) der Versuche interessant. Zur Nachvollziehbarkeit der Herstellung des Schulbezuges und entsprechenden Untersuchungen dazu in Bezug auf die wahrgenommene Relevanz (Abschnitt 9.3) sind am Ende der Kurzbeschreibung die verschiedenen schulbezogenen und klassischen Einleitungstexte und die fachdidaktischen Reflexionsfragen angegeben.

A.1.1 Schwierigkeitsgrad

Um den Schwierigkeitsgrad der Praktikumsversuche auf einem universitären Level zu halten, wird in den entwickelten Versuchen, wie im Abschnitt 6.3 beschrieben, vertieftes Hintergrundwissen über den Schulstoff hinaus vermittelt. Damit der Gesamtschwierigkeitsgrad der verschiedenen Versuche etwa gleich groß und auf Hochschulniveau ist, wurde bei der Konzeption der Versuche ein Punktesystem eingeführt, das den Schwierigkeitsgrad zwischen „leicht“ (1 Punkt), „mittel“ (2 Punkte) und „schwer“ (3 Punkte) für die sechs Kategorien Theoretische Grundidee, Messaufbau, Messtechnik, Versuchsdurchführung, Auswertung und Fehlerrechnung bewertet. Der Gesamtschwierigkeitsgrad als Summe aller Punkte hat für die Versuche im Mittel die Punktzahl 12 und unter- bzw. überschreitet nicht die 10 bzw. 14 Punkte. Liegt daher bei einem Versuch, wie zum Beispiel beim Versuch Auf- und Entladung Kondensator, der Schwerpunkt auf messtechnischen Aspekten mit dem Schwierigkeitsgrad schwer (3 Punkte), dann wird die Schwierigkeit in einer anderen Kategorie wie theoretische Grundlagen niedrig gehalten (leicht, 1 Punkt). Dadurch soll sicher gestellt werden, dass die Studierenden ihren Fokus auf den Lernschwerpunkt ausrichten können, während andere Bereiche wie im Beispiel die theoretischen Grundlagen beim Lernen eine untergeordnete Rolle spielen. Diese Einschätzungen der Schwierigkeitsgrade von Seiten der Versuchsentwicklung für die Versuche sind in den Tabellen der nachfolgenden Kurzbeschreibungen dokumentiert.

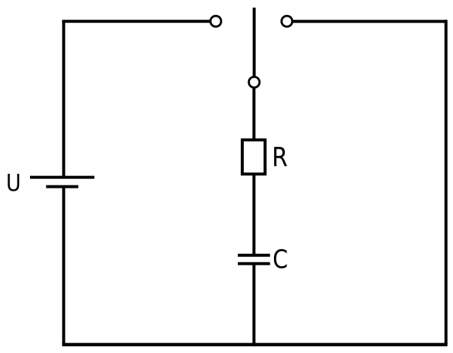
A.1.2 Auf- und Entladung Kondensator

Praktikumsteil:	Wintersemester (WS) und Sommersemester (SS)
Themenbereich:	Messtechnik
Verweis:	Die Umsetzung des Versuches mit CASSY ist inspiriert durch LD-Didactic (2014).

Versuchsziele und Aufgaben

- Aufbau einer Schaltung zur Auf- und Entladung eines Kondensators über einen Widerstand (Abb. 57)
- Messung der Spannungsverläufe am Kondensator und der Ströme während des Auf- und Entladevorganges mit dem Messsystem CASSY (LD-Didactic, o.D. a)
- Bestimmung der Zeitkonstante unter Nutzung verschiedener Auswertungsfunktionen der CASSY-Software an den vier erhaltenen Kurven (Exponentielle Kurvenanpassung, freie Anpassung, lineare Anpassung nach Logarithmierung der Werte, direktes Ablesen aus dem Graphen)
- Bestimmung der zugeführten Ladung aus dem Stromverlauf und Berechnung der Kapazität des Kondensator
- Bestimmung des Energiegehaltes des Kondensators aus der Leistung

a) b)



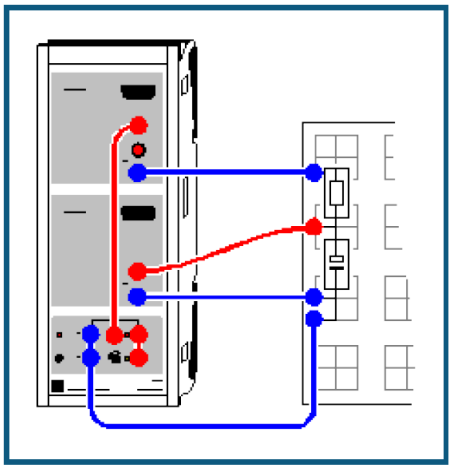


Abbildung 57: *Aufbau der Schaltung zur Auf- und Entladung eines Kondensators über einen Widerstand. a) Schaltplan (Abb. erstellt von C. Brüning), b) Aufbau mit dem CASSY Interface als Spannungsquelle, als Relais und zur Messung von Strom und Spannung (Abb. aus dem CASSY Lab Handbuch von LD-Didactic (2014), Copyright: © LEYBOLD / LD DIDACTIC GmbH / www.ld-didactic.de, Hürth)*

.

Herausforderungen

- Umsetzung der Schaltung mit Integration der Spannungsquelle und des Relais des CASSY Interfaces
- Auswahl eines geeigneten Kondensators
- Einarbeitung in die Messung von Strom und Spannung mit CASSY
- Auswahl einer sinnvollen Messzeit und eines geeigneten Abtastintervalls
- Sinnvolle Darstellung der Daten und Nutzung der verschiedenen Auswertungsfunktionen in der Software CASSY Lab (z. B. lineare, exponentielle und freie Kurvenanpassung, Integration, Rechnerfunktion)
- Bedeutung der Zeitkonstante

Schwierigkeitsgrad		
Theoretische Grundidee	leicht	1
Messaufbau	mittel	2
Messtechnik	schwer	3
Versuchsdurchführung	leicht	1
Auswertung	schwer	3
Fehlerrechnung	leicht	1
Summe		11

Tabelle 36: *Schwierigkeitsgrad des Versuches „Auf- und Entladung Kondensator“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Einführung in die Nutzung des CASSY Messsystems, welches in Schulen oft Anwendung findet
- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe II (Auf- und Entladungsvorgänge eines Kondensators)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiet	Felder - Homogenes elektrisches Feld

Schulbezogene Einleitung

„Computergestütztes Experimentieren ist ein wichtiger Teil in der physikalischen Ausbildung geworden. Hersteller von Lehrmitteln haben für schulische Zwecke Messwerterfassungssysteme mit zugehöriger Auswertesoftware entwickelt. Dazu zählen beispielsweise CASSY von LD-Didactic, PASPORT von Conatex und COBRA von

Phywe. Das hier verwendete Messsystem CASSY ist ein modulares System aus einer Vielzahl an verschiedenen Sensoren sowie Interfaces und zugehöriger Software.
Messsysteme im Unterricht decken nicht nur motivationale Aspekte für die Schülerinnen und Schüler ab, sondern sind vor allem auch eine fachlich sinnvolle Ergänzung. Stellen Sie sich vor, Sie möchten einen Versuch zum Auf- und Entladevorgang eines Kondensators im Unterricht durchführen. Bei diesem Versuch wird im Zeitbereich von Millisekunden gearbeitet. Zeitskalen unterhalb von mehreren Sekunden sind für Schülerinnen und Schüler mit den Sinnen nicht mehr erfassbar, sodass herkömmliche Zeigerinstrumente zur Demonstration nicht in Frage kommen. Um diesem Problem entsprechend zu begegnen, entscheiden Sie sich deshalb mit dem CASSY-Messsystem zu arbeiten. Das System bietet die Möglichkeit, die Messdaten für Schülerinnen und Schüler transparent aufzunehmen, entsprechend aufzulösen und auszuwerten.
Informationen zum CASSY Messsystem finden Sie im zugehörigen Handbuch (LD-Didactic, 2014), das auf OLAT hochgeladen wurde. Darin sind zudem auch Anleitungen für Versuche mit CASSY bereitgestellt. Unter anderem finden Sie dort verschiedene Varianten von Versuchen zur Auf- und Entladung eines Kondensators (LD-Didactic, 2014, ab S. 243).“

Fachdidaktische Reflexionsfragen

- Würden Sie den Versuch „Auf- und Entladung eines Kondensators mit CASSY“ als Demonstrationsversuch oder als Gruppenarbeit in einer Oberstufe durchführen? Bitte begründen Sie.
- Sie setzen in Ihrer Klasse CASSY ein, um den Schülerinnen und Schülern das Aufladen und Entladen eines Kondensators zu vermitteln. Welche typischen Probleme können die von den Schülerinnen und Schülern erstellten Diagramme aufweisen?

A.1.3 Oszillierendes Smartphone

Praktikumsteil:	Wintersemester (WS) und Sommersemester (SS)
Themenbereich:	Messtechnik
Verweis:	Der Versuch ist inspiriert durch die Smartphone Versuche von Vogt und Kuhn (2014). Von diesem Versuch existiert eine weiterentwickelte Version mit Nutzung der App *phyphox* (Staacks et al., 2018; Staacks, o.D.) anhand der Bachelorarbeit von Poeppel (2019).

Versuchsziele und Aufgaben

- Untersuchung der Messgenauigkeit der Beschleunigungssensoren des Smartphones bei Ruhelage mit der App *Physics Toolbox Suite* (R. Vieyra et al., 2015; C. Vieyra & Vieyra, o.D.)
- Federpendel: Das Smartphone wird in einer 3D-gedruckten Experimentierhülle an einer Feder aufgehängt und kann mit unterschiedlichen Gewichten beschwert werden (Abb. 58).
  - Nutzung des Smartphones zur Messung der Beschleunigung während der Federschwingung
  - Diskussion der gemessenen Kurve in Bezug auf die gemessenen Beschleunigungen und die Auslenkung des Federpendels
  - Bestimmung der Periodendauer der Schwingung anhand der Beschleunigungsmessung für unterschiedliche Zusatzmassen
  - Bestimmung der Masse des Smartphones und der Federkonstante der Feder aus den vorangegangenen Messdaten
- Fadenpendel: Das Smartphone wird in einer Experimentierhülle an zwei v-förmig angeordnete Fäden aufgehängt.
  - Diskussion der gemessenen Kurve in Bezug auf die gemessenen Beschleunigungen und die Auslenkung des Fadenpendels
  - Untersuchung des Fadenpendels als nicht-harmonischen Oszillator: Aufnahme einer Schwingung mit hohem anfänglichem Auslenkwinkel, der durch die Dämpfung über die Zeit abnimmt, und Vergleich der Periodendauer über die Zeit
  - Ordnung von drei Metallplatten nach ihrer elektrischen Leitfähigkeit, indem das Smartphone mit einem Magneten am unteren Ende der Smartphonehülle über die Platten schwingt und die Dämpfung durch die Wirbelstrombremse über den unterschiedlichen Platten verglichen wird

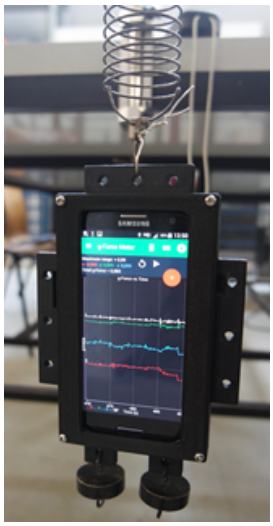


Abbildung 58: *Das Smartphone wird von einer 3D-gedruckten Experimentierhülle geschützt und kann durch die verschieden Befestigungsmöglichkeiten an der Hülle an eine Feder gehängt werden und mit Gewichten beschwert werden (Photo von J. Haacks).*

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	mittel	2
Messtechnik	schwer	3
Versuchsdurchführung	mittel	2
Auswertung	schwer	3
Fehlerrechnung	mittel	2
Summe		14

Tabelle 37: *Schwierigkeitsgrad des Versuches „Oszillierendes Smartphone“ (Erläuterungen in Abschnitt A.1.1)*

Herausforderungen

- Grobes Verständnis der Funktionsweise von Beschleunigungssensoren
- Einarbeitung in die Nutzung der Smartphone App *Physics Toolbox Suite*
- Einarbeitung in Software CASSY Lab (LD-Didactic, o.D. a) zur Darstellung und Auswertung der aufgenommenen Beschleunigungsdaten (ggf. Kurvenanpassung)
- Einarbeitung in Microsoft Excel
- Überführung der Daten der App in CASSY durch Anpassung des Datenformates
- Initialisierung von Schwingungen entlang einer Achse/Ebene ohne Verkippungen oder Drehungen
- Diskussion der gemessenen Beschleunigungen während der Schwingung
- Bestimmung der Messe des Smartphones beim Federpendel mithilfe linearer Regression und entsprechender Fehlerrechnung sowie Berücksichtigung der Federmasse

- Herangehensweise an die vergleichsweise offene Aufgabe, das Fadenpendel mit den gegebenen Materialien und vorangegangenen Erkenntnissen als nicht-harmonischen Oszillator zu veranschaulichen
- Erklärung und Nutzung des Prinzips der Wirbelstrombremse bei gegebener offener Aufgabenstellung
- Abschätzung der Möglichkeiten und Grenzen des Smartphones als Messgerät

Relevanz des Versuchsthemas für die Schule

- Diskussion der Möglichkeiten und Grenzen des Alltagsmediums Smartphone als Messgerät im Schulunterricht
- Erlernen des Umgangs mit der kostenfreien, schulrelevanten *App Physics Toolbox Suite* (oder *phyphox*) zum Auslesen und Analysieren der Daten der Smartphone Sensoren
- Vertiefung der schultypischen Versuche zum Feder- und Fadenpendel in der Sekundarstufe II

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiete	Wellen - Schwingungen und Wellen
	Felder - Elektrodynamik

Schulbezogene Einleitung

„Neben den diversen negativen Aspekten, die ein Handy mit in den Unterricht bringen kann, können durch konstruktive Anwendung des Smartphones positive Effekte wie Motivation erwirkt werden. Das Smartphone kann im Unterricht nicht nur zu Internetrecherchen und zur Dokumentation genutzt werden, sondern im Physikunterricht durch die Vielzahl an integrierten Sensoren auch als Experimentierwerkzeug angewendet werden. Zu den Sensoren zählen beispielsweise Beschleunigungssensoren, GPS-Empfänger oder ein Magnetometer. Diverse Apps ermöglichen es, diese Sensoren auszulesen, die Daten zu veranschaulichen und zu exportieren. Durch das vertraute Smartphone im Physikunterricht können die Schülerinnen und Schüler dazu motiviert und befähigt werden, auch im Alltag physikalische Phänomene zu untersuchen. In diesem Versuch sollen diese Möglichkeiten des Smartphones exemplarisch an Beschleunigungsmessungen bei verschiedenen Schwingungen aufgezeigt werden. Dieser Versuch wird mit einem Android-Smartphone und der freien App „Physics Toolbox Suite“ durchgeführt. Ähnliche Apps sind auch für andere Betriebssysteme verfügbar. Die Auswertung der Daten erfolgt mit der Software CASSY Lab 2. Diese Software gehört zu einem Messwerterfassungssystem, das von der Firma LD-Didactic für schulische Zwecke entwickelt wurde. Das Messsystem CASSY, das Sie in weiteren Versuchen kennenlernen werden, ist ein modulares System aus einer Vielzahl an verschiedenen Sensoren sowie Interfaces.“

Fachdidaktische Reflexionsfragen

- Welche Bedeutung hat ein derartiger Versuch mit einem Smartphone für die physikalische Ausbildung der Schülerinnen und Schüler?
- Welche Schwierigkeiten könnten im Physikunterricht auftreten, wenn die Schülerinnen und Schüler mit Ihren Smartphones experimentieren? Wie könnte man diesen begegnen?

A.1.4 Verformung eines Balles

Praktikumsteil: Wintersemester (WS) und Sommersemester (SS)
Themenbereich: Messtechnik

Versuchsziele und Aufgaben

- Analyse der Bewegung eines fallenden und zurückspringenden Balles mithilfe einer Hochgeschwindigkeitskamera und der Videoanalyse-Software Tracker (D. Brown & Cox, 2009; D. Brown, 2019)
- Bestimmung der Federkonstante mithilfe einer Ballpresse anhand des Hookeschen Gesetzes
- Abschätzung der Spannenergie aus der bestimmten Federkonstante und der maximalen Stauchung des Balles beim Aufprall (Abb. 59)
- Quantitative Energiebetrachtungen des Balles

Abbildung 59: *Photo des gestauchten Balles beim Aufprall auf dem Boden.*

Herausforderungen

- Einarbeitung in die Videoanalyse-Software Tracker
- Auswahl der optimalen Kameraeinstellungen (Belichtungszeit und Bildausschnitt) zur Maximierung der Bildrate
- Optimierung des Fallenlassens des Balles für einen möglichst senkrechten Fall im Bildausschnitt der Kamera
- Bestimmung der maximalen Stauchung des Balles beim Aufprall

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	leicht	1
Messtechnik	schwer	3
Versuchsdurchführung	mittel	2
Auswertung	schwer	3
Fehlerrechnung	mittel	2
Summe		13

Tabelle 38: *Schwierigkeitsgrad des Versuches „Verformung eines Balles“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Einarbeitung in die schultypische, kostenfreie Videoanalyse-Software Tracker
- Vertiefung eines Versuches zu schulischen Lerninhalten der Mittelstufe (Freier Fall, Hookesches Gesetz)
- Vertiefung des Basiskonzeptes Energie

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I und II
Sachgebiete	Energie - qualitativer und quantitativer Energiebegriff
	Mechanik - Statische Kräfte
	Mechanik - Kinematik
Basiskonzept	Energie

Schulbezogene Einleitung

„Häufig wird ein fallender und wieder hoch springender Ball im Physikunterricht eingesetzt, um mit alltagsnahem Bezug Energieumwandlungen und Energieerhaltung zu thematisieren. Dabei wird meist auf die Energieerhaltung fokussiert, indem die Rücksprunghöhe des Balles untersucht wird. Der Energieumwandlung von kinetischer in elastische Energie wird dabei relativ wenig Aufmerksamkeit gewidmet. Die elastische Verformung des aufprallenden Balles wird den Schülerinnen und Schülern als „Energiespeicher“ veranschaulicht. Mithilfe des Hookschen Gesetzes kann die Federkonstante des Balles und damit die Spannenergie näherungsweise bestimmt werden. Der Aufprall, der auf Zeitskalen von Sekundenbruchteilen stattfindet, lässt sich mithilfe einer Hochgeschwindigkeitskamera und Videoanalysesoftware untersuchen. Das hier eingesetzte Programm Tracker eignet sich für den Schulunterricht insbesondere dadurch, dass es kostenfrei zugänglich ist. Alternativ wird auch oft die freie Software VIANA eingesetzt.“

Fachdidaktische Reflexionsfragen

- Ein Iphone ermöglicht das Aufnehmen von bis zu 240 Bilder pro Sekunde. Diskutieren Sie die Vor- und Nachteile, ein Iphone für die Analyse eines fallenden Balles im Schulunterricht einzusetzen.
- Sie möchten im Unterricht den freien Fall mit unterschiedlichen Fallhöhe untersuchen lassen. Wie lässt sich dies bei einem begrenzten Zeitraum didaktisch sinnvoll umsetzen?

A.1.5 Hören

Praktikumsteil: Wintersemester (WS) und Sommersemester (SS)
Themenbereich: Messtechnik

Versuchsziele und Aufgaben

- Untersuchung der oberen und unteren Hörgrenze (Versuchsmaterial: Kopfhörer ein Schalldruckpegelmessgerät, Software: Audacity (Audacity, o.D.) und Sonic Visualizer (Cannam et al., 2010; SonicVisualiser, o.D.))
- Identifikation des Frequenzbereiches, in dem die Hörempfindlichkeit maximal ist
- Vergleich von zwei verschiedenen Kopfhörern bzgl. der unterschiedlichen Übertragungsfunktionen
- Untersuchung der spektralen Maskierung anhand zweier Töne ähnlicher Frequenz mit variierender Lautstärke
- Fourieranalyse von zwei Audiodateien (.wav und .mp3) und qualitativer Vergleich der Spektren und Spektrogramme in Hinblick auf die MP3-Komprimierung

Herausforderungen

- Einarbeitung in die Software Audacity (Toneditor) und Sonic Visualizer (Tonanalyse)
- Theorie zu Hörkurven, akustischen Maskierungseffekten als Grundlage der MP3-Komprimierung und Fourier-Analyse
- Geeignete Parametereinstellung für die Fourier-Analyse
- Identifikation der MP3-Komprimierung in den Spektrogrammen

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	leicht	1
Messtechnik	schwer	3
Versuchsdurchführung	leicht	1
Auswertung	mittel	2
Fehlerrechnung	leicht	1
Summe		11

Tabelle 39: *Schwierigkeitsgrad des Versuches „Hören“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Einarbeitung in die kostenfreien, schulrelevanten Software Audacity und Sonic Visualizer, die zur Erzeugung von Tönen und zur Tonanalyse in den Unterricht eingebunden werden kann
- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe II (Akustische Wellen, Spektren)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiete	Wellen - Wellen und Schwingungen
	Wellen - Spektren

Schulbezogene Einleitung

„Im Schulunterricht werden akustische Themen trotz ihres hohen Alltagsbezugs nicht immer ausreichend behandelt. Zum einen war in der Vergangenheit das Fehlen geeigneter Messinstrumente ein zentrales Problem der Akustik im Schulunterricht. Zum anderen besteht eine Schwierigkeit im unterschiedlich ausgeprägten Hörsinn der Schülerinnen und Schüler. Nicht jede Schülerin und jeder Schüler ist in der Lage, Veränderungen in Tonhöhe oder Lautstärke gleichermaßen wahrzunehmen.
Diesem Problem kann mit Audioanalysesoftware begegnet werden. Der Audioeditor Audacity und die Tonanalyse-Software Sonic Visualizer eignen sich als kostenfreie Programme besonders für den Einsatz im Schulunterricht.
Die Nutzung von Software im Unterricht bringt oft mit sich, dass die Schülerinnen und Schüler Schwierigkeiten haben, geeignete Einstellungen zu finden. Falsche Einstellungen können dazu führen, dass Schülerinnen und Schüler nicht erkennen können, worum es im Versuch geht, oder sogar falsche Vorstellungen entwickeln. Deshalb ist es wichtig, dass die Lehrkraft das Programm so gut wie möglich kennt.
Die Funktion beider Programme soll in diesem Versuch anhand von Audiokompression am Beispiel der MP3-Datei, die vermutlich jeden Schüler durch seinen Alltag begleitet, kennengelernt werden. Zum einen beleuchtet der Versuch dazu die erwähnte Hörwahrnehmung. Zum anderen wird auch der Effekt der Frequenz-Maskierung einbezogen, bei dem die Hörschwelle durch Erklingen eines Tones um dessen Frequenz herum angehoben wird. Obwohl die Fourieranalyse kein reguläres Schulthema ist, ist sie ein zentrale Methode zur Analyse von Schwingungen und Wellen und wird daher ebenfalls ein wesentlicher Inhalt des Versuches sein.“

Fachdidaktische Reflexionsfragen

- Wie könnten Schülerinnen und Schüler in die Software eingeführt werden? Welche Einstellungen sind dabei besonders relevant?
- Welche fehlerhaften Vorstellungen haben Schüler über mechanische Wellen wie Schallwellen?

A.1.6 Roller Coaster

Praktikumsteil:	Wintersemester (WS) und Sommersemester (SS)
Themenbereich:	Messtechnik
Verweis:	Vorarbeiten zur Entwicklung des Versuches wurden im Rahmen der Abschlussarbeiten von Schaadt (2014) undSchuster (2018) durchgeführt.

Versuchsziele und Aufgaben

- Analyse der Bewegung eines Wagens auf einer Roller Coaster Bahn mit einer schiefen Ebene mithilfe einer Smartphonekamera und der Videoanalyse-Software Tracker (D. Brown & Cox, 2009; D. Brown, 2019)
- Bestimmung der Widerstandsbeiwerte eines Wagens mit verschiedenen Profilen mittels Kurvenanpassung des Geschwindigkeitsverlaufes
- Ordnung zweier Metallplatten nach ihrer elektrischen Leitfähigkeit anhand der Abbremsung eines vorbeifahrenden Wagens mit einem befestigten Magneten (Abb. 60)

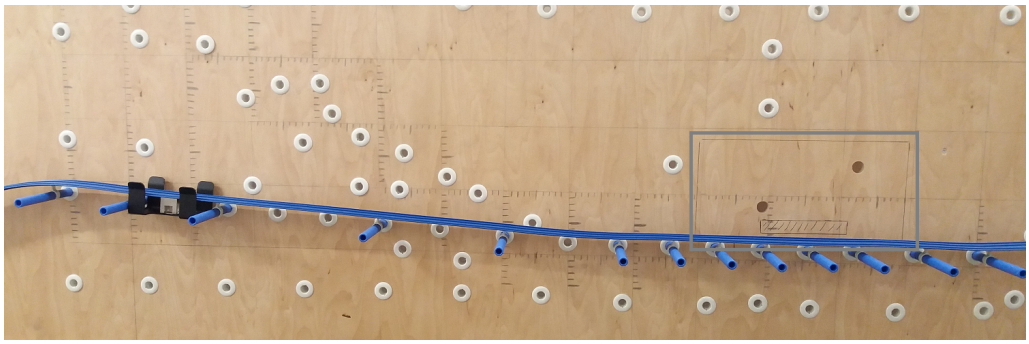


Abbildung 60: *Aufbau des Versuches Roller Coaster an einer Holzwand mit verschiedene Bohrungen für unterschiedliche Bahnen und einer Markierung für die Position der Metallplatten (Photo von Schuster, 2018).*

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	leicht	1
Messtechnik	schwer	3
Versuchsdurchführung	leicht	1
Auswertung	schwer	3
Fehlerrechnung	mittel	2
Summe		12

Tabelle 40: *Schwierigkeitsgrad des Versuches „Roller Coaster“ (Erläuterungen in Abschnitt A.1.1)*

Herausforderungen

- Einarbeitung in die Videoanalyse-Software Tracker
- Kurvenanpassung am gemessenen Geschwindigkeitsverlaufes in Tracker zur Bestimmung des Widerstandsbeiwertes
- Auswahl der optimalen Kameraeinstellungen am eigenen Smartphone
- Erklärung und Nutzung des Prinzips der Wirbelstrombremse

Relevanz des Versuchsthemas für die Schule

- Einarbeitung in die schultypische, kostenfreie Videoanalyse-Software Tracker
- Vertiefung eines Versuches zu schulischen Lerninhalten der Mittelstufe (Geschwindigkeit, Beschleunigte Bewegung, Reibung)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I und II
Sachgebiet	Mechanik - Beschleunigte Bewegungen
	Mechanik - Dynamik
	Felder - Elektrodynamik

Schulbezogene Einleitung

„Die Kinematik und die Integration dieser in die Dynamik sind wesentliche Elemente des Physikunterrichts. Vielen Schülerinnen und Schülern fällt es jedoch schwer Diagramme der Geschwindigkeit oder Beschleunigung über die Zeit mit der zugehörigen Bewegung zusammenzubringen. Die Analyse von Bewegungen anhand von Videomaterial kann die fehlende Brücke schlagen und ist inzwischen auch Teil der Fachanforderungen. Schülerinnen und Schülern können die Geschwindigkeits- oder Beschleunigungsgraphen der Bewegung mithilfe eines Videos selbst ermitteln, indem sie in den Einzelbildern des Videos die Position des bewegten Objekts markieren. Diese Form der Bewegungsanalyse hat zum Beispiel den Vorteil, dass Momentaufnahmen der Bewegung und die zugehörige Position im resultierenden Geschwindigkeits- oder Beschleunigungsgraphen gleichzeitig betrachtet und gemeinsam interpretiert werden können. Die Software Tracker beispielsweise eignet sich für die Analyse von Videos im Schulunterricht besondere dadurch, dass viele nützliche Funktionen zur Bewegungsanalyse bietet und kostenfrei für alle zugänglich ist.“

Fachdidaktische Reflexionsfragen

- Ist das Programm Tracker ein geeignetes Hilfsmittel für den Physikunterricht? Für welche Themen, die im Physikunterricht behandelten werden, könnte sich die Verwendung von Tracker als sinnvoll erweisen?
- Welches didaktische Potential hat der Roller Coaster bezüglich der Verwendung im Physikunterricht?

A.1.7 Zerfallsgesetz

Praktikumsteil:	Wintersemester (WS)
Themenbereich:	Atomphysik
Verweis:	Die Idee des Versuches ist inspiriert durch die Untersuchung von Bierschaumzerfall als Analogie für radioaktiven Zerfall von von Theyßen (2009).

Versuchsziele und Aufgaben - Versuchsteil Würfelzerfall

- Atomkerne werden durch 100 12-seitige Würfel symbolisiert. Das Eintreten eines bestimmten Würfelereignisses steht für einen Zerfallsprozess. Die Würfel werden wiederholt gewürfelt. Nach jedem Wurf werden die „zerfallenen" Würfel aussortiert und die Anzahl notiert. Dieser Schritt wird mit den übrig gebliebenen Würfeln wiederholt. Die Messreihe wird mehrfach wiederholt. Die Wahrscheinlichkeit des Zerfallsprozesses, die Anzahl der Würfel und Würfe werden so gewählt, dass der Würfelzerfall möglichst gut eine exponentielle Zerfallskurve veranschaulichen kann.
- Graphischer Vergleich der Zerfallskurve aus dem Würfelexperiment mit der theoretischen Zerfallskurve
- Bestimmung der Halbwertszeit
- Beurteilung, ob sich das Würfelexperiment zur Veranschaulichung des radioaktiven Zerfalls als statistischer Prozess eignet

Versuchsziele und Aufgaben - Versuchsteil Bierschaumzerfall

- Messung des zeitlichen Zerfalls von Bierschaum in einem Messzylinder durch manuelle Aufnahme von Messwerten der oberen und unteren Schaumkanten in geeigneten zeitlichen Abständen oder mittels Videoanalyse
- Analyse der zeitlichen Veränderung des Schaumvolumens und des Volumens der Flüssigkeit im Schaum in Hinblick auf exponentiellen Zerfall (ggf. mit der Videoanalyse-Software Tracker (D. Brown & Cox, 2009; D. Brown, 2019))
- Berechnung der mittleren Lebensdauer und der Halbwertszeit
- Beurteilung, ob sich der Zerfall von Bierschaum als Analogie zum radioaktiven Zerfall eignet

Herausforderungen

- Würfelzerfall: begründete Auswahl einer geeigneten Zerfallswahrscheinlichkeit und Menge an Würfeln zur optimalen Veranschaulichung radioaktiven Zerfalls
- Bierschaumzerfall: Interpretation der Ergebnisse in Bezug auf die Mechanismen des Schaumzerfalls
- geeignete Darstellung und Auswertung der (exponentiellen) Abnahmen

- Vergleich und Beurteilungen der Ergebnisse in Bezug auf die Analogie

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	leicht	1
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	mittel	2
Summe		11

Tabelle 41: *Schwierigkeitsgrad des Versuches „Zerfallsgesetz“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Thematisierung der „Richtlinie zur Sicherheit im Schulunterricht“ (Kultusministerkonferenz, 1994) und der Strahlenschutzverordnung zum Umgang mit radioaktiven Präparaten im Schulunterricht
- Quantitative experimentelle Prüfung der in der Schule vorkommenden Analogie des Bierschaumzerfalls für den radioaktiven Zerfall
- Kritische Beurteilung der Eignung verschiedener Analogien zum radioaktiven Zerfalls

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I
Sachgebiet	Atom- und Kernphysik - Radioaktiver Zerfall

Schulbezogene Einleitung

„Für den Umgang mit radioaktiven Präparaten im Schulunterricht gibt es sehr strenge Vorschriften. Die „Richtlinie zur Sicherheit im Schulunterricht“ (Kultusministerkonferenz, 1994) ist eine Empfehlung der Kultusministerkonferenz und gibt entsprechende Hinweise zum Schutz der zum Teil minderjährigen Schülerinnen und Schüler. Die Vorschriften zu radioaktiven Präparaten basieren auf der Strahlenschutzverordnung und umfassen diverse Aspekte. Darunter werden Informationen zu einzuhaltenden Grenzwerten der Strahlenexposition, die Lagerung der radioaktive Präparate in speziellen Schutzbehälter, die Instanz von Strahlenschutzbeauftragten und vielem mehr gegeben.
Wie soll man Schülerinnen und Schülern den radioaktiven Zerfall nun erklären, wenn man die entsprechenden Vorschriften an der Schule nicht erfüllen kann und keine Versuche mit radioaktiven Präparaten zeigen darf?

Im Unterricht werden für die Verdeutlichung von exponentiellen Prozessen aus diesem Grund oft Analogien hinzugezogen. Dazu zählen beispielsweise statistische Experimente wie das Werfen von Münzen oder Würfeln. Aber wie gut eignen sich Analogien wirklich? Nicht selten wird auch der Zerfall von Bierschaum als Analogon betrachtet, obwohl bei genauerem Hinsehen ein exakter exponentieller Zerfall des Bierschaums theoretisch gar nicht plausibel ist. Folgt der Zerfall von Bierschaum experimentell dem Zerfallsgesetz? Oder geht es hier rein um den motivationalen Charakter durch das alkoholfreie (!) Bier? Untersuchen Sie es selbst und testen Sie, welche Motivation auch von Würfeln und Bierschaum ausgehen kann!"

Fachdidaktische Reflexionsfragen

- Welche Vor- und Nachteile hat der Würfelzerfall als Analogversuch zum radioaktivem Zerfall im Physikunterricht?
- Wie können Sie den Bierschaumzerfall im Unterricht einsetzen und was sollten Sie dabei beachten?

Klassische Einleitung

„Viele natürliche Prozesse folgen einer exponentiellen Gesetzmäßigkeit. Die Anzahl von Bakterien in einem Biotop beispielsweise kann exponentiell abnehmen oder wachsen. Auch in der Wirtschaft wird man mit exponentiellem Zuwachs anhand von Kapital durch Zinseszins konfrontiert.
Mathematisch besitzen all diese Prozesse eine Äquivalenz und lassen sich durch folgende Differentialgleichung beschreiben.

$$\frac{dN}{dt} = \lambda N$$

Die zeitliche Änderung einer Größe N ist damit gegeben als Produkt der Größe multipliziert mit einer Konstante λ, die je nach Vorzeichen als Wachstums- oder Zerfallskonstante bezeichnet wird. Die nachfolgende Lösung dieser Differentialgleichung beschreibt die exponentielle Ab- oder Zunahme der Ursprungsgröße N_0 nach der Zeit t.

$$N(t) = N_0 \cdot \mathrm{e}^{\lambda t}$$

Auch in der Physik ist die Exponentialfunktion ein mathematisches Grundwerkzeug, das zur Beschreibung von Prozessen herangezogen wird. Die natürliche Umwandlung eines radioaktiven Kerns in einen andern unter der Aussendung von α-Teilchen (Helium-Kerne), β-Teilchen (Elektronen oder Positronen) bzw. γ-Strahlung (Photonen) ist ebenso ein statistischer Prozess und lässt sich mit dem exponentiellen Zerfallsgesetz beschreiben."

A.1.8 Franck-Hertz-Versuch

Praktikumsteil:	Wintersemester (WS)
Themenbereich:	Atomphysik
Verweis:	Die Umsetzung des Versuches mit CASSY ist inspiriert durch den Hersteller LD-Didactic (2014).

Versuchsziele und Aufgaben

- Quecksilber
  - Aufnahme von Franck-Hertz-Kurven (I-U-Kennlinie von Auffängerstrom und Beschleunigungsspannung) für Quecksilber bei systematischer Variation der Steuergrößen des Experiments (Heiztemperatur der Quecksilber-Röhre, Emissionsgitterspannung und Gegenspannung) mit dem Messwerterfassungssystem CASSY (LD-Didactic, o.D. a)
  - Aufnahme einer optimalen Franck-Hertz-Kurve mit geeigneten Steuergrößen
  - Qualitative Erläuterungen der Franck-Hertz-Kurve
  - Quantitative Analyse der diskreten Energieaufnahme der Atome durch Elektronenstöße anhand der Minima der Franck-Hertz-Kurve und Bestimmung der emittierten Wellenlänge
- Neon
  - Aufnahme einer Franck-Hertz-Kurve für Neon mit optimalen Parametern
  - Analyse der diskreten Energieaufnahme der Atome durch Elektronenstöße anhand der Minima der Kurve und Bestimmung der Wellenlänge der Spektrallinien
  - Beschreibung und Erläuterung der beobachtbaren Leuchterscheinungen unter Berücksichtigung der vorangegangenen Ergebnisse und dem Termschema von Neon

a) b)

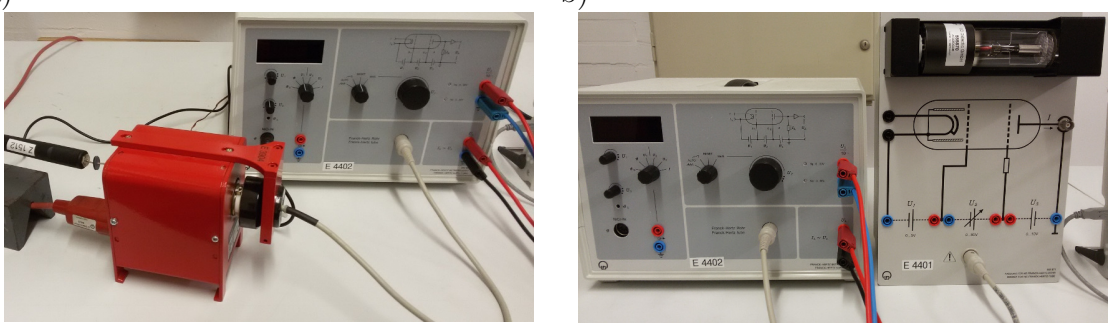


Abbildung 61: *Franck-Hertz-Versuch mit a) Quecksilber (Quecksilber-Franck-Hertz-Röhre im Heizofen links und Betriebsgerät rechts) und b) Neon (Betriebsgerät links und Neon-Franck-Hertz-Röhre mit zugehörigem Schaltbild rechts.*

Herausforderungen

- Theoretische Überlegungen zu den physikalischen Prozessen und der Franck-Hertz-Kurve/Strom-Spannungs-Kennlinie bei a) einer Vakuumröhre, b) einem Gas in der Röhre, bei dem nur elastische Stöße auftreten, und c) bei inelastischen Stößen mit einem Gas wie Quecksilber oder Neon.
- Darstellung und Auswertung der Franck-Hertz-Kurven in CASSY Lab
- Erläuterung des Einfluss der systematisch variierten Steuergrößen (Heiztemperatur der Quecksilber-Röhre, Emissionsgitterspannung und Gegenspannung) auf die Franck-Hertz-Kurven
- Erklärung der orangenen Leuchterscheinungen unter Einbeziehung des Termschemas für Neon

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	mittel	2
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	leicht	1
Summe		12

Tabelle 42: *Schwierigkeitsgrad des Versuches „Franck-Hertz-Versuch“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung eines zentralen Versuches des Physikunterrichts der Oberstufe
- Die Verwendung schultypischer Aufbauten und gezielter Variation von Parametern soll vertiefend auf mögliche Frage- und Problemstellungen im Schulunterricht vorbereiten.
- Einsatz der schultypischen Messwerterfassungssystems CASSY

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiet	Quanten - Quantenphysikalisches Atommodell

Schulbezogene Einleitung

„Der Franck-Hertz-Versuch gilt als eindrucksvolles Experiment zur Bestätigung der Quantenphysik. Der Versuch ist daher auch besonders beliebt in der Physikausbildung. Im Unterricht spielt der Franck-Hertz-Versuch eine zentrale Rolle und wird

ausgiebig thematisiert. Der Versuch ist somit auch regelmäßig Gegenstand von Abschlussprüfungen. Trotzdem wird der Versuch selten korrekt erklärt. In vielen Fällen ist das quantitative Ergebnis für die diskrete Energieaufnahme zwar richtig, aber der Weg dahin ist nicht konsistent mit dem Schaltbild. Selbst einige Lehrbücher enthalten elementare Fehler. Sie könnten zum Beispiel einem ähnlichen Lehrbuchtext wie nachfolgendem begegnen.

„Beim Franck-Hertz-Versuch werden Quecksilber-Atome in einer evakuierten Röhre mit Elektronen beschossen (Abb. 62). Durch Erhöhung der Beschleunigungsspannung U_b, erhält man den Graphen für den Auffängerstrom $I_A(U_b)$ in Abbildung 63. Der Graph zeigt Maxima im Abstand von 4.9 V. Die Maxima des Auffängerstroms erklären sich durch die Energieabgabe der Elektronen an die Quecksilber-Atome durch Stöße. Die Atome nehmen offensichtlich nur Energien von 4.9 eV auf.“

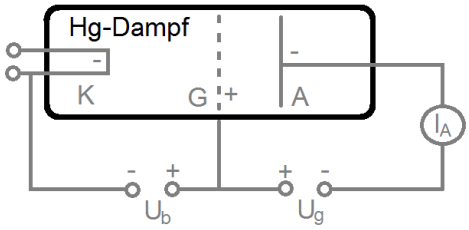


Abbildung 62: *Schaltbild für Franck-Hertz-Versuch. Die Elektronen treten aus einer geheizten Kathode (K) aus und werden in einem elektrischen Feld mit der Spannung U_b zu einem Gitter (G) beschleunigt. Zwischen Gitter (G) und Anode (A) werden den Elektronen anschließend durch das Gegenfeld mit der Spannung U_g wieder abgebremst. Elektronen mit ausreichender Energie erreichen die Anode und tragen zum Auffängerstrom I_A bei.*

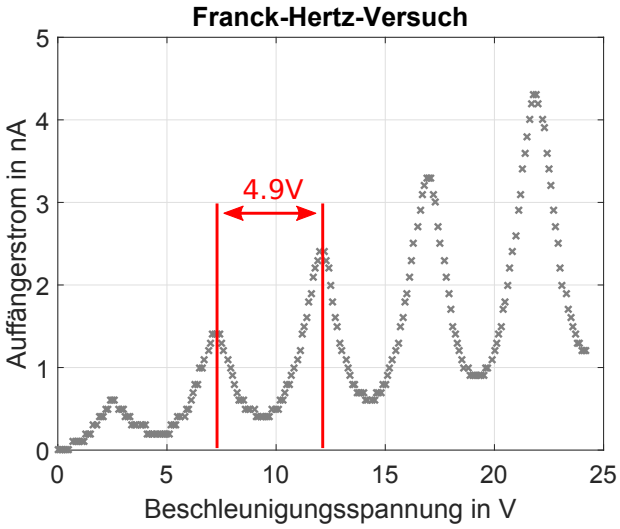


Abbildung 63: *Auffängerstrom I_A aufgetragen über der Beschleunigungsspannung U_b. Abstand der Maxima beträgt im Mittel 4.9 V.*

Haben Sie die Unstimmigkeit zwischen Schaltbild und Erklärung entdeckt?
Bei einem Maximum des Auffängerstroms gelangen dem Schaltplan nach viele Elektronen mit ausreichender Energie von der Kathode zur Anode. Warum sollte dies ein Indiz für eine Energieabgabe des Elektrons sein?
Kennen Sie die richtige Erklärung für den Franck-Hertz-Versuch?“

Fachdidaktische Reflexionsfragen

- In beiden vorgestellten Versuchsaufbauten ist der zugrunde liegende physikalischen Effekt des Stoßes von beschleunigten Elektronen mit Gasatomen gleich. Welche Vor- und Nachteile bieten die Frack-Hertz-Röhren mit Quecksilber und Neon im Vergleich dennoch für den Schulunterricht?
- Die Franck-Hertz-Kurven können sowohl mit einem Messwerterfassungssystem wie CASSY, aber auch mit einem Oszilloskop oder mit Drehspulmessinstrumenten bzw. Digitalmultimetern aufgenommen werden. Welche Vor- und Nachteile bieten die Messgeräte in der Schulpraxis?

Klassische Einleitung

„1913 entwickelte Nils Bohr ein neues Atommodell. Er erweiterte das Rutherford'sche Atommodell, das von einem schweren, positiv geladene Atomkern und darum befindlichen negative geladenen Elektronen ausgeht, um die Struktur der Elektronen. Bohr postulierte, dass die Elektronen den Atomkern auf festen, geschlossenen Bahnen umkreisen, sodass diskrete Energiestufen für Atome existieren. Er erklärte die von Johann Jakob Balmer 30 Jahre zuvor beobachteten Spektrallinien von Wasserstoff dadurch, dass beim Wechsel eines Elektrons von einer Bahn bzw. Energiestufe in eine andere elektromagnetische Strahlung mit der Energiedifferenz absorbiert oder emittiert wird. James Franck und Gustav Hertz untersuchten zeitgleich, aber unabhängig von den Bohrschen Überlegungen die Stoßionisation von Quecksilber-Atomen. Der Versuch zeigte durch die Anregung von Atomen durch Stöße mit beschleunigten Elektronen, dass diskrete Energieniveaus in Atomen existieren. Der nach ihnen benannte Franck-Hertz-Versuch gehört damit zu den klassischen, experimentellen Belegen der Quantentheorie. 1925 bekamen Franck und Hertz für diesen Versuch den Nobelpreis für Physik.“

A.1.9 Photoeffekt

Praktikumsteil:	Wintersemester (WS)
Themenbereich:	Atomphysik
Verweis:	Eine Vorarbeit zur Entwicklung des Versuches wurden im Rahmen der Bachelorarbeit von Simon (2015) durchgeführt.

Versuchsziele und Aufgaben

- Bestimmung des Planckschen Wirkungsquantums h auf Basis des Photoeffekts mithilfe der Gegenfeldmethode und verschiedenfarbiger a) Laser und b) LEDs (Abb. 64)
- Untersuchung des Einflusses der Intensität des Lasers auf Gegenfeld und Photostrom
- Vergleich der Ergebnisse für h für den Versuchsaufbau mit LEDs und mit Lasern

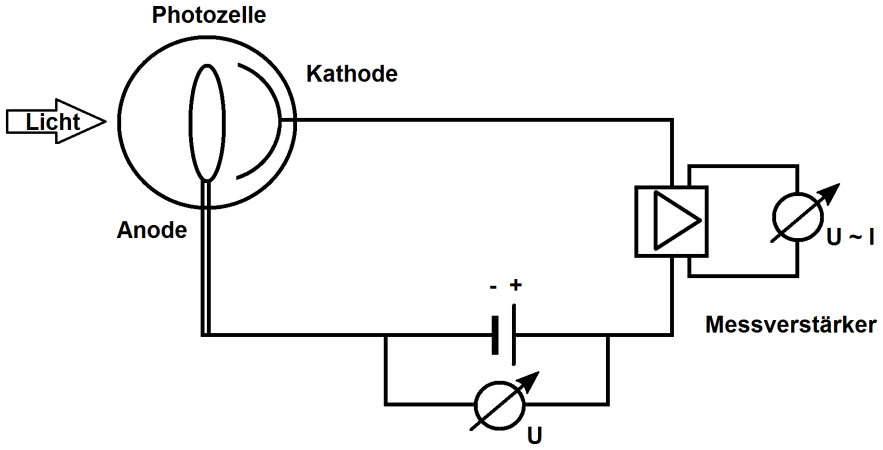


Abbildung 64: *Schaltplan zum Versuch Photoeffekt zur Bestimmung des Planckschen Wirkungsquantums mittels Gegenfeldmethode. Eine Photozelle mit beschichteter Kathode und ringförmiger Anode wird mit Licht verschiedenfarbiger Laser und LEDs beleuchtet. Aufgrund des Photoeffektes werden Elektronen aus der Kathode ausgelöst. Die Spannung im Gegenfeld wird so eingestellt, dass gerade kein Stromfluss mehr zwischen Kathode und Anode unter Zuhilfenahme eines Messverstärkers gemessen wird.*

Herausforderungen

- Erklärung der Messmethode
- Aufbau des Versuches anhand das Schaltplans (Abb. 64)
- Geeignete Nutzung des Messverstärkers
- Identifikation und Reduzierung von Störeinflüssen

- Diskussion der Vor- und Nachteile der Laser und LEDs zur Bestimmung von h (z. B. Nachteil der LEDs: nicht monochromatisch)
- Bestimmung von h mittels linearer Regression und entsprechender Fehlerrechnung

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	schwer	3
Messtechnik	schwer	3
Versuchsdurchführung	mittel	2
Auswertung	leicht	1
Fehlerrechnung	mittel	2
Summe		13

Tabelle 43: *Schwierigkeitsgrad des Versuches „Photoeffekt“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe II (Teilcheneigenschaften von Licht, Photoeffekt, Plancksches Wirkungsquantum)
- Schulpraktische Versuchsumsetzung mit verschiedenfarbigen LEDs im Vergleich zum klassischen Aufbau mit einer Quecksilber-Hochdrucklampe und Prisma

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiet	Quanten - Teilcheneigenschaften des Lichts

Schulbezogene Einleitung

„Die Natur des Lichts hat schon viele Menschen zutiefst fasziniert. In der Antike gab es beispielsweise die Vorstellung von „Sehstrahlen“. Inzwischen versteht die moderne Quantenphysik das Licht als Welle genauso wie als Teilchen. Doch der Weg zum Welle-Teilchen-Dualismus war recht lang und voller beschwerlicher Uneinigkeit. Experimente belegten anhand von Reflexion, Interferenz, Beugung, Polarisation und Brechung den Wellencharakter von Licht. Hypothesen von Max Planck und Albert Einstein zur quantisierten Energieübertragung von elektromagnetischer Strahlung und dem Teilchencharakter von Licht ließen die Diskussion über das Wesen der Lichts wieder stark aufleben. Millikan konnte diese Thesen experimentell nachweisen, indem er den photoelektrischen Effektes mit der Gegenfeldmethode analysierte. Daher wird dieses fundamentale Experiment weitverbreitet im Physikunterricht als Befund für den Teilchencharakter eingesetzt, um den Welle-Teilchen-Dualismus des

Lichts einzuführen.
Der herkömmliche Versuch zur Untersuchung des photoelektrischen Effektes mit der Gegenfeldmethode in der Schule nutzt das Licht einer Quecksilber-Hochdrucklampe, das je nach Aufbau durch Prismen oder Filter und durch Blenden und Linsen auf die Kathode der Photozelle fällt. Laser und LED-Lampen bergen neue technische Möglichkeiten, um den Versuch für Schülerinnen und Schüler deutlich übersichtlicher zu gestalten.
Untersuchen Sie selbst, ob trotz einer didaktische Reduktion auf die wesentlichen Elemente Lichtquelle und Photozelle die experimentelle Genauigkeit bei der Bestimmung des Planckschen Wirkungsquantums gegeben ist.“

Fachdidaktische Reflexionsfragen

- Eignet sich der Versuch zur Untersuchung des photoelektrischen Effektes mit der Gegenfeldmethode zur Einführung des Welle-Teilchen-Dualismus im Physikunterricht?
- Der herkömmliche Versuch zur Untersuchung des photoelektrischen Effektes mit der Gegenfeldmethode nutzt das Licht einer Quecksilber-Hochdrucklampe, das je nach Aufbau durch Prismen oder Filter und durch Blenden und Linsen auf die Kathode der Photozelle fällt. Diskutieren Sie die Vor- und Nachteile des hier vorgestellten Versuches mit LEDs bzw. Lasern im Vergleich zum herkömmlichen Versuch?

Klassische Einleitung
„Das Wesen des Lichts hat schon viele Menschen fasziniert. In der Antike gab es beispielsweise die Vorstellung von „Sehstrahlen“. Auch in der Physik herrschte bereits viel Uneinigkeit über die Natur des Lichts. 1690 beschrieb Huygen die Welleneigenschaften des Lichts. Nur ein paar Jahre später wiederum betrachtete Newton Licht als Teilchen. Experimente belegten anhand von Reflexion, Interferenz, Beugung, Polarisation und Brechung den Wellencharakter von Licht. Doch war der Streit damit nicht beigelegt. Denn Licht - verstanden als Welle - konnte nicht die quantisierte Energieübertragung von elektromagnetischer Strahlung erklären, welche Max Planck 1900 bei der Analyse der Wärmestrahlung des schwarzen Körpers entdeckte. Albert Einstein nutzte fünf Jahre später Plancks Arbeit bei der Analyse des photoelektrischen Effektes. Während Planck lediglich den Austausch von Energien durch Strahlung als quantisiert betrachtete, stellte Albert Einstein die Lichtquantenhypothese auf, wonach die elektromagnetische Strahlung selbst quantisiert ist. Der wieder aufgebrachte Welle-Teilchen-Dualismus war mit der klassischen Physik nicht direkt vereinbar. Deshalb erhielt Einstein erst Jahre später (1921) den Nobelpreis für die Theorie des Lichtquants. Der experimentelle Nachweis des photoelektrischen Effektes mithilfe der Gegenfeldmethode durch Millikan war daher ein historisch bedeutsames Experiment.“

A.1.10 Rutherford'sche Streuung

Praktikumsteil:	Wintersemester (WS)
Themenbereich:	Atomphysik
Verweis:	Vorarbeiten und weiterführende Untersuchungen zur Entwicklung und Optimierung des Versuches wurden im Rahmen der Bachelorarbeiten von Geertz (2016) und Brauer (2017) und Marten (2018) durchgeführt.

Hinweis: *Dieser Versuch wurde ausführlicher in Abschnitt 6.6 dieser Arbeit vorgestellt. Dort ist auch der Aufbau des Versuches abgebildet (Abb. 10) und der Schwierigkeitsgrad in den einzelnen Teilbereichen aufgeführt (Tab. 6). An diese Kurzbeschreibung ist daher auch die ausführliche Versuchsanleitung des Versuches zur Rutherford'schen Streuung nachgestellt.*

Versuchsziele und Aufgaben

- Kalibrierung des Luftkissentisches als zentrales Element des Versuchsaufbaus, um möglichst Reibungsfreiheit zu erreichen (Abb. 10)
- Videoanalyse der Abschussgeschwindigkeit mit der Videoanalyse-Software Tracker (D. Brown & Cox, 2009; D. Brown, 2019) eines Magnetpucks (Analogie: Alpha-Teilchen), der mittels einer Abschussvorrichtung zentral auf einen Luftkissentisch geschossen wird
- Quantitative Energiebetrachtungen für einen Magnetpuck (Analogie: Alpha-Teilchen), der zentral auf einen in der Mitte des Luftkissentisches montierten Kernmagneten (Analogie: Goldkern) geschossen wird, mithilfe von Abschussgeschwindigkeit und minimalem Abstand zwischen Magnetpuck und Kernmagnet
- Untersuchung der Streukurve eines Magnetpucks (Analogie: Alpha-Teilchen), der dezentral auf den Kernmagneten (Analogie: Goldkern) geschossen wird, in Bezug auf den Verlauf und die Beschleunigung des Pucks bzw. die Kraft auf den Puck
- Bestimmung der Streuwinkel für verschiedene Stoßparameter und Prüfung der Gültigkeit des von Rutherford aufgestellten Zusammenhangs dieser Größen für das Magnetfeld
- Beurteilung der Analogie zum Rutherford'schen Streuversuch hinsichtlich ihrer Möglichkeiten und Grenzen

Herausforderungen

- Vergleich des Rutherford'schen Versuches im Coulomb-Potential eines Goldatomkerns mit dem durchgeführten Streuversuch im magnetischen Potential in Hinsicht auf Entsprechungen und Unterschiede

- Vergleich der Kraft und potentiellen Energie zwischen den zwei magnetischen Dipolen mit der Coulomb-Kraft zwischen zwei Ladungen
- Geometrische Überlegung zur Streuung eines Teilchens an einem Kern und Übertragung dessen auf den Versuchsaufbau
- Beurteilung der Ergebnisse in Bezug auf die Analogie
- Sorgfältige Ausrichtung des Luftkissentisches
- Reproduzierbare Handhabung der Abschussvorrichtung für präzise Messungen
- Auswahl der optimalen Kameraeinstellungen (Belichtungszeit)
- Nutzung spezieller Auswertefunktionen der Videoanalyse-Software Tracker

Relevanz des Versuchsthemas für die Schule

- Quantitative experimentelle Analyse eines Analogieversuches zum Rutherford'schen Streuversuch, der im Physikunterricht der Oberstufe bei der historischen Entwicklung des Atommodells theoretisch betrachtet wird
- Kritische Beurteilung von Möglichkeiten und Grenzen der Analogie
- Vertiefung der Kenntnisse der schultypischen, kostenfreien Videoanalyse-Software Tracker

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiet	Quanten - Quantenphysikalisches Atommodell

Schulbezogene Einleitung
„siehe nachfolgende Versuchsbeschreibung“

Fachdidaktische Reflexionsfragen

- Welche Vorteile bieten Analogie-Versuche wie dieser zum Rutherford'schen Streuversuch für den Schulunterricht?
- Beurteilen Sie, ob bzw. in welchem Rahmen dieser Analogie-Versuch für Schülerinnen und Schüler der Sekundarstufe II geeignet ist.

Klassische Einleitung
„Zu den bekanntesten Vorstellungen über den Aufbau des Atoms, die im Laufe der Entwicklung der Atommodelle entstanden sind, gehören unter anderem das Thomsonsche und das Rutherford'sche Atommodell. In der Vorstellung nach Thomson von 1903 ist das Atom eine kugelförmige, homogene positive Ladungsverteilung mit eingestreuten negativ geladenen Elektronen. Dieses Rosinenkuchenmodell von J. J. Thomson war jedoch nicht mit den experimentellen Befunden von E. Rutherford und seinen Kollegen H. Geiger und E. Marsden vereinbar. Sie untersuchten 1911

die Streuung von α-Teilchen, die sie auf an dünne Goldfolien schossen. Dabei beobachteten Sie deutlich größere Streuwinkel der α-Teilchen mit bis zu 180° als nach dem Thomson Modell zu erwarten waren. Rutherford konnte seine Ergebnisse mit einem positiv geladenen Atomkern erklären. Der Atomkern vereint dabei nahezu die gesamte Masse im Zentrum des Atoms und ist umgeben von Elektronen in einer Atomhülle.
Die positiven α-Teilchen erfahren in diesem Versuch eine abstoßende Coulomb-Kraft durch das Goldatom. Schießt man einen magnetischen Dipol mit Richtung des magnetischen Momentes senkrecht zur Schussebene auf einen weiteren gleich ausgerichteten Dipol, so wirken ebenfalls abstoßende Kräfte. Während die Coulomb-Kraft von $1/r^2$ abhängt, ist die Kraftwirkung zwischen den zwei magnetischen Dipolen abhängig von $1/r^4$. In diesem Versuch geht es darum, ob ein Streuversuch mit zwei magnetischen Dipolen vergleichbare Ergebnisse liefert wie die Rutherford'sche Untersuchung der Streuung. Zudem wird der Frage nachgegangen, ob Rutherford das Potential hätte bestimmen können, wenn er die Bahnkurve der α-Teilchen untersuchen hätte können.“

Versuchsanleitung - Rutherford'scher Streuversuch

[Stand: WS1819]

Zu Themen aus der Atom- und Kernphysik können bzw. dürfen einige Versuche nicht im Physikunterricht der Schule durchgeführt werden. Aus diesem Grund greift man nicht selten auf Analog-Experimente zurück, um die mikroskopische Welt der Teilchen sichtbar und (be-)greifbar zu machen. So ist es auch bei einem der bedeutendsten Experimenten der Physik: Die Entdeckung des Atomkerns durch Ernest Rutherford. Rutherford schoss α-Teilchen auf eine Goldfolie und untersuchte die Streuung dieser an den Atomkernen.
Kircher et al. (2015) stellen in ihrem Buch zur Physikdidaktik im Kapitel zu Analogien einen Versuch zur Rutherfordstreuung vor. Dabei wird ein an einem Faden aufgehängter, elektrisch geladener Tischtennisball langsam auf eine gleichartig geladene Kugel zu gependelt. Bei kleinen Abständen kann die Abstoßung des Tischtennisballes durch die Kugel beobachtet werden. Welche Gemeinsamkeiten ergeben sich zum Rutherford-Experiment? Formale Analogien ergeben sich offensichtlich durch die Übertragbarkeit des geladenen Tischtennisballes auf das α-Teilchen und der fest positionierten, geladenen Kugel auf den Atomkern. Kircher et al. (2015) kritisieren die Analogie jedoch auf der Handlungsebene mit der Frage: „Was ist im Analogversuch vom „Schießen“ übrig?“. In der Tat hat das Verfahren Alphateilchen auf Atomkerne zu schießen mit dem beschriebenen Versuch nicht viel gemein.
Gibt es einen Analog-Versuch, bei dem auch die Handlungsebene des Rutherford-Experimentes aufgegriffen werden kann? In diesem Versuchsaufbau werden Sie auf einem Luftkissentisch Magnetpucks als Analogon für das α-Teilchen auf einen fest positionierten Magneten, der für den Atomkern steht, schießen. Wo liegen hier die Grenzen und Möglichkeiten der Analogie? Inwieweit sind die Abstoßungen im magnetischem Feld und im elektrischen Feld qualitativ und quantitativ vergleichbar?

Versuchsziele

In diesem Versuch soll der Rutherford'sche Streuversuch mit einem Streuversuch mit zwei magnetischen Dipolen verglichen werden. Es sollen Energiebetrachtungen mithilfe der mittleren Abschussgeschwindigkeit des Magnetpucks und des minimalen Abstandes des Magnetpucks zum Kernmagneten beim zentralen Stoß angestellt werden. Die Kraft zwischen den magnetischen Dipolen soll anhand der Bahnkurve untersucht werden. Es soll die Verteilung der Streuwinkel des Magnetpucks abhängig vom Stoßparameter im Vergleich zum Rutherford'schen Versuch untersucht werden. Es soll beurteilt werden, ob der Streuversuch im Magnetfeld sich als Analogie-Experiment zum Rutherford'schen Steuversuch im Schulunterricht eignet.

Versuchsvorbereitung

Nutzen Sie die Literaturtipps und die Kontrollfragen zusätzlich zur Versuchsanleitung als Unterstützung bei der Versuchsvorbereitung. Die Hausaufgaben dienen zur Gegenüberstellung des Rutherford'schen Versuches im Coulomb-Potential und diesem Versuch im magnetischen Potential. Im Besonderen kann die Literatur zu Tracker Ihnen bei der Vorbereitung der Videoauswertung helfen.

Literaturhinweise

Demtröder, W. (2016b). *Experimentalphysik 3 - Atome Moleküle und Festkörper* [5. Auflage; Seite 64 f: Kapitel 2.8.5 Rutherford'sches Atommodell, Kapitel 2.8.6 Rutherford'sche Streuformel]. Springer-Verlag.

Ministerium für Schule und Berufsbildung Schleswig-Holstein. (2019). Fachanforderungen Physik [Stand: Februar 2019].

Schulphysikwiki. (o.D.). *Bewegungsanalyse mit einem Video (Tracker)* [letzter Zugriff: 02.02.16]. http://schulphysikwiki.de/index.php/Bewegungsanalyse_mit_einem_Video_(Tracker)

Supermagnete. (o.D. b). *Sicherheitsabstände* [letzter Zugriff: 21.05.16]. http://www.supermagnete.de/faq/distance

Tracker. (o.D.). *Manual für Tracker - Freie Software für Bewegungsanalysen* [letzter Zugriff: 17.03.16]. http://physlets.org/tracker/tracker_help.pdf

Kontrollfragen

1. Mit welchem Ziel haben Rutherford und seine Kollegen den Streuversuch durchgeführt?
2. Wie lief der Streuversuch von Rutherford ab?
3. Was hat Rutherford beobachtet?
4. Wie sieht das Atommodell von Rutherford aus?
5. Wie funktioniert der Luftkissentisch?
6. Wie erstellt man in Tracker automatisiert eine Bahnkurve?
7. Wie kann man in Tracker einen Maßstab für die Entfernung festlegen?
8. Wie kann man in Tracker Winkel bestimmen?
9. Wie kann man in Tracker Diagramme analysieren, um beispielweise Mittelwerte und Standardabweichungen einer Messgröße zu bestimmen?
10. Wie kann man in Tracker Modellfunktionen an Kurven anfitten?
11. Was versteht man unter Belichtungszeit?

12. Was ist mit Verzeichnung bei der Nutzung von Weitwinkel-Objektiven gemeint?

Hausaufgaben

Hausaufgabe 1) - Fachanforderungen
Ordnen Sie diesen Versuch zum Thema Atommodelle in die Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) ein:

☐ Sekundarstufe I ☐ Sekundarstufe II

Sachgebiet: _____

Hausaufgabe 2) - Vergleich Rutherford'scher Versuch und Streuversuch im Magnetfeld
Vergleichen Sie den Rutherford'schen Versuch im Coulomb-Potential eines Goldatomkerns mit dem hier durchgeführten Streuversuch im magnetischen Potential in Hinsicht auf Entsprechungen und Unterschieden.

Hausaufgabe 3) - Kraft, Potential und potentielle Energie

a) Die Kraft $\vec{F}_M$, die von einem magnetischen Dipol 1 mit der magnetischen Flussdichte $vecB_1$ auf einen anderen Dipol 2 ausgeübt wird, ergibt sich aus

$$\vec{F}_M = \nabla \left(\vec{m}_2 \cdot \vec{B}_1 \right).$$

$$\vec{F}_M = \frac{3\mu_0}{4\pi r^4} \left[\vec{m}_2(\vec{m}_1 \cdot \vec{r}_n) + \vec{m}_1(\vec{m}_2 \cdot \vec{r}_n) + \vec{r}_n(\vec{m}_1 \cdot \vec{m}_2) - 5\vec{r}_n(\vec{m}_1 \cdot \vec{r}_n)(\vec{m}_2 \cdot \vec{r}_n) \right]$$

In diesem Versuch sind die Magneten so ausrichtet, dass beide magnetische Momente $\vec{m}_{1,2}$ der Dipole parallel zueinander und senkrecht zur Ebene des Luftkissentisches und somit auch senkrecht zum Einheitsvektor $\vec{r}_n$ von Dipol 1 zu 2 ausgerichtet sind. Wie vereinfacht sich dadurch die Gleichung für diesen Versuch? Vergleichen Sie die Kraft $\vec{F}_M$ zwischen zwei magnetischen Dipolen mit der Coulomb-Kraft $\vec{F}_C$ zwischen zwei Ladungen.

b) Wie hängen Kraft F und Potential V zusammen?

c) Wie hängen Kraft F und potentielle Energie E_{pot} zusammen?

d) Tragen Sie die fehlenden Exponenten in die freien Boxen $\square$ in den Formeln in der nachfolgenden Tabelle ein. Vervollständigen Sie die Formeln für die zugehörigen Konstanten $k_{C,M}$. r sei der Abstand zwischen den Ladungen q bzw. magnetischen Dipolen mit dem magnetischen Moment m.

	Rutherford'scher Versuch	Streuversuch im Magnetfeld
Kraft	$F_C = \frac{k_C}{r^{\square}}$	$F_M = \frac{k_M}{r^{\square}}$
Potential des Kerns	$V_C = \frac{k_C}{q \cdot r^{\square}}$	$V_M = \frac{k_M}{3m \cdot r^{\square}}$
Potentielle Energie	$E_{pot,C} = \frac{k_C}{r^{\square}}$	$E_{pot,M} = \frac{k_M}{3r^{\square}}$
Konstante	$k_C =$	$k_M =$

Hausaufgabe 4) - Winkelkorrektur

Wie man sich anhand der geometrischen Überlegungen in Abbildung 65) verdeutlichen kann, ist der gemessene Ablenkwinkel ϑ in Punkt (x_p, y_p) nicht gleich dem Streuwinkel φ. Formulieren Sie nachfolgenden Zusammenhang für den Streuwinkel φ mithilfe der beobachtbaren Größen R, a, b und ϑ um.

$$cos(\varphi) = \frac{(x_p + a)}{R'} = \frac{(x_p + a)}{\sqrt{(x_p + a)^2 + (y_p - b)^2}}$$

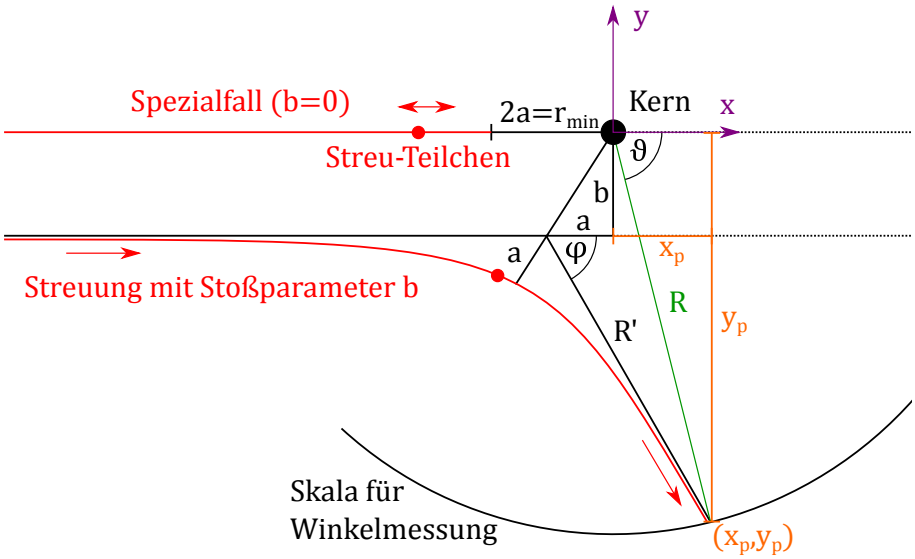


Abbildung 65: *Geometrische Überlegung zur Streuung eines Teilchens (rot) an einem Kern (schwarz). ϑ sei der Messwinkel, φ der Streuwinkel, b der Stoßparameter und r_{min} der minimale Abstand des Teilchens zum Kern beim zentralen Stoß (Abb. nach Geertz, 2016).*

Messgeräte und Versuchsaufbau

Materialien

- Luftkissentisch mit Gebläse
- Videokamera an Aufhängung über dem Tisch
- PC mit Videoanalyse-Software Tracker und LabView Aufnahmeprogramm
- Abschussvorrichtung
- Neodym-Magnet mit Befestigung für die Tischmitte („Kernmagnet“)
- Magnetpuck
- Winkelskala
- Wasserwaage
- Maßwerkzeug
- Waage

Versuchsaufbau

Hauptbestandteil dieses Versuches ist der Luftkissentisch (Abb. 68). Das Gebläse sollte für den Versuch auf Einstellung 6 betrieben werden. Zentral über dem Luftkissentisch ist eine Kamera angebracht, um die Bewegung der Pucks als Video aufzunehmen. Bildrate sowie Belichtungszeit der Kamera können im Aufnahmeprogramm eingestellt werden. An einer Seite des Luftkissentisches ist eine seitlich verschiebbare Abschussvorrichtung für den Magnetpuck angebracht (Abb. 67). Der Kernmagnet kann in der Mitte des Tisches befestigt werden (Abb. 66). Auf den Luftkissentisch kann eine Winkelskala aus Holz aufgelegt werden.

Hinweise für den sicheren Umgang mit den Magneten

Magnetisches Feld

Diese Magnete erzeugen ein weit reichendes, starkes Magnetfeld. Sie können unter anderem Fernseher und Laptops, Computer-Festplatten, Kreditkarten und EC-Karten, Datenträger, mechanische Uhren, Hörgeräte und Lautsprecher beschädigen. Halten Sie Magnete von allen Geräten und Gegenständen fern, die durch starke Magnetfelder beschädigt werden können. Der Hersteller gibt auf seiner Website (Supermagnete, o.D. b) Empfehlungen zu Sicherheitsabständen.

Metall-Splitter

Neodym-Magnete sind spröde. Wenn ein Magnet mit einem Gegenstand kollidiert, kann er zersplittern. Scharfkantige Splitter können meterweit weg geschleudert werden und Ihre Augen verletzen. Vermeiden Sie Kollisionen von Gegenständen mit den starken Magneten. Vermeiden Sie das Tragen von magnetischem Schmuck. Tragen Sie bei der Handhabung mit den Magneten eine Schutzbrille. Achten Sie darauf, dass umstehende Personen ebenfalls geschützt sind oder Abstand halten.

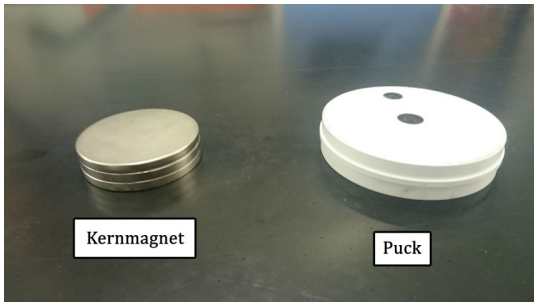


Abbildung 66: *Kernmagnet (links) und Magnetpuck (rechts) (Photo von Geertz, 2016).*

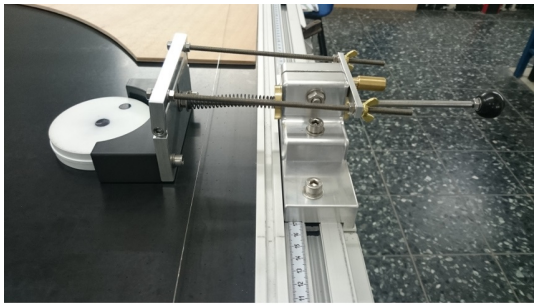


Abbildung 67: *Abschussvorrichtung mit Magnetpuck (Photo von Geertz, 2016).*

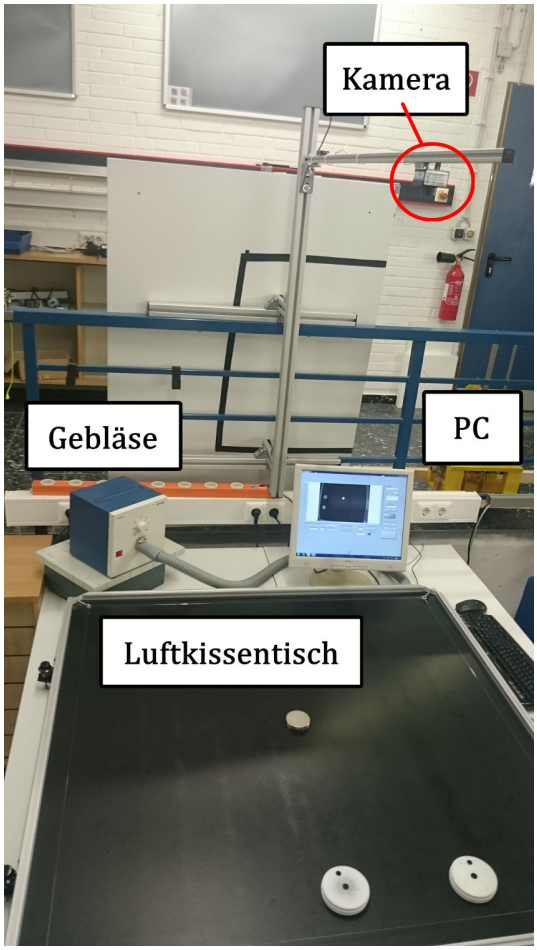


Abbildung 68: *Ein Gebläse sorgt für die Luftzufuhr am Luftkissentisch. Über dem Tisch ist eine Kamera befestigt, sodass am PC Videos von den Puckbewegungen aufgenommen und ausgewertet werden können (Photo von Geertz, 2016).*

Aufgaben

Aufgabe 1) - Ausrichtung des Luftkissentisches
Vor Beginn des Messungen muss der Luftkissentisch sorgfältig justiert werden. Richten Sie den Luftkissentisch möglichst horizontal aus, indem die Schraubfüße der Dreipunktauflage entsprechend einstellen. Nutzen Sie dazu eine Wasserwaage. Überprüfen Sie die Justierung mit einem Puck.

Aufgabe 2) - Abschussgeschwindigkeit
Untersuchen Sie, ob die Abschussvorrichtung den Puck mit reproduzierbarer Richtung und Geschwindigkeit abschießt. Finden Sie eine geeignete reproduzierbare Handhabung der Abschussvorrichtung. Schießen Sie den Magnetpuck mehrfach zentral auf die freie Fläche des Tisches und nehmen Sie die Bewegung mit der Kamera auf. Wählen Sie dafür eine geeignete Einstellung der Kamera aus. Werten Sie das Video der Bewegung mit Tracker aus. Bestimmen Sie die mittlere Abschussgeschwindigkeit v und die Abschussrichtung des Pucks sowie deren Messunsicherheiten.

Aufgabe 3) - Energiebetrachtungen
Rutherford konnte anhand von Energiebetrachtungen den minimalen Abstand r_{min} des α-Teilchens zum Goldatom beim zentralen Stoß berechnen und so den Kernradius des Goldatoms eingrenzen. Anders als bei Rutherford kann in diesem Streuversuch der Minimalabstand zwischen Magnetpuck und Kernmagnet beim zentralen Stoß direkt bestimmt werden.

a) Befestigen Sie den Kernmagneten in der Mitte des Luftkissentisches. Finden Sie die Position der Abschussvorrichtung, an welcher der Puck zentral auf den Kernmagneten geschossen wird. Beschreiben Sie die Bewegung des Pucks und betrachten Sie die Energieumwandlung dabei.

b) Bestimmen Sie mithilfe einer Videoauswertung möglichst genau den Minimalabstand r_{min} zwischen Puck und Kernmagnet beim zentralen Stoß.

c) Nutzen Sie die Energieerhaltung zur Bestimmung der Konstante k_M. Bestimmen Sie dazu die kinetische Energie E_{kin} des abgeschossenen Pucks mithilfe der Abschussgeschwindigkeit v. Aus dem bestimmten Minimalabstand r_{min} zwischen Puck und Kernmagnet können Sie auf die maximale potentielle Energie E_{pot} schließen.

Aufgabe 4) - Untersuchung der Kraft auf den Magnetpuck

Verändern Sie den Stoßparameter b so, dass der Magnetpuck im Wirkungsbereich des Kernmagneten von seiner Bahn abgelenkt wird. Tracken Sie die Bahn des Pucks. Legen Sie das Koordinatensystem so, dass der Ursprung in der Mitte des Kernmagneten liegt und eine Achse im Mittelpunkt des Magnetpucks mitwandert. Im Reiter Koordinaten kann dazu die Festlegung des Winkels des Koordinatensystems aufgehoben werden. Veranschaulichen Sie die getrackte Bahnkurve anhand eines Screenshots und beschreiben Sie diese. Diskutieren Sie den Verlauf der Geschwindigkeit des Magnetpucks $\vec{v}_r$ in radiale Richtung. Bestimmen Sie die Beschleunigung $\vec{a}_r$ in radiale Richtung in Abhängigkeit vom Abstand zum Kernmagneten r.
Untersuchen Sie, ob die Beschleunigung $\vec{a}_r$ bzw. somit auch die Kraft $\vec{F}_r$ auf den Dipol in radiale Richtung von $\frac{1}{r}$, $\frac{1}{r^2}$, $\frac{1}{r^3}$ oder $\frac{1}{r^4}$ abhängt. Fitten Sie dazu im Analyse-Tool in Tracker entsprechende Modellfunktionen an die Kurve. Bestimmen Sie den Proportionalitätskonstante k_M aus der Anpassung. Vergleichen Sie Ihre Ergebnisse mit Hausaufgabe 3) und Aufgabenteil 3.

Aufgabe 5) - Verteilung der Streuwinkel

Rutherford konnte die Bahn des α-Teilchen nicht wie hier direkt beobachten und analysieren. Zusammen mit seinen Kollegen hatte er die Häufigkeit der gemessenen Streuwinkel φ festgestellt. Die Verteilung des Streuwinkels stimmte mit der theoretischen Verteilung für eine Punktladung mit einem Potential abhängig von $\frac{1}{r}$ überein. Rutherford konnte die Bahn der α-Teilchen daraus theoretisch als Hyperbeln ermitteln. Für die Bahnkurven gilt
$cot(\frac{\varphi}{2}) = \frac{b}{a} = \frac{2b}{r_{min}}$.

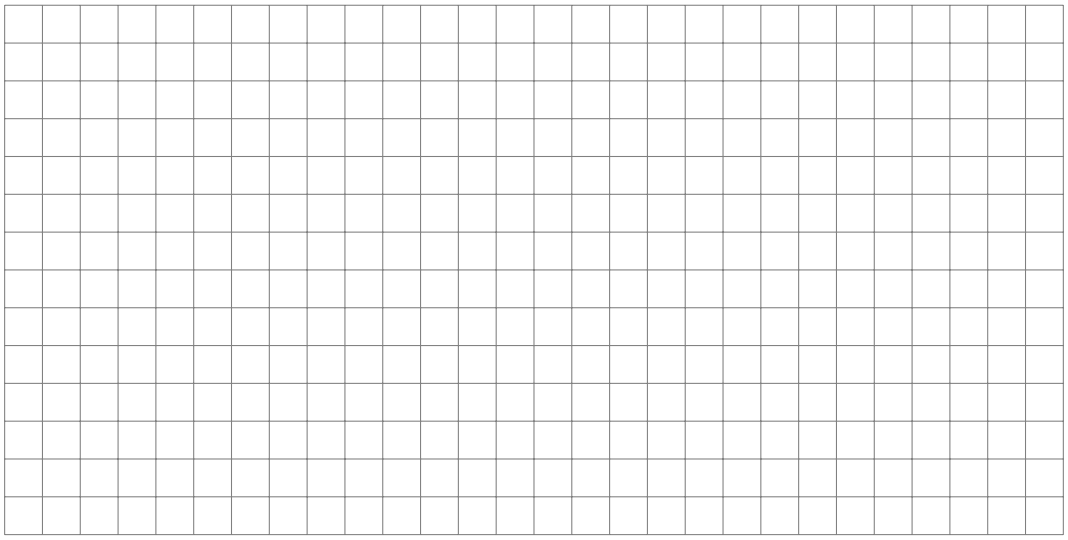
Nutzen Sie die Winkelskala auf dem Tisch, um für 10 verschiedene Stoßparameter b den Messwinkel ϑ zu messen. Führen Sie für jeden Stoßparameter b mehrere Messungen durch und bestimmen Sie eine Messunsicherheit für den Messwinkel ϑ. Berücksichtigen Sie die Einschränkung der Messung durch die Lücke in der Winkelskala für die Abschussvorrichtung. Auch bei Rutherford gab es eine Austrittsöffnung für die α-Teilchen, in der keine Messungen vorgenommen werden konnten. Rechnen Sie den Messwinkel ϑ in den Streuwinkel φ um (siehe Hausaufgabe 4)). Tragen Sie $\varphi(b)$ auf. Werten Sie $cot(\frac{\varphi}{2})$ in Abhängigkeit von b graphisch aus, um zu überprüfen, ob der Rutherford'sche Zusammenhang zwischen Streuwinkel φ und Stoßparameter b auch für den Streuversuch im Magnetfeld gilt.

Aufgabe 6) - Vergleich zum Rutherford'schen Streuversuch

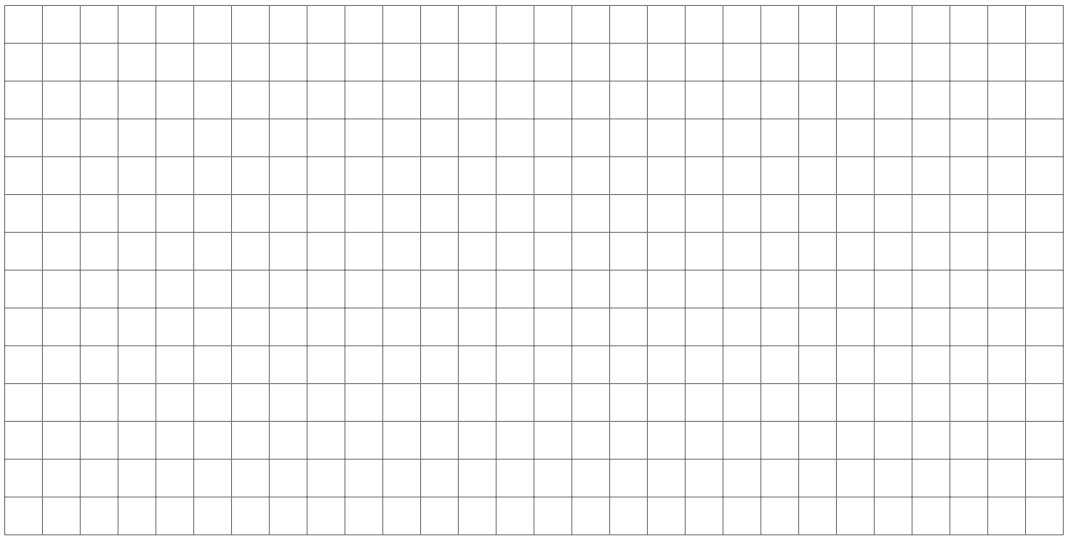
Ergänzen Sie nach der Durchführung dieses Versuches den Vergleich des Rutherford'schen Streuversuches mit der Streuung im Magnetfeld aus Hausaufgabe 2). Zeigen Sie die Grenzen des Streuversuches im Magnetfeld auf.

Fachdidaktische Reflexionsfragen

1. Welche Vorteile bieten Analogie-Versuche wie dieser zum Rutherford'schen Streuversuch für den Schulunterricht?



2. Beurteilen Sie, ob bzw. in welchem Rahmen dieser Analogie-Versuch für Schülerinnen und Schüler der Sekundarstufe II geeignet ist.



A.1.11 Interferenz am Smartphone

Praktikumsteil:	Wintersemester (WS)
Themenbereich:	Optik
Verweis:	Eine Vorarbeit zur Entwicklung des Versuches wurden im Rahmen der Bachelorarbeit von Gersner (2016) durchgeführt.

Versuchsziele und Aufgaben

- Untersuchung der Interferenz am Transmissionsgitter bei senkrechtem Lichteinfall und Bestimmung der Gitterkonstante
- Untersuchung der Interferenz am Smartphone-Display als Reflexionsgitter bei senkrechtem Lichteinfall
- Bestimmung der Gitterkonstante des Smartphone-Displays
- Untersuchung der Interferenz am Smartphone-Display bei schrägem Lichteinfall mit unterschiedlichen Winkeln
- Bestimmung des Fehlers der Gitterkonstante, der durch Vernachlässigung des Einfallswinkels entstehen kann
- Rückschlüsse aus Ergebnisse zum Aufbau des Smartphone-Displays

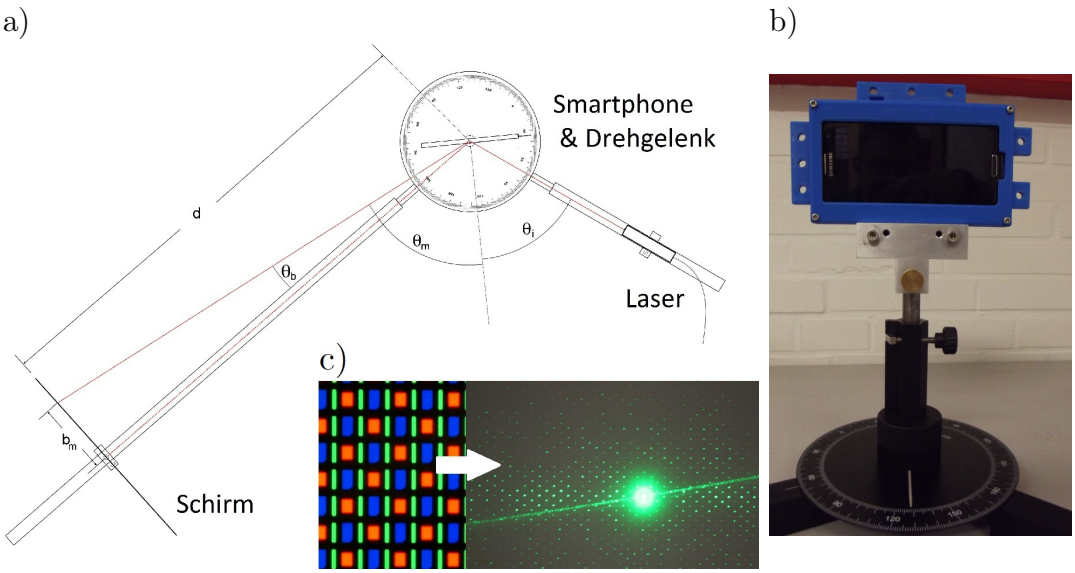


Abbildung 69: *Aufbau zur Untersuchung der Interferenz am Smartphone-Display bestehend aus a) einer optischen Bank mit einem Drehgelenk, an deren einem Ende der Laser und deren anderem Ende ein Schirm befestigt ist (Abb. von Gersner, 2016). b) Das Smartphone befindet sich in einer 3D-gedruckten Experimentierhülle, die mithilfe eines Adapters auf dem Drehgelenk der optischen Bank befestigt werden kann. c) Mikroskopaufnahme des Smartphone-Displays mit den roten, grünen und blauen Subpixeln (links) und das resultierende Interferenzbild bei Nutzung des Displays als Reflexionsgitter (Abb. t.T. von Gersner, 2016).*

Herausforderungen

- Theoretische Betrachtung der Interferenz am Gitter mit schrägem Lichteinfall
- Manuelle Bedienung der Kamera mit entsprechender Belichtungszeit zur Aufnahme von analysierbaren Photos der Interferenzmuster
- Vermessung der photographierten Interferenzbilder mit der Software Tracker (D. Brown & Cox, 2009; D. Brown, 2019)
- Diskussion des Aufbaus des Smartphone-Displays

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	mittel	2
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	schwer	3
Fehlerrechnung	mittel	2
Summe		14

Tabelle 44: *Schwierigkeitsgrad des Versuches „Interferenz am Smartphone“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Integration des Alltagsmediums Smartphone als Experimentiermaterial im Physikunterricht
- Einsatz der schultypischen, kostenfreien Videoanalyse-Software Tracker zur Analyse von Photos
- Vertiefung eines Versuches zu schulischen Lerninhalten der Oberstufe (Interferenz)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiet	Wellen - Welleneigenschaften des Lichts

Schulbezogene Einleitung

„In der Schule wird Interferenz am Gitter gerne an einer CD vertieft. Dieser alltägliche Gegenstand wirkt wie ein Reflexionsgitter. Aus dem Gitterspektrum kann man die Speicherkapazität der CD bestimmen. Aber wie lange wird die CD noch ein gewohntes Objekt aus dem Alltag der Schülerinnen und Schüler sein? Ist Ihnen schonmal aufgefallen, dass Sie ein anderes Reflexionsgitter mit sich tragen: Das Display Ihres Smartphones! Somit kann das Smartphone im Unterricht nicht nur

zu Dokumentationszwecken dienen und motivationale Aspekte einschließen, sondern auch als Element eines optischen Experiments genutzt werden.
Aber wie ist so ein Display überhaupt aufgebaut? Wie hängt das Interferenzmuster mit dem Aufbau des Displays zusammen? Lassen sich die Erfahrungen am Gitter wie zum Beispiel die Gittergleichung auf das Smartphone-Display übertragen? Kann man eine Art Gitterkonstante bestimmen?
Wie kann man das Smartphone als Versuchsobjekt in einen Aufbau auf der optischen Bank integrieren? Die in diesem Versuch eingesetzte Experimentierhülle ermöglicht das Einsetzen des Smartphones in einen Stativhalter. Steht kein 3D-Drucker mit entsprechender Software zur Verfügung, um eine solche Hülle herzustellen, kann das Smartphone auch auf das Pult gelegt werden. Der Laser wird dann schräg auf das Smartphone gerichtet, sodass das Interferenzbild an der Wand oder einem Schirm abgebildet wird. Doch wie ändern sich die Zusammenhänge für den nicht senkrechten Lichteinfall?
All diese Fragen gehen über die Fachanforderungen der Physik (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) hinaus und bieten eine interessante Vertiefung zu dem Thema. Finden Sie die Antworten heraus!“

Fachdidaktische Reflexionsfragen

- In welchem Rahmen und Umfang lässt sich der Versuch im Physikunterricht umsetzten?
- Nennen Sie je ein Beispiel wie die drei prozessbezogene Kompetenzen Erkenntnisgewinnung und Fachmethoden, Kommunikation sowie Bewertung und Reflexion in diesem Versuch gefördert werden.

Klassische Einleitung
„Wird ein Punkt auf einer Wellenfront betrachtet, so ist dieser nach dem Huygen'schen Prinzip der Ausgangspunkt einer neuen Elementarwelle mit gleicher Frequenz wie die ursprüngliche Elementarwelle. Durch Überlagerung aller Elementarwellen ergibt sich die Wellenfront. Trifft eine Welle teilweise auf ein Hindernis, so kann an der Kante des Hindernisses eine Elementarwelle beobachtet werden, welche sich hinter das Hindernis ausbreitet. Man spricht von Beugung.
Mithilfe der Huygen'schen Elementarwellen lässt sich das Beugungsbild eines Lichtbündels am Einfachspalt erklären. Die Wellenfront im Spalt kann als Ausgangslage vieler Elementarwellen aufgefasst werden, die sich überlagern. Die Überlagerung mehrerer kohärenter Wellen wird als Interferenz bezeichnet. Die Amplituden der Wellen können sich dabei nach dem Superpositionsprinzip konstruktiv oder destruktiv addieren. Bei der Betrachtung von mehreren Spalten ergibt sich eine Kombination aus dem Beugungsbild der einzelnen Spalte für sich und dem Interferenzbild, das durch Überlagerung mehrere Lichtbündel aus dem Mehrfachspalt entsteht.
Interferenz tritt an allen Objekten auf, die eine parallele und äquidistante Struktur aufweisen, welche größer als die Wellenlänge des Lichtes ist. Somit ist Interferenz

nicht nur am Einfachspalt, an Doppelspalten oder optischen Gittern beobachtbar. Interferenz tritt auch in alltäglichen Gegenständen auf. Die Interferenz kann an dünnen Schichten wie einem Ölfilm auf einer Wasseroberfläche auftreten. Ein Beispiel für ein Reflexionsgitter aus dem Alltag ist das Display eines Smartphones, welches in diesem Versuchstag untersucht wird. Die Beugung und Interferenz an diesem Gitter wird bei senkrechtem Lichteinfall betrachtet und um den schrägen Lichteinfall erweitert.“

A.1.12 LEGO-Interferometer

Praktikumsteil:	Wintersemester (WS)
Themenbereich:	Optik
Verweis:	Dieser Versuch ist inspiriert von den optischen Experimenten aus LEGO-Bauteilen der Gruppe *myphotonics* von Prof. Dr. Imlau an der Universität Osnabrück (myphotonics, o.D. Linke, 2017).

Versuchsziele und Aufgaben

- Aufbau und Justierung des Michelson-Interferometers aus LEGO[15]-Bausteinen (Abb. 70)
- Bestimmung des Brechungsindex von Luft, in dem ein evakuiertes Rohrstück in einem der Interferometerarme belüftet wird

a) b)

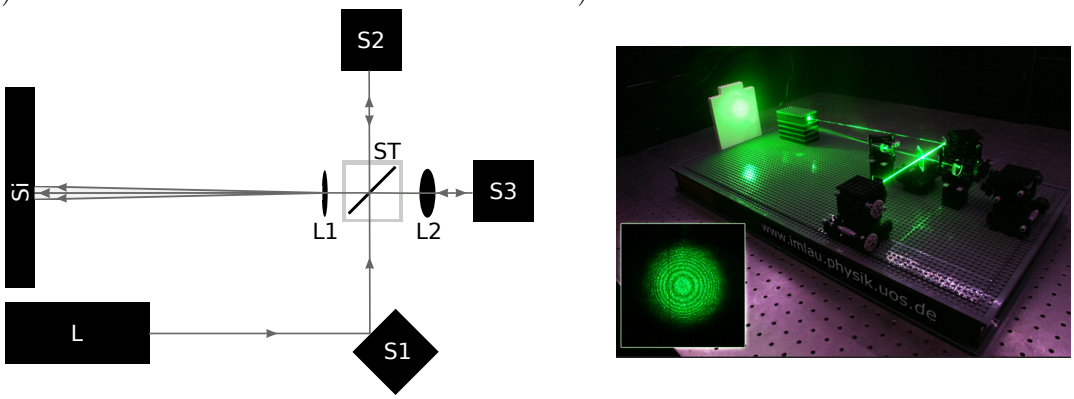


Abbildung 70: *a) Schematischer Aufbau des LEGO-Interferometers mit Laser (L), Schirm (Si), Spiegeln (S), Strahlteiler (ST) und Linsen (L) sowie Skizzierung der Strahlverläufe (Abb. erstellt von M. Marsand). b) Michelson-Interferometer aus LEGO®-Bausteinen. BMBF-Projekt: myphotonics.eu. Quelle: M. Imlau, Universität Osnabrück (myphotonics, o.D.).*

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	schwer	3
Messtechnik	mittel	2
Versuchsdurchführung	schwer	3
Auswertung	leicht	1
Fehlerrechnung	leicht	1
Summe		12

Tabelle 45: *Schwierigkeitsgrad des Versuches „LEGO-Interferometer“ (Erläuterungen in Abschnitt A.1.1)*

[15] LEGO® ist eine eingetragene Marke, welche dieses Projekt nicht unterstützt.

Herausforderungen

- Sorgfältige Justierung des empfindlichen Interferometers aus LEGO
- Erklärung der Beobachtungen (z. B. ringförmiges Interferenzmuster)
- Beurteilung der Eignung und Genauigkeit des LEGO-Aufbaus

Relevanz des Versuchsthemas für die Schule

- LEGO als kostengünstiges Versuchsmaterial als Alternative zu teuren optischen Bauteilen
- Vertiefung eines Versuches zu schulischen Lerninhalten der Oberstufe (Welleneigenschaften des Lichts, Interferenz)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiet	Wellen - Welleneigenschaften des Lichts

Schulbezogene Einleitung

„Schulen stehen oft nur begrenzt Mittel für Ausstattung zur Verfügung. Daher muss man im Physikunterricht nicht selten erfinderisch werden oder in Richtung Open-Source denken. Haben Sie selbst schon alternative Versuchsmaterialien für Experimente zweckentfremdet?

Das Projekt „myphotonics“ an der Universität Osnabrück (myphotonics, o.D.) beschäftigt sich seit 2013 der Umsetzung von optischen Experimenten mit Hilfe von unterschiedlichen Baukastensystemen wie zum Beispiel LEGO. Die Baupläne und Video-Anleitungen sind im Open-Source-Gedanken auf der Webseite der Forschungsgruppe veröffentlicht (myphotonics, o.D.). Das Michelson-Interferometer ist der erste realisierte Aufbau. Bis auf den Laser und die zugehörigen Optiken besteht der Aufbau des Interferometers komplett aus LEGO-Steinen. Fallen Ihnen Experimente ein, die sich mit Bausteinen realisieren lassen?

Die Kosten für dieses Interferometer aus LEGO-Steinen sind fast um den Faktor 10 geringer als bei professionellen Systemen. Somit werden Laboraufbauten für den Schulunterricht erschwinglich. Aber erfüllt der Aufbau aus LEGO wirklich die vorausgesetzten mechanischen Eigenschaften für präzise Messungen mit dem Interferometer?

Ein weiterer Vorteil des Einsatzes der bekannten Bausteine ist der Begeisterungsfaktor. Die Arbeitsgruppe um Prof. Imlau konnte zeigen, dass der motivationale Charakter des LEGO-Aufbaus den Lerneffekt positiv beeinflusst. Hätten Sie nicht auch Spaß daran, einen Versuchsaufbau mit LEGO selbst auszuprobieren?“

Fachdidaktische Reflexionsfragen

- Wie lässt sich das Zusammensetzen der einzelnen Bauteile aus Lego und das Aufbauen eines Michelson-Interferometers effektiv und motivierend im Schulunterricht realisieren?
- Welches Vorwissen sollten sich Schülerinnen und Schüler im Vorfeld eines Versuches mit dem Michelson-Interferometer angeeignet haben?

Klassische Einleitung

„Bis Ende des 19. Jahrhunderts nahmen Naturwissenschaftler an, dass der uns umgebende Raum mit einem ruhenden Stoff sehr geringer Dichte genannt Äther gefüllt sei. Die Theorie des Äthers ermöglichte die Erklärung der Ausbreitung von Licht und anderen elektromagnetischen Wellen. Das Michelson-Interferometer erlangte Bekanntheit als Albert Michelson und Edward Morley die Äthertheorie experimentell belegen wollten. Im Michelson-Morley-Experimente sollte die Geschwindigkeit der Erde relativ zum Äther bestimmt werden. Das Experiment wurde mit einem Interferometer durchgeführt, das ausreichende Genauigkeit bot, um unterschiedliche Lichtgeschwindigkeiten indirekt zu messen. Im Interferometer wird ein Laserstrahl in zwei Teilstrahlen aufgeteilt, sodass diese unterschiedliche Strecken durchlaufen, anschließend wieder zusammengeführt werden und interferieren. Die zwei Teilstrahlen wurden in Bewegungsrichtung der Erde um die Sonne bzw. in Bewegungsrichtung des Ätherwindes und senkrecht dazu ausgerichtet. Allerdings konnten die beiden Wissenschaftler keine Relativgeschwindigkeit der Erde zu einem Äther feststellen, sodass die Gültigkeit der Ätherhypothese in Frage gestellt werden musste. Albert Einstein bot mit seiner speziellen Relativitätstheorie eine alternative Möglichkeit die Ausbreitung elektromagnetischer Wellen zu erklären.“

A.1.13 Augenmodell

Praktikumsteil:	Wintersemester (WS)
Themenbereich:	Optik
Verweis:	Eine Vorarbeit zur Entwicklung des Versuches wurden im Rahmen der Examensarbeit von Weißbach (2013) durchgeführt. Der Versuch ist zudem inspiriert durch den Versuch „Geometrische Optik / Auge“ aus dem adressatenspezifischen Physikpraktikum für Mediziner von Theyßen (2000).

Versuchsziele und Aufgaben

- Untersuchung des Linsensystems des menschlichen Auges anhand eines Modells (Abb. 71)
- Untersuchung des Einflusses der Hornhaut (Brechung an einer gekrümmten Grenzfläche)
- Untersuchung des Einflusses des Kammerwassers (Brechungsgesetz für unterschiedliche Medien)
- Prüfung der Gültigkeit der Linsengleichung und des Abbildungsmaßstabes mit der akkomodationsfähigen Linse
- Bestimmung des Nahpunktes im Modell und am eigenen Augen
- Untersuchung von Fehlsichtigkeiten und deren Korrektur

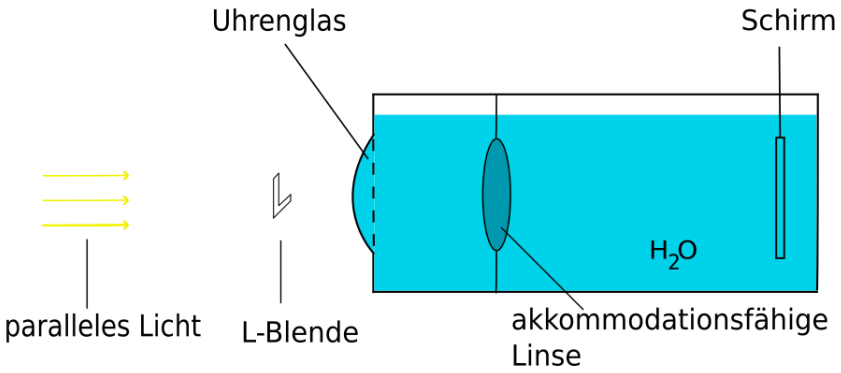


Abbildung 71: *Das Augenmodell besteht aus einem mit Wasser (Kammerwasser) befülltem länglichen transparenten Behälter (Augenkörper), an dessen vorderes Ende ein Uhrenglas (Hornhaut) eingelassen ist. Im Modell befindet sich eine akkomodationsfähige Linse (Augenlinse) und ein Schirm (Netzhaut) befindet (Abb. erstellt von C. Brüning).*

Herausforderungen

- Erzeugung von parallelem Licht mit einer Richtleuchte mit Kondensor und einer Sammellinse
- Einfluss des Uhrenglases (Hornhaut) in Kombination mit dem Wasser (Kammerwasser)
- Variablenkontrolle bei Prüfung von Linsengleichung und Abbildungsmaßstab
- Beurteilung der Güte des Modells für das menschliche Auge

Schwierigkeitsgrad		
Theoretische Grundidee	leicht	1
Messaufbau	mittel	2
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	leicht	1
Summe		10

Tabelle 46: *Schwierigkeitsgrad des Versuches „Augenmodell“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Thematisierung von fächerübergreifender Vernetzung (Physik und Biologie)
- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe I (geometrische Optik)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I
Sachgebiete	Optik - Ausbreitung des Lichts
	Optik - Lichtbrechung und optische Abbildungen

Schulbezogene Einleitung

„Sie kommen aus dem Unterricht ins Lehrerzimmer. Dort spricht ein Kollege aus der Biologie mit einer Kollegin aus der Physik.

Biologie: *„Hey, sag mal, ein Schüler erwähnte gestern, dass das Auge nächste Woche auch Thema in der Physik sein wird."*

Physik: *„Ja, das Thema Auge bietet die Möglichkeit einfach alles Gelernte noch einmal zusammenzubringen - Lichtausbreitung, Brechung, Bildentstehung, optische Geräte wie Linsen und so.“*

Biologie: *„Wir haben letzte Einheit auch über das Auge gesprochen. Hauptsächlich über den Aufbau und die lichtempfindlichen Zellen. Über das optische System konnte ich nicht so viel im Detail sagen.“*

Physik: *„Dafür kann ich nur wenig über diese lichtempfindlichen Zellen sagen... Ergänzt sich dann ja bestens.“*

Biologie: *„Wie machst Du das denn genau? Das Auge ist ja schon sehr komplex aufgebaut mit der Hornhaut, der Augenflüssigkeit, der Augenlinse und dem Glaskörper.“*

Physik: *„In der Physik wird das oft über ein sehr vereinfachtes Linsensystem nachgestellt. All die Elemente, die Du da gerade aufgezählt hast, werden über eine einzelne Sammellinse zusammengefasst. Das Bild eines beleuchteten Gegenstandes wird dann auf einem Schirm abgebildet, der die Netzhaut darstellt.“*

Biologie: *„Oh, das ist natürlich sehr vereinfacht. Wie vermittelt man den Schülerinnen und Schülern denn dabei die Anpassung der Augenlinse an die Entfernung des Gegenstandes? Tauscht man die Linse aus?“*

Physik: *„Das ist eine Möglichkeit. Eine andere Methode ist es, den Abstand zwischen Linse und Schirm für verschiedene Entfernungen des Gegenstandes zu verändern.“*

Biologie: *„Die Augenlinse auszutauschen kommt der Realität dabei ja zumindest etwas näher. Bei der Akkommodation wird die Augenlinse ja durch einen Muskel verändert. Aber die Länge des Augenkörpers ist im Auge fest. Diese zu verändern impliziert ja völlig falsche Vorstellungen. Ein zu langer oder kurzer Augenkörper ist ja vielmehr die Ursache für Fehlsichtigkeiten.“*

Was schließen Sie aus diesem Gespräch? Gibt es eine Möglichkeit, wie man das klassische, physikalische Modell des menschlichen Auges realitätsnäher gestalten kann? Wie kann man den Schülerinnen und Schülern die Anpassungsfähigkeit der Linse, den Einfluss der Hornhaut und der Augenflüssigkeit experimentell veranschaulichen? In diesem Praktikumsversuch werden Sie Antworten finden.“

Fachdidaktische Reflexionsfragen

- Warum lohnt es sich, dem organisatorischen Aufwand zum Trotz das Thema Auge in der Schule in Abstimmung oder sogar gemeinsam mit dem Fach Biologie zu unterrichten?
- Warum ist es didaktisch sinnvoll, das Augenmodell nicht mit einem Laser zu beleuchten, sondern einen beleuchteten Gegenstand zu verwenden?

Klassische Einleitung

„Das Auge ist eines unserer wichtigsten Sinnesorgane und ein erstaunliches Zusammenwirken von Physik und Biologie. Es besteht aus einem Organ mit einem Linsensystem. Das optische System ermöglicht das Abbilden von Gegenständen auf der Netzhaut, auf der sich lichtempfindliche Zellen befinden. In der Physik wird das Linsensystem des menschlichen Auges oft als einzelne Sammellinse vereinfacht zusammengefasst. Dies wird dem Strahlengang durch Hornhaut, Augenflüssigkeit, Augenlinse und Glaskörper allerdings im Detail nicht gerecht. Die Krümmung der Augenlinse ist über einen Muskel veränderbar, sodass Gegenstände in unterschiedlichen Entfernungen auf der Netzhaut scharf gestellt werden können. Die Iris mit der Pupille wirkt wie eine Blende mit variabler Öffnung, die die Intensität des einfallenden Lichtes regelt. Die lichtempfindlichen Zellen ermöglichen die Wahrnehmung von Farben im Wellenlängenbereich von 380 nm bis 780 nm. Bei Fehlfunktionen des Auges kann das Verständnis der geometrischer Optik genutzt werden, um durch Erweiterung des Linsensystems durch Brillen Fehlsichtigkeiten zu korrigieren.“

A.1.14 Reflexion, Brechung und Polarisation

Praktikumsteil: Wintersemester (WS)
Themenbereich: Optik

Versuchsziele und Aufgaben

- Reflexion und Brechung: Nachweis des Reflexions- und das Brechungsgesetz mithilfe eines Lasers und einem halbkreisförmigen Plexiglas-Körper, sowie Bestimmung des kritischen Winkels, ab dem Totalreflexion stattfindet, und des Brechungsindexes des Plexiglases
- Polarisation: Überprüfung des Gesetzes von Malus mithilfe eines Lasers, eines Polarisators und dem CASSY Lux-Sensor (LD-Didactic, o.D. a)
- Polarisation: Untersuchung der optischen Aktivität von Zuckerlösungen und Quantifizierung des Zusammenhangs zwischen Drehwinkel der Polarisationsrichtung und Konzentration der Zuckerlösung, sowie Bestimmung der Zuckerkonzentration einer Limonade daraus (Abb. 72)

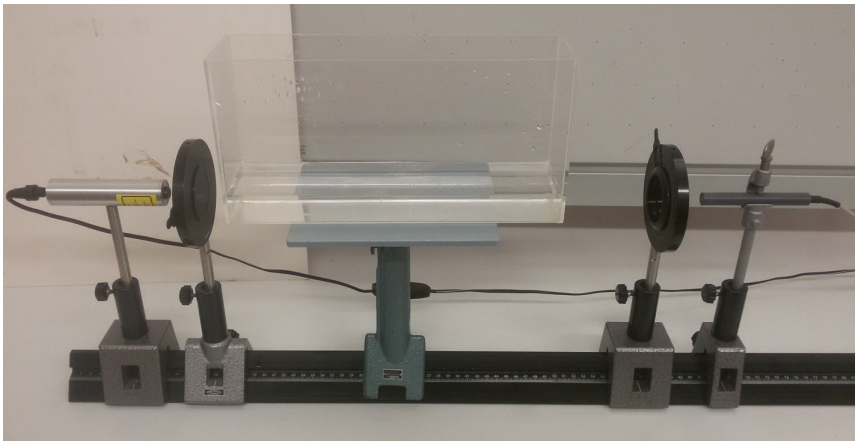


Abbildung 72: *Aufbau zum Versuchsteil zur optischen Aktivität von Zucker: Auf der optischen Bank ist ein Laser auf einen optionalen Polarisator gerichtet. Rechts daneben befindet sich ein transparentes Gefäß mit einer Zuckerlösung. Hinter dem Gefäß trifft der Laserstrahl auf einen Analysator und anschließend auf den CASSY Lux-Sensor.*

Herausforderungen

- Theorie zur Polarisation
- Nutzung des CASSY Lux-Sensors
- Systematische Variablenvariation im Teamwork
- Freie Kurvenanpassung und Interpretation der Parameter für die Prüfung des Gesetzes von Malus

Schwierigkeitsgrad		
Theoretische Grundidee	leicht	2
Messaufbau	mittel	2
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	mittel	2
Summe		12

Tabelle 47: *Schwierigkeitsgrad des Versuches „Reflexion, Brechung und Polarisation“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe I (Reflexion und Brechung) und II (Polarisation)
- Einsatz des schultypischen Messwerterfassungssystems CASSY mit dem Lux-Sensor

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I und II
Sachgebiete	Optik - Reflexion an ebenen Flächen
	Optik - Lichtbrechung und optische Abbildungen
	Wellen - Welleneigenschaften des Lichts

Schulbezogene Einleitung

„Während Reflexion und Brechung schon lange Teile des Lehrplanes sind, wurde das Thema Polarisation nun in den aktuellen Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) neu aufgenommen. Ob die neue 3D-Technik im Kino das Thema Polarisation in das Interessenfeld der Schülerinnen und Schüler (sowie Lehrkräfte) gerückt hat?
Wie kann man den Schülerinnen und Schülern Polarisation als nicht sofort sichtbare Eigenschaft einer Welle nun näher bringen? Zur Veranschaulichung der Polarisation von elektromagnetischen Wellen werden oft Analogien zu mechanischen Wellen gezogen. Karlheinz Meier von der Universität Heidelberg (Meier, 2011) zeigt in einem seiner Motivationsvideos für den Physikunterricht eine Welle in einem Gummischlauch, der durch einen Schlitz senkrecht bzw. parallel zur Auslenkung läuft. Das Drahtgittermodell in Abbildung 73, durch das nur mechanische Wellen mit entsprechender Ausrichtung durch passen, erfüllt ebenfalls diese Analogie. Lässt sich die Polarisation mit einem Gitter auch mit Schallwellen veranschaulichen? Die Antwort können Sie sich selbst geben, wenn Sie sich Schall als Longitudinalwelle vor Augen

rufen. Bei Demonstrationsversuchen in der Lehre werden oft Drahtgitter zur Polarisation von Mikrowellen eingesetzt. Diese Polarisatoren bestehen aus parallelen, leitenden Metalldrähten, die je nach Ausrichtung das elektrische Feld durchlassen oder reflektieren. Aber warum können sogenannte Drahtgitter-Polarisatoren nicht bei sichtbarem Licht eingesetzt werden? Da Abstand und Durchmesser der Drähte klein sein müssen gegenüber der Wellenlänge, sind die erforderlichen Gitterabstände für geringe Wellenlängen nicht mehr zu realisieren. In diesem Fall werden sogenannte Polarisationsfolien genutzt. Wie genau funktionieren diese Polarisationsfilter? Und welche Phänomene sind bei polarisiertem sichtbaren Licht denn nun beobachtbar? Finden Sie es in diesem Versuch heraus!"

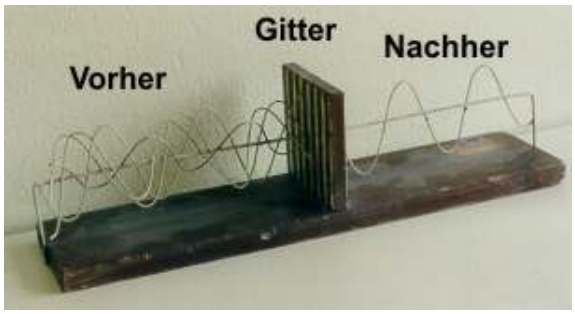


Abbildung 73: *Analogie der Polarisation zu mechanischen Wellen mit Drahtmodell (Abb. von der Joachim Herz Stiftung, www. leifiphysik.de, (Joachim Herz Stiftung, o.D.)).*

Fachdidaktische Reflexionsfragen

- Wäre es sinnvoll Polarisation im Physikunterricht mit einem Besuch im 3D-Kino einzuleiten? Welchen Einstieg in das Thema würden Sie wählen?
- Welche Schwierigkeit kann bei den Schülern und Schülerinnen auftauchen, wenn Polarisation mithilfe von Gittern an Mikrowellen veranschaulicht wird?

Klassische Einleitung

„Elektromagnetische Wellen sind Transversalwellen, deren elektrische und magnetische Feldstärke senkrecht zueinander und senkrecht zur Ausbreitungsrichtung schwingen. Diese elektromagnetischen Wellen wechselwirken mit Materie. Das elektrische Feld der Welle regt die Atomelektronen des Mediums zu Schwingungen an. Dadurch ändert sich der Abstand zwischen negativ geladenem Elektron und dem positiv geladenem Atomkern und somit auch das elektrische Feld zwischen den beiden. Man spricht von elektrischen Dipolschwingungen. Dies führt zur Aussendung von Sekundärstrahlung. In einem unendlich ausgedehnten, homogenen, isotropen Medium haben die Sekundärwellen hauptsächlich die gleiche Ausbreitungsrichtung wie die der anregenden Welle, da seitwärts oder rückwärts gerichtete Strahlung zu Null interferiert. Trifft eine ebene Welle auf eine Grenzfläche zwischen zwei begrenzte Medien

mit unterschiedlichen Brechungsindizes, so wird die Interferenz gestört. Dadurch resultiert an den Grenzflächen eine „Rückwärtsstrahlung“ der einfallenden Welle, die als Reflexion bezeichnet wird, und eine Richtungsänderung der „Vorwärtsstrahlung“, bei der man von Brechung spricht. Die Intensitäten der reflektierten und gebrochenen Strahlen sind abhängig von der Polarisation der Welle. Eine elektromagnetische Welle heißt beispielsweise linear polarisiert, wenn die elektrische Feldstärke in jedem Raumpunkt in der gleichen Ebene schwingt. Durch Reflexion können zudem auch Wellen mit bevorzugter Polarisationsrichtung erzeugt werden.“

A.1.15 Thermohaus

Praktikumsteil:	Wintersemester (WS) und Sommersemester (SS)
Themenbereich:	Wärmelehre
Verweis:	Eine Vorarbeit zur Entwicklung des Versuches wurden im Rahmen der Bachelorarbeit von Siebert (2016) durchgeführt.

Versuchsziele und Aufgaben

- Untersuchung des Strahlungsverhaltens unterschiedlich gestalteter Wände eines geheizten Modellhauses mit einer Infrarot-Kamera (Abb. 74)
  - Untersuchung des zeitlichen Temperaturverlaufes einer Wand mit unterschiedlich dickem Isoliermaterial
  - Untersuchung der geometrischen Einflüsse von Außen- und Innenecken
  - Untersuchung der räumlichen Auflösung der Infrarot-Kamera anhand unterschiedlich großer Defekte in einer Wand
  - Untersuchung des Emissions- bzw. Reflexionsverhalten unterschiedlicher Teilflächen (Metall, Glas, Holz, Gips, ...)
- Interpretation von realen Thermogrammen anhand der gewonnenen Erkenntnisse

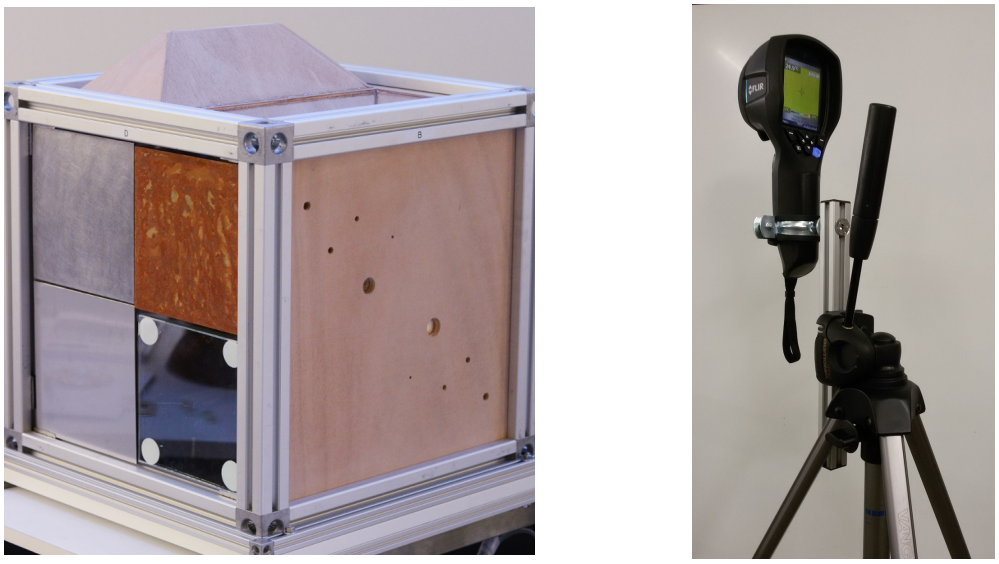


Abbildung 74: *Aufbau zum Versuch Thermohaus bestehend aus einem aufheizbares Modellhaus mit unterschiedliche gestalteten Wandflächen (links) und der Infrarot-Kamera im Stativ (rechts) (Photos von Siebert, 2016).*

Herausforderungen

- Funktion der Infrarot-Kamera
- Einsatz der freien Auswerte-Software für Thermogramme von FLIR
- Diskussion der Möglichkeiten und Schwierigkeiten bei der Auswertung von thermographischen Messungen (z. B. Einfluss Emissionsgrad, geometrische Strukturen, räumliche Auflösung)

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	mittel	2
Messtechnik	schwer	3
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	leicht	1
Summe		13

Tabelle 48: *Schwierigkeitsgrad des Versuches „Thermohaus“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Einarbeitung in die Nutzung einer Infrarot-Kamera mit entsprechender Software, die als moderne Messtechnik immer mehr auch in Schulen Einzug erhält
- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe I (Wärmetransport, Wärmestrahlung)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I
Sachgebiet	Wärme - Temperatur und Wärmetransport

Schulbezogene Einleitung

„Thermographie ist zur Zeit in aller Munde. Alle kennen die typischen blau-roten Falschfarbenbilder und sind zugleich verwundert und begeistert, über die Möglichkeit kontaktlos die Temperatur von Objekten zu vermessen. Und dann ist das Ergebnis auch noch ganz intuitiv zu verstehen, zeigt die Aufnahme doch direkt eine räumliche Temperaturverteilung. Typische Anwendungsbeispiele sind im privaten Umfeld die Analyse der Wärmedämmung von Häusern. Im industriellen Rahmen wird die Thermographie z. B. zur Fehlersuche in elektrischen Schaltkreisen eingesetzt. Sogar in der Schule kommen bereits die ersten Wärmebildkameras zum Einsatz. Entsprechende Aufsätze für Smartphones und günstige Einstiegsmodelle machen es möglich.

Es ist also nur eine Frage der Zeit, wann Sie die erste Wärmebildkamera in Händen halten werden und damit Experimente im Physikunterricht durchführen wollen. Doch wie funktionieren diese Kameras eigentlich? Darf man die Bilder genauso interpretieren wie gewöhnliche Fotografien? Sind die erhaltenen Temperaturverteilungen wirklich so intuitiv zu interpretieren, wie man auf dem ersten Blick meint? Ist die Physik von Infrarotstrahlung und sichtbarem Licht gleich oder gibt es Unterschiede? Darf man aus dem Bild im Sichtbaren direkt auf das Infrarotbild schließen? Ist die Temperaturmessung immer korrekt oder gibt es Messfehler bzw. systematische Abweichungen? Vielleicht ist die Analyse von Wärmebildern ja viel komplizierter als gedacht? Dieser Versuch gibt ihnen die Möglichkeit, genau diesen Fragen anhand von thermographischen Messungen an einem Modellhaus nachzugehen.“

Fachdidaktische Reflexionsfragen

- Wie würden Sie eine IR-Kamera in der Sekundarstufe I einführen?
- In welche Versuche können Sie die bereits vorgestellte IR-Kamera in der Schule einbinden? Welche Vorteile sehen Sie darin?

Klassische Einleitung

„Die Temperatur eines Körpers ist ein Maß für die kinetische Energie der ungeordneten Bewegung der Atome und Moleküle. Aufgrund der Wärmebewegung seiner Teilchen strahlt der Körper elektromagnetische Wellen mit temperaturabhängigen Frequenzen aus. Diese werden als Wärmestrahlung bezeichnet. Wärmestrahlung ist eine Form des Energietransportes, sodass andersherum elektromagnetische Wellen thermische Schwingungen in einem Körper anregen können. Das Plancksche Gesetz bildet die Grundlage der Beschreibung dieses Zusammenhanges von Temperatur und Wärmestrahlung. Dieser Zusammenhang wird bei Strahlungsthermographie ausgenutzt, um aus der Wärmestrahlung des Körpers im Infrarotbereich auf die Oberflächentemperatur einer Körpers schließen zu können. Thermographische Aufnahmen von Körpern werden mit Kameras mit infrarotempfindlichen Sensoren gemacht.“

A.1.16 Wärmeäquivalent

Praktikumsteil:	Wintersemester (WS) und Sommersemester (SS)
Themenbereich:	Wärmelehre
Verweis:	Der Versuch ist inspiriert durch den Hersteller der Experimentiermaterialien LD-Didactic (o.D. b).

Versuchsziele und Aufgaben

- Mechanische Arbeit: (Aufbau und Durchführung ziehe Abb.75)
  - Prüfung und Diskussion, ob die dem System zugeführte mechanische Arbeit durch das Kurbeln gleich der Zunahme der inneren Energie in Form von Wärme bestimmt über die Temperatur ist
  - Bestimmung der spezifischen Wärmekapazität von Aluminium und Wasser
- Elektrische Arbeit: (Abb.76)
  - Prüfung und Diskussion, ob die dem System zugeführte elektrische Arbeit über den Heizdraht gleich der Zunahme der inneren Energie in Form von Wärme bestimmt über die Temperatur ist
  - Bestimmung der spezifischen Wärmekapazität von Aluminium
- Vergleich der Ergebnisse der beiden Verfahren

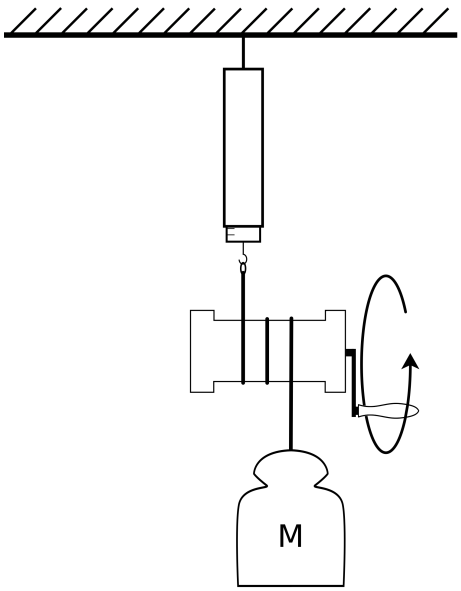


Abbildung 75: *Mechanischer Versuchsteil: Einer von drei metallischen Test-Zylindern wird in eine Kurbel-Konstruktion mit Umdrehungszähler eingehängt. Um den Test-Zylinder wird mit mehreren Wicklungen eine Schnur geschlungen, die mit einem Gewicht von 5 kg belastet ist. Das andere Ende ist an einer Federwaage befestigt. Der Metallzylinder wird mithilfe der Kurbel gedreht. Dabei hebt sich das Gewichtsstück auf eine konstanten Höhe. Der Temperaturanstieg des Zylinders wird in regelmäßigen Abständen mit einem Thermometer bestimmt (Abb. erstellt von M. Marsand nach Phywe, o.D.).*

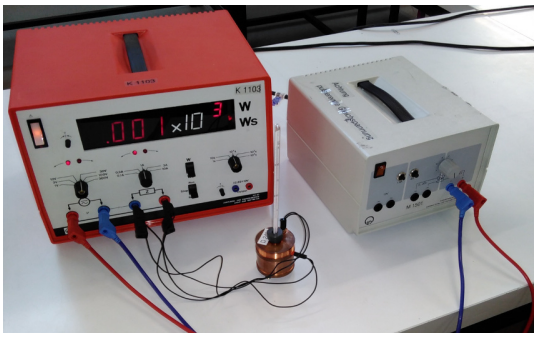


Abbildung 76: *Elektrischer Versuchsteil: Der Testzylinder wird über eine Heizwicklung aufgewärmt. Gleichzeitig wird die dafür aufgewendete elektrische Arbeit über einen Leistungsmesser bestimmt und die Temperatur des Zylinders gemessen.*

Herausforderungen

- Mechanischer Versuchsteil: Kräftebetrachtung und Berechnung der mechanischen Arbeit
- Auswertung mittels lineare Regression und zugehörige Fehlerrechnung
- Methodenvergleich

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	mittel	2
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	mittel	2
Summe		12

Tabelle 49: *Schwierigkeitsgrad des Versuches „Wärmeäquivalent“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe I (Temperatur, Wärme, Energie)
- Diskussion der Schülerfehlvorstellung zu den Begrifflichkeiten Temperatur und Wärme
- Vertiefung des Basiskonzeptes Energie

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I
Sachgebiete	Wärme - Temperatur und Wärmetransport
	Energie - Qualitativer und Quantitativer Energiebegriff
Basiskonzept	Energie

Schulbezogene Einleitung

„Eine Schülerin und ein Schüler stören mit immer lauterem Getuschel dem Physikunterricht. Die beiden Unruhestifter werden von der Lehrkraft darauf angesprochen und reagieren wie folgt:

„Wir verstehen das einfach nicht! Diesen Unterschied zwischen Temperatur und Wärme. Die Temperatur ist ein Maß für die Energie der Teilchen. Ok. Wärme ist auch Energie, die man transportieren kann. Ok. Aber das soll nicht das gleiche sein!? Wenn etwas warm wird, dann erhöht sich die Temperatur, oder nicht?“

„Genau, das verstehe ich auch nicht. Wenn ich die Hände aneinander reibe und die Hände warm werden, teilt sich die Energie dann in Wärme und Temperatur auf? Sie haben ja gesagt, dass das nicht dasselbe ist... Und die Energie soll ja erhalten bleiben, die kann sich dann ja nicht einfach verdoppeln!“

Könnten Sie den Schülerinnen und Schülern helfen, die Begriffe zu unterscheiden? Was ist denn nun Wärme? Und wie hängt sie mit der Temperatur zusammen? Und wenn Sie weiter denken: Was ist dann innere Energie? Und was passiert in Bezug auf Energien genau, wenn Sie die Hände aufeinander reiben? Dieser Versuch kann Ihnen helfen firm mit diesen Begriffen zu werden.“

Fachdidaktische Reflexionsfragen

- Wie könnten Sie auf die Schülerfrage „Wo ist denn der Unterschied zwischen Temperatur und Wärme?“ eingehen?
- Wie können die vier Aspekte des Basiskonzeptes Energie aus den neuen Fachanforderungen beim Thema Wärme aufgegriffen werden?

Klassische Einleitung

„Energie ist eine der wichtigsten physikalischen Zustandsgrößen. Sie ist als die Fähigkeit eines Systems definiert, Arbeit zu verrichten. Energie kann in unterschiedlichen Formen erscheinen, zwischen Systemen transportiert werden und ist umwandelbar. Wird an einem Körper beispielsweise mechanische Arbeit verrichtet, wird mindestens ein Teil der mechanischen Energie in Wärme umgewandelt. Wärmeenergie wird zwischen Körpern bzw. Systemen mit unterschiedlichen Temperaturen übertragen. Die Wärmemenge sowie die Wärmekapazität des Körpers bestimmen, wie groß die Änderung der Temperatur des Körper ist. Energie ist in abgeschlossenen Systemen eine Erhaltungsgröße. In der Thermodynamik findet die Energieerhaltung im ersten Hauptsatz der Thermodynamik eine spezifizierte Formulierung. Demnach kann die innere Energie eines Systems nur verändert werden, indem Energie durch Arbeit bzw. Wärme in das System eingebracht wird.“

A.1.17 Wärmepumpe

Praktikumsteil:	Wintersemester (WS) und Sommersemester (SS)
Themenbereich:	Wärmelehre
Verweis:	Der Versuch zur Wärmepumpe ist eine lehramtsspezifische Adaption eines Versuches aus dem Anfängerpraktikum für 1-Fach-Physiker von Dr. Victor de Manuel Gonzales an der Christian-Albrechts-Universität zu Kiel. Der Versuch ist inspiriert durch den Hersteller der Wärmepumpe LD-Didactic (o.D. c).

Versuchsziele und Aufgaben

- Messung der Temperatur des Kaltwasserbehälters, die Temperatur des Warmwasserbehälters sowie die Leistung P des Kompressors bei Betrieb der Wärmepumpe (Abb. 77) über ca. 25 min a) ohne Wärmezufuhr auf der Verdampferseite und b) bei konstant gehaltener Temperatur auf der Verdampferseite
- Graphische Auftragung der Temperaturen und Bestimmung der Leistungszahl der Kompressionswärmepumpe
- Untersuchung des Einflusses der Temperaturdifferenz zwischen Verdampfungs- und Verflüssigerseite auf die Leistungszahl der Kompressionswärmepumpe

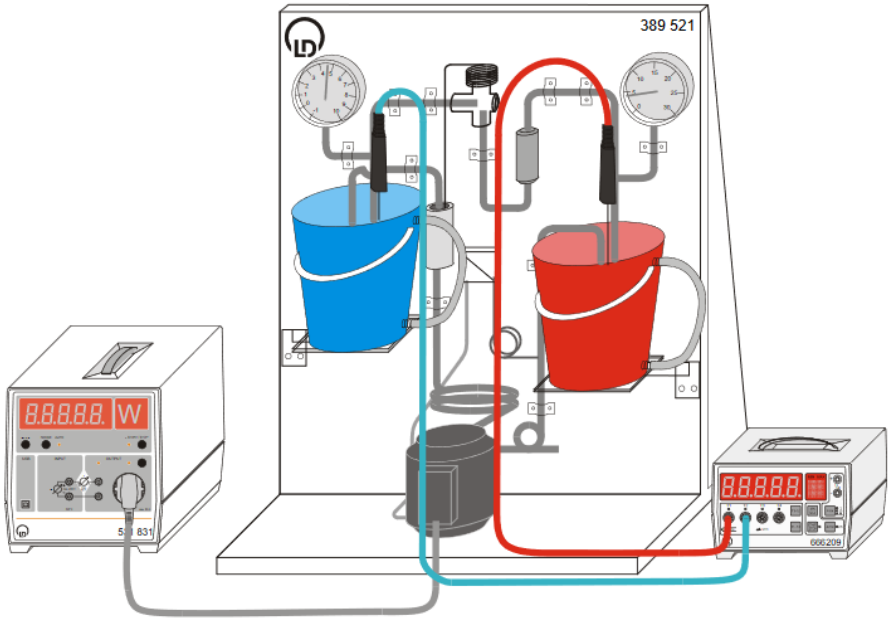


Abbildung 77: *Aufbau zum Versuch Wärmepumpe bestehend aus einer Kompressionswärmepumpe mit zwei Wasserreservioren auf der Verdampferseite links und der Verflüssigerseite rechts. Mithilfe von zwei Temperaturfühlern wird die Temperatur des Wasserreserviors bestimmt. Mithilfe eines Leistungsmessgerätes wird die Leistung des Kompressors bestimmt (Abb. vom Hersteller der Wärmepumpe LD-Didactic (o.D. c), Copyright: © LEYBOLD / LD DIDACTIC GmbH / www.ld-didactic.de, Hürth)*

Herausforderungen

- Verständnis des Kreisprozesses, der Funktion der Wärmepumpe und der Bedeutung der Leistungszahl

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	mittel	2
Messtechnik	leicht	1
Versuchsdurchführung	leicht	1
Auswertung	mittel	2
Fehlerrechnung	leicht	1
Summe		10

Tabelle 50: *Schwierigkeitsgrad des Versuches „Wärmepumpe“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe I (Temperatur, Wärme, Energie)
- Vertiefung des Basiskonzeptes Energie

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I
Sachgebiete	Wärme - Temperatur und Wärmetransport
	Energie - Qualitativer und Quantitativer Energiebegriff
Basiskonzept	Energie

Schulbezogene Einleitung

„Die Basiskonzepte der Physik werden in den Fachanforderungen des Landes Schleswig-Holstein (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) als didaktisches Instrument zum Aufbaus einer vernetzten Wissensbasis innerhalb der verschiedenen Sachgebiete der Physik vorgestellt. Zum kumulativen Kompetenzaufbau bei den Schülerinnen und Schülern werden in der Physik die Basiskonzepte Energie, Materie, System und Wechselwirkung genutzt. Das Basiskonzept Energie wird dabei in die vier Teilgebiete Energieformen, Energieumwandlung bzw. -transport, Energieentwertung und Energieerhaltung unterteilt.

Doch wie sieht dieses theoretische Basiskonzept Energie nun angewendet auf ein physikalisches Beispiel aus? Der Versuch zum Wirkungsgrad einer Wärmepumpe eignet sich gut, sich die Aspekte des Basiskonzeptes Energie zu verdeutlichen. Welche Energieformen treten auf? Inwieweit verändern sich Erscheinungsform oder -ort? Welcher Energieanteil ist nutzbar? Wo wird das Energieerhaltungsprinzip deutlich?

Inwiefern sind auch die anderen Basiskonzepte zu Materie, System und Wechselwirkung im Versuchsbeispiel vertreten?“

Fachdidaktische Reflexionsfragen

- Wie können die vier Aspekte des Basiskonzeptes Energie aus den Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) in diesem Versuch zum Thema Wirkungsgrad einer Wärmepumpe aufgegriffen werden?
- Nennen Sie beispielhaft wie Sie methodisch vorgehen könnten, um eine Vernetzung in Hinsicht auf das Basiskonzept Energie zwischen dem Versuch zur Wärmepumpe und einem zurückliegenden, beispielsweise mechanischen Versuch zu erreichen? Welches Problem kann dabei im Schulunterricht auftauchen?

Klassische Einleitung

„Unter idealen, thermodynamischen Maschinen versteht man periodisch arbeitende Maschinen, die unter Aufwand oder Gewinn von mechanischer Arbeit, Energie in Form von Wärme zwischen zwei Wärmebädern transportieren. Eine Wärmekraftmaschine entnimmt einem Wärmebad höherer Temperatur Wärme und wandelt einen Teil in mechanische Arbeit (Abb. 78). Nach der Kelvinschen Aussage bleibt jedoch stets ein Rest Wärme, der an ein Wärmebad geringerer Temperatur abgegeben werden muss (2. Hauptsatz der Thermodynamik). Der maximal erreichbare Leistungsgrad einer Wärmekraftmaschine ist der des Carnotschen Kreisprozesses.
Eine Maschine, die Energie von einem Wärmebad geringerer Temperatur zu einem Wärmebad höherer Temperatur transportiert bezeichnet man aus Sicht des ersten Wärmebades als Kältemaschine und aus Sicht des zweiten als Wärmepumpe. Wärmepumpen funktionieren meist nach folgendem Prinzip: Auf der Verdampferseite wird das flüssige Betriebsmittel der Wärmepumpe durch Herabsetzen des Druckes zum Verdampfen gezwungen. Die hierzu nötige Verdampfungswärme wird der Umgebung entnommen, so dass es zu einer Temperaturerniedrigung kommt. Das nun gasförmige Betriebsmittel wird durch einen Kompressor abgesaugt und unter Erhöhung des Druckes auf die Verflüssigerseite transportiert. Der Druck ist hier derart, dass das Betriebsmittel kondensiert. Die hierbei wieder abgegebene Verdampfungswärme führt zu einer Erwärmung.“

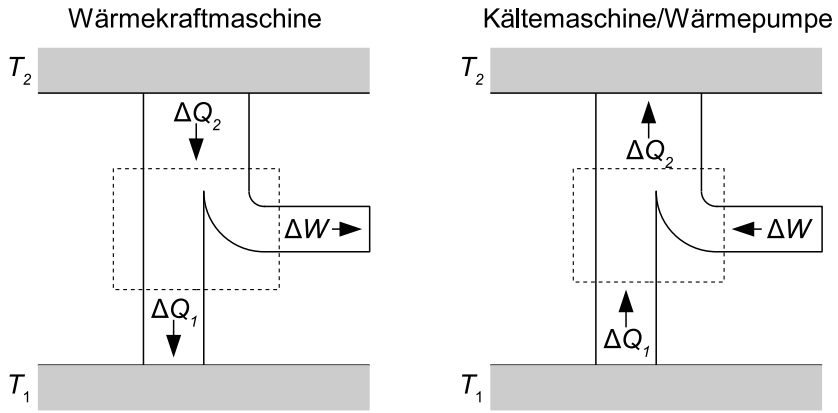


Abbildung 78: *Schematische Zeichnung einer idealen Wärmekraftmaschine und Wärmepumpe. T sei dabei die Temperatur des Reserviors mit $T_2 > T_1$, ΔQ die Wärmemenge und ΔW die Arbeit (Abb. aus der Anleitung des Anfängerpraktikums von Dr. Victor de Manuel Gonzales).*

A.1.18 Schwarzer Strahler

Praktikumsteil: Wintersemester (WS) und Sommersemester (SS)
Themenbereich: Wärmelehre

Versuchsziele und Aufgaben

- Planung und Durchführung einer Messung zur Überprüfung der Hypothese, dass sich eine Glühbirne wie ein schwarzer Strahler verhält
- Vergleich der Glühbirne mit einem schwarzen Strahler
- Diese offene Aufgabenstellung kann unter Nutzung des gegebenen CASSY-USB-Spektrometers (LD-Didactic, o.D. a) folgendermaßen bearbeitet werden:
  - Vermessung des Spektrums einer Glühbirne bei unterschiedlichen Helligkeiten
  - Qualitativer und z. T. quantitativer Vergleich der Spektren der Glühbirne mit dem Planckschen Strahlungsspektren, mit dem Wienschen Verschiebungsgesetz und dem Stefan-Boltzmann-Gesetz
  - Vergleich des Spektrums der Glühbirne mit dem Spektrum des Sonnenlichts

Herausforderungen

- Eigenständiger Versuchsplanung zur Bearbeitung der offenen Aufgabenstellung
- Identifikation der Eigenschaften der Eigenschaften von schwarzen Strahlern anhand der Strahlungsgesetze
- Einarbeitung in die Nutzung des CASSY Spektrometers und die zugehörige Software CASSY Spectra Lab

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	mittel	2
Messtechnik	schwer	3
Versuchsdurchführung	mittel	2
Auswertung	schwer	3
Fehlerrechnung	leicht	1
Summe		14

Tabelle 51: *Schwierigkeitsgrad des Versuches „Schwarzer Strahler“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Nutzung des CASSY Spektrometers und der zugehörigen Software CASSY Spectra Lab
- Verknüpfung und Vertiefung von schulischen Lerninhalten der Sekundarstufe I (Wärmestrahlung) und II (Spektren)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I und II
Sachgebiete	Wärme - Temperatur und Wärmetransport
	Wellen - Spektren

Schulbezogene Einleitung

„In einer Sitzung der Kolleginnen und Kollegen aus dem Fach Physik, soll über die diesjährige Nutzung des Budgets für die Physiksammlung entschieden werden. Unter anderem schlägt eine Person aus dem Kollegenkreis die Anschaffung eines Hohlraumstrahlers mit Heizofen zur Untersuchung der Eigenschaften eines schwarzen Körper vor. Eine Reaktion darauf ist folgende:
„Das ist zum einen viel zu teuer und zum anderen auch komplett unnötig. Man kann als schwarzen Strahler doch auch einfach eine leuchtende Glühbirne verwenden.“
Was würden Sie dazu sagen?“

Fachdidaktische Reflexionsfragen

- In welche motivierenden Kontexte können Sie den Versuch einbinden?
- Mit welchem Vorwissen können Sie rechnen?

Klassische Einleitung

„Als schwarzen Körper bezeichnet man in der Physik einen idealen Strahler. Die Strahlung, die auf einen schwarzen Strahler trifft, wird vollständig absorbiert und wieder emittiert. Die Annäherung eines schwarzen Strahlers mittels eines Hohlraums nimmt eine wichtige Rolle bei der Untersuchung von Wärmestrahlung ein. Dazu wird ein Hohlkörper aus strahlungsundurchlässigem Material mit einer Öffnung, die im Vergleich zum Volumen des Hohlraums sehr klein ist, genutzt. Die durch die kleine Öffnung eintreffende Strahlung wird im inneren des Hohlkörpers wiederholt reflektiert, teilweise absorbiert und emittiert, sodass die Strahlung mit hoher Wahrscheinlichkeit vollständig absorbiert wird. Dadurch stellt sich ein thermisches Gleichgewicht ein. Die austretende Strahlung ist somit charakteristisch für die Temperatur der Wände des Hohlkörpers. Nicht selten wird auch die Glühbirne als alternative Annäherung für einen schwarzen Körper diskutiert.“

A.1.19 Newtonsche Axiome

Praktikumsteil: Sommersemester (SS)
Themenbereich: Mechanik

Versuchsziele und Aufgaben

- Kalibrierung der Luftkissenbahn als zentrales Element des Versuchsaufbaus, um möglichst Reibungsfreiheit zu erreichen
- Prüfung des Trägheitsgesetzes (Newton I) durch Messung und Vergleich der Geschwindigkeiten eines Gleiters an verschiedenen Positionen auf der Luftkissenbahn mittels Lichtschranken
- Prüfung des Aktionsprinzips (Newton II) durch Messung der Weg-Zeit-Kurven eines beschleunigten Gleiters mittels Bewegungsmesswandler (Abb. 79) und Bestimmung der Beschleunigungen und beschleunigten Massen
- Prüfung des Reaktionsprinzips (Newton III) anhand zwei aufeinander stoßenden Gleitern mit Kraftsensoren

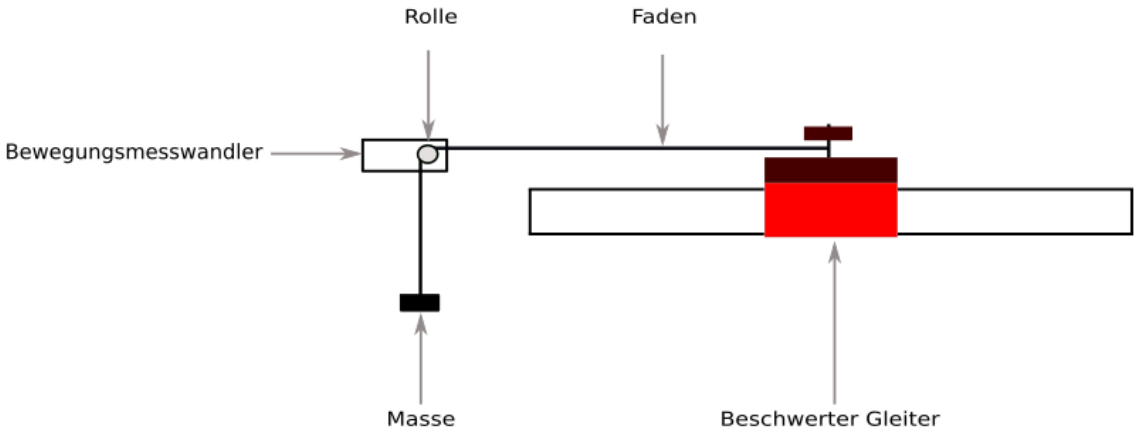


Abbildung 79: *Aufbau zum Versuch Newtonsche Axiome zur Prüfung des Aktionsprinzips (Newton II) bestehend aus einer Luftkissenbahn mit Gleiter. Der Gleiter ist an einem Faden befestigt, der über die Rolle des Bewegungsmesswandlers umgelenkt wird und am Ende mit einem Gewicht beschwert ist (Abb. erstellt von C. Brüning).*

Herausforderungen

- Präzises Messen durch sorgfältige Ausrichtung des Ausbaus, insbesondere der Luftkissenbahn und der Kraftsensoren
- Zielgerichteter Einsatz und Umgang mit verschiedenen CASSY Sensoren (LD-Didactic, o.D. a) (Lichtschranken, Bewegungsmesswandler, Kraftsensoren)
- Bestimmung der Geschwindigkeit und Beschleunigung anhand der gemessenen Weg-Zeit-Kurve durch Kurvenanpassung und Differenzierung

- Auswahl einer geeigneten Abtastrate in der CASSY Software insbesondere bei starkem numerischem Rauschen beim Differenzieren

Schwierigkeitsgrad		
Theoretische Grundidee	leicht	1
Messaufbau	mittel	2
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	leicht	1
Summe		10

Tabelle 52: *Schwierigkeitsgrad des Versuches „Newtonsche Axiome“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung eines Versuches zu schulischen Lerninhalten der Mittelstufe (Bewegung und Kraft, Geschwindigkeit und Beschleunigung)
- Diskussion der Vorteile und Bedingungen der Nutzung einer Luftkissenbahn für die quantitative Prüfung der Newtonschen Axiome im Physikunterricht
- Einsatz des schultypischen Messwerterfassungssystems CASSY mit verschiedenen Sensoren

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I und II
Sachgebiete	Mechanik - Geschwindigkeit und Beschleunigte Bewegungen Mechanik - Kinematik und Dynamik

Schulbezogene Einleitung

„Probleme der Mechanik nehmen sowohl in der Sekundarstufe I als auch in der Sekundarstufe II einen zentralen Platz im Lehrplan ein. Während in der Sekundarstufe I eine qualitative Betrachtung und Begriffsbildung (Was ist eine Kraft? Was versteht man unter Geschwindigkeit? Was ist Energie?) im Vordergrund steht, geht es in der Sekundarstufe II um eine quantitative und damit mathematische Betrachtung. Leider beginnen genau hier die Probleme beim Experimentieren. Nehmen wir einmal die Newtonschen Axiome: Das Trägheitsgesetz besagt, dass ein Körper auf den keine äußere Kraft einwirkt sich geradlinig und gleichförmig weiterbewegt oder in Ruhe verbleibt. Der zweiten Teil ist aus der Alltagserfahrung direkt einsichtig, die erste Aussage bezüglich der ewig konstanten Geschwindigkeit ist jedoch problematisch. So haben wir doch den Schüler und Schülerinnen in der Sekundarstufe I vermittelt, dass

es Umwandlung und Erhaltung von Energie aber kein Perpetuum Mobile (weder 1. noch 2. Art) gibt. Und nun fordern wir gerade dies bei Newton? Die Krux liegt in der Bedingung eines kräftefreien Systems. Beim Experimentieren und im Alltag treten immer Reibungskräfte auf, d.h. wir brauchen für den experimentellen Nachweis ein Experiment in dem (so gut wie) keine Reibungskräfte auftreten. Dazu kommen noch die Probleme der Unterscheidung von Translations- und Rotationsbewegungen. Letztere wollen wir gerne ausschließen, da sie in der mathematischen Beschreibung zu aufwändig sind. Aber wie realisiert man einen Wagen bei dem sich nichts dreht und der zugleich „reibungsfrei" ist? In vielen Schulbüchern wird einem als Lösung die Luftkissenbahn vorgeschlagen. Doch kann diese wirklich die hohen Anforderungen erfüllen? Genau dies sollen Sie im heutigen Versuch herausfinden."

Fachdidaktische Reflexionsfragen

- In welche motivierenden Kontexte können Sie diesen Versuch einbinden, wenn Sie ihn in der Schule durchführen würden?
- Ist „Schummeln" erlaubt? Ist es zulässig, dass die Lehrkraft die Luftkissenbahn unerwähnt schief aufbaut, wenn dadurch Reibungsfreiheit erreicht werden kann?

Klassische Einleitung

„Die drei grundlegenden Gesetze der Mechanik wurden im 17. Jahrhundert von Newton in der Schrift Philosophiae Naturalis Principia Mathematica formuliert. Die Newtonschen Axiome beschreiben die Bewegungen von Körpern und damit einhergehend den Kraftbegriff.
Im ersten Axiom schreibt Newton jedem massebehafteten Körper eine Trägheit zu. Diese bewirkt, dass ein Körper in seinem Zustand der Ruhe oder der gleichförmigen Bewegung verbleibt, solange keine Kraft auf den Körper wirkt. Das zweite Newtonsche Axiom ist das Aktionsprinzip. Demnach werden Beschleunigungen durch Kräfte verursacht. Anders formuliert bedeutet das, dass die zeitliche Änderung des Impulses eines beweglichen Körpers proportional zur einwirkenden Kraft und in Richtung der Kraft ist. Das Reaktionsprinzip ist das dritte Newtonsche Axiom und wird oft als actio = reactio kurzgefasst. Es besagt, dass Kraftwirkungen zweier wechselwirkender Körper stets gleich und entgegen gerichtet sind.
In diesem Versuch sollten die Newtonschen Axiome überprüft werden. Um geradlinige Bewegungen und Stöße nahezu reibungsfrei zu untersuchen, wird eine Luftkissenbahn eingesetzt."

A.1.20 Reibung

Praktikumsteil: Sommersemester (SS)
Themenbereich: Mechanik

Versuchsziele und Aufgaben

- Anwendung und Vergleich unterschiedlicher Methoden zur Bestimmung von Haftreibungskoeffizienten: Gleiter mit unterschiedlichen Kontaktflächen auf einer horizontalen Ebene (Abb. 80) oder einer schiefen Ebene
- Untersuchung des Einflusses der Masse und der Größe der Kontaktflächen auf die Reibungskraft und die Reibungskoeffizienten
- Bestimmung von Gleitreibungskoeffizienten auf der horizontalen Ebene mithilfe eines Laserentfernungssensors und CASSY (LD-Didactic, o.D. a)
- Vergleich der Haft- und Gleitreibungskoeffizienten

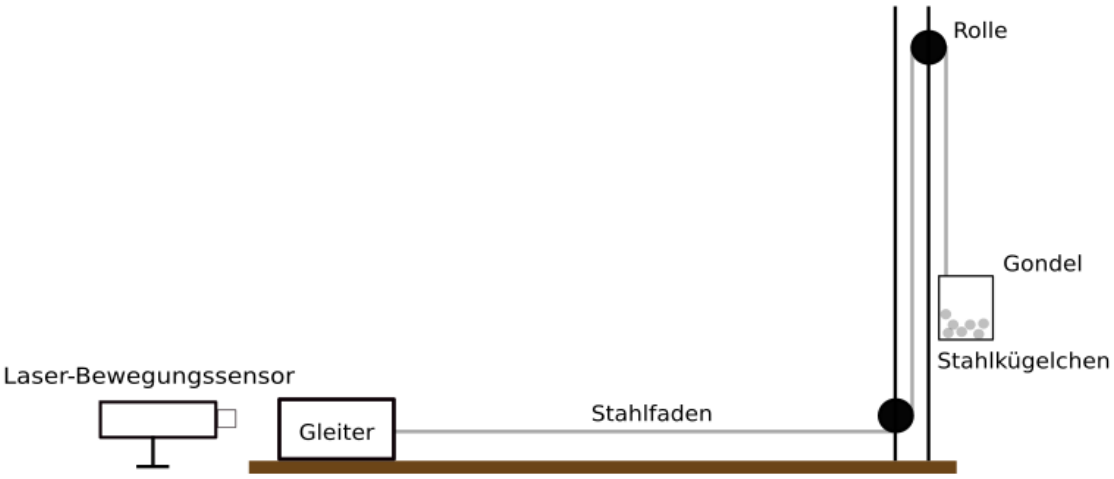


Abbildung 80: *Aufbau zum Versuch Reibung zur von Haft- und Gleitreibungskoeffizienten eines Gleiters auf einer horizontalen Ebene. Der Gleiter ist an einem Faden befestigt, der über zwei Rollen umgelenkt wird und am Ende mit einer Gondel mit unterschiedlich vielen Stahlkügelchen beschwert werden kann. Zur Bestimmung der Gleitreibungskoeffizienten steht ein Laser-Entfernungsmesser zur Verfügung (Abb. erstellt von C. Brüning).*

Herausforderungen

- Aufstellung der Kräftegleichgewichte für die unterschiedlichen Methoden
- Präzise Messungen (z. B. mit fester Startposition für den Gleiter)
- Nutzung des Laserentfernungssensors mit CASSY
- Diskussion der Einflussfaktoren auf die Reibung, insbesondere der Schülerfehlvorstellungen zur Kontaktfläche

- Diskussion der Genauigkeit und der Praktikabilität der verschiedenen Messmethoden

Schwierigkeitsgrad		
Theoretische Grundidee	leicht	1
Messaufbau	leicht	1
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	mittel	2
Summe		10

Tabelle 53: *Schwierigkeitsgrad des Versuches „Reibung“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung eines Versuches zu schulischen Lerninhalten der Mittelstufe (Reibung)
- Einsatz der schultypischen Messwerterfassungssystems CASSY mit dem Laserentfernungssensor
- Diskussion von Schülerfehlvorstellungen zur Reibung
- Diskussion der Methoden in Bezug auf Eignung für den Schulunterricht

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I und II
Sachgebiete	Mechanik - Beschleunigte Bewegungen
	Mechanik - Dynamik

Schulbezogene Einleitung

„Wenn man mit Schülerinnen und Schülern den Begriff Kraft in der Sekundarstufe I erarbeitet, dann ist eines der wesentlichen Lernziele, dass der Kraft zwei Eigenschaften zugesprochen werden: Richtung und Stärke. Während die Begriff den Schülerinnen und Schülern intuitiv sofort einleuchten, ergeben sich jedoch bei der Betrachtung mehrerer Kräfte schnell Schwierigkeiten. Warum sich die Beträge der Kräfte nicht einfach addieren sollen, ist vielen Schülerinnen und Schülern nicht klar. In diesem Zusammenhang betrachtet man häufig die Kräftezerlegung an der schiefen Ebene. Hier wirken Gravitation, Normalkraft und Reibungskraft. Dabei kommt der Reibungskraft eine besondere Rolle zu. Deren Existenz ist den Schülerinnen und Schülern zwar aus dem Alltag bekannt, aber, was Reibung eigentlich ist und wie man diese beschreibt ist ihnen unbekannt. Erstaunlich ist, dass dies auch im weiteren Verlauf der Physikausbildung in der Schule so bleibt. Reibung wird entweder

für Rechnungen vernachlässigt oder bei energetischen Betrachtungen benutzt, um die gemessenen Abweichungen von der Energieerhaltung zu erklären. Dabei bleibt der Prozess der Reibung und die Beschreibung der Reibungskräfte aber immer eine Unbekannte. Sie können sich also sicher sein, dass Sie früher oder später von einer Schülerin oder einem Schüler gefragt werden, was Reibung ist. Wissen Sie es? Wie ist es zum Beispiel mit der Richtung der Reibungskraft? Für Gravitations- oder Coulombkräfte können Sie sofort eine vektorielle Gleichung aufschreiben. Können Sie dies auch für die Reibungskraft? $\vec{F_R} = \mu \vec{F_g}$ ist sicherlich falsch. In diesem Versuch können Sie Antworten auf Ihre Fragen zum Thema Reibung finden.“

Fachdidaktische Reflexionsfragen

- Eine fehlerhafte Vorstellung von Schülerinnen und Schülern zur Reibung ist, dass die Reibung von der Fläche abhängig sei. Wie lässt sich diese fehlerhafte Schülervorstellung korrigieren? Wie kann man dabei auch Variablenkontrollstrategie einbringen?
- Welche Lernziele fachlicher und methodischer Art setzen Sie, wenn Sie diesen vereinfachten Versuch in der Mittelstufe durchführen?

Klassische Einleitung

„Bei Kontakt eines Körpers mit einer festen Oberfläche tritt Reibung auf. Die Kräfte, die zwischen den berührenden Oberflächen wirken, werden als Reibungskräfte bezeichnet. Die Reibungskräfte ist proportional zur Normalkraft, die auf den Körper wirkt: $F_R = \mu F_N$. Reibungskräfte sind immer so gerichtet, dass sie einer Bewegung entgegenwirken und diese hemmen oder verhindern. Der Proportionalitätsfaktor μ hängt von vielen Faktoren ab. Zum einen unterscheidet er sich deutlich für einen ruhenden und einen bewegten Körper (Haft - und Gleitreibung), zum anderen geht aber auch die Beschaffenheit der Oberfläche ein. Zum Beispiel haben raue Oberflächen in der Regel einen höheren Reibungskoeffizienten als glatte. Auch verschiedene Materialien haben deutlich unterschiedliche Reibungskoeffizienten.
In physikalischen Betrachtungen werden Reibungskräfte oft nicht berücksichtigt. Das liegt zum einen daran, dass sie als vernachlässigbar klein angenommen werden. Zum anderen kommt hinzu, dass Reibungskräfte in den meisten Fällen quantitativ schwer zu bestimmen sind. Die genauen Vorgänge auf atomarer Ebene bei Reibungsprozessen sind auch heute noch nicht vollständig geklärt. Daher sind theoretische Berechnung von Reibungskoeffizienten nicht möglichen. Reibungskoeffizienten werden stattdessen empirisch bestimmt. Folglich sind Literaturwert für Reibungskoeffizienten von Materialpaarungen auch nur begrenzt aussagekräftig.
In viel praktischen Anwendungen spielt Reibung allerdings eine große Rolle. Ohne Reibungseffekten zwischen der Fußsohle und dem Boden würde das Gehen schwer fallen. Ein Auto zu fahren wäre ohne Reibung schwer steuerbar.“

A.1.21 Fallender Doppler-Effekt

Praktikumsteil:	Sommersemester (SS)
Themenbereich:	Mechanik
Verweis:	Der Versuch ist inspiriert durch die Smartphone Versuche von Müller et al. (2010).

Versuchsziele und Aufgaben

- Bestimmung der Erdbeschleunigung anhand der auftretende Frequenzverschiebung des Tons einer frei fallenden Schallquelle
- Auswertung eines gebremsten Falles

Herausforderungen

- Auswahl einer geeigneten Sendefrequenz
- Qualitative und quantitative Auswertung der Spektrogramme des beobachteten Signals
- Theorie zur Fourier-Analyse
- Geeignete Parametereinstellung für die Fourier-Analyse
- Einarbeitung in die Software Audacity (Toneditor, Audacity, o.D.) und Sonic Visualizer (Tonanalyse, Cannam et al., 2010; SonicVisualiser, o.D.)

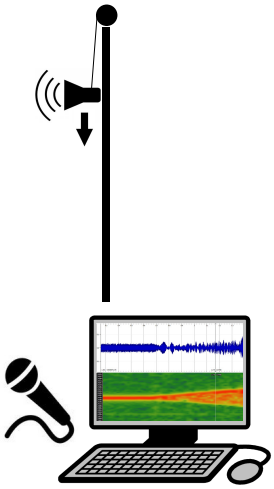


Abbildung 81: *Aufbau zum Versuch Fallender Doppler-Effekt bestehend aus einem frei fallenden Bluetooth Lautsprecher, der einen Ton konstanter Frequenz emittiert und einem Mikrophon nahe des Auftreffpunktes.*

Relevanz des Versuchsthemas für die Schule

- Nutzung der kostenfreien, schulrelevanten Software Audacity (Toneditor) und Sonic Visualizer (Tonanalyse)
- Verknüpfung und Vertiefung von schulischen Lerninhalten der Oberstufe: Freier Fall, Doppler-Effekt und Spektren
- Diskussion zur schulpraktischen Umsetzung des Versuches

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	leicht	1
Messtechnik	mittel	2
Versuchsdurchführung	leicht	1
Auswertung	mittel	2
Fehlerrechnung	mittel	2
Summe		11

Tabelle 54: *Schwierigkeitsgrad des Versuches „Fallender Doppler-Effekt“ (Erläuterungen in Abschnitt A.1.1)*

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiete	Mechanik - Kinematik
	Wellen - Spektren

Schulbezogene Einleitung

„Stellen Sie sich folgende Situation vor: Sie behandeln mit Ihrem Profilkurs Physik gerade das Thema Schwingungen und Wellen. In diesem Zusammenhang möchten Sie nun den Dopplereffekt einführen und auch quantitativ behandeln. Als Einstieg haben sich entschieden, einen Kontextbezug zu wählen.
Da das typische Beispiel eines vorbeifahrenden Krankenwagens mit Blaulicht und Martinshorn im besten Fall ein zufälliges und seltenes Ereignis ist, wollen Sie den Schülerinnen und Schülern eine Videoaufnahme eines Krankenwagens vorspielen. Dabei bewegt sich der Krankenwagen zunächst auf die Kamera zu und entfernt sich anschließend von ihr. Die Schülerinnen und Schüler sollen dann ihre Beobachtungen bzw. ihre Höreindrücke schildern. Ziel soll eine qualitative Beschreibung des Dopplereffektes sein. Als Abschluss der Einheit würden Sie jetzt aber gerne noch eine quantitative Analyse anschließen. Da Ihnen die Geschwindigkeit des Krankenwagens nicht bekannt ist und sich auch wegen des Perspektivenwechsels der Kamera nicht bestimmen lässt, scheidet eine Analyse des Videos aus. Sie möchten daher die Situation mit Hilfe eines Lautsprechers und eines Mikrophons nachstellen.
Hierbei stellen sich Ihnen eine Reihe von Problemen, die es zu lösen gilt:

- *Wie erreicht man eine kontrollierbare und reproduzierbare Geschwindigkeit des Senders?*
- *Welche Frequenz sollte man beim Sender wählen, um einen sichtbaren Messeffekt zu bekommen?*
- *Wie wertet man die Tonaufnahme aus?*

Der heutige Versuch ermöglicht Ihnen genau diese Fragen an einem professionellen Aufbau zu untersuchen. Sicherlich ist Ihnen schnell klar, wie man eine technisch einfachere Variante in der Schule realisieren kann. Viel Spaß beim Experimentieren!“

Fachdidaktische Reflexionsfragen

- Mit welchen Versuchsmaterialien können Sie diesen Versuch einfach in den Unterricht integrieren, wenn Ihnen kein Bluetooth-Lautsprecher und keine Fallkonstruktion zur Verfügung steht?
- Worin bestehen Vorteile und Schwierigkeiten des kumulativen Lernens durch innerfachliche Vernetzung von Lerninhalten wie beispielsweise dem freier Fall und dem Doppler-Effekt?

Klassische Einleitung

„Die Ausbreitung von Wellen in einem Medium wird klassisch durch gekoppelte Oszillatoren modelliert. Hierbei fungiert ein Oszillator als Erreger von dem aus die Störung sich im System ausbreitet. Die Detektion der Welle erfolgt durch eine Amplitudenmessung eines anderen Oszillators. In diesem Modell sind sowohl Erreger als auch Empfänger (Detektor) ortsfest. Wenn sich einer der beiden jedoch relativ zu dem anderen bewegt, ändert sich die Amplitudenverteilung des Wellenfeldes. Insbesondere die Wellenlänge ist nicht mehr räumlich isotrop, sondern in Bewegungsrichtung kürzer und entgegen der Bewegungsrichtung länger, man spricht vom Dopplereffekt. Im Rahmen dieses Versuche soll der Dopplereffekt am Beispiel von Schallwellen quantitativ untersucht werden.“

A.1.22 Stöße ausgedehnter Körper

Praktikumsteil: Sommersemester (SS)
Themenbereich: Mechanik

Versuchsziele und Aufgaben

- Kalibrierung des Luftkissentisches als zentrales Element des Versuchsaufbaus, um möglichst Reibungsfreiheit zu erreichen
- Prüfung des Reflexionsgesetzes anhand des Stoßes eines ausgedehnten Pucks mit einer festen Wand (Abb. 82)
- Prüfung der Erhaltungssätze von Energie und Impuls bei Stößen ausgedehnter Pucks im Kontakt und im magnetischen Wechselwirkungspotential (zentraler Stoße zweier sich anziehender Magnetpucks, dezentraler Stoß zweier sich abstoßender Pucks, dezentraler Stoß von zwei normalen Pucks unterschiedlicher Masse)

Herausforderungen

- Sorgfältige Ausrichtung des Luftkissentisches
- Auswahl der optimalen Kameraeinstellungen (Belichtungszeit)
- Nutzung spezieller Auswertefunktionen der Videoanalyse-Software Tracker (Abb. 82) (D. Brown & Cox, 2009; D. Brown, 2019)
- Erkennen der Vorteile der Magnetpucks zur Realisierung von zentralen Stößen (anziehend) und kontaktfreien Stößen (abstoßend)
- Erklärung des Rotationsanteils des ausgedehnten Pucks nach dem Stoß mit der Wand
- Einbezug der Rotationsenergie in die Energiebetrachtungen

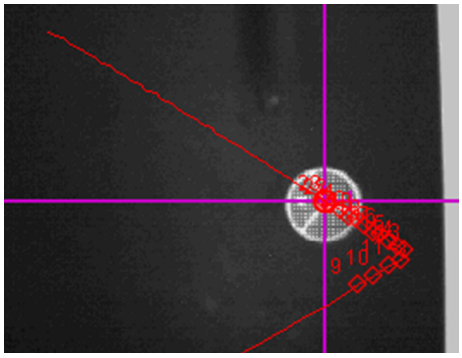


Abbildung 82: *Analyse der Bewegung des Pucks nach dem Stoß mit der Wand mithilfe der Videoanalyse-Software Tracker.*

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	leicht	1
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	schwer	3
Fehlerrechnung	leicht	1
Summe		12

Tabelle 55: *Schwierigkeitsgrad des Versuches „Stöße ausgedehnter Körper“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung der Kenntnisse der schultypischen, kostenfreien Videoanalyse-Software Tracker
- Vertiefung eines Versuches zu schulischen Lerninhalten, die in der Oberstufe als theoretisches Beispiel oder Simulation angeführt werden (Bewegungsanalysen im 2D-Raum unter Einbezug von Stößen und Rotationsbewegungen)

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I und II
Sachgebiete	Energie - qualitativer und quantitativer Energiebegriff
	Mechanik - Dynamik
	Felder - Bewegungen in radialsymmetrischen Feldern
Basiskonzept	Energie

Schulbezogene Einleitung

„Wenn Sie in Standard-Lehrwerken der Physik zum Stichwort „Stoß“ nachschlagen, dann finden Sie eine Unterscheidung von elastischen und inelastischen Stößen. Zur Berechnung werden die Energie- und Impulserhaltung verwendet. Doch wenn Sie abends in der Kneipe am Billardtisch stehen, kommen Ihnen Zweifel, dass dies die ganze Wahrheit ist. Impulserhaltung erscheint optional und ein Stoß mit der Bande erfüllt quasi nie die Beziehung „Einfallwinkel = Ausfallwinkel“. Aber gerade darin liegt der Reiz des Spiels. Aus physikalischer Sicht ist das natürlich in höchstem Maße besorgniserregend und schreit nach einer Erklärung. Intuitiv ist Ihnen wahrscheinlich sofort klar, dass die Ursache in der Ausdehnung der Körper zu suchen ist. Doch wie funktioniert das genau? Im Rahmen dieses Versuches sollen Sie daher den elastischen Stoß ausgedehnter Körper mit Hilfe der Videoanalyse quantitativ untersuchen, damit Sie beim Billardspielen auf der Klassenfahrt auf jeden Fall eine gute Figur abgeben (zumindest theoretisch).“

Fachdidaktische Reflexionsfragen

- Welche Vor- und Nachteile hat die Nutzung eines Luftkissentisches (2D) im Vergleich zu einer Luftkissenbahn (1D) im Physikunterricht zur Veranschaulichung und Analyse von Stößen?
- Diskutieren Sie, welche Vor- und Nachteile die Nutzung einer Stoß-Simulation im Physikunterricht gegenüber einem Demonstrationsversuch hat.

Klassische Einleitung

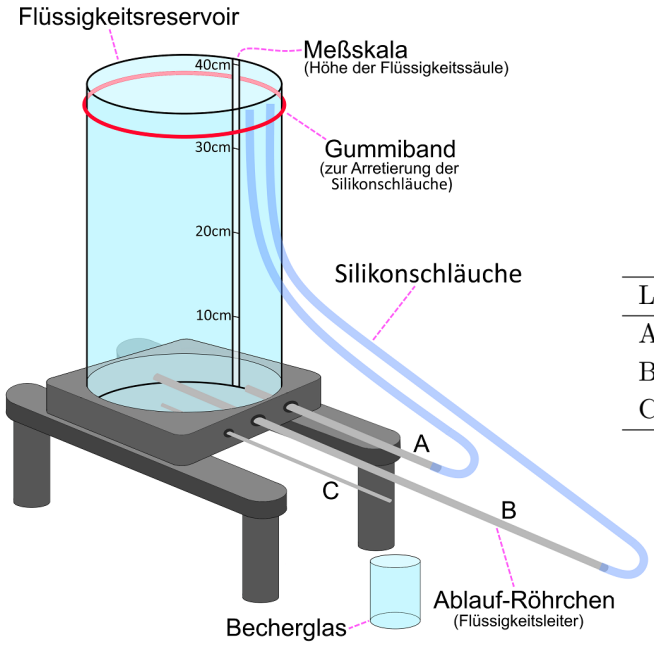
„Man spricht bei kurzzeitigen Wechselwirkungen zwischen zwei oder mehreren Körpern, die Änderungen des Impulses und der kinetischen Energie bewirken, von Stößen. Zur Berechnung der Änderung werden die Energie- und Impulserhaltung verwendet. Man unterscheidet elastische und inelastische Stöße.
Der Luftkissentisch ermöglicht es Ihnen, Stöße in 2D zu untersuchen. Dies bringt besondere Herausforderungen mit sich. Zum einen verlangt der Versuch eine vektorielle Betrachtung der Impulserhaltung. Zum anderen zeigen sich Besonderheiten, die durch die Ausdehnung der Pucks resultieren.“

A.1.23 Wasserstromkreis

Praktikumsteil:	Sommersemester (SS)
Themenbereich:	Elektrizitätslehre
Verweis:	Die Umsetzung des Versuches ist inspiriert durch den Versuch „Strömungsmechanik / Blutkreislauf“ aus dem adressatenspezifischen Physikpraktikum für Mediziner von Theyßen (2000).

Versuchsziele und Aufgaben

- Messung der Volumenstromstärke und Druckdifferenzen für vier verschiedene Füllhöhen im Flüssigkeitsreservior und für vier Konstellationen der Flüssigkeitsleiter (Abb. 83)
- Untersuchung des Zusammenhangs von Volumenstromstärke, Druckdifferenz und Strömungswiderstand analog zum Ohmschen Gesetz
- Untersuchung des Einflusses von Durchmesser und Länge des Flüssigkeitsleiters sowie turbulenten Strömungen auf den Strömungswiderstand und Vergleich mit elektrischem Widerstand
- Untersuchung der Reihen- und Parallelschaltungen von Strömungswiderständen und Vergleich mit elektrischen Widerständen
- Beurteilung der Analogie Wasserstromkreis in der Elektrizitätslehre hinsichtlich ihrer Möglichkeiten und Grenzen



Leiter	Durchmesser	Länge
A	$d = 5.1\,mm$	$l = 255\,mm$
B	d	$2l$
C	$0.5d$	l

Abbildung 83: *Aufbau zum Versuch Wasserstromkreis bestehend aus einem zylinderförmigen Flüssigkeitsreservoir für Glycerinlösung mit Ablauf-Röhrchen A, B und C, deren Durchmesser und Länge systematisch, wie in der Tabelle dargestellt, variieren (Abb. erstellt von S. Wyprich).*

Herausforderungen

- Verknüpfung der theoretischen Grundlagen zum elektrischem Stromkreis und zur Strömungslehre (z. B. Gesetz von Hagen-Poiseuille als Ohmsches Gesetz für laminare Strömungen)
- Gegenüberstellung von elektrischem Stromkreis und Wasserstromkreis mithilfe der Zuordnung der analogen Größen und Bauteile
- Systematische Variation von Variablen bei einer gut geplanten und sorgfältig durchführten Messung (z. B. bei der Volumenentnahme: Minimierung von Messunsicherheiten und gleichzeitig näherungsweise Konstanthaltung der Druckdifferenz)
- Auswertung mittels linearer Regression und zugehöriger Fehlerbetrachtung inkl. Abschätzung von Messunsicherheiten und Fehlerfortpflanzung
- Vergleiche und Beurteilungen der Ergebnisse in Bezug auf die Analogie

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	leicht	1
Messtechnik	leicht	1
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	mittel	2
Summe		10

Tabelle 56: *Schwierigkeitsgrad des Versuches „Wasserstromkreis“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Quantitative experimentelle Analyse einer in der Schule häufig genutzten Analogie des Stromkreises
- Untersuchung des Analogiebereiches und kritische Beurteilung von Möglichkeiten und Grenzen der Analogie

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	I
Sachgebiete	Elektrizitätslehre - Einfache elektrische Stromkreise
	Elektrizitätslehre - Stromstärke und Spannung

Schulbezogene Einleitung

„Wenn man einen Sachverhalt erklären will, dann benutzt man häufig Phrasen wie „Das ist genauso wie beim ...“ oder „Das kennen Sie schon vom ...“. Bei diesen Vergleichen spricht man in der Fachdidaktik von Analogien (Kircher et al., 2015). Die Idee bei einer Analogie ist, dass man eine Ähnlichkeit zwischen zwei Objekten oder Sachstrukturen ausnutzt, um den Lernprozess zu unterstützen. Die Voraussetzung ist dabei allerdings, dass einem der Analogbereich so vertraut sein muss, dass man sich mit diesem Wissen den Lernbereich leichter erschließen kann. Ein weiterer entscheidender Aspekt bei der Verwendung von Analogien ist, dass Analog- und Lernbereich ähnlich aber nicht gleich sind. Dies hat zur Folge, dass eine Ähnlichkeit der Sachstruktur zwischen zwei Bereichen der Physik nur bis zu einem gewissen Grad vorliegt. In diesem Sinne bergen Analogien sowohl Chancen als auch Risiken. In der Praxis ist eine der am häufigsten in der Schule verwendete Analogie, die zwischen Wasserströmungen und elektrischen Strömen. Hierbei verwendet man die Wasserströmung in Rohren als Analogbereich, um elektrische Ströme in Leitern zu erklären. Hierbei stellt sich nun die Frage, wie gut kennen Sie denn die Physik der Wasserströmung in Rohren? Wie ähnlich sind sich Wasserströmung und elektrischer Strom? Gibt es z. B. ein Ohmsches Gesetz oder Kirchhoffsche Gesetze für Wasserstromkreise? Wie weit geht die Analogie zwischen diesen beiden Gebieten? Finden Sie es in diesem Versuch heraus!“

Fachdidaktische Reflexionsfragen

- In welcher Klassenstufe und mit welchem Ziel würden Sie die Analogie Wasserstromkreis einsetzen? Mit welchen Schwierigkeiten würden Sie dabei rechnen?
- Inwieweit können und sollten Sie Ihren Schülern die Grenzen der Wasserstromkreisanalogie aufzeigen?

Klassische Einleitung

„In einem vereinfachten Bild hängt der Volumendurchsatz bei der Strömung einer Flüssigkeit durch ein Rohr nur vom Durchmesser des Rohres (d.h. seiner Querschnittsfläche) und der Strömungsgeschwindigkeit ab. Diese Beschreibung ist für ein Suprafluid auch zutreffend. In der Praxis sind Flüssigkeiten aber nicht suprafluid. Ihre Viskosität und der Kontakt mit der Rohrwandung sorgen für ein radiales Geschwindigkeitsprofil mit einer geringen Strömungsgeschwindigkeit außen und einer maximalen Strömungsgeschwindigkeit in der Mitte des Rohres. Folglich ist der Volumendurchsatz auch nicht proportional zur Querschnittsfläche des Rohres. Dieses Phänomen wird für laminare Strömungen quantitativ in guter Näherung durch das Gesetz von Hagen-Poiseuille beschrieben.
Zwangsläufig bedeutet das auch, dass sich Wasserströmungen in Rohren anders verhalten als elektrische Ströme in Leitern. In diesem Versuch geht es daher darum, diese Unterschiede quantitativ zu untersuchen.“

A.1.24 Feldeffekttransistor

Praktikumsteil:	Sommersemester (SS)
Themenbereich:	Elektrizitätslehre
Verweis:	Der Versuch zum Feldeffekttransistor ist inspiriert durch die entsprechenden Versuche des Elektronik-Grundpraktikums für 1-Fach-Physiker von Prof. Dr. Dietmar Block an der Christian-Albrechts-Universität zu Kiel.

Versuchsziele und Aufgaben

- Aufnahme der Eingangskennlinien eines selbstsperrenden MOSFET und Bestimmung der Steilheit und der Gate-Threshold-Spannung des FET
- Aufnahme der Eingangskennlinien eines selbstleitenden JFET
- Aufnahme des Ausgangskennlinienfeldes des selbstsperrenden MOSFET in CASSY (LD-Didactic, o.D. a) und Bestimmung der Arbeitgerade
- Anwendung des Feldeffekttransistors als lichtgesteuerter Schalter mit einem lichtabhängigen Widerstand

Herausforderungen

- Aufbau der entsprechenden Schaltungen
- Systemische Betrachtung des Feldeffekttransistors
- Verständnis des Spannungsteilers
- Nutzung des Power CASSY als programmierbare Spannungs- oder Stromquelle mit integrierter Strom- oder Spannungsmessung

Schwierigkeitsgrad		
Theoretische Grundidee	schwer	3
Messaufbau	schwer	3
Messtechnik	mittel	2
Versuchsdurchführung	schwer	3
Auswertung	mittel	2
Fehlerrechnung	leicht	1
Summe		14

Tabelle 57: *Schwierigkeitsgrad des Versuches „Feldeffekttransistor“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung des Basiskonzeptes System anhand der systemischen Betrachtung des Feldeffekttransistors
- Einsatz der schultypischen Messwerterfassungssystems CASSY mit Power CASSY

Einordnung des Versuchsthemas in die Fachanforderungen

Basiskonzept System

Schulbezogene Einleitung

*„Auf der schleswig-holsteinischen Landestagung der MNU (Deutschen Vereins zur Förderung **m**athematisch-**n**aturwissenschaftlichen **U**nterrichts), die jedes Jahr Mitte September in Kiel stattfindet, werden Sie Zeuge der folgenden Diskussion:*

Kollege A: *„Sag mal, machst Du eigentlich noch mit Deinen Schülern den Transistor?“*

Kollegin B: *„Nein, zumindest den bipolaren Transistor mache ich nicht mehr. Zu kompliziert! Ich bespreche stattdessen den FET“*

Kollege C: *„Echt? So richtig mit Schaltung und allem drum und dran oder nur "Guck mal, Lampe an - Lampe aus!“*

Kollegin B: *„Jein, ich bespreche den FET zwar nicht im Detail, d.h. seinen internen Aufbau, aber ich bespreche schon die typischen Eigenschaften eines FET und wie man diese zum Schalten nutzen kann. Ich finde das insofern ganz spannend, weil man am Beispiel des FET sehr schön eine systemische Betrachtung einüben kann. Was ist Ursache, was ist Wirkung, ...“*

Kollege A: *„Das ist ja spannend! Ich würde das ja auch gerne machen, aber bei uns funktioniert der Versuch nicht. Da kann man einstellen, was man will, die Lampe leuchtet immer.“*

Kollege C: *„Das ist bei uns auch so. Mein Kollege meint, der FET ist wahrscheinlich kaputt und leider gibt es bei den Lehrmittelherstellern keinen Ersatz. Das Modell, was wir haben, hat mal ein Kollege zusammen gelötet und der ist jetzt in Rente.“*

Kollegin B: *„Seid ihr sicher, dass der kaputt ist? Was habt ihr denn für FETs? Es gibt ja verschiedene Typen und die muss man unterschiedlich beschalten...“*

Was ist Ihr Fazit aus der Diskussion? Stimmt das, was gesagt wurde? Könnten Sie beurteilen, ob Sie den FET richtig beschaltet haben? Kann man wirklich den FET systemisch betrachten? Wie müsste man das experimentell umsetzen? Dieser Prak-

tikumsversuch gibt Ihnen Antworten auf diese Fragen“

Fachdidaktische Reflexionsfragen

- Wie können Sie das „Basiskonzept System“ aus den Fachanforderungen (Ministerium für Schule und Berufsbildung Schleswig-Holstein, 2019) durch den FET im Physikunterricht aufgreifen?
- Wie könnte man den FET im Schulunterricht erarbeiten? Welche didaktischen Reduktionen sind sinnvoll?

Klassische Einleitung

„Der Feldeffekttransistor (FET) ist ein elektronisches Bauelement, welches die leistungslose Steuerung von Strömen ermöglicht. Das Funktionsprinzip basiert auf der Steuerung der freien Ladungsträgerkonzentration an einem pn-Übergang mit Hilfe der angelegten Spannung. Erhöht man die Ladungsträgerkonzentration steigt die Leitfähigkeit. Auf diese Weise kann man die Leitfähigkeit um mehrere Größenordnungen variieren. Anders als bei einer Diode fließt jedoch nicht der Strom durch den pn-Übergang, sondern in einem „separaten“ Kanal (JFET). Der pn-Übergang selbst wird in Sperrrichtung betrieben und ist deshalb so hochohmig, dass der Stromfluss vernachlässigbar ist. Folglich ist das Produkt aus Strom und Steuerspannung (also die Leistung) ebenfalls vernachlässigbar klein. Man spricht daher von einer leistungslosen Steuerung. Man unterscheidet verschiedene Typen von FETs. Beim MOSFET ist Aufbau und Funktion etwas komplizierter, aber im Kern sorgt auch hier eine Steuerspannung für eine Änderung der Ladungsträgerkonzentration in einem Kanal eines dotierten Halbleitermaterials. Im Rahmen dieses Versuches sollen sie die wesentlichen Kennlinien eines FETs vermessen. Sie bilden die Grundlage um im zweiten Teil des Versuches die Anwendung des FET als Schalter kennenlernen zu können.“

A.1.25 Lorentzkraft

Praktikumsteil: Sommersemester (SS)
Themenbereich: Elektrizitätslehre

Versuchsziele und Aufgaben

- Messung der Lorentzkraft auf einen stromdurchflossenen Leiter im Magnetfeld
- Messung und Berechnung der magnetischen Flussdichte
- Untersuchung des Einflusses von unterschiedlichen Leiterschleifen und Magnetabständen

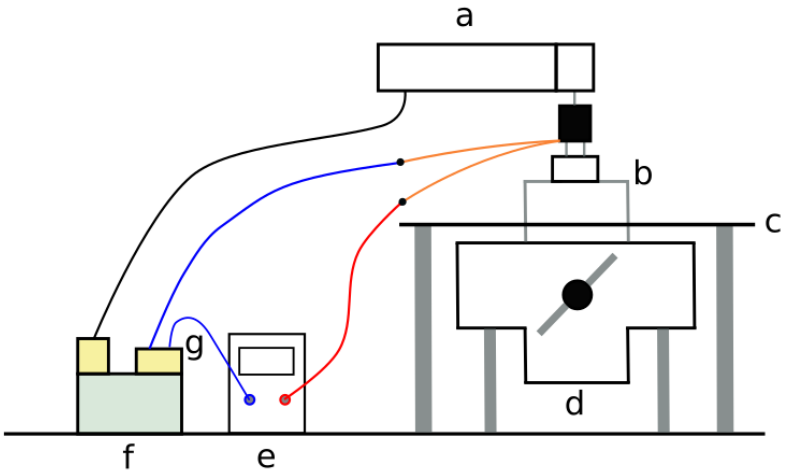


Abbildung 84: *Aufbau zum Versuch Lorentzkraft bestehend aus den Elementen: a) Kraftsensor, b) Leiterschleife, c) Plexiglasabschirmung, d)Schraubstock mit Riesenmagneten, e) Netzgerät, f) CASSY Interface, g) CASSY 30-Ampere-Box (Abb. erstellt von C. Brüning)*

Herausforderungen

- Präzises Messen durch sorgfältige Ausrichtung des Kraftsensors und der Hall-Sonde
- Zielgerichteter Einsatz und Umgang mit verschiedenen CASSY Sensoren (Hall-Sonde, Kraftsensor) (LD-Didactic, o.D. a)
- Bestimmung der magnetischen Flussdichte aus Lorentzkraft und Leiterschleifenstrom durch lineare Regression in CASSY
- Systematische Parametervariation zur Untersuchung der Einflussgrößen

Schwierigkeitsgrad		
Theoretische Grundidee	leicht	1
Messaufbau	mittel	2
Messtechnik	mittel	2
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	mittel	2
Summe		11

Tabelle 58: *Schwierigkeitsgrad des Versuches „Lorentzkraft“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Quantitative Durchführung und Vertiefung des Versuches zur Lorentzkraft, der im Schulunterricht der Oberstufe oft nur qualitativ demonstriert wird
- Einsatz der schultypischen Messwerterfassungssystems CASSY mit Kraftsensor und Hall-Sonde
- Diskussion zur schulpraktischen Umsetzung des Versuches

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiet	Felder - Bewegungen in Magnetfeldern

Schulbezogene Einleitung

„Damit Schülerinnen und Schüler Bewegungen von Ladungsträgern in Magnetfeldern, Induktion und technische Anwendungen wie zum Beispiel den Transformator verstehen können, müssen Sie die Lorentzkraft kennen. Im Schulunterricht wird die Lorentzkraft klassischer Weise qualitativ über eine Leiterschleife im Hufeisenmagneten veranschaulicht. Die Ursache-Vermittlung-Wirkungs-Beziehung die zur Vorhersage der Richtung der Lorentzkraft eingesetzt wird, führt in der Regel bei den Schülerinnen und Schülern zu wilden Handverrenkungen. In vielen Fällen bleibt es in der Schule jedoch bei dieser rein qualitativen Betrachtung, da für eine quantitative Messung sowohl ein starkes homogenes Magnetfeld, als auch hohe Stromstärken erforderlich sind, um makroskopische Lorentzkräfte zu erzeugen. Dies führt jedoch dazu, dass den Schülerinnen und Schülern kein Gefühl für die Stärke der Lorentzkraft vermittelt wird. Ein Versuch zur Messung der Stärke der Lorentzkraft und zur Abhängigkeit der Lorentzkraft von den verschiedenen Parametern (Fläche, Stromstärke, Feldstärke) wäre sicherlich ein Gewinn. Der klassische Versuchsaufbau verwendet einen Hufeisenmagnet, eine Leiterschlaufe und eine Stromwaage. Letztere ist erforderlich, die kleinen Kräfte überhaupt messbar zu machen. Doch muss die Messung so kompliziert sein? Gibt es keine Möglichkeit mit zwei „anständigen“ Magneten und einer Leiterschleife die Kraft direkt zu messen? Probieren Sie es aus!

Wir geben Ihnen ein paar Leiterschleifen, einen CASSY Kraftsensor und die stärksten und größten Permanentmagnete, die es derzeit auf dem Markt zu kaufen gibt. Glauben Sie nicht? Auf der Webseite des Herstellers (Supermagnete, o.D. a) finden Sie in einem eindrucksvollen Video, was man mit diesen Magneten machen kann und was man besser sein lässt.“

Fachdidaktische Reflexionsfragen

- Wie würden Sie diesen Versuch vereinfacht im Unterricht der Sekundarstufe II umsetzen?
- Welche Argumente sprechen dafür, die 3-Finger-Regel für beide Hände einzuführen, welche dagegen?

Klassische Einleitung

„Die Bewegung eines elektrisch geladen Teilchens in einem Magnetfeld bewirkt eine Kraft auf das Teilchen. Diese wurde nach dem niederländischen Mathematiker und Physiker Hendrik Anton Lorentz benannt. Die Lorentzkraft $\vec{F}$ ist proportional zur Geschwindigkeit $\vec{v}$ des Teilchens in der Ebene senkrecht zum Magnetfeld sowie zur Stärke des Magnetfeldes $\vec{B}$ und der Ladung des Teilchen q. Ferner ist diese Kraft in der Ebene senkrecht zum Magnetfeld ebenfalls senkrecht zur Geschwindigkeit orientiert. Es gilt $\vec{F}_L = q\vec{v} \times \vec{B}$.
In diesem Versuch soll die Stärke der Lorentzkraft experimentell ermittelt werden. Zu diesem Zweck wird die Kraft gemessen, die auf stromdurchflossene Leiterschleifen in einem homogenen Magnetfeld wirkt. Ziel der Messungen ist der quantitative Nachweis der oben genannten Proportionalitäten.“

A.1.26 Fieldplotter

Praktikumsteil:	Sommersemester (SS)
Themenbereich:	Elektrizitätslehre
Verweis:	Der Versuch zum Fieldplotter ist eine lehramtsspezifische Adaption eines Versuches aus dem Anfängerpraktikum für 1-Fach-Physiker von Dr. Victor de Manuel Gonzales an der Christian-Albrechts-Universität zu Kiel.

Versuchsziele und Aufgaben

- Zeichnung einer Auswahl unterschiedlicher vorgegebener Anordnungen von Punktladungen, Kondensatoren (mit und ohne Isolator), einer Elektronenlinse und eines Feldemissions-Mikroskops mit einem Leitlackstift auf einem mit Kohlenstoff imprägniertem Widerstandspapier
- Bestimmung der Äquipotentiallinien und Feldlinien anhand einer Kompensationsmessung mit einer Wheatstoneschen Brückenschaltung (Abb. 85)
- Bestimmung des Betrags des elektrisches Feldes entlang der Symmetrielinie aus den gemessenen Äquipotentiallinien für eine Anordnung

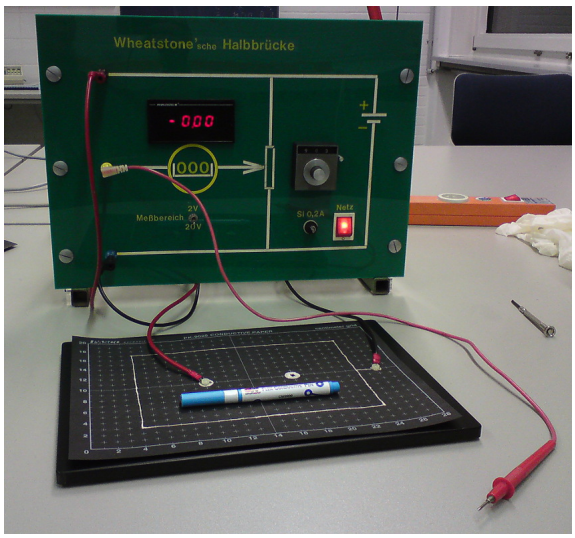


Abbildung 85: *Die Messung der Äquipotentiallinien wird mit einer Kompensationsmessung mit einer Wheatstoneschen Brückenschaltung durchgeführt. Um eine Äquipotentiallinie aufzunehmen, wird ein Potential über das Potentiometer eingestellt und mit einer Prüfspitze auf dem Widerstandspapier mit der mit Leitlack gezeichneten Anordnung gesucht. Die Äquipotentiallinie ist genau dann gefunden, wenn durch das Messgerät kein Strom mehr fließt. Dieser Punkt wird auf dem Widerstandspapier markiert.*

Herausforderungen

- Verständnis des Messprinzips (Kompensationsmessung, Wheatstonesche Brückenschaltung)
- Erläuterungen der Phänomene am Rand des Widerstandspapiers

Schwierigkeitsgrad		
Theoretische Grundidee	mittel	2
Messaufbau	mittel	2
Messtechnik	schwer	3
Versuchsdurchführung	mittel	2
Auswertung	mittel	2
Fehlerrechnung	leicht	1
Summe		12

Tabelle 59: *Schwierigkeitsgrad des Versuches „Fieldplotter“ (Erläuterungen in Abschnitt A.1.1)*

Relevanz des Versuchsthemas für die Schule

- Vertiefung eines Versuches zu schulischen Lerninhalten der Sekundarstufe II (Kondensator, Dielektrikum, Elektrisches Feld und Potential)
- Übertragbarkeit des Versuches auf den Schulunterricht zur Veranschaulichung von Feldlinien

Einordnung des Versuchsthemas in die Fachanforderungen

Sekundarstufe	II
Sachgebiet	Felder - Homogenes Elektrisches Feld

Schulbezogene Einleitung

„Die Darstellung von Feldern ist im Schulunterricht ein wichtiger Aspekt. Gerade weil der Feldbegriff so abstrakt ist und die Vorstellung vielen Schülerinnen und Schülern Probleme bereitet, wäre es hier sehr wünschenswert, wenn die Schülerinnen und Schüler die Feldverteilung im Rahmen eines Schülerversuches selber vermessen könnten. Doch da gibt es ein kleines Problem: Elektrische Feldverteilungen um geladene Körper werden üblicherweise mit Grieskörnern in Glycerin oder Papierschnipsel bei Bandgeneratoren visualisiert. Diese Versuche erfordern jedoch den Einsatz von Hochspannung und sind deshalb nur als Demonstrationsversuche durchführbar. Hinzu kommen die Einschränkungen, dass bei diesen Versuchen Feldlinien aufgrund von Störeinflüssen nur ungenau abgebildet werden. Ursachen sind die Gravitation, die Viskosität des Glycerin und die Dipolladungen der Grieskörner. Alternativ könnte man daher auf die Idee kommen, Simulationen zu verwenden, um Schülern den Verlauf von Feld- und Äquipotentiallinien zu verdeutlichen. Hierbei geht aber ein großer Teil der Beweiskraft eines realen Experimentes verloren. Gibt es einen Ausweg aus diesem Dilemma? Kann man einen Schülerversuch zur Messung elektrischer Feldlinien realisieren? Der Versuch „Fieldplotter“ basiert auf der Spannungsmessung auf Widerstandspapier. Die Ladungsverteilungen werden dabei durch Leitlack realisiert, welcher mit Hilfe von Stiften auf das Widerstandspapier aufgetragen wird. Wenn man an diese leitenden Flächen eine Spannungsquelle anschließt,

können Schülerinnen und Schüler mit einer Messbrücke die Potentialverteilung auf dem Papier vermessen. Doch ist das das Gleiche, wie eine Kraft auf eine Testladung im Feld zu messen? Finden Sie heraus, ob dieser Aufbau wirklich eine äquivalente Feldverteilung liefert, wie Sie sie im Vakuum bzw. in Luft erwarten würden und ob er als Schülerversuch tauglich ist. Vielleicht steckt ja auch hier der Teufel im Detail!?“

Fachdidaktische Reflexionsfragen

- Welche fehlerhaften Vorstellungen haben Schüler vom physikalischen Feld? Wie lässt sich diese korrigieren?
- Ist der Fieldplotter-Versuch sinnvoll in der Schule einsetzbar? Bitte begründen Sie. In welchem Rahmen und in welcher Klassenstufe würden Sie diesen Versuch durchführen?

Klassische Einleitung

„Dieser Versuch befasst sich mit dem Zusammenhang zwischen einer gegebenen Ladungsdichte ρ und dem daraus resultierenden elektrischen Feld $\vec{E}$, wie er aus der ersten Maxwell-Gleichung

$$\mathrm{div}\vec{E} = \frac{\rho}{\varepsilon_0} \tag{1.1}$$

im Vakuum hervorgeht. Der Operator div steht in diesem Fall für die Divergenz von $\vec{E}$, ε_0 ist die elektrische Feldkonstante. Diese Differentialgleichung lässt sich allerdings nur für relativ einfache Ladungsverteilungen geschlossen und in Spezialfällen näherungsweise lösen. Kompliziertere Anordnungen können daher oft nur noch numerisch berechnet werden. Einfacher können die elektrischen Felder solcher Ladungsverteilungen experimentell bestimmt werden. Dabei werden sie nicht im Vakuum, sondern in Medien untersucht.“

A.2 Skalendokumentation

A.2.1 Wahrgenommene Relevanz

Dieser Anhang ergänzt die Skalendokumentation der wahrgenommenen Relevanz aus dem Abschnitt 8.2.2.

Überblick wahrgenommene Relevanz

Skala:	Wahrgenommene Relevanz
Subskalen:	-
Anzahl Items:	8
Ratingskala:	5-stufige Likert-Skala (1 = „*trifft gar nicht zu*“, 5 = „*trifft völlig zu*“)
Messzeitpunkte:	versuchsbegleitend (für 2x12 Praktikumsversuche)
Literatur:	Schiefele et al. (2002)
Querverweis:	Abschnitt 8.2.2

Kenngrößen

Wahrgenommene Relevanz

Versuch	M	SD	SE	Min	Max	n	α	$\bar{r}_{i(t-i)}$
Gesamtdatensatz	3.65	0.66	0.02	1.25	5.00	1805	.81	.52
Augenmodell	4.09	0.52	0.06	3.00	5.01	72	.86	.61
Newtonsche Axiome	3.97	0.48	0.05	2.90	5.46	78	.82	.55
Reibung	3.86	0.42	0.05	2.70	4.59	80	.78	.50
Wasserstromkreis	3.84	0.57	0.07	1.88	4.93	74	.83	.55
Thermohaus	3.84	0.53	0.07	2.30	5.45	65	.81	.53
Oszillierendes Smartphone	3.80	0.59	0.06	1.93	4.88	89	.83	.55
Lorentzkraft	3.76	0.38	0.04	2.72	4.83	76	.74	.45
Stöße ausgedehnter Körper	3.69	0.41	0.05	2.47	4.54	75	.86	.60
Roller Coaster	3.68	0.56	0.19	2.98	4.54	9	.85	.62
Brechung, Reflexion, Polarisation	3.67	0.53	0.06	2.20	4.86	71	.84	.57
Verformung eines Balles	3.66	0.52	0.06	2.30	4.93	77	.74	.45
Fallender Doppler-Effekt	3.66	0.47	0.05	2.05	4.68	78	.83	.57
Zerfallsgesetz	3.66	0.58	0.07	1.45	4.70	73	.82	.55
Wärmepumpe	3.60	0.49	0.06	2.11	4.55	67	.83	.56
Auf- und Entladung Kondensator	3.59	0.48	0.05	2.18	4.55	88	.74	.45
Franck-Hertz-Versuch	3.57	0.56	0.07	2.16	5.32	71	.78	.49
Photoeffekt	3.57	0.48	0.06	1.95	4.58	72	.76	.48
Wärmeäquivalent	3.57	0.43	0.05	2.49	4.67	68	.85	.60
Schwarzer Strahler	3.57	0.53	0.06	2.09	4.40	68	.78	.48
Fieldplotter	3.56	0.48	0.06	1.95	4.61	74	.79	.50
Hören	3.54	0.57	0.06	1.93	4.70	88	.82	.54
Rutherford'sche Streuung	3.46	0.49	0.06	2.12	4.83	72	.82	.54
Lego-Interferometer	3.42	0.64	0.08	1.59	4.82	71	.82	.55
Interferenz am Smartphone	3.39	0.56	0.07	1.84	4.56	73	.84	.58
Feldeffekttransistor	3.13	0.72	0.08	0.94	4.68	76	.87	.63

Tabelle 60: *Deskriptive Statistik für die* ***wahrgenommene Relevanz****. In der ersten Zeile der Tabelle sind die Werte für den Gesamttest angegeben und in den darauffolgenden Zeilen für die einzelnen Praktikumsversuche mit Zentrierung. Für die Deskription der Datensätze sind Mittelwert (M), Standardabweichung (SD), Standardfehler (SE), Wertebereich auf einer 5-teiligen Likert-Skala (Min, Max), Stichprobengröße (n), Cronbachs α und die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$) angegeben.*

Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
WR_1	Das Wissen, das an diesem Praktikumstag vermittelt wurde, ist für meine späteren beruflichen Tätigkeiten ziemlich wichtig.	.77	.61
WR_2	Die Themen dieser Lehrveranstaltung halte ich für einen wesentlichen Teil meiner beruflichen Qualifikation.	.78	.55
WR_3	Der Inhalt des Versuches wurde anhand von Beispielen aus der Praxis erklärt.	.80	.42
WR_4	Bei diesem Versuch wurde die praktische Bedeutsamkeit der behandelten Themen deutlich gemacht.	.79	.50
WR_5	In diesem Praktikum werden Bezüge zwischen Theorie und Praxis hergestellt.	.78	.54
WR_6	Den heutigen Versuch halte ich für bedeutsam für den Schulunterricht. *[neu formuliert]*	.77	.62
WR_7	Bei diesem Versuch wurde ein Bezug zum Unterrichten von Physik in der Schule hergestellt. *[neu formuliert]*	.78	.57
WR_8	Der heutige Praktikumsinhalt ist nicht relevant für meine zukünftige Tätigkeit als Lehrkraft. *[neu, negativ formuliert]*	.81	.38

Tabelle 61: *Itemanalyse exemplarisch für die Gesamtskala zur wahrgenommenen Relevanz als Ergänzung zu den Erläuterungen in Abschnitt 8.2.2. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .81$.*

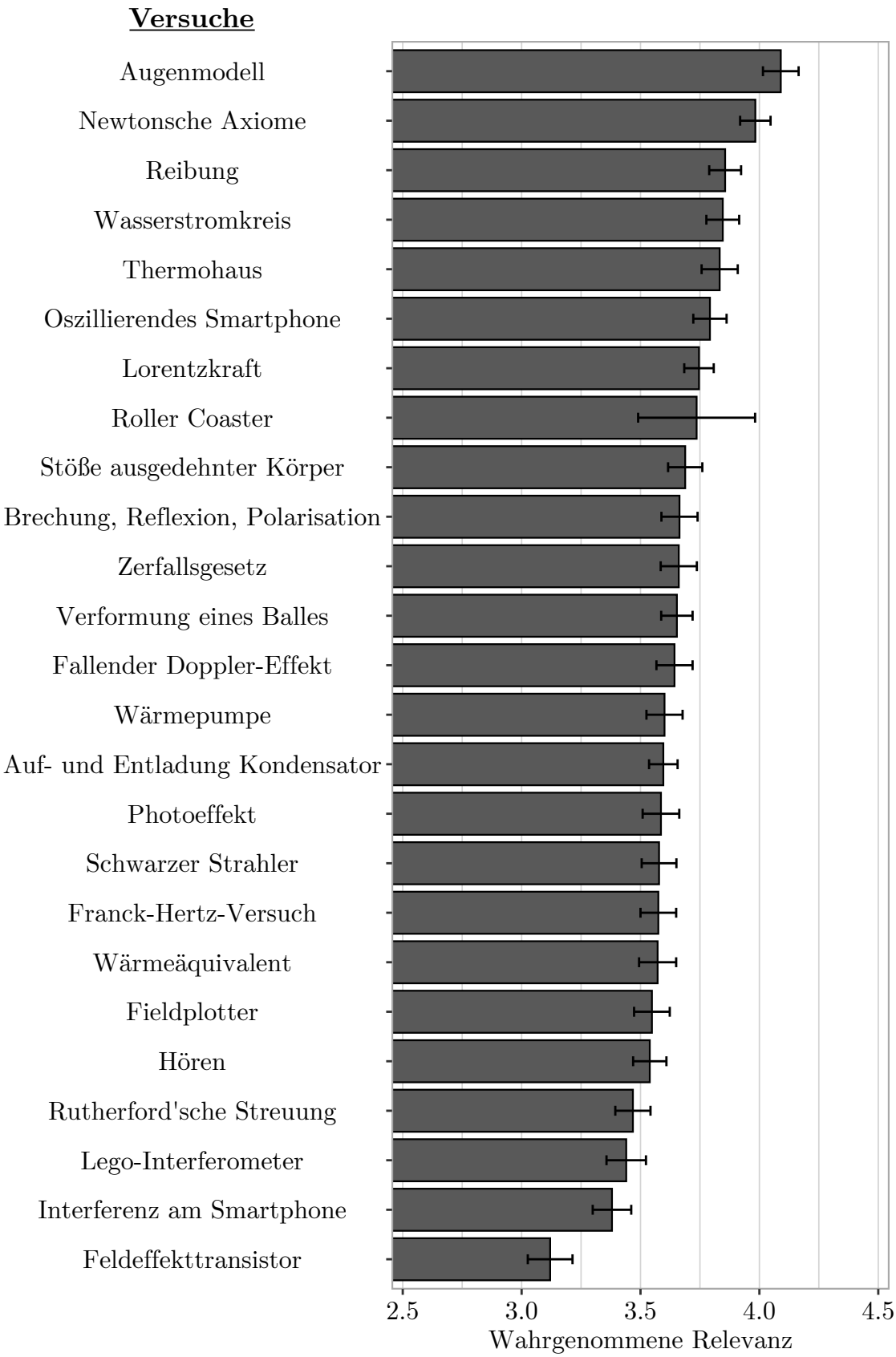


Abbildung 86: *Darstellung der **nicht zentrierten** wahrgenommenen Relevanz (5-teilige Likert-Skala) für die einzelnen Praktikumsversuche ($60 \leq n \leq 89$, $n_{RollerCoaster} = 9$). Das Balkendiagramm zeigt Mittelwerte und Standardfehler.*

A.2.2 Versuchseigenschaften

Als Ergänzung zu Abschnitt 8.2.3 wird in diesem Anhang ein Überblick über das Testinstrument gegeben. In den Tabellen 62 und 63 sind die Items der Versuchseigenschaften dargestellt. Tabelle 64 gibt einen Überblick über ergänzend hinzu genommene Versuchseigenschaften (Schulbezug, Bearbeitungsdauer und Schwierigkeitsgrad).

Überblick Versuchseigenschaften

Skala:	Versuchseigenschaften
Subskalen:	-
Anzahl Items:	27
Ratingskala:	3-stufige Likert-Skala (1 = „*trifft nicht zu*“, 2 = „*teils / teils*“, 3 = „*trifft zu*“)
Messzeitpunkte:	versuchsbegleitend (für 2x12 Praktikumsversuche)
Literatur:	Eigenentwicklung
Querverweis:	Abschnitt 8.2.3

Kategorie 1	Itemcode	Itemkürzel	Itemwortlaut
allgemein	VS_anschaulich	anschaulich	Der Versuch ist anschaulich.
	VS_abwechslungsreich	abwechslungsreich	Der Versuch ist abwechslungsreich.
	VS_aktuell	aktuell	Der Versuch ist aktuell.
	VS_ueberraschend	ueberraschend	Der Versuch ist überraschend.
	VS_abstrakt	abstrakt	Der Versuch ist abstrakt.
	VS_anspruchsvoll	anspruchsvoll	Der Versuch ist anspruchsvoll.
	VS_vernetzt	vernetzt	Der Versuch ist interdisziplinär / vernetzt unterschiedliche Fachbereiche.
Anleitung	Anl_klar	Anleitung (klar strukturierte Anleitung)	Die Anleitung ist klar strukturiert.
	Anl_Ziel	Anleitung (klares Versuchsziel)	Das Versuchsziel ist anhand der Anleitung klar erkennbar.
Kontext	Ktxt_Schulalltag	Kontext zum Schulalltag	Es wird ein Kontext zum Schulalltag hergestellt.
	Ktxt_Lebensalltag	Kontext zum Lebensalltag	Es wird ein Kontext zum Lebensalltag hergestellt.
Vorbereitung	Vorb	effektive Vorbereitung	Der Versuch ließ sich effektiv mit den gegebenen Literaturhinweisen, der Anleitung und den Aufgaben vorbereiten.
Material	VSM_Schulsamm	Versuchsmaterial in Schulsammlung	Das Versuchsmaterial ist in einer gut ausgestatteten Schulsammlung vorhanden.
	VSM_beschaffbar	Versuchsmaterial leicht beschaffbar	Das Versuchsmaterial ist in vereinfachter Form für die Schule leicht beschaffbar oder herstellbar.
	VSM_modern	moderne Versuchsmaterialien	Das Versuchsmaterial ist modern / innovativ.
	VSM_ungewoehnlich	ungewöhnliche Versuchsmaterialien	Das Versuchsmaterial ist ungewöhnlich.
	VSM_Quali	Versuchsmaterial von guter Qualität	Das Versuchsmaterial ist von guter Qualität.
	VSM_empfindlich	empfindliche Versuchsmaterialien	Das Versuchsmaterial ist sehr empfindlich / leicht zu beschädigen.

Tabelle 62: *Der erste Teil der 27 Items des neu entwickelten Instruments zur Erhebung der Eigenschaften von Praktikumsversuchen. Das Itemkürzel wird zur Übersicht in den Auswertungen verwendet. Kategorie 1 gibt dabei den Bezug auf die Elemente und Phasen eines Praktikumsversuches wie zum Beispiel Durchführung an. Kategorie 2 bezieht sich übergeordneter auf Strukturelle Aspekte / Zugang zum Versuch (Struktur), Fachliche Anforderungen auf universitärer Ebene (Uni), schulpraktische und fachdidaktische Aspekte (Schule), Aspekte zu Motivation und Interesse im Allgemeinen (Motivation). (Fortsetzung in Tabelle 63)*

Kategorie 1	Itemcode	Itemkürzel	Itemwortlaut
Aufbau	Aufbau	eigenständiger Versuchsaufbau	Der Versuch ist von uns mit vorhandenen Materialien eigenständig aufgebaut worden.
Aufgaben	Aufg_klar	Anleitung (verständliche Aufgabenstellung)	Die Versuchsaufgaben sind verständlich formuliert.
	Aufg_eng	eng geführte Versuchsaufgaben	Die Versuchsaufgaben sind eng geführt.
	Aufg_trick	Aufgaben erfordert „trickreiche“ Idee	Die Bearbeitung der Versuchsaufgaben erfordert eine „trickreiche“ Idee.
Durchführung	Durchf_Geschick	Geschick fordernde Durchführung	Die Durchführung des Versuches erfordert praktische Geschicklichkeit.
	Durchf_repetitiv	repetitive Versuchsdurchführung	Die Durchführung des Versuches enthält Dopplungen / ist repetitiv.
	Durchf_Ideen	Einbringung eigener Ideen möglich	Die Durchführung des Versuches bietet Raum für eigene Ideen und selbstständiges Arbeiten.
Auswertung	Ausw_quanti	viel quantitative Auswertung	Die Auswertung des Versuches hat einen hohen quantitativen Anteil.
	Ausw_eindeutig	eindeutig interpretierbare Ergebnisse	Die Auswertung des Versuches liefert eindeutig interpretierbare Ergebnisse.

Tabelle 63: *Fortsetzung von Tabelle 62: Zweiten Teil der 27 Items des neu entwickelten Instruments zur Erhebung der Eigenschaften von Praktikumsversuchen. Das Itemkürzel wird zur Übersicht in den Auswertungen verwendet. Kategorie 1 gibt dabei den Bezug auf die Elemente und Phasen eines Praktikumsversuches wie zum Beispiel Durchführung an. Kategorie 2 bezieht sich übergeordneter auf Strukturelle Aspekte / Zugang zum Versuch (Struktur), Fachliche Anforderungen auf universitärer Ebene (Uni), schulpraktische und fachdidaktische Aspekte (Schule), Aspekte zu Motivation und Interesse im Allgemeinen (Motivation).*

Itemcode	Itemkürzel	Itemwortlaut	Antwortmöglichkeiten
	Schwierigkeitsgrad	Bitte bewerten Sie die Schwierigkeit des Versuches in den nachfolgend gelisteten Teilbereichen und als Gesamtversuch.	5-teilige Likert-Skala (sehr leicht bis sehr schwer)
schwierig_theo	Theorie	Theoretische Grundidee	
schwierig_aufbau	Aufbau	Messaufbau	
schwierig_tech	Messtechnik	Messtechnik	
schwierig_durchf	Durchführung	Versuchsdurchführung	
schwierig_ausw	Auswertung	Auswertung	
schwierig_fehler	Fehlerrechnung	Fehlerrechnung	
schwierig_alles	Gesamtschwierigkeit	Gesamteinschätzung des Versuches	
	Zeitdauer	Wieviel Zeit haben Sie für den Versuch benötigt?	Freitext Stundenanzahl
t_vorb	Vorbereitung	Vorbereitung	
t_vs	Durchführung	Versuchstag im Labor	
t_nachb	Nachbereitung	Nachbereitung inkl. Erstversion Protokoll	
	Schulbezug		
inSchule	Schulerinnerung	Wurde während Ihrer eigenen Schulzeit ein ähnlicher Versuch im Physikunterricht durchgeführt?	Ja / Nein / Ich kann mich nicht erinnern.

Tabelle 64: *Überblick über die zusätzlichen Fragen und deren Antwortmöglichkeiten im Fragebogen zu den Versuchseigenschaften in Hinsicht auf Schwierigkeitsgrad, Bearbeitungsdauer und Schulbezug.*

A.2.3 Interesse

Die Beschreibungen des Erhebungsinstrumentes zum situationalen Interesse aus dem Abschnitt 11.1.1 werden in diesem Abschnitt um Details ergänzt.

Überblick Interesse

Konstrukt:	Interesse
Subskalen:	Interesse an Physik und am Unterrichten von Physik
Anzahl Items:	12 (6+6)
Ratingskala:	5-stufige Likert-Skala (1 = „*trifft gar nicht zu*“, 5 = „*trifft völlig zu*“)
Messzeitpunkt:	Prä- und Post-Test
Literatur:	Keller (u.a. 2011)
Querverweis:	Abschnitt 11.1.1

Kenngrößen

Skala	n	M	SD	SE	Min	Max	α
Interesse							
Physik	88	3.75	0.51	0.05	2	5	.65
Physikunterricht	88	4.33	0.53	0.06	3	5	.88

Tabelle 65: *Deskriptive Statistik für das Interesse an Physik und am Unterrichten von Physik exemplarisch zum Messzeitpunkt Post 1 nach dem ersten Praktikumssemester. Für die Deskription der Datensätze sind Stichprobengröße (n), Mittelwert (M), Standardabweichung (SD), Standardfehler (SE), der Wertebereich (Min, Max) und Cronbachs α angegeben.*

Interesse an Physik (Post 1)

Itemcode	**Itemwortlaut**	$\alpha_{ohneItem}$	$r_{i(t-i)}$
INT_PHY_FI_1	Vorlesungen und Praktika im Fach Physik gehören zu meinen Lieblings-Lehrveranstaltungen.	.64	.30
INT_PHY_FI_2	Die Inhalte der Physik-Lehrveranstaltungen sind mir manchmal eher gleichgültig. *[negativ formuliert]*	.62	.37
INT_PHY_FI_3	Im Allgemeinen freue ich mich auf Physik-Lehrveranstaltungen.	.55	.55
INT_PHY_SI_1	Ich kann mich für Physik begeistern.	.59	.50
INT_PHY_SI_2	Außerhalb des Physikstudiums befasse ich mich nicht mit Physik. *[negativ formuliert]*	.66	.28
INT_PHY_SI_3	Wenn ich genügend Zeit hätte, würde ich mich mit bestimmten physikalischen Themen, auch unabhängig von Prüfungsanforderungen, intensiver beschäftigen.	.60	.41

Tabelle 66: *Itemanalyse der Skala zum Interesse an Physik exemplarisch zum Messzeitpunkt Post 1 nach dem ersten Praktikumssemester als Ergänzung zu den Erläuterungen in Abschnitt 11.1.1. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .65$.*

Interesse am Unterrichten von Physik (Post 1)

Itemcode	**Itemwortlaut**	$\alpha_{ohneItem}$	$r_{i(t-i)}$
INT_UNT_1	Ich erkläre gerne physikalische Zusammenhänge.	.85	.71
INT_UNT_2	Physik zu unterrichten besitzt für mich eine große Bedeutung.	.87	.64
INT_UNT_3	Ich freue mich darauf, SchülerInnen Physik beizubringen.	.85	.74
INT_UNT_4	Ich überlege gern, wie man physikalische Zusammenhänge anschaulich erklären kann.	.87	.60
INT_UNT_5	Das Vermitteln physikalischer Inhalte mach mir Spaß.	.84	.82
INT_UNT_6	Ich sehe meiner zukünftigen Tätigkeit als PhysiklehrerIn freudig entgegen.	.87	.64

Tabelle 67: *Itemanalyse der Skala zum Interesse am Unterrichten von Physik exemplarisch zum Messzeitpunkt Post 1 nach dem ersten Praktikumssemester als Ergänzung zu den Erläuterungen in Abschnitt 11.1.1. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu* $\alpha = .88$.

A.2.4 Situationales Interesse

Dieser Anhang ergänzt die Beschreibungen des Erhebungsinstrumentes zum situationalen Interesse aus dem Abschnitt 11.1.1.

Überblick situationales Interesse

Konstrukt:	situationales Interesse
Subskalen:	Catch (Aufmerksamkeit, Emotion)
	Hold (Wert, Epistemisch)
Anzahl Items:	16 (8+8)
Ratingskala:	5-stufige Likert-Skala
	(1 = „*trifft gar nicht zu*“, 5 = „*trifft völlig zu*“)
Messzeitpunkt:	versuchsbegleitend (für 2x12 Praktikumsversuche)
Literatur:	Lewalter und Knogler (2014), siehe auch Knogler (2014)
Querverweis:	Abschnitt 11.1.1

Kenngrößen

Situationales Interesse - Catch

Versuch	M	SD	SE	Min	Max	n	α	$\bar{r}_{i(t-i)}$
Gesamtdatensatz	3.50	0.85	0.02	1.00	5.00	1805	.93	.77
Fallender Doppler-Effekt	3.81	0.57	0.06	2.48	5.08	78	.93	.76
Brechung, Reflexion, Polarisation	3.78	0.66	0.08	1.91	4.94	71	.94	.80
Thermohaus	3.72	0.67	0.08	2.00	5.65	65	.91	.72
Photoeffekt	3.69	0.61	0.07	2.32	4.98	72	.93	.77
Lego-Interferometer	3.66	0.90	0.11	1.07	6.03	71	.95	.81
Schwarzer Strahler	3.65	0.66	0.08	1.97	5.25	68	.92	.74
Augenmodell	3.64	0.69	0.08	1.71	4.93	72	.95	.83
Lorentzkraft	3.62	0.56	0.06	1.95	4.70	76	.93	.75
Franck-Hertz-Versuch	3.62	0.79	0.09	1.68	6.03	71	.95	.80
Fieldplotter	3.58	0.80	0.09	1.10	5.15	74	.94	.79
Hören	3.57	0.86	0.09	1.72	5.49	88	.95	.81
Rutherford'sche Streuung	3.53	0.64	0.08	1.46	4.69	72	.94	.78
Verformung eines Balles	3.52	0.67	0.08	1.58	4.62	77	.91	.72
Stöße ausgedehnter Körper	3.50	0.61	0.07	2.18	4.87	75	.89	.67
Oszillierendes Smartphone	3.49	0.77	0.08	1.03	4.97	89	.94	.80
Roller Coaster	3.48	0.76	0.25	2.08	4.37	9	.94	.81
Newtonsche Axiome	3.44	0.59	0.07	2.10	4.58	78	.91	.71
Zerfallsgesetz	3.44	0.67	0.08	1.85	4.90	73	.92	.73
Reibung	3.42	0.61	0.07	2.03	4.70	80	.93	.76
Wärmeäquivalent	3.31	0.64	0.08	1.86	4.41	68	.94	.80
Wasserstromkreis	3.30	0.73	0.09	1.31	4.73	74	.93	.77
Auf- und Entladung Kondensator	3.26	0.70	0.07	1.28	4.98	88	.93	.76
Wärmepumpe	3.18	0.69	0.08	1.43	4.65	67	.92	.73
Interferenz am Smartphone	3.15	0.78	0.09	1.23	5.21	73	.93	.77
Feldeffekttransistor	3.07	0.92	0.11	1.06	4.86	76	.96	.84

Tabelle 68: *Deskriptive Statistik für die **Catch-Komponente des situationalen Interesses**. In der ersten Zeile der Tabelle sind die Werte für den Gesamttest angegeben und in den darauffolgenden Zeilen für die einzelnen Praktikumsversuche mit Zentrierung. Für die Deskription der Datensätze sind Mittelwert (M), Standardabweichung (SD), Standardfehler (SE), Wertebereich auf einer 5-teiligen Likert-Skala (Min, Max), Stichprobengröße (n), Cronbachs α und die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$) angegeben.*

Situationales Interesse - Hold

Versuch	M	SD	SE	Min	Max	n	α	$\bar{r}_{i(t-i)}$
Gesamtdatensatz	3.30	0.68	0.02	1.00	5.00	1805	.85	.58
Thermohaus	3.55	0.57	0.07	2.12	5.00	65	.84	.59
Brechung, Reflexion, Polarisation	3.44	0.49	0.06	1.87	4.50	71	.85	.60
Fallender Doppler-Effekt	3.43	0.50	0.06	1.94	4.49	78	.84	.57
Franck-Hertz-Versuch	3.41	0.59	0.07	2.03	5.95	71	.85	.58
Schwarzer Strahler	3.41	0.51	0.06	2.32	4.86	68	.82	.54
Fieldplotter	3.37	0.58	0.07	1.37	5.12	74	.88	.65
Lorentzkraft	3.36	0.42	0.05	2.00	4.20	76	.82	.53
Stöße ausgedehnter Körper	3.35	0.41	0.05	2.14	4.20	75	.78	.49
Photoeffekt	3.35	0.42	0.05	2.34	4.74	72	.80	.52
Augenmodell	3.35	0.54	0.06	2.16	4.56	72	.89	.67
Reibung	3.34	0.50	0.06	2.07	4.38	80	.80	.52
Oszillierendes Smartphone	3.33	0.55	0.06	1.30	4.62	89	.82	.54
Newtonsche Axiome	3.31	0.43	0.05	2.20	4.32	78	.82	.53
Lego-Interferometer	3.27	0.58	0.07	1.99	5.45	71	.85	.60
Wärmepumpe	3.27	0.50	0.06	1.97	4.38	67	.86	.60
Roller Coaster	3.26	0.59	0.20	2.25	3.87	9	.87	.62
Hören	3.24	0.71	0.08	1.18	4.70	88	.87	.64
Verformung eines Balles	3.23	0.43	0.05	2.10	4.14	77	.77	.48
Wasserstromkreis	3.22	0.58	0.07	1.22	4.28	74	.84	.58
Rutherford'sche Streuung	3.22	0.44	0.05	1.91	3.97	72	.85	.58
Wärmeäquivalent	3.19	0.50	0.06	2.05	4.21	68	.86	.61
Zerfallsgesetz	3.18	0.49	0.06	2.15	4.42	73	.84	.58
Feldeffekttransistor	3.16	0.80	0.09	0.85	4.95	76	.92	.73
Auf- und Entladung Kondensator	3.13	0.48	0.05	1.80	4.12	88	.83	.55
Interferenz am Smartphone	3.12	0.55	0.06	1.50	4.25	73	.88	.65

Tabelle 69: *Deskriptive Statistik für die **Hold-Komponente des situationalen Interesses**. In der ersten Zeile der Tabelle sind die Werte für den Gesamttest angegeben und in den darauffolgenden Zeilen für die einzelnen Praktikumsversuche mit Zentrierung. Für die Deskription der Datensätze sind Mittelwert (M), Standardabweichung (SD), Standardfehler (SE), Wertebereich auf einer 5-teiligen Likert-Skala (Min, Max), Stichprobengröße (n), Cronbachs α und die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$) angegeben.*

Subskala	Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
Catch	SITINT_CA_AUFM_1	Die Inhalte des Versuches haben mein Interesse geweckt. *[neu formuliert]*	.92	.78
	SITINT_CA_AUFM_2	Der Versuch hat meine Neugier geweckt.	.93	.75
	SITINT_CA_AUFM_3	Der Versuch konnte meine Aufmerksamkeit fesseln.	.92	.81
	SITINT_CA_AUFM_4	Der Versuch war langweilig. *[negativ umformuliert]*	.93	.73
	SITINT_CA_EMO_1	Die Inhalte des Versuches finde ich spannend.	.93	.77
	SITINT_CA_EMO_2	Der Versuch hat mir Spaß gemacht.	.92	.79
	SITINT_CA_EMO_3	Den Versuch finde ich faszinierend. *[neu formuliert]*	.93	.75
	SITINT_CA_EMO_4	Der Versuch war unterhaltsam.	.92	.78
Hold	SITINT_HO_EPI_1	Über Teile des heutigen Praktikums möchte ich gerne mehr erfahren.	.81	.57
	SITINT_HO_EPI_2	Ich bin während des Praktikumstages auf Themen gestoßen, zu denen ich gerne mehr Informationen hätte.	.81	.61
	SITINT_HO_EPI_3	Durch den Versuch haben sich Fragen ergeben, auf die ich gerne eine Antwort hätte.	.82	.51
	SITINT_HO_EPI_4	Ich würde gerne mehr Versuche zu diesem Inhalt durchführen. *[neu formuliert]*	.80	.62
	SITINT_HO_WERT_1	Das Thema des heutigen Versuches ist für mich persönlich relevant.	.81	.60
	SITINT_HO_WERT_2	Die Inhalte des Versuches erachte ich für mich persönlich als wichtig.	.81	.62
	SITINT_HO_WERT_3	Die Beschäftigung mit den Inhalten des Versuches war für mich nützlich.	.81	.57
	SITINT_HO_WERT_4	Die heute bearbeiteten Inhalte sehe ich nicht als bedeutsam an. *[negativ, neu formuliert]*	.83	.38

Tabelle 70: *Itemanalyse exemplarisch für die Gesamtskala des situationalen Interesses in den beiden Subskalen Catch und Hold als Ergänzung zu den Erläuterungen in Abschnitt 11.1.1. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha_{Catch} = .93$ und $\alpha_{Hold} = .85$.*

A.2.5 Selbstwirksamkeit

Dieser Anhang bietet einen Überblick über das Erhebungsinstrument zur Selbstwirksamkeit zum Experimentieren im Schulunterricht ergänzend zu den Beschreibungen aus Abschnitt 11.1.2.

Überblick Selbstwirksamkeit

Konstrukt:	Selbstwirksamkeitserwartung
Subskalen:	in Bezug auf Planung von Experimenten und Durchführung von Experimenten im Unterricht
Anzahl Items:	16 (8+8)
Ratingskala:	5-stufige Likert-Skala (1 = „*trifft gar nicht zu*“, 5 = „*trifft völlig zu*“)
Messzeitpunkt:	Prä- und Post-Test
Literatur:	Rabe et al. (2012) und Meinhardt (2018)
Querverweis:	Abschnitt 11.1.2

Kenngrößen

Skala	n	M	SD	SE	Min	Max	α
Selbstwirksamkeit							
Planung	88	3.47	0.51	0.05	2	5	.74
Durchführung	86	3.52	0.56	0.06	2	5	.84

Tabelle 71: *Deskriptive Statistik für die Selbstwirksamkeit in Bezug auf Planung und Durchführung von Experimenten im Physikunterricht exemplarisch zum Messzeitpunkt Post 1 nach dem ersten Praktikumssemester. Für die Deskription der Datensätze sind Stichprobengröße (n), Mittelwert (M), Standardabweichung (SD), Standardfehler (SE), der Wertebereich (Min, Max) und Cronbachs α angegeben.*

Selbstwirksamkeit - Planung (Post 1)

Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
SWE_EX_P1	Ich kann in meiner Unterrichtsplanung zu den Lernzielen passende Experimente aufbauen, auch wenn die Physiksammlung schlecht ausgestattet ist.	0.71	0.43
SWE_EX_P2	Ich kann bei meiner Unterrichtsplanung ein Experiment gegebenenfalls so variieren, dass ich es in einer Übungsphase einsetzen kann, auch wenn ich es bisher nur als Einstiegsexperiment genutzt habe.	0.69	0.52
SWE_EX_P3	Ich kann Schülerexperimente so zusammenstellen, dass die Fähigkeiten meiner Schülerinnen und Schüler auf verschiedenen Niveaus gefördert werden, auch wenn ich bei der Planung unter Zeitdruck stehe.	0.72	0.39
SWE_EX_P4	Ich kann ein Experiment planen, das meine Schülerinnen und Schüler begeistert, auch wenn sie sich sonst wenig für Physik interessieren.	0.72	0.37
SWE_EX_P5	Ich kann für ein physikalisches Experiment begründet entscheiden, ob es didaktisch sinnvoller ist, es als Demonstrations- oder Schülerexperiment einzuplanen, auch wenn ich das Experiment noch nicht eingesetzt habe.	0.73	0.33
SWE_EX_P6	Ich kann in meiner Unterrichtsvorbereitung ein Experiment planen, welches meine Schülerinnen und Schüler dazu anregt, physikalische Fragestellungen zu entwickeln, auch wenn ich dieses Experiment neu entwickeln muss.	0.68	0.56
SWE_EX_P7	Ich kann mehrere Experimente so zusammenstellen, dass bei der Auswertung unterschiedliche Möglichkeiten des Umgangs mit Messdaten deutlich werden, auch wenn ich diesbezüglich keine Unterrichtsvorschläge kenne.	0.70	0.45
SWE_EX_P8	Ich kann bei der Unterrichtsplanung didaktisch begründet entscheiden, ob ein Experiment mit Hilfe von schultypischen Experimentiersets oder mit Alltagsgegenständen durchgeführt werden soll, auch wenn ich die Lerngruppe noch nicht lange kenne.	0.72	0.40

Tabelle 72: *Itemanalyse der Skala zur Selbstwirksamkeit in Bezug auf Planung von Experimenten im Physikunterricht exemplarisch zum Messzeitpunkt Post 1 nach dem ersten Praktikumssemester als Ergänzung zu den Erläuterungen in Abschnitt 11.1.2. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .74$.*

Selbstwirksamkeit - Durchführung (Post 1)

Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
SWE_EX_D1	Ich kann physikalische Experimente an interessante Impulse meiner Schülerinnen und Schüler anpassen, auch wenn ich das vorher nicht geplant hatte.	0.82	0.66
SWE_EX_D2	Ich kann ein Demonstrationsexperiment für meine Schülerinnen und Schüler nachvollziehbar durchführen, auch wenn es sich um eine komplexe Versuchsanordnung handelt.	0.83	0.54
SWE_EX_D3	Ich kann beim Experimentieren spontan mit den Schülerinnen und Schülern über das Wechselspiel von Theorie und Experiment reflektieren, auch wenn ich den Anlass nicht vorgesehen hatte.	0.82	0.66
SWE_EX_D4	Ich kann auf unvorhergesehene Verständnisschwierigkeiten meiner Schülerinnen und Schüler beim Experimentieren so reagieren, dass sie selbstständig weiterarbeiten können, auch ohne einfach einen Lösungsweg vorzugeben.	0.82	0.63
SWE_EX_D5	Ich kann ein Experiment, das im Physikunterricht nicht auf Anhieb funktioniert, zum Laufen bringen, auch wenn ich unter Zeitdruck stehe.	0.84	0.50
SWE_EX_D6	Ich kann unerwartete Messwerte aus einem Demonstrationsexperiment spontan als Lernanlass für meine Schülerinnen und Schüler nutzen, auch ohne ünpassende"Werte einfach zu übergehen.	0.84	0.50
SWE_EX_D7	Ich kann ein Experiment so inszenieren, dass meine Schülerinnen und Schüler motiviert sind mitzuarbeiten, auch wenn es sich um eine unbeliebte Randstunde handelt.	0.84	0.50
SWE_EX_D8	Ich kann meine Schülerinnen und Schüler bei der Planung ihres experimentellen Vorgehens unterstützen, auch wenn sie im Physikunterricht ihren eigenen Fragestellungen nachgehen.	0.82	0.65

Tabelle 73: *Itemanalyse der Skala zur Selbstwirksamkeit in Bezug auf Durchführung von Experimenten im Physikunterricht exemplarisch zum Messzeitpunkt Post 1 nach dem ersten Praktikumssemester als Ergänzung zu den Erläuterungen in Abschnitt 11.1.2. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .84$.*

A.2.6 Selbsteinschätzungen Lernzuwachs

Dieser Anhang ergänzt die Skalendokumentation der Selbsteinschätzungen des Lernzuwachses des Fachwissens und der experimentellen Kompetenz anhand des Praktikumsqualitätstests (PraQ-Test) aus den Abschnitten 11.1.3 und 11.1.2.

Überblick Selbsteinschätzungen Lernzuwachs

Skalen:	Lernzuwachs Fachwissen und experimentelle Kompetenz
Subskalen:	Fachwissen: *Theorie und Praxis* Experimentelle Kompetenz: *Erwartungen formulieren, deklarative Gerätekenntnis, Versuchsplanung und Geräte zusammenstellen, Versuchsanordnung aufbauen, Messung durchführen, Messungen dokumentieren, Rohdaten graphisch / tabellarisch darstellen, mit Problemen und Fehlern umgehen, Daten interpretieren*
Anzahl Items:	Fachwissen: 8 (2 pro Subskala) Experimentelle Kompetenz: 40 (3 - 7 pro Subskala)
Ratingskala:	6-stufige Likert-Skala (1 = „*trifft gar nicht zu*“, 6 = „*trifft völlig zu*“)
Messzeitpunkte:	Post-Test
Literatur:	(Rehfeldt, 2018a, 2018b)
Querverweis:	Abschnitte 11.1.3 und 11.1.2

Kenngrößen

Skala	n	M	SD	SE	Min	Max	α
Zuwachs Fachwissen	87	4.90	0.60	0.06	2	6	.87
Zuwachs exp. Kompetenz	82	4.64	0.67	0.07	2	6	.97

Tabelle 74: *Deskriptive Statistik für die Selbsteinschätzungen des Fachwissenszuwachs und des Zuwachses an experimenteller Kompetenz (PraQ-Test) exemplarisch zum Messzeitpunkt Post 1. Für die Deskription der Datensätze sind Stichprobengröße (n), Mittelwert (M), Standardabweichung (SD), Standardfehler (SE), der Wertebereich (Min, Max) und Cronbachs α der Gesamtskala angegeben.*

Fachwissenszuwachs (SE)				
Subskala	Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
		Aufgrund dieses Praktikums...		
Theorie	fwtheo.theo	...kann ich wichtige Theorien zu dem behandelten Thema wiedergeben.	.84	.74
	fwtheo.ueberblick	...kann ich einen Überblick über die meisten behandelten Themen geben.	.86	.61
	fwtheo.anschaulich	...kann ich behandelte, komplizierte Sachverhalte anschaulich darstellen.	.85	.63
	fwtheo.typisch	...sehe ich mich nun in der Lage, eine typische Fragestellung des behandelten Gegenstandsbereiches zu bearbeiten.	.84	.73
Praxis	fwpraxis.quali	...kann ich die Qualität experimenteller Arbeitsweisen besser beurteilen.	.86	.60
	fwpraxis.sachverhalte	...verstehe ich die im Praktikum behandelten Sachverhalte besser.	.85	.66
	fwpraxis.theoexp	...kann ich Bezüge zwischen Theorie und Experiment herstellen.	.86	.59
	fwpraxis.fehler	...kenne ich verschiedene Fehlerquellen, die beim Experimentieren auftreten können.	.87	.50

Tabelle 75: *Itemanalyse der Gesamtskala zur Selbsteinschätzung (SE) des Fachwissenszuwachs (PraQ-Test) exemplarisch zum Messzeitpunkt Post 1 als Ergänzung zu den Erläuterungen in Abschnitt 11.1.3. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .87$.*

Zuwachs experimentelle Kompetenz (SE) - Teil 1				
Subskala	Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
		Aufgrund dieses Praktikums...		
Erwartungen formulieren	experw.erw	...kann ich konkrete Erwartungen an den Ausgang eines Experiments formulieren.	.96	.57
	experw.ausgang	...kann ich nun besser einschätzen, wie ein Experiment 'ausgehen' wird (Endergebnis).	.96	.54
	experw.konkret	...sind meine Erwartungen an die experimentellen Ergebnisse eines Versuchs konkreter geworden.	.96	.58
	experw.theo	...kann ich besser einschätzen, welche theoretischen Aspekte ich mir vor dem Experimentieren anschauen muss.	.96	.54
	experw.durchfuehrung	...kann ich die einzelnen Durchführungsschritte eines Experiments besser planen.	.96	.55
Gerätekenntnis (deklarativ)	expgerdekla.eignung	...kann ich die Eignung von Geräten für das Experimentieren besser einschätzen.	.96	.52
	expgerdekla.vor	...erkenne ich besser, welches Gerät ich vor mir habe.	.96	.56
	expgerdekla.welche	...weiß ich besser, welche Geräte ich konkret für das Experiment benötigen werde.	.96	.58

Tabelle 76: *Die Tabelle zeigt den ersten von fünf Teilen der Itemanalyse der Gesamtskala zur Selbsteinschätzung (SE) des Zuwachses an experimenteller Kompetenz (PraQ-Test) exemplarisch zum Messzeitpunkt Post 1 als Ergänzung zu den Erläuterungen in Abschnitt 11.1.2. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .97$.*

Zuwachs experimentelle Kompetenz (SE) - Teil 2				
Subskala	Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
		Aufgrund dieses Praktikums...		
Versuchsplanung und Geräte zusammenstellen	expverger.mess	...weiß ich in der Planung besser, welche Messgrößen ich während des Experimentierens verändern muss.	.96	.68
	expverger.geraet	...kann ich zum Experiment gehörende Gerätschaften besser auswählen.	.96	.63
	expverger.montieren	...kann ich die einzelnen Geräte besser montieren.	.96	.70
	expverger.problem	...kann ich beim Experimentieren besser auf unerwartete Probleme reagieren.	.96	.61
	expverger.eigen	...kann ich Experimente besser nach meinem eigenen Versuchsplan durchführen.	.96	.53
Versuchsanordnung aufbauen	expverauf.auf	...kann ich Versuchsanordnungen besser aufbauen.	.96	.64
	expverauf.test	...kann ich das Funktionieren von Versuchsanordnungen besser testen.	.96	.70
	expverauf.anbringen	...kann ich Bestandteile des Versuchsaufbaus besser anbringen.	.96	.65
	expverauf.fehler	...erkenne ich Aufbaufehler besser.	.96	.54

Tabelle 77: *Die Tabelle zeigt den zweiten von fünf Teilen der Itemanalyse der Gesamtskala zur Selbsteinschätzung (SE) des Zuwachses an experimenteller Kompetenz (PraQ-Test) exemplarisch zum Messzeitpunkt Post 1 als Ergänzung zu den Erläuterungen in Abschnitt 11.1.2. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .97$.*

Zuwachs experimentelle Kompetenz (SE) - Teil 3				
Subskala	Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
		Aufgrund dieses Praktikums...		
Messung durchführen	expmess.umgang	...kann ich besser mit Messgeräten umgehen.	.96	.70
	expmess.vollstaendig	...kann ich eine vollständige Datenaufnahme besser gewährleisten.	.96	.67
	expmess.reihenfolge	...bin ich besser in der Lage, eine Datenaufnahme Schritt für Schritt in logischer Reihenfolge zu gewährleisten.	.96	.75
	expmess.verfahren	...beherrsche ich mehr Messverfahren als vorher.	.97	.46
Messungen dokumentieren	expmessdok.proto	...kann ich das Messprotokoll nachvollziehbarer erstellen.	.96	.69
	expmessdok.doku	...bin ich besser in der Lage, eine vollständige Dokumentation der Datenaufnahme zu gewährleisten.	.96	.74
	expmessdok.welche	...habe ich einen besseren Überblick, welche Daten im Laborbuch dokumentiert werden müssen.	.96	.51

Tabelle 78: *Die Tabelle zeigt den dritten von fünf Teilen der Itemanalyse der Gesamtskala zur Selbsteinschätzung (SE) des Zuwachses an experimenteller Kompetenz (PraQ-Test) exemplarisch zum Messzeitpunkt Post 1 als Ergänzung zu den Erläuterungen in Abschnitt 11.1.2. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .97$.*

Zuwachs experimentelle Kompetenz (SE) - Teil 4				
Subskala	Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
Rohdaten graphisch / tabellarisch darstellen	exproh.tabell	...kann ich Rohdaten besser tabellarisch darstellen.	.96	.57
	exproh.graph	...kann ich Rohdaten besser graphisch darstellen.	.96	.68
	exproh.auswerten	...bin ich besser in der Lage, die Messdaten auszuwerten.	.96	.77
	exproh.musterdaten	...erkenne ich Zusammenhänge innerhalb meiner gewonnenen Daten besser.	.96	.72
	exproh.geeignet	...fällt es mir leichter, geeignete Darstellungsformen (Grafiken, Tabellen, Schemata) zu erstellen.	.96	.66
	exproh.repertoire	...nutze ich mehr unterschiedliche Darstellungsformen für die Darstellung meiner Ergebnisse.	.97	.54
		Aufgrund dieses Praktikums...		
mit Problemen und Fehlern umgehen	expprob.unsicherheit	...kann ich besser mit Messunsicherheiten umgehen.	.96	.56
	expprob.fehler	...beachte ich experimentelle Fehlerquellen adäquater.	.96	.52
	expprob.bericht	...kann ich Messunsicherheiten besser im Protokoll berichten.	.96	.54
	expprob.erkennen	...erkenne ich besser mögliche Fehlerquellen des Experiments.	.96	.69

Tabelle 79: *Die Tabelle zeigt den vierten von fünf Teilen der Itemanalyse der Gesamtskala zur Selbsteinschätzung (SE) des Zuwachses an experimenteller Kompetenz (PraQ-Test) exemplarisch zum Messzeitpunkt Post 1 als Ergänzung zu den Erläuterungen in Abschnitt 11.1.2. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .97$.*

Zuwachs experimentelle Kompetenz (SE) - Teil 5				
Subskala	Itemcode	Itemwortlaut	$\alpha_{ohneItem}$	$r_{i(t-i)}$
Daten interpretieren	expdat.mustergraph	...erkenne ich Zusammenhängein den grafischen Darstellungen meiner Daten besser.	.96	.69
	expdat.interpretieren	...kann ich besser Messwerte interpretieren.	.96	.78
	expdat.frage	...fällt es mir leichter, die Ergebnisse des Versuchs auf die angedachte Fragestellung zu beziehen.	.96	.70
	expdat.erwartung	...bin ich besser in der Lage, die Ergebnisse des Versuchs mit meinen anfänglichen Erwartungen zu vergleichen.	.96	.74
	expdat.schluss	...kann ich besser Schlussfolgerungen aus meinen Ergebnissen ziehen.	.96	.78
	expdat.quali	...kann ich die Qualität meiner Ergebnisse besser einschätzen.	.96	.75

Tabelle 80: *Die Tabelle zeigt den fünften von fünf Teilen der Itemanalyse der Gesamtskala zur Selbsteinschätzung (SE) des Zuwachses an experimenteller Kompetenz (PraQ-Test) exemplarisch zum Messzeitpunkt Post 1 als Ergänzung zu den Erläuterungen in Abschnitt 11.1.2. Neben den Wortlauten der Items sind als Kennwerte der Itemanalyse Cronbachs α ohne dieses Item und die Trennschärfe des Items $r_{i(t-i)}$ angegeben. Cronbachs α für die Gesamtskala ergibt sich zum Vergleich zu $\alpha = .97$.*

A.2.7 Fachwissenstest

Dieser Anhang dient der Ergänzung der Skalendokumentation des Fachwissenstest, über die in Abschnitt 11.1.3 ein Überblick gegeben wurde. Die einzelnen Aufgaben des Fachwissenstests sind im digitalen Anhang B zu finden.

Überblick Fachwissenstest

Skala:	Fachwissen Physik
Subskalen:	Themenbereiche: Messtechnik, Atomphysik, Optik, Wärmelehre, Mechanik, Elektrizitätslehre
Anzahl Aufgaben:	48 (8 pro Themenbereich bzw. 2 pro Praktikumsversuch)
Ratingskala:	Testscore hauptsächlich anhand von Entscheidungsfragen mit dichotomem Antwortformat und Auswahlaufgaben mit mehr als zwei Antwortmöglichkeiten (Multiple-Choice)
Messzeitpunkte:	Prä- und Post-Test
Literatur:	KiL (Kröger, 2019), FCI (Hestenes et al., 1992), BWOS (Mešić et al., 2016), BEMA (Ding et al., 2006), NEPS (Blossfeld, 2019), CSEM (Maloney et al., 2001), TUG-K (Beichner, 1994)
Querverweis:	Abschnitt 11.1.3

Kenngrößen

		M	SE	n	α	$\bar{r}_{i(t-i)}$	min	max
VS-nah	Prä	.50	.01	88	.63	.22	-.23	.55
	Post	.69	.01	88	.61	.20	-.17	.56

Tabelle 81: *Kennwerte für den **versuchsnahen Datensatz** (VS-nah) des Fachwissenstest ergänzend zu den Kennwerten des Gesamtdatensatzes in Tabelle 19 in Abschnitt 11.1.3. Dabei wird unterschieden zwischen dem Messzeitpunkt vor (Prä) und nach (Post) dem Praktikumssemester. Für die Deskription des Datensatzes sind Mittelwert (M) in anteiligem Testwert, Standardfehler (SE) und Stichprobengröße (n) angegeben. Als Kenngrößen für die Analyse des Testinstrumentes ist neben der Reliabilität (Cronbachs α) auch die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$) der Items sowie deren Extremwerte (min, max).*

MZP	Subskalen	Deskription				Itemtrennschärfe		
		M	SE	n	α	$\bar{r}_{i(t-i)}$	min	max
Prä	Messtechnik	.51	.02	89	.27	.10	-.08	.34
	Atomphysik	.57	.02	73	.22	.14	-.20	.37
	Optik	.37	.02	73	.34	.15	-.08	.33
	Wärmelehre	.40	.02	68	.48	.21	-.05	.38
	Mechanik	.55	.02	83	.53	.25	.08	.46
	E-Lehre	.33	.02	84	.29	.16	-.16	.52
Post	Messtechnik	.59	.02	88	.36	.14	-.08	.30
	Atomphysik	.72	.02	73	.39	.10	-.11	.41
	Optik	.56	.02	73	.49	.22	-.03	.40
	Wärmelehre	.56	.02	68	.43	.20	.05	.46
	Mechanik	.72	.02	82	.38	.16	-.08	.36
	E-Lehre	.43	.02	81	.14	.08	-.11	.28

Tabelle 82: *Kennwerte für die Subskalen des gesamten Fachwissenstest. Dabei wird unterschieden zwischen dem Messzeitpunkt (MZP) vor (Prä) und nach (Post) dem Praktikumssemester. Für die Deskription des Datensatzes sind Mittelwert (M) in anteiligem Testwert, Standardfehler (SE) und Stichprobengröße (n) angegeben. Als Kenngrößen für die Analyse des Testinstrumentes ist neben der Reliabilität (Cronbachs α) auch die mittlere Trennschärfe ($\bar{r}_{i(t-i)}$) der Items sowie deren Extremwerte (min, max).*

Lösungswahrscheinlichkeit und Bearbeitungsgrad

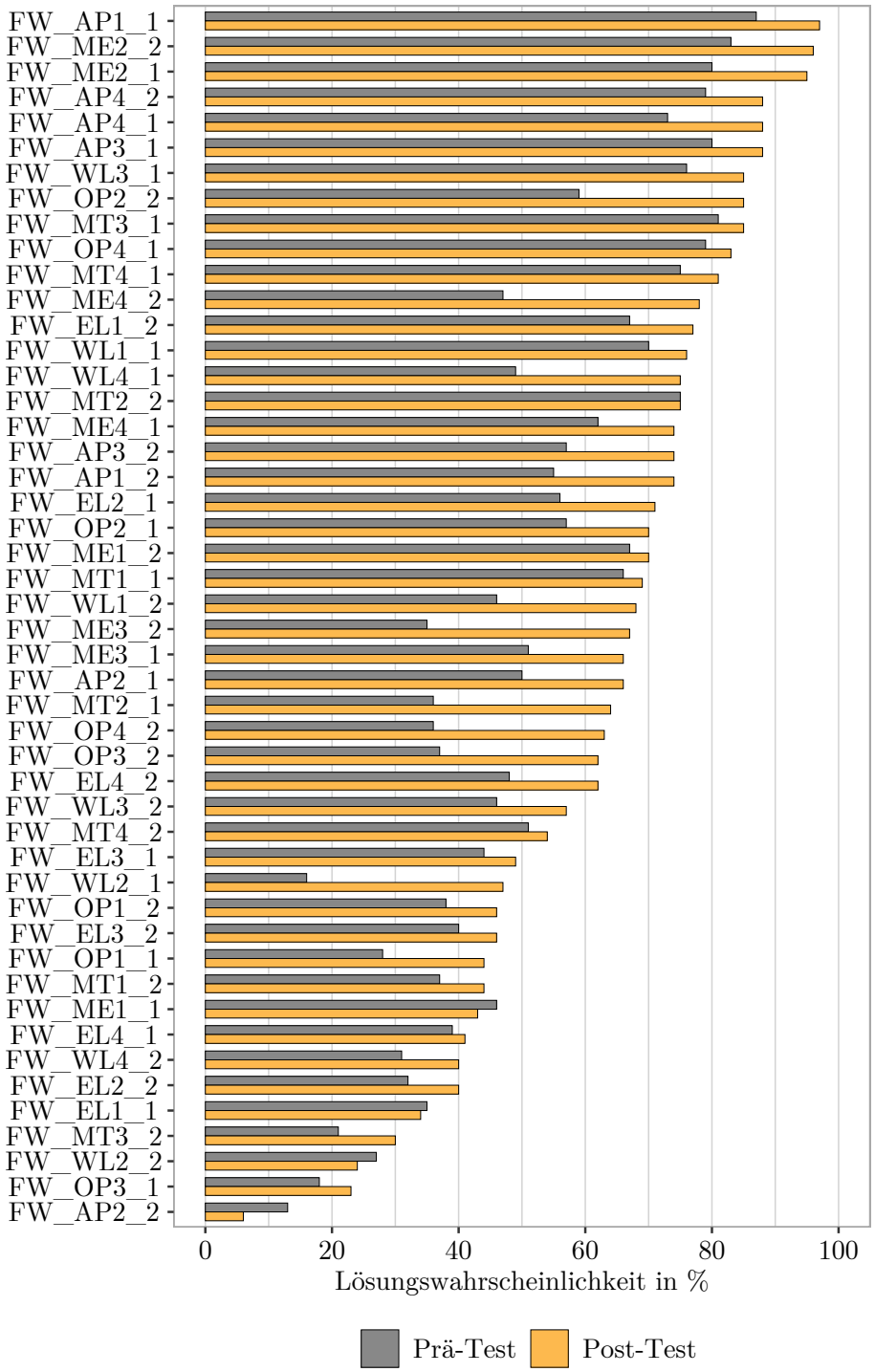


Abbildung 87: *Lösungswahrscheinlichkeit der einzelnen Fachwissensaufgaben für den Prä- und Post-Test im Vergleich. Die 48 Aufgaben sind nach absteigender Lösungswahrscheinlichkeit und damit zunehmender Aufgabenschwierigkeit im Post-Test sortiert. Die Fachwissensaufgaben sind im digitalen Anhang anhand der Itemkodierung zu finden. Die Aufgaben wurden im Mittel von $n = 78 \pm 1$ Studierenden beantwortet.*

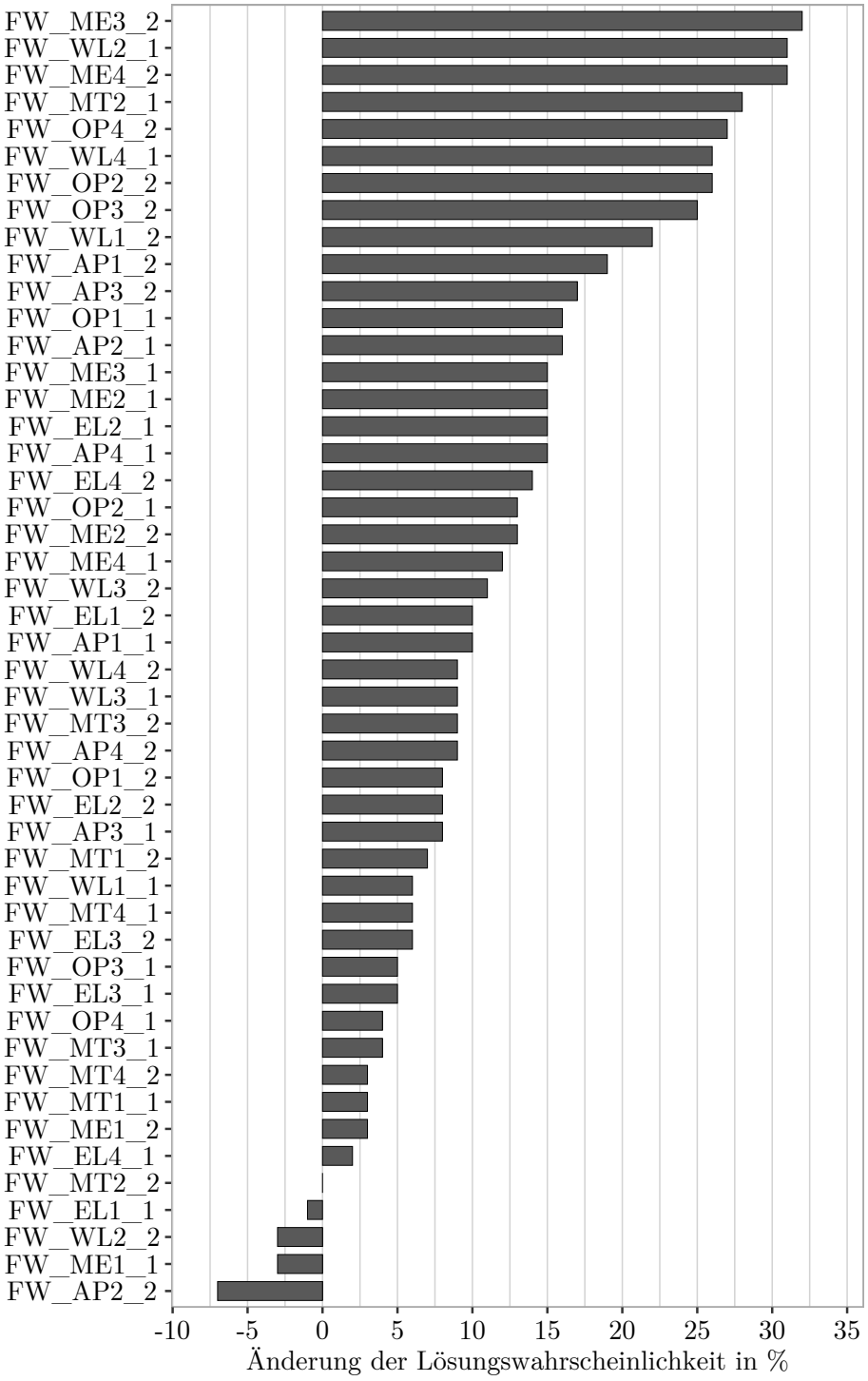


Abbildung 88: *Änderung der Lösungswahrscheinlichkeit der einzelnen Fachwissensaufgaben vom Prä-Test zum Post-Test im Vergleich. Die 48 Aufgaben sind nach absteigender Änderung der Lösungswahrscheinlichkeit sortiert. Die Fachwissensaufgaben sind im digitalen Anhang anhand der Itemkodierung zu finden. Die Aufgaben wurden im Mittel von* $n = 78 \pm 1$ *Studierenden beantwortet.*

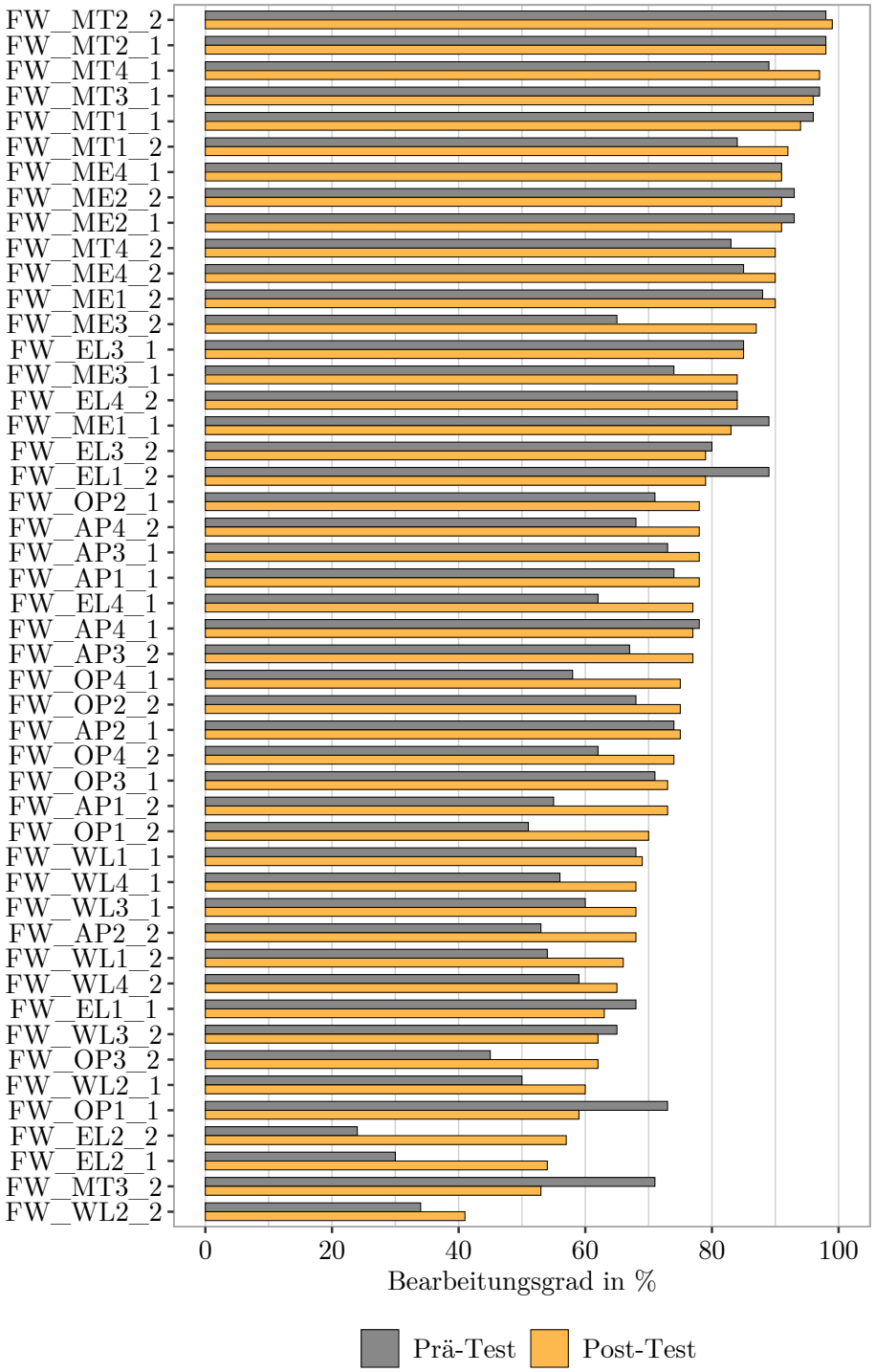


Abbildung 89: *Bearbeitungsgrad der einzelnen Fachwissensaufgaben für den Prä- und Post-Test im Vergleich. Die 48 Aufgaben sind nach absteigendem Bearbeitungsgrad und damit implizit zunehmender Aufgabenschwierigkeit im Post-Test sortiert. Die Fachwissensaufgaben sind im digitalenAnhang anhand der Itemkodierung zu finden. Die Aufgaben wurden im Mittel von* $n = 78 \pm 1$ *Studierenden beantwortet.*

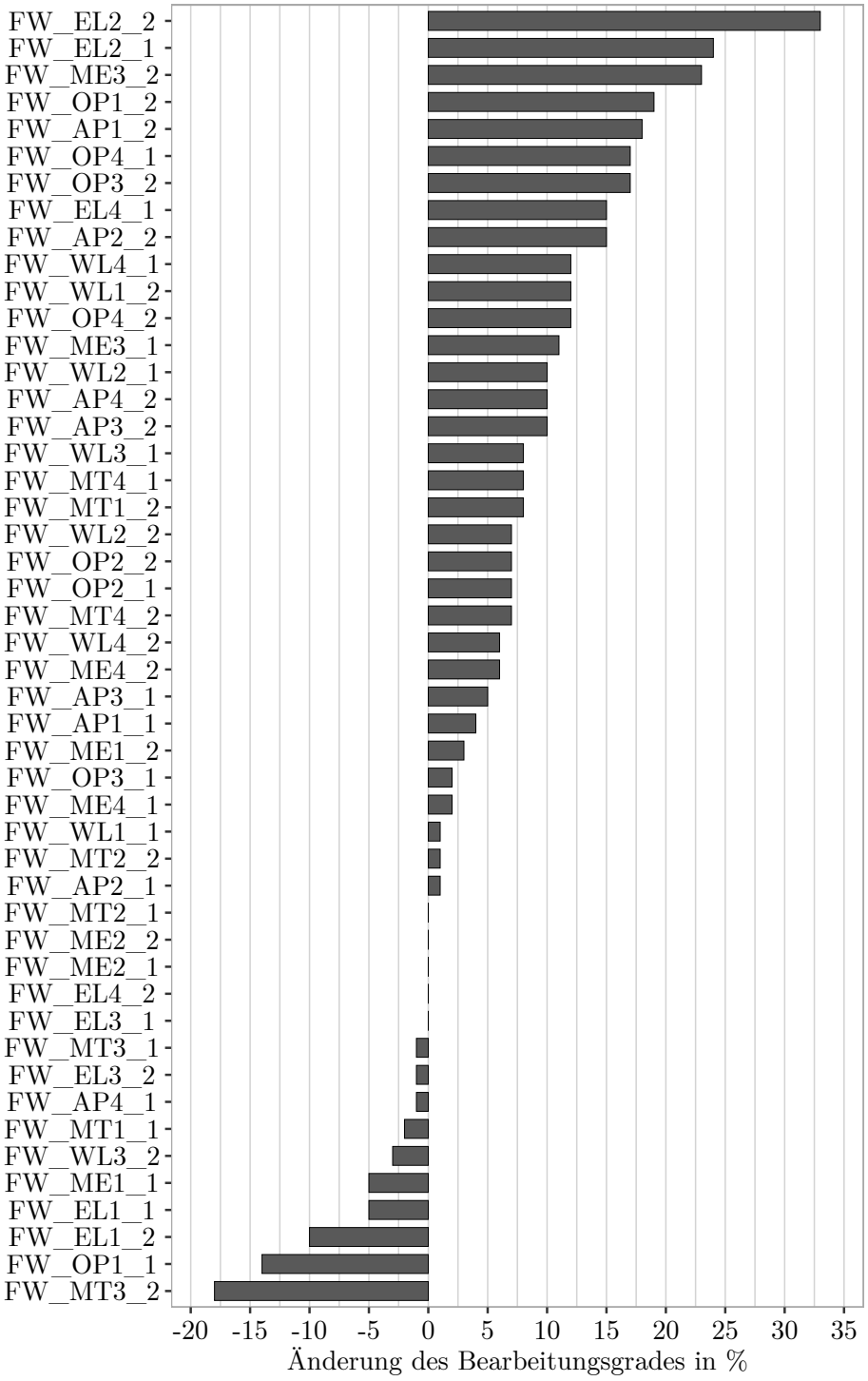


Abbildung 90: *Änderung des Bearbeitungsgrades der einzelnen Fachwissensaufgaben vom Prä-Test zum Post-Test im Vergleich. Die 48 Aufgaben sind nach absteigender Änderung des Bearbeitungsgrades sortiert. Die Fachwissensaufgaben sind im digitalen Anhang anhand der Itemkodierung zu finden. Die Aufgaben wurden im Mittel von* $n = 78 \pm 1$ *Studierenden beantwortet.*

Publikationen und Tagungsbeiträge

Publikationen

Andersen, J., Block, D., Neumann, K. & Wendlandt, H. (2018). Lehramtsausbildung Physik 2.0 - Vernetzung von Fach, Fachdidaktik und schulpraktischen Aspekten. In B. Brouër, A. Burda-Zoyke, J. Kilian & I. Petersen (Hrsg.), *Vernetzung in der Lehrerinnen-und Lehrerbildung: Ansätze, Methoden und erste Befunde aus dem LeaP-Projekt an der Christian-Albrechts-Universität zu Kiel* (S. 23–35). Waxmann Verlag.

Vorträge

Andersen, J., Block, D. & Neumann, K. Entwicklung und Evaluierung eines spezifischen Anfängerpraktikums für Lehramtsstudierende im Fach Physik. In: Vortrag. Klausurtagung zur Lehrkräftebildung LeaP@CAU und CAU-LiB, Sankelmark. 2016.

Andersen, J., Block, D. & Neumann, K. Entwicklung und Evaluierung eines spezifischen Anfängerpraktikums für Lehramtsstudierende im Fach Physik. In: Vortrag. DPG-Schule 2017 der Arbeitsgruppe Physikalische Praktika der Deutschen Physikalischen Gesellschaft, Bad Honnef. 2017.

Andersen, J., Block, D. & Neumann, K. Evaluation eines spezifischen Physikpraktikums fürs Lehramtsstudium. In: Vortrag. Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik 2017, Regensburg. 2017.

Poster

Andersen, J., Block, D. & Neumann, K. Entwicklung und Evaluierung eines spezifischen Anfängerpraktikums für Lehramtsstudierende im Fach Physik. In: Poster. Tag der Lehre an der Christian-Albrechts-Universität zu Kiel, Kiel. 2016.

Andersen, J., Block, D. & Neumann, K. Entwicklung und Evaluierung eines spezifischen Anfängerpraktikums für Lehramtsstudierende im Fach Physik. In: Poster. Frühjahrstagung der Deutschen Physikalischen Gesellschaft (DPG) 2016, Hannover. 2016.

Andersen, J., Block, D. & Neumann, K. Entwicklung und Evaluierung eines spezifischen Anfängerpraktikums für Lehramtsstudierende im Fach Physik. In: Poster. Klausurtagung zur Lehrkräftebildung LeaP@CAU und CAU-LiB, Leck. 2017.

Andersen, J., Block, D. & Neumann, K. Evaluierung eines spezifischen Anfängerpraktikums für Lehramtsstudierende im Fach Physik. In: Poster. Frühjahrstagung der lehrerbildenden Projekt der Christian-Albrechts-Universität zu Kiel, Kiel. 2018.

Andersen, J., Block, D. & Neumann, K. Was macht einen Versuch relevant für Lehramtsstudierende? In: Poster. Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik 2018, Kiel. 2018.

Andersen, J., Block, D. & Neumann, K. Organisationsform, Praktikumsversuche und Evaluation des adressatenspezifischen Anfängerpraktikums für Lehramtststudierende im Fach Physik. In: Poster. DPG-Schule 2019 der Arbeitsgruppe Physikalische Praktika der Deutschen Physikalischen Gesellschaft, Bad Honnef. 2019.

Block, D., Andersen, J. & Neumann, K. Entwicklung eines spezifischen Physikpraktikums fürs Lehramtsstudium. In: Poster. Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik 2017, Regensburg. 2017.

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich bei dieser Arbeit unterstützt haben. Mein besonderer Dank gilt dabei Prof. Dr. Dietmar Block und Prof. Dr. Knut Neumann. Ich danke beiden sehr für die Möglichkeit, in diesem vielseitigen Projekt zu promovieren und viele wertvolle Erfahrungen zu gewinnen. Ich bin sehr dankbar für die konstruktiven und lehrreichen Diskussionen, bei denen nie der Blick für das Wesentliche verloren ging. Dabei war sehr bereichernd, dass mir dieses besondere Betreuer-Duo sowohl bei fachlichen als auch bei fachdidaktische Fragen kompetent weitergeholfen hat. Ich bedanke mich sehr dafür, dass beide mir in diesem Projekt mit Rat und Tat zur Seite gestanden haben.

Ein ganz besonders großer Dank geht auch an Frank Brach, Mario Knüppel, Martin Laß, Michael Poser und Volker Rohwer sowie an die Teams der zentralen Werkstatt des Physik-Institutes und der Werkstatt am IPN. Ohne diese tatkräftige Unterstützung hätte ich das Praktikum nicht in dieser Form und in dieser kurzen Zeit von nur einem Jahr entwickeln können. Alle haben sich die Zeit genommen, um technische Umsetzungsmöglichkeiten mit mir zu diskutieren und auch große Ideen, wie z. B. einen 4 m hohen Fallturm, schnellstmöglich zu realisieren. Der Praktikumsbetrieb wäre zudem ohne die kurzfristigen Reparaturen nicht so reibungslos verlaufen.

Ganz besonderes Glück habe ich auch damit, dass ich mit dem IEAP und dem IPN sogar zwei motivierende und unterstützende Arbeitsumfelder mit tollen Kolleginnen und Kollegen habe. Besonders dankbar möchte ich an dieser Stelle Dr. Andrea Bernholt, Dr. Simon Grund, Dr. Melanie Keller, Dr. Stefan Sorge und Holger Wendlandt nennen, die mir bei Fragen sehr weitergeholfen haben. Ein großer Dank auch an alle, die an der Expertenrunde zur Sammlung der Versuchseigenschaften teilgenommen haben, Teile der Arbeit auf Rechtschreibung geprüft haben, Aspekte meiner Arbeit mit mir diskutiert haben und/oder mir wertvolle Anregungen gegeben haben. An dieser Stelle möchte ich besonders Oguz Han Asnaz, Alice Hesse, Katharina Hilger, Dr. Hendrik Jung, Dr. Markus Kubsch, Dustin Schiering, Kerstin Sgonina, Dr. Benjamin Tadsen und Dr. Jochen Wilms nennen. Zudem möchte ich mich bei Mirijam Bihl, Lulu Hoffmeister, Linda Sommer, Stephanie Thiedemann und Ulrike Weyerke sowie den Teams der EDV am IEAP und am IPN bedanken, die bei allen Herausforderungen mit Verwaltung bzw. PC schnell und unkompliziert geholfen haben.

Mein Dank gilt außerdem der Programminitiative CAU-LiB, die dieser Arbeit den Rahmen gegeben hat und viele Möglichkeiten für anregenden Austausch mit anderen Projekten und Fachbereichen geboten hat. In diesem Kontext möchte ich mich auch ganz herzlich bei Prof. Dr. Dr. h.c. Ilka Parchmann bedanken, die stets großes Interesse an meiner Arbeit gezeigt hat.

Außerdem möchte ich meinen Dank allen Hilfskräften in diesem Projekt aussprechen, die Praktikumsversuche auf ihre Tauglichkeit getestet haben, in vielen Stunden die Fragebögen sorgfältig digitalisiert haben und/oder mich bei der Betreuung des Praktikums unterstützt haben.

Mein Dank geht auch an alle Lehramtsstudierenden im Praktikum, die die Vielzahl an Fragebögen ausgefüllt haben. Außerdem danke ich für die direkte konstruktive Kritik, die zur Weiterentwicklung des Lehramtspraktikums beigetragen hat.

Nicht zuletzt und von ganzem Herzen möchte ich mich bei meiner Familie und meinen Freunden, bei meiner Herde, bedanken. Jeder war auf seine ganz eigene Art in dieser Zeit für mich da. Da waren viele offene Ohren, motivierende Worte und Zuspruch, Schokolade, Überraschungspakete und Postkarten, Aale, Radtouren und Spaziergänge, Eis, Rechtschreibprüfungen, Shiatsu, Yoga und so vieles mehr, was mich unterstützt hat.

Danke

Eidesstattliche Erklärung

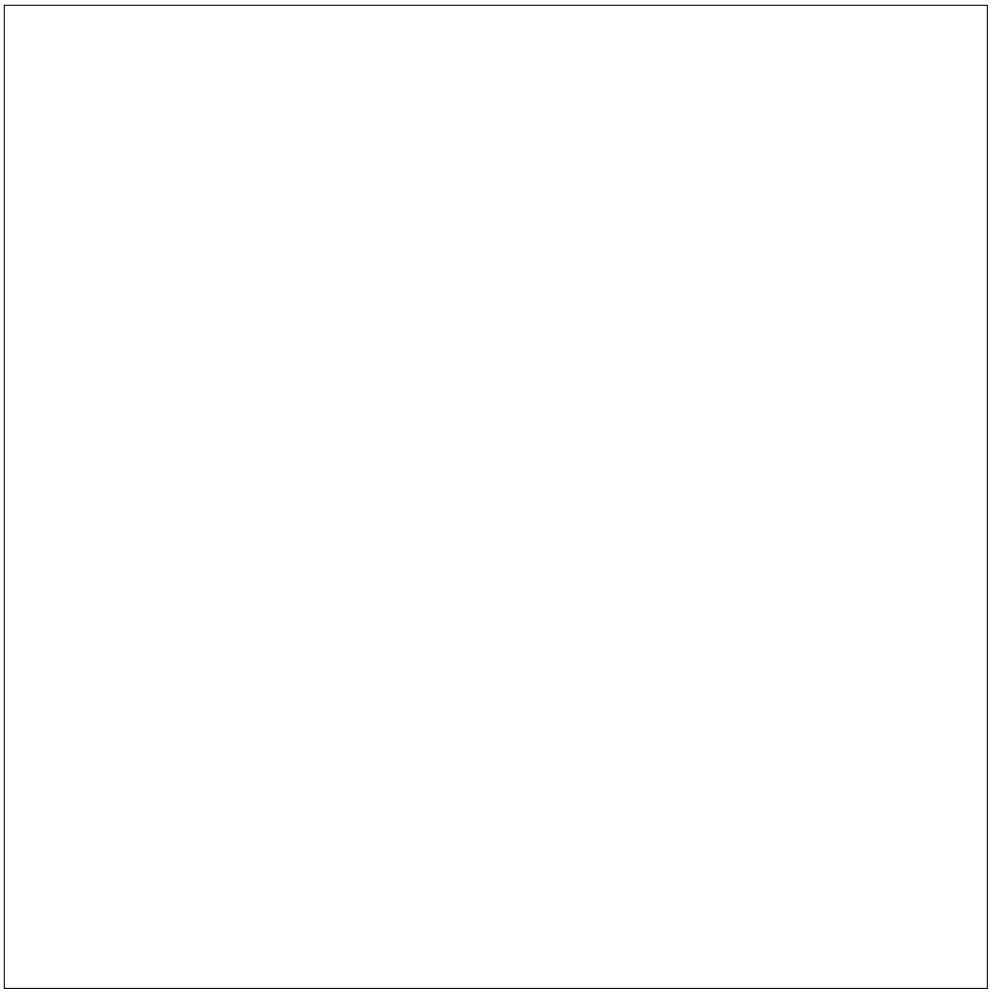
Hiermit erkläre ich, dass die vorliegende Dissertation – abgesehen von der Beratung durch meine Betreuer – nach Inhalt und Form meine eigene und nur mit den angegebenen Hilfsmitteln verfasste Arbeit ist. Die vorliegende Dissertation wurde weder ganz noch zum Teil einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt, veröffentlicht oder zur Veröffentlichung eingereicht. Teile der Befunde der vorliegenden Dissertation wurden bereits in Form eines Sammelbandbeiträges publiziert. Die vorliegende Arbeit ist unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden. Ich erkläre ferner, dass mir kein akademischer Grad entzogen wurde.

Kiel, den

Jasmin Andersen

B Digitaler Anhang

Folgender Anhang ist digital auf dem Datenträger unten zu finden bzw. kann am Institut für Experimentelle und Angewandte Physik der Christian-Albrechts-Universität zu Kiel im Bereich Lehramtspraktika angefragt werden:

B.1 Skalendokumention Fachwissenstest (Teil 2)

B.2 Versuchsbeschreibungen