

ZEITSCHRIFT
FÜR
EMPIRISCHE
HOCHSCHULFORSCHUNG



JG 4

HEFT 2/2020

STEFANIE BRENNING, ELKE WOLF MINT-Projekte für Schülerinnen an Hochschulen. Analyse des Wirkungsmechanismus und Meta-Evaluation der empirischen Evidenz

JOHANNES SCHULT Test- und notenbasierte Erfolgsprognose im wirtschaftswissenschaftlichen Masterstudium: Befunde aus dem Nationalen Bildungspanel

MAX KNAUT, MATTHIAS ZIEGLER Akzeptanz und Nutzung psychologischer Verfahren zur Auswahl von Promovierenden

KATHARINA THIES, ROBERT KORDTS-FREUDINGER Wie gehen Hochschullehrende mit ihren Emotionen in der Lehre um? Eine Experience-sampling-Befragung zum situativen Emotionsausdruck von Freude, Angst und Ärger



IMPRESSUM

Zeitschrift für empirische Hochschulforschung (ZeHf)

HERAUSGEBERINNEN

Prof. Dr. Joachim Grabowski, Institut für Psychologie, Leibniz Universität Hannover

PD Dr. Otto Hüther, INCHER-Kassel

Prof. Dr. Elke Wild, Leiterin der Arbeitseinheit Pädagogische Psychologie, Universität Bielefeld

WISSENSCHAFTLICHER BEIRAT

Prof. Dr. Stefan Fries, Fakultät für Psychologie und Sportwissenschaft, Abteilung Psychologie, Arbeitseinheit

Psychologie der Erziehung und Bildung, Universität Bielefeld; Prof. Dr. Anke Hanft, Institut für Pädagogik, Universität

Oldenburg; Prof. Dr. Michael Hölscher, Lehrstuhl für Hochschul- und Wissenschaftsmanagement, Deutsche Universität

für Verwaltungswissenschaften Speyer; Prof. Dr. Monika Jungbauer-Gans, Deutsches Zentrum für Hochschul- und

Wissenschaftsforschung, Leibniz Universität Hannover; Prof. Dr. Georg Krücken, Internationales Zentrum für Hoch-

schulforschung, Universität Kassel; Prof. Dr. Stephan Laske, Institut für Organisation und Lernen, Universität Innsbruck

ERSCHEINEN UND BEZUGSBEDINGUNGEN

Die Zeitschrift erscheint 2 x jährlich mit einem Jahresumfang von rd. 240 Seiten (Print und Online).

Einzelheft Print: 33,00 €

Jahresabonnement Print: Institutionen 87,00 €; Privat: 54,00 €

Jahresabonnement Online: Institutionen 131,00 €, Privat: 65,00 €

Jahresabonnement Print+Online: Institutionen 131,00 €; Privat 65,00 € Mit einem Online- oder einem Print+Online-Abonnement haben Sie freien Zugriff auf das gesamte Volltext-Archiv der ZeHf, solange Ihr Abonnement besteht.

Print-Preise jeweils zzgl. Versandkosten.

Die Bezugspreise enthalten die gültige Mehrwertsteuer. Abonnements verlängern sich automatisch um ein Jahr.

Abonnement-Kündigungen bitte schriftlich an den Verlag. Kündigungsfrist bis drei Monate zum Jahresende.

Das digitale Angebot finden Sie auf: <https://zehf.budrich-journals.de>.

BESTELLUNGEN BITTE AN DEN BUCHHANDEL ODER AN

Verlag Barbara Budrich, Stauffenbergstr. 7, D-51379 Leverkusen-Opladen

Tel.: (+49) (0)2171 79491 50 Fax: (+49) (0)2171 79491 69, info@budrich.de

www.budrich.de • www.budrich-journals.de • www.shop.budrich.de

Jg. 4, Heft 2 (2020)

© 2020 Verlag Barbara Budrich, Opladen • Berlin • Toronto

Die Zeitschrift sowie alle darin enthaltenen Beiträge sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Namentlich gekennzeichnete Beiträge geben nicht in jedem Fall die Meinung der Redaktion wieder.

Aktuelle Mediadaten/Anzeigenpreisliste: ZEHF21 vom 01.02.2021

Druck: Elanders GmbH, Waiblingen, printed in Europe

Satz: Beate Glaubitz, Leverkusen

Umschlaggestaltung: Bettina Lehfeldt, Kleinmachnow – www.lehfeldtgraphic.de

ISSN 2367-3044

ISSN Online 2367-3052

Inhalt

Editorial	109
MINT-Projekte für Schülerinnen an Hochschulen. Analyse des Wirkungsmechanismus und Meta-Evaluation der empirischen Evidenz <i>Stefanie Brenning, Elke Wolf</i>	111
Test- und notenbasierte Erfolgsprognose im wirtschaftswissenschaftlichen Masterstudium: Befunde aus dem Nationalen Bildungspanel <i>Johannes Schult</i>	130
Akzeptanz und Nutzung psychologischer Verfahren zur Auswahl von Promovierenden <i>Max Knaut, Matthias Ziegler</i>	144
Wie gehen Hochschullehrende mit ihren Emotionen in der Lehre um? Eine Experience-sampling-Befragung zum situativen Emotionsausdruck von Freude, Angst und Ärger <i>Katharina Thies, Robert Kordts-Freudinger</i>	165

Liebe Leserinnen und Leser,

das neueste Heft der ZeHf erscheint corona-bedingt etwas später als geplant, liefert aber erneut ein Kaleidoskop an hoch spannenden Befunden.

Denn es werden, wie schon in früheren Heften, unter je eigenen Fragestellungen unterschiedliche Etappen einer erfolgreichen „hochschulischen Laufbahn“ in den Blick genommen, die ja schon bei der schulischen Interessensfindung beginnen kann, sich über das erfolgreiche Durchlaufen des Studiums und der Promotions- bzw. Post-Doc-Phase fortsetzt und schließlich in die Tätigkeit als Professorin oder Professor mündet.

Der Beitrag von Stefanie Brenning und Elke Wolf geht dabei der ungebrochen wichtigen Frage nach, ob MINT-Projekte für *Schülerinnen* zu einem gesteigerten Interesse an einem entsprechenden Studiengang führen und damit perspektivisch zu einer Steigerung der Frauenquote in diesen Studiengängen beitragen können. Zu diesem Zweck werden von den Programmveranstaltern selbst vorgenommene Evaluationsbemühungen unter dem Blickwinkel ihrer Aussagekraft analysiert.

Aus Sicht von Hochschulen nicht minder bedeutsam ist die anschließend behandelte Frage, ob Tests über HZB-Noten hinaus zur Vorhersage erfolgreicher Studienabschlüsse beitragen. Johannes Schult liefert hierzu neuere, auf Daten von *Studierenden* der Wirtschaftswissenschaften basierende Auswertungen.

In der Regel stehen Studierende nach erfolgreichem Masterabschluss vor der Entscheidung, ob sie eine Promotion oder unmittelbar eine Berufstätigkeit in der Praxis anstreben. Welche Auswahlverfahren Hochschulen im Wettbewerb um „exzellente“ *NachwuchswissenschaftlerInnen* vorsehen und wie diese von den BewerberInnen angenommen werden, ist Gegenstand des dritten Beitrags von Max Knaut und Matthias Ziegler.

Der vierte, von Katharina Thies und Robert Kordts-Freudinger vorgelegte Beitrag fokussiert schließlich das emotionale Erleben von *Hochschullehrenden*. Er zeigt auf, inwiefern im Rahmen von Lehrveranstaltungen konkrete Gefühle wie Freude, Angst oder Ärger empfunden und diese Emotionen gegenüber den Studierenden zum Ausdruck gebracht werden.

Neben der Diversität der adressierten Themen resultiert der Reiz der gebündelten Beiträge aus der Vielfalt der gewählten methodischen Zugänge. Diese reichen von einer Meta-Evaluation vorliegender Daten zum Nutzen von Schülerinnenprogrammen über die Nutzung von NEPS-Daten und die quantifizierende Auswertung real geführter Auswahlverfahren bis hin zu einer Experience-sampling-Erhebung, die eine Differenzierung zwischen Emotionsschwankungen innerhalb und zwischen Personen erlaubt.

Alle Beiträge haben im Rahmen des in der ZeHF verfolgten Peer-review-Verfahrens die jeweils hinzugezogenen GutachterInnen überzeugt, denen wir an dieser Stelle noch einmal ausdrücklich danken möchten. Im Ergebnis verdienen die hier zusammengestellten Artikel eine hohe Aufmerksamkeit und versprechen allen Leserinnen und Lesern eine anregende Lektüre.

Ihr Herausgabeteam

MINT-Projekte für Schülerinnen an Hochschulen. Analyse des Wirkungsmechanismus und Meta- Evaluation der empirischen Evidenz

Stefanie Brenning, Elke Wolf

Zusammenfassung: An deutschen Hochschulen für angewandte Wissenschaften existieren zahlreiche Angebote, um Schülerinnen für Studiengänge und Berufe im MINT-Bereich zu begeistern. Wissenschaftlich gesicherte Erkenntnisse über deren Wirksamkeit fehlen jedoch weitgehend. Ausgehend vom aktuellen Forschungsstand entwickeln wir eine mehrstufige Wirkungslogik dieser Maßnahmen und analysieren anschließend die Methodik der uns vorliegenden 13 Evaluationen mithilfe einer Meta-Evaluation. Die Befunde zeigen, dass diese Evaluationen von MINT-Projekten nur sehr eingeschränkt der Erfolgsmessung dienen, da sich die schriftlichen Fragebögen größtenteils auf kurzfristige Erfolgsindikatoren beschränken und die Untersuchungsdesigns nicht den Ansprüchen einer Kausalanalyse genügen.

Schlüsselwörter: Wirkungslogik, MINT-Projekte, Meta-Evaluation, Programmevaluation, Gleichstellung, Hochschule

STEM projects for female pupils at Universities of Applied Sciences. Analysis of impact mechanisms and meta-evaluation of empirical evidence.

Abstract: Universities of Applied Sciences in Germany offer numerous programs for female pupils in order to spark their interest in studying and working in STEM disciplines. Yet, scientifically verified impact assessments of these programs are still largely missing. Proceeding from the current state of research, we develop a multi-level impact logic of these programs and subsequently analyze the applied methods in 13 evaluations that are accessible to us using a meta-evaluation. Our results show that these evaluations of STEM projects are constrained in their ability to measure success because their questionnaires mostly focus on very short-term success indicators and their research designs don't fulfill the requirements of causal analyses.

Keywords: impact logic, STEM projects, meta-evaluation, program evaluation, gender equality, higher education

1 Einleitung

Der Anteil weiblicher Studierender an Hochschulen und Universitäten und weiblicher Fachkräfte im MINT¹-Bereich liegt nach wie vor weit hinter dem Anteil ihrer männlichen Mitstreiter. Während der Frauenanteil in MINT-Studiengängen zum Wintersemester 2018/19 insgesamt knapp ein Drittel beträgt (30.9%), liegt dieser in den Ingenieurwissenschaften und der Informatik nur etwas über einem Fünftel (23.6% bzw. 21.4%) (Statistisches Bundesamt, 2019). Der Frauenanteil unter den erwerbstätigen MINT-Akademikerinnen und -Akademikern betrug im Jahr 2015 21.5%, zugleich war der Anteil erwerbstätiger weiblicher MINT-Fachkräfte an allen erwerbstätigen MINT-Fachkräften mit 11% noch deutlich niedriger (Anger, Koppel & Plünnecke, 2018, S. 14f.). Dies steht in Kontrast zu den im April 2018 486,600 offenen Stellen im MINT-Bereich, die nicht besetzt werden konnten (Anger et al., 2018, S. 50).

Die Gründe der Unterrepräsentanz von Frauen im MINT-Bereich sind vielfältig. Neben dem Hindernis, als eine von wenigen Frauen in einem männerdominierten Umfeld zu studieren und/oder zu arbeiten, zeigt sich auch, dass geschlechterstereotype Berufsbilder, falsche Vorstellungen von MINT-Berufen und eine geringere Selbsteinschätzung junge Frauen von der Wahl eines MINT-Berufes abhalten (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung [BLK], 2002; Ertl, Luttenberger & Paechter, 2014; Pöllmann-Heller & Rudolph, 2019; Wentzel, 2008; Weinhardt, 2017).

Die Bundesregierung initiiert und unterstützt seit einigen Jahren zahlreiche Programme, um Frauen für MINT-Berufe, insbesondere im Bereich der Informationstechnologie, zu gewinnen (z.B. Komm, mach MINT; Girls' Day; MINT-E-Plattform; Initiative Klischeefrei).

An Hochschulen und Universitäten liegt der Schwerpunkt der Aktivitäten zur Gewinnung von Frauen für ingenieur- und naturwissenschaftliche Studiengänge im Bereich der Kontext- und Begleitmaßnahmen für Schülerinnen (Löther & Girlich, 2011). Hierzu zählen in erster Linie Informationsveranstaltungen und Workshops zur Forschung und Anwendung im MINT-Bereich. Ziel dieser Angebote ist es in der Regel, das Interesse an ingenieur- und naturwissenschaftlichen Fragestellungen zu erhöhen, die Studien- und Berufsfindung mit Informationsangeboten zu unterstützen und die Einschätzung der eigenen Fähigkeiten zu verbessern. Angesichts der Fülle der Maßnahmen und den nicht unerheblichen Ressourcen, die hierfür investiert werden, stellt sich dringend die Frage nach den Wirkungen dieser Maßnahmen.

Aber wann ist ein MINT-Projekt erfolgreich? Was kann bewirkt werden und wie kann dies gemessen werden? Auch wenn die Maßnahmen darauf abzielen, den Frauenanteil in den MINT-Berufen zu erhöhen, wird die alleinige Teilnahme an einem MINT-Projekt die Studien- und Berufswahlentscheidung nicht monokausal beeinflussen. Die Studien- wie auch die Berufswahl ist komplex und von vielen Faktoren abhängig, auf die spezielle Angebote für Mädchen keinen Einfluss nehmen können. Dies gilt beispielsweise für die frühe geschlechtsspezifische Sozialisation, die vorwiegend innerhalb der Familie und durch die Eltern stattfindet. Gleichwohl ist Sozialisation ein Lernprozess, der nicht im Kindesalter abgeschlossen ist, sondern die kontinuierliche Auseinandersetzung mit der sozialen Umwelt

1 MINT bezeichnet die Fächergruppe der Mathematik, Informatik, Naturwissenschaften und Technik.

umfasst (Kray, 2019; Siegler, Eisenberg, De Loache & Saffran, 2016). In späteren Lebensphasen können dann unterschiedliche Personen und Institutionen, wie die Hochschule und damit verbundene Fördermaßnahmen, zur Sozialisation beitragen (Hurrelmann, Bauer, Grundmann & Walper, 2015). Demnach kann erwartet werden, dass die Teilnahme an einem MINT-Projekt einer Hochschule für Angewandte Wissenschaften für Schülerinnen die veränderbaren Determinanten der Studien- und Berufswahl positiv beeinflusst und relevante Hürden bei der Entscheidung für den MINT-Bereich reduziert. Hierzu zählen u.a. die Selbsteinschätzung ihrer Fähigkeiten sowie das Wissen über verschiedene MINT-Berufe. Die Wirksamkeit einer Maßnahme lässt sich demnach anhand jener empirisch relevanten Determinanten der Studien- und Berufswahl messen, welche durch die Teilnahme an einer Maßnahme verändert werden können.

Löther und Girlich (2011) zeigen allerdings in ihrer Meta-Analyse u.a., dass in Evaluationen von MINT-Schülerinnen-Projekten die Zufriedenheit der Teilnehmerinnen meistens als das zentrale Erfolgskriterium verwendet wird. Nur wenige Evaluationsstudien zu Fördermaßnahmen im MINT-Bereich für Mädchen messen den Erfolg anhand von Einstellungsänderungen (vgl. u.a. Stöger, Duan, Schirner, Greindl & Ziegler, 2013; Mokhonko, Nickolaus & Windaus, 2014). Darüber hinaus beschränken sich die meisten Studien auf eine einmalige Befragung am Ende der Projekte. Belastbare empirische Befunde zur Effektivität der MINT-Schülerinnen-Projekte liefern Stöger et al. (2013), Hiller (2014), Kenkmann (2005) und Zeisberg (2013), die in Prä-Post-Befragungen positive Effekte auf relevante Studien- und Berufswahl-Determinanten nachweisen. So zeigen Stöger et al. (2013) mit ihrer Evaluation eines Online-Mentoring-Programms für Schülerinnen von 11 bis 18 Jahren, dass die Teilnehmerinnen bei sechs von sieben Determinanten Treatment-Effekte aufweisen. Die Schülerinnen verfügten nach Abschluss des Mentoring-Programms im Vergleich zu einer Kontrollgruppe u.a. über ein größeres Wissen über Studiengänge und Berufe in MINT. Aber auch bei der Intention, nach Abschluss der Schule ein MINT-Studium aufzunehmen, einem der längerfristig orientierten Erfolgsindikatoren, zeigten sich positive Effekte. Mokhonko (2016) untersucht die Effekte von Schülerlaboren an Gymnasien für die siebte bis zehnte Klasse mit Durchführungsdauern zwischen einer bis 18 Stunden auf das fachspezifische Fähigkeitsselbstkonzept, die fachspezifischen und beruflichen Interessen sowie die naturwissenschaftlich-technischen Bildungs- und Berufsvorhaben. Insgesamt zeigten sich nur punktuell positive und insgesamt geringe Effekte, z.B. stieg bei Schülerinnen der siebten Klasse das Interesse am intellektuell-forschenden Bereich nach Teilnahme am Schülerlabor signifikant an. Es fällt weiterhin auf, dass sich auch für Maßnahmen längerer Dauer wider Erwarten keine positiven Fördereffekte feststellen lassen konnten. Löther und Girlich (2011, S. 96) kommen zu der Schlussfolgerung, dass die existierenden Evaluationen oftmals der „Entwicklung und Verbesserung des Programms sowie Legitimation innerhalb der Hochschule und gegenüber den Geldgebern“ dienen und nur sehr begrenzt zur objektiven Überprüfung der Wirksamkeit der Projekte und des tatsächlichen Effekts auf die nachfolgende Studien- und Berufswahl der Teilnehmerinnen im MINT-Bereich beitragen.

Insbesondere für die Projekte an Hochschulen für angewandte Wissenschaften (HAW), die aufgrund der hohen Durchlässigkeit beim Studiengang, den Kooperationen mit Unternehmen und dem hohen MINT-Anteil besonders gute Bedingungen zur Gewinnung von MINT-Fachkräften bieten, steht eine Analyse der Evaluationen von Schülerinnen-Projekten

zur Gewinnung von Frauen im MINT-Bereich noch aus. Wir konzentrieren uns daher auf MINT-Fördermaßnahmen für Schülerinnen, die an deutschen HAW durchgeführt werden.²

Ziel dieser Studie ist es, existierende Evaluationen von MINT-Schülerinnen-Projekten anhand ausgewählter Indikatoren zu analysieren. Wir gehen in mehreren Schritten vor. Zu Beginn beschreiben wir den Wirkungsmechanismus der Maßnahmen anhand eines Logic Charts und identifizieren verschiedene Indikatoren auf den drei Wirkungsebenen Output, Outcome und Impact. Diese theoretische und empirische Herleitung der intendierten Effekte kann zukünftig auch zur Konzeption von Projekten und deren Evaluationen genutzt werden. Uns dient die Wirkungslogik im Folgenden der Klassifizierung der in den Evaluationen gemessenen Erfolgsindikatoren. Der empirische Schwerpunkt unserer Untersuchung ist in Abschnitt 3 dargestellt: Wir analysieren die Methodik der vorliegenden Evaluationen anhand der existierenden Standards der Evaluationsforschung, die sich auf Basis der vorliegenden Evaluationsberichte untersuchen lassen. Hierzu zählen die Auswahl der Erfolgsindikatoren sowie die Validität der Wirkungsanalyse. Somit können die Zuverlässigkeit der Erkenntnisse bewertet und Verbesserungspotenziale für die Planung und Durchführung zukünftiger Evaluierungen aufgezeigt werden (Scriven, 1969; Widmer, 1996; Patton, 1997).

Von einer inhaltlichen (Meta-)Analyse wird aufgrund der eingeschränkten Vergleichbarkeit der Erfolgsindikatoren und den mediokren Befunden der Meta-Evaluation abgesehen. Die eingeschränkte Aussagekraft und geringe Anzahl an Evaluationen verdeutlichen jedoch unweigerlich die Dringlichkeit, die Wirkungsmessung von MINT-Projekten zu thematisieren und zu professionalisieren.

2 Die Wirkungslogik von MINT-Projekten für Schülerinnen

Bei der Herleitung einer Wirkungslogik konzentrieren wir uns auf Maßnahmen für Schülerinnen, die deren MINT-Interesse in erster Linie durch eigene Aktivitäten wecken wollen und ihnen Einblicke in das Studium an HAW und angewandte Forschungsfragen gewähren, da der potentielle Wirkungsmechanismus dieser Projekte sehr ähnlich ist.

Auf Basis einer Umfrage an allen HAW zeigen Brenning und Wolf (2019), dass die Angebote zur Studien- und Berufsorientierung für Schülerinnen sehr vielfältig und unterschiedlich sind. Dennoch verfolgen alle MINT-Projekte für Schülerinnen die Ziele, die Teilnehmerinnen mithilfe von ein- oder mehrtägigen Veranstaltungen über naturwissenschaftlich-technische Fragestellungen, Themen und Studiengänge zu informieren, sie an MINT-Fragestellungen heranzuführen und letztendlich für eine Studien- und Berufswahl im MINT-Bereich zu begeistern. Die Schülerinnen nehmen dafür an Workshops, Vorlesungen und Informationsveranstaltungen teil oder führen Gespräche mit Studierenden, Lehrenden,

2 Das dieser Analyse zugrundeliegende Verbundvorhaben „MINT-Strategien 4.0 – Strategien zur Gewinnung von Frauen für MINT-Studiengänge an Hochschulen für angewandte Wissenschaften“ wird mit Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 01FP1714 und 01FP1715 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen. Wir danken Andrea Löther (GESIS – Leibniz-Institut für Sozialwissenschaften, Köln) für zahlreiche und konstruktive Diskussionen sowie zwei anonymen Gutachter/innen für wertvolle Kommentare zu unserem Manuskript.

Hochschulmitarbeiterinnen und Hochschulmitarbeitern oder Frauen, die im MINT-Bereich arbeiten. Zumeist besichtigen sie Hochschulräumlichkeiten wie Labore und Hörsäle und lernen Studiengänge aus dem MINT-Bereich kennen. Teils bearbeiten die Schülerinnen konkrete Fragestellungen, experimentieren mit naturwissenschaftlich-technischen Phänomenen und machen sich mit der angewandten Forschung an HAW vertraut. Viele Fördermaßnahmen an HAW kooperieren auch mit Unternehmen aus ihrer Region, indem die Schülerinnen Unternehmen besuchen und Einblicke in MINT-Berufsbilder und den Arbeitsalltag, z.B. einer Ingenieurin, erhalten können (Brenning & Wolf, 2019).

Die Wirkungslogik der Schülerinnen-Projekte beschreiben wir mithilfe eines Logic Charts, dessen Wirkungsebenen im darauffolgenden Kapitel zur Untersuchung der Evaluationsberichte dienen. Das logische Modell wurde von der W. K. Kellogg Foundation als Evaluierungswerkzeug entwickelt, um Maßnahmen effektiv zu planen, zu implementieren und zu evaluieren. Ein Logic Chart visualisiert die Beziehungen zwischen Aktivitäten, Programminhalten sowie kurz-, mittel- und langfristigen Wirkungen (W. K. Kellogg Foundation, 2004, S. III). Unsere Wirkungsanalyse der Fördermaßnahmen für Schülerinnen baut auf dem Logic Chart von Löther und Girlich (2011) auf, welches das allgemeine logische Modell auf die Funktions- und Wirkungsweisen von Maßnahmen zur Studien- und Berufsorientierung für Schülerinnen anwendet.³

Die Wirkungen einer Maßnahme lassen sich typischerweise in drei Ebenen einteilen: Output, Outcome und Impact (Döring & Bortz, 2016). Für eine umfassende Bewertung einer Maßnahme sollten Evaluationen die Wirkungen auf allen drei Ebenen des Logic Charts (s. Abbildung 1) erfassen.

Der Output, die unmittelbaren Ergebnisse einer Maßnahme, umfasst im Allgemeinen Indikatoren der tatsächlich erbrachten Leistungen und Angebote einer Intervention, der Inanspruchnahme der Leistungen sowie die Zufriedenheit mit den Angeboten. In unserem Logic Chart stellt der Output (1.) die Akzeptanz der Maßnahme bei den Teilnehmerinnen und somit eine Grundvoraussetzung für den Erfolg einer Maßnahme dar. Dies zeigt sich daran, ob die angebotenen Plätze ausgebucht sind, wie auch in der Bewertung der Veranstaltungen, der Inhalte und der gesamten Fördermaßnahme (1.1), in der Bewertung der beteiligten Personen, z.B. der Organisatorinnen und Organisatoren sowie der Referentinnen und Referenten (1.2), und anhand einer potenziellen Weiterempfehlung des Schülerinnen-Projekts wie auch der Frage, ob die Teilnehmerinnen selbst wieder an einem MINT-Projekt für Schülerinnen teilnehmen würden (1.3).

Die zweite Stufe des logischen Modells, der Outcome (2.), beschreibt alle kurzfristigen Wirkungen, die vom Output bei der Zielgruppe verursacht werden (Döring & Bortz, 2016). Wir berücksichtigen hier jene empirisch relevanten Determinanten der Studien- und Berufswahl, welche durch die Teilnahme an einer Maßnahme verändert werden können. Hierzu zählen die potenzielle Steigerung des Interesses der Schülerinnen am MINT-Bereich (2.1), die Unterstützung bei der Studien- und Berufsfindung (2.2), die Erhöhung der Selbstwirksamkeit und Selbsteinschätzung (2.3) sowie die Reduktion von Geschlechterste-

3 Wir begrenzen das Logic Chart jedoch auf die geplanten bzw. gewünschten Ergebnisse und Wirkungen der Maßnahmen (Output, Outcome und Impact), da der Fokus unserer Meta-Evaluation auf der Bewertung der Erfolgsindikatoren der Maßnahmen liegt und eine Kenntnis des Inputs der Programme keine notwendige Voraussetzung zum Verständnis der Meta-Evaluation der Evaluationen darstellt.

reotypen (2.4). Anhand einer Befragung unter den Koordinator*innen von MINT-Schülerinnen-Projekten an HAW (Brenning & Wolf, 2019) zeigt sich, dass sich diese Outcome-Indikatoren vollständig mit den Zielstellungen von MINT-Schülerinnen-Projekten decken und somit sehr gut geeignet sind, die Erfolge dieser Projekte zu beschreiben.

Grundsätzlich ist das Interesse an Fragestellungen und Phänomenen der MINT-Disziplinen (2.1) ein wichtiger Einflussfaktor für die spätere Studien- und Berufswahl (siehe u.a. Lent, Brown & Hackett, 1994; Taskinen, Schütte & Prenzel, 2013; Obergruber, 2018). Mädchen machen jedoch bereits in der Grundschule weniger Erfahrungen mit Technik als Jungen, da sie in ihrer frühen Kindheit im familiären und weiteren sozialen Umfeld seltener an Aktivitäten mit Technik- oder Forschungsbezug herangeführt werden. Sie entwickeln in der Folge weniger Interesse am technischen Bereich (Solga & Pfahl, 2009). Taskinen, Schütte und Prenzel (2013) machen deutlich, dass u.a. das individuelle Interesse an Naturwissenschaften Einfluss auf das Interesse von Schülerinnen an naturwissenschaftlichen Berufen hat.

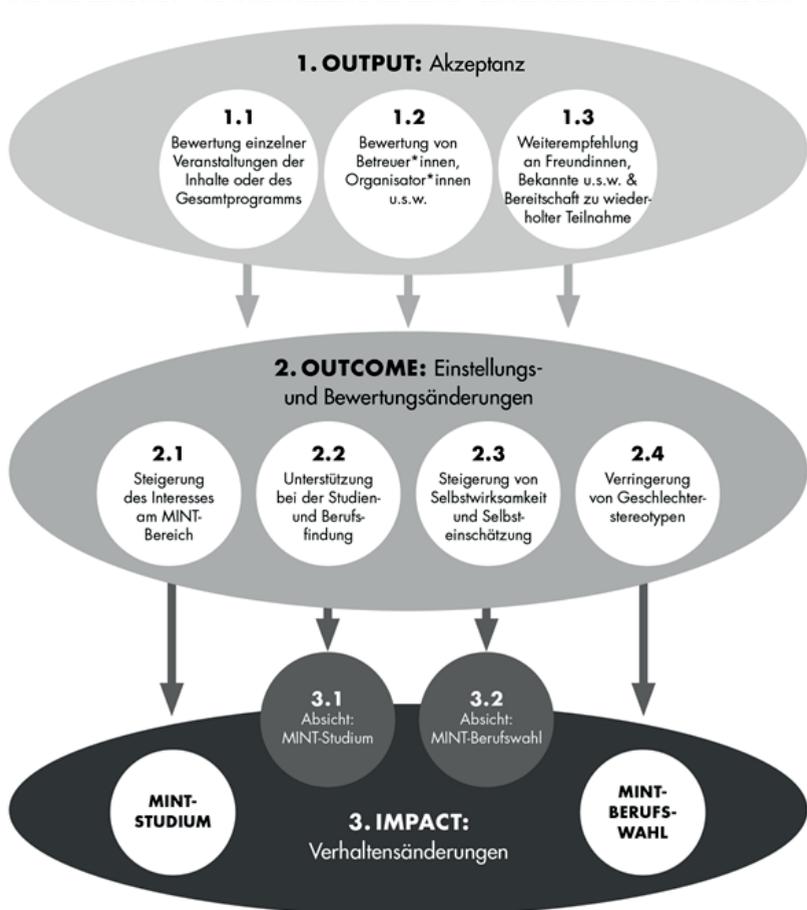


Abbildung 1: Logic Chart – Darstellung der Wirkungsweisen von Maßnahmen zur Studien- und Berufsorientierung für Schülerinnen

Wissenslücken über mögliche Berufsbilder im MINT-Bereich stellen ebenfalls eine relevante Hürde bei der Entscheidung für einen MINT-Beruf dar (siehe hierzu die umfangreiche Darstellung in Matthes, 2019). Aufgrund einer geschlechtsspezifischen Sozialisation kann davon ausgegangen werden, dass Mädchen im Durchschnitt seltener Kontakt zu naturwissenschaftlich-technischen Themenbereichen haben (Taskinen, 2010; Jones, Howe & Rua, 2000; Prenzel, Lankes & Minsel, 2000). Informationen über Studiengänge, Berufe und Arbeitgeber im MINT-Bereich unterstützen junge Frauen bei der Studien- und Berufsfindung (2.2). Diese Informationen können jedoch auch nicht-intendierte Effekte für die Studien- und Berufswahl im MINT-Bereich haben, da sich einige Schülerinnen aufgrund der neu gewonnenen Informationen auch explizit gegen einen MINT-Beruf entscheiden könnten.

Eine wichtige Rolle spielen ebenso die Selbsteinschätzung und die Selbstwirksamkeit der Schülerinnen (2.3): Die Ergebnisse von Weinhardt (2017) zeigen, dass sich Schülerinnen im Vergleich zu Schülern im Fach Mathematik selbst bei gleichen Noten schlechter einschätzen. Darüber hinaus zeigt Correll (2001, 2004), dass Mädchen ihre Erfolge in Mathematik und Naturwissenschaften eher auf Glück und Anstrengung zurückführen, während Jungen den Erfolg als Konsequenz ihrer Kompetenz betrachten. Mädchen entscheiden sich demnach tendenziell eher für ein naturwissenschaftliches Fach, wenn sie ihre mathematischen Fähigkeiten positiv einschätzen.

Auch die Selbstwirksamkeit, nach Bandura (1997) die Überzeugung, bestimmte Handlungen erfolgreich ausführen zu können, hat Effekte für die Studien- und Berufswahlentscheidungen von Schülerinnen. Wenn diese grundlegende Überzeugung, z.B. den Anforderungen eines (MINT-)Studiums gewachsen zu sein, bei Schülerinnen in geringerem Maße vorhanden ist, da sie weniger auf ihre Kompetenzen vertrauen (Correll, 2001, 2004), könnte dies nach Banduras Theorie die Wahrscheinlichkeit einer Studien- und Berufswahl im MINT-Bereich senken. Empirische Evidenz dieses Zusammenhangs findet sich u.a. bei Bandura, Barbaranelli, Caprara und Pastorelli (2001) sowie Sheu et al. (2010).

Des Weiteren haben Geschlechterstereotypen (2.4) einen Effekt auf die spätere Studien- und Berufswahl: Zum einen können negative Stereotype hinsichtlich der mathematischen und analytischen Fähigkeiten von Frauen die entsprechenden Leistungen von Schülerinnen mindern, wie es in der Literatur zur Stereotypenbedrohung beschrieben wird (s. u.a. Spencer, Steele & Quinn, 1999). Als Folge davon kann wiederum das Interesse von Mädchen und Frauen am MINT-Bereich sinken (Shapiro & Williams, 2012) und zu einer schlechteren Einschätzung der eigenen Fähigkeiten führen. Zum anderen werden viele MINT-Disziplinen mit Männlichkeit assoziiert und Kompetenzen im MINT-Bereich bei Frauen als unattraktiv und unweiblich wahrgenommen (Ertl et al., 2014; Kessels & Hannover, 2006). In Anlehnung an die Theorie der kognitiven Dissonanz von Festinger (1957) strebt jedes Individuum nach Konsistenz der eigenen Meinungen und Einstellungen und versucht Dissonanzen zwischen nicht miteinander vereinbaren Kognitionen, z.B. eigenen und konträren Ansichten, zu reduzieren. Wenn zwischen dem eigenen geschlechterspezifischen Selbstkonzept und den Stereotypen über Mädchen und Frauen im MINT-Bereich Diskrepanzen vorliegen, distanzieren sich Frauen tendenziell von diesem Berufsbild und wählen daher eher „mädchentypische“ Studien- und Berufsfelder, die nicht im MINT-Bereich liegen (Kessels, 2015), um somit die als unangenehm erlebte kognitive Dissonanz zu reduzieren.

Veränderungen dieser Studien- und Berufswahldeterminanten können unmittelbar nach der Teilnahme an einem Programm gemessen werden. Kurzfristig ist ein Programm also erfolgreich, wenn die beschriebenen Indikatoren auf der Outcome-Ebene durch die Teilnahme an dem Programm positiv beeinflusst werden. Da die Hürden für die Aufnahme eines MINT-Studiums individuell sehr unterschiedlich sind, ist davon auszugehen, dass ein Projekt umso erfolgreicher ist, je mehr positive Teileffekte auf die unterschiedlichen Indikatoren erzielt werden, da somit sehr unterschiedliche Barrieren überwunden werden können. Auch wenn sich aus der Messung der Einstellungen keine direkte Kausalität zur späteren Studien- und Berufswahl herleiten lässt, kann eine in gewünschter Richtung erfolgte Veränderung einer oder mehrerer relevanter Determinanten durchaus die Wahrscheinlichkeit erhöhen, ein MINT-Studium aufzunehmen oder einen MINT-Beruf zu ergreifen.

Auf der dritten Wirkungsebene unseres Logic Charts (Abbildung 1) misst der Impact die langfristigen Wirkungen (auf gesellschaftlicher Ebene) (Döring & Bortz, 2016), die durch verändertes Verhalten oder eine Veränderung der privaten oder beruflichen Situation der Teilnehmerinnen erzeugt werden. Der langfristige Erfolg einer Maßnahme zeigt sich anhand der tatsächlichen Verhaltensänderungen, die ein MINT-Projekt zur Folge haben kann.

Ob die erwünschte Wirkung einer Fördermaßnahme für Schülerinnen, den Frauenanteil in naturwissenschaftlich-technischen Studiengängen zu steigern, erreicht wird, lässt sich idealerweise mithilfe einer Untersuchung der späteren Studien- und Berufswahl der Teilnehmerinnen im MINT-Bereich ermitteln. Die methodische Umsetzung ist jedoch in der Praxis aufgrund fehlender Ressourcen und der Schwierigkeit, die Schülerinnen nach mehreren Jahren zu befragen, kaum realisierbar. In Anbetracht der vielfältigen Einflussfaktoren der Studien- und Berufsentscheidungen wäre damit jedoch noch nicht geklärt, wie relevant die Teilnahme an einem bestimmten MINT-Projekt tatsächlich für diese Entscheidungen war.

Alternativ kann der Impact der MINT-Schülerinnen-Projekte anhand der Intention, ein MINT-Studium (3.1) oder einen MINT-Beruf (3.2) aufzunehmen, gemessen werden. Die Theorie des geplanten Verhaltens von Ajzen (1991) beschreibt ein Modell, welches das Verhältnis zwischen einer Verhaltensintention und dem tatsächlichen Verhalten unter spezifischen Bedingungen erklärt. Ajzen (1991) argumentiert, dass die Verhaltensintention wiederum von Einstellungen bestimmt wird, die durch Erfahrungen und äußere Faktoren beeinflusst werden. Das bedeutet, Verhaltensintentionen sind indirekt durch exogene Ereignisse veränderbar und stehen in direktem Zusammenhang zum tatsächlichen Verhalten. Kim und Hunter (1993) zeigen in ihrer Meta-Evaluation, dass Einstellungen 50% der Varianz bei den bestimmten Verhaltensintentionen erklären und diese wiederum 30% der Varianz im tatsächlichen Verhalten erklären. Auch Armitage und Conner (2001) kommen im Rahmen ihrer Meta-Evaluation von 185 verschiedenen Studien zu dem Ergebnis, dass Intentionen und selbstberichtete Vorhersagen gute Prädiktoren von Verhalten darstellen. Auch wenn die Theorie des geplanten Verhaltens darauf abzielt, den kurzfristigen Zusammenhang zwischen Intention und Handlung zu erklären, zeigen Randall und Wolff (1994), dass der starke Korrelationskoeffizient auch bis zu einer zeitlichen Distanz von 15 Jahren erhalten bleibt. Stöger et al. (2013) nutzen in ihrer Evaluation eines einjährigen Online-Mentoring-Programms für elf- bis 18-jährige Schülerinnen die Studienwahlintentionen einer Versuchs-

und Kontrollgruppe als einen der Erfolgsindikatoren für die Wirksamkeit des Programms. Auch Lent et al. (1994) argumentieren, dass sich die Theorie des geplanten Verhaltens auf die Studien- und Berufswahlentscheidung anwenden lässt, d.h. auch diese Entscheidung wird durch die vorherige Verhaltensintention beeinflusst. Eine in gewünschter Richtung erfolgte Veränderung der Studien- und Berufswahlintentionen erhöht somit die Wahrscheinlichkeit einer bestimmten Studien- und Berufswahl und kann folglich als Erfolg der MINT-Fördermaßnahme interpretiert werden.

Zur Messung der kurzfristigen Effekte von MINT-Projekten können somit die beeinflussbaren Determinanten der Studien- und Berufswahl (Outcome-Indikatoren 2.1 bis 2.4) sowie die Intention einer bestimmten Bildungsentscheidung herangezogen werden. Langfristig zeigt sich der Erfolg anhand eines veränderten Studien- und Berufswahlverhaltens, wobei nicht von einem monokausalen Zusammenhang zu einem einzelnen MINT-Projekt ausgegangen werden kann. Der Blick auf die verschiedenen Teileffekte begegnet darüber hinaus der Kritik von Müller und Albrecht (2016), dass Evaluationen, die sich nur auf den Impact konzentrieren, keine Erkenntnisse über die Wirkungsweise der untersuchten Maßnahmen generieren.

3 Meta-Evaluation der Evaluationen von MINT-Projekten

Im Folgenden analysieren wir die existierenden Evaluationsstudien von MINT-Projekten für Schülerinnen an HAW hinsichtlich ihrer Methodik. In unsere Meta-Evaluation beziehen wir jegliche Evaluationen von MINT-Projekten ein, die explizit für Mädchen konzipiert sind, einen eindeutigen Bezug zu MINT-Studienfächern und -Berufen aufweisen, mindestens einen halben oder mehrere Tage bis zu einem halben Jahr dauern und in den letzten 15 Jahren an einer HAW durchgeführt wurden. Ausgeschlossen wurden Evaluationsberichte von Projekten, die sowohl männliche als auch weibliche Studierende betrachten und deren Analysen und Ergebnisse nicht nach Geschlechtern differenzieren. Des Weiteren haben wir Angebote exkludiert, die nur aus einer kurzen Informationsveranstaltung bestehen.

Insgesamt liegen uns 13 Evaluationsberichte von zehn Einzelmaßnahmen für Schülerinnen vor.⁴ Sieben dieser Maßnahmen werden von den einzelnen Hochschulen selbst durchgeführt. Drei Projekte sind Verbundprojekte, bei denen sich die Hochschulen an Initiativen von Ministerien und Verbänden beteiligen oder mit anderen Hochschulen kooperieren. Alle Angebote, von denen Evaluationsberichte vorliegen, führen direkt nach Abschluss des Projekts eine schriftliche Befragung der teilnehmenden Schülerinnen durch. Elf der 13 Evaluationen wurden hochschulintern durchgeführt.

Die evaluierten Projekte richten sich an Schülerinnen der siebten Klasse bis zum letzten Schuljahr oder Mädchen im Übergang zum Studium und Beruf. Neun der zehn Einzelmaß-

4 Um auch nicht publizierte Evaluationen zu berücksichtigen, haben wir eine umfassende Recherche zu MINT-Projekten für Schülerinnen an HAW durchgeführt. Anschließend wurden die Projektkoordinatorinnen und Projektkoordinatoren von 98 Schülerinnen-Projekten per E-Mail angeschrieben und Projektwebseiten durchsucht, um Berichte bisher durchgeführter Evaluationen zu erhalten (Brenning & Wolf, 2019). Alle Evaluationsberichte wurden uns entweder per E-Mail von Projektkoordinator*innen zugesendet oder standen auf der Webseite des jeweiligen MINT-Angebots zum Download zur Verfügung.

nahmen kooperieren auch mit Unternehmen, um Schülerinnen Einblicke in bestimmte Berufsfelder zu geben. Nicht alle Projekte äußern sich explizit zu den angestrebten Zielen. Eine Bewertung anhand der von den Projekten selbst gesetzten Ziele wäre demnach nicht umfassend genug. Wenn jedoch konkrete Ziele benannt werden, so sind diese immer in den Outcome-Indikatoren unseres Logic Charts inkludiert. Trotz der großen Heterogenität der Projekte kann dieses Wirkungsschema somit grundsätzlich für die Analyse der Evaluationsberichte herangezogen werden.

Wir bewerten die Evaluationen anhand der Güte und Vielfalt der Erfolgsindikatoren sowie der Validität der Wirkungsanalyse.⁵ Eine klassische Meta-Evaluation lässt sich aufgrund der Datenlage und -qualität leider nicht durchführen: die geringe Anzahl an Evaluationen, insbesondere belastbarer Evaluationen, sowie die hohe Heterogenität innerhalb der Fragestellungen erschweren die Vergleichbarkeit auf quantitativer Ebene.

3.1 Güte und Vielfalt der Erfolgsindikatoren

Zur Bewertung der verwendeten Erfolgsindikatoren werden die Fragen der Evaluationsstudien den drei Stufen des Logic Chart zugeordnet. Idealerweise enthalten Evaluationen Informationen zu allen Wirkungsebenen. Als unmittelbares Ergebnis einer Fördermaßnahme für Schülerinnen erfasst der Output die Akzeptanz des Projekts bei den Teilnehmerinnen. Diese lässt sich mithilfe der Bewertung der unterschiedlichen Dimensionen eines Programms durch die teilnehmenden Schülerinnen messen (Löther & Girlich, 2011). Wir ordnen alle Informationen zur Inanspruchnahme und Fragen zur Empfehlung des Programms sowie zur Bewertung der Organisation und Inhalte der Angebote dieser Wirkungsebene zu (siehe 1.1 bis 1.3 in Abbildung 1). Fragen nach der Zufriedenheit der Teilnehmerinnen, welche ein wichtiger Indikator für die Akzeptanz einer Fördermaßnahme ist, werden ebenfalls dem Output zugeordnet.

Eine Schlüsselrolle für die Aussagekraft einer Evaluation spielen die Fragen zum Outcome. Nur wenn die Teilnehmerinnen ein gesteigertes Interesse an MINT-Berufen entwickeln, nützliche Unterstützung bei der Studien- und Berufswahl erhalten, ihre Kompetenzen oder zumindest die Bewertung ihrer Kompetenzen aufbessern oder Geschlechterstereotype hinsichtlich Frauen in MINT-Berufen abbauen, kann ein MINT-Projekt die spätere Studien- und Berufswahl in gewünschter Weise beeinflussen (s. 2.1 bis 2.4 in Abbildung 1). Alle Fragen, die sich auf diese Veränderungen von Einstellungen und Bewertungen der Teilnehmerinnen beziehen, ordnen wir demnach der Outcome-Ebene zu.

Häufig enthalten die Evaluationen auch Fragen zum persönlichen Nutzen des Projekts für die Teilnehmerinnen. Sofern sich diese Fragen nicht konkret auf einzelne Outcome-Indikatoren beziehen, bleibt unklar, ob der persönliche Nutzen der Schülerinnen dazu beitragen kann, dass sie mit höherer Wahrscheinlichkeit ein Studienfach oder einen Beruf im MINT-Bereich wählen werden. Der Nutzen der Teilnahme könnte letztlich darin bestehen, dass die Schülerinnen nach Abschluss des Projekts wissen, dass sie *keinen* MINT-Beruf ergreifen wollen oder dass sie einen persönlichen Nutzen in dem neu erworbenen Wissen se-

5 Weitere Qualitätskriterien von Evaluationsstudien, wie beispielsweise Informationen zur Kompetenz der Evaluatoren, der Interaktion zwischen den Partnern des Evaluationsprozesses oder der Implementierung des Prozesses (DeGEval, 2016) können hier nicht analysiert werden, da ein Großteil der Evaluationsberichte hierzu keine Angaben machen.

hen (um beispielsweise ihre schulischen Leistungen zu verbessern). Somit werten wir allgemeine Fragen nach dem Nutzen nicht als Indikatoren für den Outcome der Maßnahme.

Analyse der Evaluationen: Da keine der vorliegenden Evaluationen die tatsächliche spätere Studien- oder Berufswahl der Schülerinnen beobachtet, kann der Impact nur anhand der Intention approximiert werden. Einige Evaluationen erfragen beispielsweise die aktuelle Neigung, einen Studiengang oder Beruf im MINT-Bereich aufzunehmen, was als Indikator für die spätere Entscheidung interpretiert werden kann, da sie die Intention zur Studien- und Berufswahl im MINT-Bereich erfassen und damit tatsächliche Verhaltensänderungen zumindest annähernd beschreiben (s. Abschnitt 2). Demnach werten wir Fragen nach den Studien- und Berufszielen als Proxy für den Impact einer Maßnahme.

Tabelle 1: Klassifizierung der Fragen

	Anzahl der Fragen pro Evaluation (Durchschnitt)	Anzahl (Anteil) der Evaluationsberichte ohne Fragen zur jeweiligen Wirkungsebene
Output	2.42	2 (15%)
Outcome	3.07	2 (15%)
Impact	1.40	5 (38%)
Insgesamt	6.89	–

Quelle: Eigene Erhebung im Rahmen des Forschungsvorhabens „MINT-Strategien 4.0“. Grundgesamtheit: 13 Evaluationsberichte.

Tabelle 1 zeigt die Zuordnung der genutzten Fragen zu den drei Wirkungsebenen. Die Evaluationen der Schülerinnen-Projekte nutzten im Mittel ungefähr sieben Fragen zur Messung des Erfolgs. Im Durchschnitt bezogen sich 2.42 Fragen pro Evaluation auf die Bewertung und Akzeptanz der konkreten Maßnahmen, und 3.07 Fragen zielten darauf ab, die verschiedenen Indikatoren auf der Outcome-Ebene (also der Determinanten der Studien- und Berufswahl) zu messen. Immerhin 1.40 Fragen pro Fragebogen waren dazu geeignet, den langfristigen Impact zu messen. Angesichts der teilweise sehr kurzen Dauer der Schülerinnen-Projekte ist es nicht verwunderlich, dass der Schwerpunkt der Fragen nicht im Bereich der Impact-Indikatoren liegt. Gleichwohl enthält eine umfassende Evaluation auch Fragen zu den tatsächlichen Verhaltensänderungen bzw. der Verhaltensintention, was bei knapp 40% der hier analysierten Evaluationen nicht der Fall ist. Das heißt, der Großteil der vorliegenden Evaluationen von Fördermaßnahmen für Schülerinnen konzentriert sich auf eine Abfrage der Einstellungsänderungen bei den Teilnehmerinnen (z.B. Steigerung des Interesses am MINT-Bereich). Intentionen zur Verhaltensänderung, also ob die Teilnehmerinnen aufgrund der Teilnahme an der Fördermaßnahme planen, eine Studien- und Berufswahl zugunsten des MINT-Bereichs zu treffen, werden jedoch am wenigsten untersucht: Fünf von 13 Berichten bzw. 38% der Evaluationsberichte erheben keine Informationen, die den Impact des jeweiligen MINT-Projekts messen, wohingegen nur 15% der Berichte keine Fragen zum Outcome stellen. Evaluationen, die keine Fragen zum Output stellen (15%), verwehren sich dadurch die Möglichkeit der inhaltlichen Nachbesserung und Anpassung ihres MINT-Projekts auf die Zielgruppe der Schülerinnen.

3.2 Validität der Wirkungsanalyse

Ein MINT-Projekt für Schülerinnen gilt dann als erfolgreich, wenn aufgrund der Teilnahme an der Maßnahme die Wahrscheinlichkeit steigt, dass die Schülerinnen einen Beruf oder ein Studium im MINT-Bereich aufnehmen und danach in diesem Beruf arbeiten. Wie in Abschnitt 2 dargestellt, kann die erhöhte Wahrscheinlichkeit anhand von veränderten Determinanten der Studien- und Berufswahl und einer stärkeren Intention, ein MINT-Studium aufzunehmen oder einen MINT-Beruf zu ergreifen, gemessen werden.

$$\delta = Y_{i1}(X_{i1}, Z_1, T = 1) - Y_{i1}(X_{i1}, Z_1, T = 0)$$

Der kausale Treatment-Effekt δ wird also als Differenz des Erfolgsindikators Y_{i1} (z.B. die Intention, ein MINT-Studium oder einen MINT-Beruf zu ergreifen, oder eine andere Outcome-Variable, gemessen zum Zeitpunkt $t = 1$) bei Teilnahme ($T = 1$) und bei Nichtteilnahme ($T = 0$) an einem Schülerinnen-Projekt gemessen. X_{i1} beschreibt weitere individuelle Einflussfaktoren zum Zeitpunkt $t = 1$ (z.B. die Schulnoten oder die geschlechtsspezifische Sozialisation) und Z_1 allgemeine Determinanten (zum Beispiel das Lohnniveau oder das Prestige eines Berufs) auf die persönliche Berufswahl. Die Herausforderung der Evaluation liegt nun darin, dass eine Person i nie in beiden Zuständen gleichzeitig beobachtet werden kann (Wooldridge, 2013). Zur Überwindung dieser Problematik lassen sich verschiedene Evaluationsdesigns nutzen, die sich unterschiedlich gut zur Kausalitätsanalyse eignen und jeweils Vor- und Nachteile in Planung, Durchführung und Aussagekraft mit sich bringen. Im Folgenden stellen wir die Möglichkeiten und Restriktionen verschiedener Untersuchungsdesigns dar und bewerten anschließend die Validität der in den vorliegenden Evaluationsberichten verwendeten Designs anhand dieser Klassifizierung. Wir beginnen mit der idealtypischen Wirkungsmessung von MINT-Schülerinnen-Projekten und diskutieren dann die eher pragmatischen Herangehensweisen.

Das experimentelle Untersuchungsdesign ist methodisch am besten zur Ursache-Wirkungs-Messung geeignet. Der kontrafaktische Zustand $Y_{i1}(X_{i1}, Z_1, T = 0)$ wird hierbei mit Hilfe einer Kontrollperson $Y_{j1}(X_{j1}, Z_1, T = 0)$ bzw. Kontrollgruppe gemessen. Diese Methode erzeugt in der Praxis jedoch nur dann valide Ergebnisse, wenn sich die Personen nicht hinsichtlich der relevanten personenspezifischen Merkmale unterscheiden ($X_{i1} = X_{j1}$). Da die Teilnehmerinnen der MINT-Projekte vermutlich aber bereits eine höhere Affinität zu diesem Fächerspektrum aufweisen als Mädchen, die nicht an einer Maßnahme teilnehmen, könnte ein einfacher Vergleich der Teilnehmerinnen mit Nicht-Teilnehmerinnen den Effekt einer Maßnahme überschätzen. Um sicherzustellen, dass sich die beiden Gruppen nicht in wesentlichen Eigenschaften voneinander unterscheiden und letztlich vergleichbar sind, müssten die Teilnehmerinnen per Zufallsauswahl auf die Gruppen verteilt werden, was im Falle der Schülerinnen-Projekte aufgrund ethischer Bedenken nicht umsetzbar ist. Theoretisch stellen mehrmalige Befragungen von Vergleichs- und Kontrollgruppe (Difference-in-difference-Ansatz) zwar eine Möglichkeit dar, die Unterschiede bei beobachtbaren und nicht-beobachtbaren Determinanten zu eliminieren und die so geschätzten Effekte präzise und unverzerrt zu messen (Wooldridge, 2013). In der Praxis ist dieses Vorgehen jedoch sehr aufwendig. Darüber hinaus können die Ergebnisse dennoch durch kleine Stichprobengrößen, den Hawthorne-Effekt oder die Unmöglichkeit von „verdeckten“ Treatments sowie Placeboeffekte (Colbjørnsen, 2003; Deaton & Cartwright, 2018) verzerrt sein.

Das quasi-experimentelle Design, das der Logik eines Experiments folgt, aber bei der Auswahl der Treatment- und Kontrollgruppe keine Randomisierung vornimmt, ist das Untersuchungsdesign mit der nächstbesten Aussagekraft zur kausalen Effektmessung. Die Auswirkungen einer fehlenden Zufallsaufteilung lassen sich zwar bei großen Stichproben mithilfe multivariater statistischer Verfahren und/oder dem Matching der Teilnehmerinnen der Versuchs- und Kontrollgruppe mildern, jedoch besteht dennoch das Risiko der Verzerrung der Effekte des Treatments durch nicht gemessene Drittvariablen (Diekmann 2012). Darüber hinaus finden MINT-Projekte typischerweise in kleinen Gruppen statt, so dass große Stichproben eher unwahrscheinlich sind. Alternativ könnte im Falle kleiner Stichproben das sequenzielle Treatment angewendet werden, bei dem die Teilnehmerinnen anderer/späterer Maßnahmen (z.B. Schülerinnen der Warteliste der Maßnahme) als Kontrollgruppe dienen (Duflo, Glennerster & Kremer, 2007). Dies setzt jedoch voraus, dass es eine Warteliste gibt oder die Teilnehmerinnen der Schülerinnen-Projekte relativ langfristig vorher bekannt sind.

Vorexperimentelle Designs verzichten auf den Vergleich mit einer Kontrollgruppe und befragen die Teilnehmerinnen entweder vor und nach der Teilnahme am Schülerinnen-Projekt oder nur nach dem Projekt (Diekmann, 2012). Da bei einer einmaligen Befragung jedoch keine Veränderungen der Determinanten der Studien- und Berufswahl gemessen werden können, lassen sich auf diesem Weg keine kausalen Effekte nachweisen⁶. Ein Vorher-Nachher-Vergleich ohne Kontrollgruppe kann dagegen nur dann den kausalen Effekt einer Maßnahme auf die Determinanten der Berufswahl messen, wenn der Outcome nach Nichtteilnahme $Y_{i1}(X_{i1}, Z_1, T = 0)$ identisch ist mit dem Outcome vor Nichtteilnahme $Y_{i0}(X_{i0}, Z_0, T = 0)$. Theoretisch kann aber nicht ausgeschlossen werden, dass sich während des Beobachtungszeitraums die Studien- und Berufswahlentscheidungen insgesamt gewandelt haben und die Veränderungen, die bei den Teilnehmerinnen beobachtet werden, nicht durch das Projekt, sondern durch Veränderungen der persönlichen Einflussfaktoren X_i oder allgemeine Trends (Veränderungen der Determinanten Z) ausgelöst wurden.⁷ Je kürzer der Beobachtungszeitraum und stabiler die sonstigen Determinanten Z der Berufswahl sind, desto geringer ist jedoch diese Problematik.

Da die Schülerinnen-Projekte zumeist auf einen oder wenige Tage begrenzt sind, erscheint der Nutzen eines Kontrollgruppendedesigns (mit ein- oder mehrmaliger Befragung) bei der Evaluation von MINT-Schülerinnen-Projekten eher fraglich. Darüber hinaus sind die persönlichen Einflussfaktoren X_i (z.B. die persönlichen Kompetenzen, die geschlechtsspezifische Sozialisation der Schülerinnen und dazugehörige Einflüsse wie die Einstellungen von Freunden und Familie) zumeist zeitlich sehr stabil, so dass während des relativ kurzen Beobachtungszeitraums keine gravierenden Veränderungen zu erwarten sind und ein einfacher Vorher-Nachher-Vergleich durchaus zu vergleichsweise verlässlichen Ergeb-

6 Folgt man den Befunden von Müller (2016), zeigt sich jedoch, dass eine retrospektive Selbsteinschätzung nur unter bestimmten Bedingungen eine valide Messung der kontrafaktischen Situation darstellt.

7 Dies wäre beispielweise der Fall, wenn sich während des Beobachtungszeitraums die Löhne in den MINT-Berufen drastisch erhöhen, die Vereinbarkeit von Familie und Beruf offensichtlich verbessern und die Arbeitslosenquote in den frauendominierten Berufen signifikant erhöhen würde, so dass sich Frauen aufgrund dieser Veränderungen ohnehin deutlich häufiger für MINT-Berufe entscheiden würden.

nissen führen kann. Dieses Untersuchungsdesign lässt sich auch in der Praxis ohne größeren Aufwand umsetzen und stellt daher eine gute und pragmatische Lösung dar.

Nichtsdestotrotz müssen die Befragungsergebnisse mit Vorsicht interpretiert werden. So erhöht die häufig direkt nach der Teilnahme an der Maßnahme stattfindende Befragung der Schülerinnen die Wahrscheinlichkeit sozial erwünschter Antworten. Dies gilt insbesondere dann, wenn die Teilnehmerinnen die Fragebögen in Anwesenheit der Projektkoordinatorinnen und Projektkoordinatoren ausfüllen, deren Projekt sie möglicherweise nicht schlecht bewerten möchten. Auch wenn die Schülerinnen dies nicht intendieren, führt der Hawthorne-Effekt, der durch die besondere Aufmerksamkeit entsteht, die den Teilnehmerinnen zuteilwird, bei einer Prä-Post-Befragung zu einer Überschätzung der gemessenen Effekte (Colbjørnsen, 2003). Diese Verzerrung obliegt jedoch allen Befragungen und kann nur durch die Gestaltung einer professionellen Befragungssituation gemildert werden. Zum anderen wäre es denkbar, dass die Wirkung der Fördermaßnahme direkt nach der letzten Veranstaltung größer ist als zu einem späteren Zeitpunkt, so dass insbesondere die langfristigen Effekte eventuell überschätzt werden.

Aus unserer Sicht liefern somit Vorher-/Nachher-Befragungen auch ohne den Vergleich mit einer Kontrollgruppe durchaus valide Informationen, um Veränderungen in den Einstellungen und Intentionen der Teilnehmerinnen von MINT-Projekten erfassen zu können.

Analyse der Evaluationen: In den vorliegenden Evaluationen werden nur zwei verschiedene Untersuchungsdesigns angewendet. Unsere erste Kategorie umfasst alle Einmalbefragungen am Ende des Projekts. In Kategorie 2 fallen Evaluationen, deren Befragte sowohl vor als auch nach dem Projekt Fragebögen beantworteten. Evaluationen aus der Kategorie 1 erlauben aus wissenschaftlicher Sicht keine Schlussfolgerungen über die kausale Wirkung der Maßnahme.

Von den 13 vorliegenden Evaluationsberichten führte der Großteil der Fördermaßnahmen einmalige Befragungen zum Ende des MINT-Projekts durch (Kategorie 1). Nur drei der analysierten Evaluationen beschreiben, dass Vorher-Nachher-Befragungen (Kategorie 2) genutzt wurden, und werden damit zumindest den minimalen Anforderungen an eine Evaluation eines MINT-Schülerinnen-Projekts gerecht. Die Validität der Wirkungsanalyse der existierenden Evaluationen ist in der Summe demnach eher begrenzt. Angesichts der deutlich komplexeren Datenerhebung und der begrenzten Vorteile hinsichtlich der Validität ist es nicht verwunderlich, dass es unseres Wissens keine Evaluation mit Hilfe eines Kontrollgruppendesigns gibt.

4 Zusammenfassung und kritische Reflexion

Hochschulen für angewandte Wissenschaften versuchen durch gezielte Angebote für Schülerinnen den Frauenanteil in MINT-Studiengängen zu erhöhen. Angesichts der bundesweit nur sehr geringfügigen Erhöhung der Frauenanteile und der beträchtlichen personellen und finanziellen Ressourcen für die Durchführung dieser Aktivitäten stellt sich dringend die Frage nach der Wirksamkeit dieser Maßnahmen.

Auch wenn kein direkter kausaler Zusammenhang zwischen der Teilnahme an einer Maßnahme und der späteren Berufswahl aufgrund der Komplexität der Entscheidung zu

erwarten ist, so sollten die Schülerinnen-Projekte zumindest dazu beitragen, die Hürden zur Aufnahme eines MINT-Studiums oder -Berufs abzubauen und/oder die Determinanten dieser Bildungsentscheidung positiv beeinflussen. Ob die Schülerinnen-Projekte tatsächlich einen Einfluss auf die relevanten und veränderbaren Einflussfaktoren der Studien- und Berufswahl haben, lässt sich allerdings anhand der uns vorliegenden Evaluationsberichte nicht beantworten. Unsere Analysen zeigen, dass die bisherigen Evaluationen sowohl hinsichtlich der Güte und Vielfalt der Erfolgskriterien als auch der Validität der Wirkungsanalyse verbesserungsfähig sind. Die Bewertung eines Projekts anhand einer Selbstauskunft der teilnehmenden Schülerinnen zur Zufriedenheit mit dem Projekt (Output) lässt keine Schlüsse über die Effekte auf die Determinanten der Studien- und Berufswahl zu. Hierzu müssten grundsätzlich mehr Fragen zu den verschiedenen Outcome-Indikatoren gestellt werden. Auch von der Möglichkeit, den langfristigen Impact anhand der Intention der Studien- und Berufswahl zu approximieren, wird sehr selten Gebrauch gemacht. Darüber hinaus lassen sich kausale Effekte auf Einstellungen der Schülerinnen nicht mithilfe der meistens verwendeten Einmalbefragung am Ende des Projekts feststellen.

Auch wenn die Datenbasis unserer Meta-Evaluation sehr gering ist, gehen wir davon aus, dass die Ergebnisse ein realistisches Bild der vorhandenen Evaluationen abgeben. Aus der Kommunikation mit Projektverantwortlichen hat sich gezeigt, dass einige Projektverantwortliche und/oder Evaluierende, z.B. aus Sorge um Datenschutz und Wahrung der Anonymität der Teilnehmerinnen, ihre Evaluationsberichte nicht zur Verfügung stellen wollten. Insbesondere im Hinblick darauf, dass die Projekte häufig von derselben Person durchgeführt und evaluiert werden, steigt der Anreiz für die Verantwortlichen, unvorteilhafte Ergebnisse nicht zu veröffentlichen. Da wissenschaftlich profundere Studien vermutlich eher von externen Wissenschaftlerinnen und Wissenschaftlern durchgeführt werden, ist nicht unbedingt zu erwarten, dass uns diese Evaluationen systematisch vorenthalten wurden.

Die relativ kleine Anzahl an auswertbaren Evaluationen schmälert aus unserer Sicht auch nicht die Erkenntnisse aus dieser Meta-Evaluation. Ganz im Gegenteil, sie verdeutlicht die Relevanz unserer Forschung und macht das große Potenzial im Bereich der Evaluationen von MINT-Projekten für Schülerinnen an HAW deutlich. Es mangelt nicht nur an der Anzahl empirischer Studien, sondern auch an der wissenschaftlichen Fundierung der Evaluationsansätze. Angesichts dieser Wissenslücke stehen viele Hochschulen vor der offenen Frage, wie sie ihre knappen Ressourcen für die Anwerbung von Frauen in MINT-Studiengängen am effektivsten nutzen können. Darüber hinaus könnten aussagekräftige Evaluationen wichtigen Input für die Weiterentwicklung von Maßnahmen liefern.

Das von uns entwickelte Wirkungsmodell, welches explizit die drei Wirkungsebenen von MINT-Angeboten für Schülerinnen an HAW systematisch darstellt, sowie die Diskussion der verschiedenen Wirkungsanalysen können Projektverantwortlichen auch bei der Konzeption einer eigenen Evaluation dienlich sein. Darüber hinaus ist ein „Werkzeugkasten“ mit praxistauglichen Anleitungen zur Durchführung von Evaluationen hilfreich, um methodisches Wissen aus der Wissenschaft an die Praxis weiterzugeben und somit die Erstellung belastbarer Evaluationen von laufenden und zukünftigen MINT-Projekten zu fördern (siehe z.B. Nickolaus & Mokhonko, 2016; Nationales MINT Forum, 2018; sowie Brenning & Wolf, 2021).

Literatur

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-t](https://doi.org/10.1016/0749-5978(91)90020-t)
- Anger, C., Koppel, O. & Plünnecke, A. (2018). *MINT-Frühjahrsreport 2018. MINT – Offenheit, Chancen, Innovationen*. Verfügbar unter: [https://www.arbeitgeber.de/www/arbeitgeber.nsf/res/MINT-Fruehjahrensreport_2018.pdf/\\$file/MINT-Fruehjahrensreport_2018.pdf](https://www.arbeitgeber.de/www/arbeitgeber.nsf/res/MINT-Fruehjahrensreport_2018.pdf/$file/MINT-Fruehjahrensreport_2018.pdf)
- Armitage, C. J. & Conner, M. (2001). Efficacy of the theory of planned behaviour. A meta-analytic review. *British Journal of Social Psychology*, 40(4), 471–499. <https://doi.org/10.1348/014466601164939>
- Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York: Freeman. <https://doi.org/10.4135/9780857025753.n178>
- Bandura, A., Barbaranelli, C., Caprara, G. V. & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72(1), 187–206. <https://doi.org/10.1111/1467-8624.00273>
- BLK – Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (2002). Frauen in den ingenieur- und naturwissenschaftlichen Studiengängen. Bericht der BLK vom 2. Mai 2002. *Materialien zur Bildungsplanung und zur Forschungsförderung, Heft 100*. Bonn.
- Brenning, S. & Wolf, E. (2019). *What works? Eine Meta-Evaluation von Evaluationen von MINT-Projekten für Schülerinnen und Studentinnen*. Verfügbar unter: https://www.oth-regensburg.de/fileadmin/media/fakultaeten/s/forschung_projekte/mint/MINT-Strategien_4_0_Meta-Evaluation_EW_SB.pdf
- Brenning, S. & Wolf, E. (2021). *Wirkung messen. Handbuch zur Evaluation von MINT-Projekten für Schülerinnen*. Hochschule München.
- Colbjørnsen, T. (2003). Der Hawthorne-Effekt oder die Human-Relations-Theorie. Über die experimentelle Situation und ihren Einfluss. In S. U. Larsen & E. Zimmermann (Hrsg.), *Theorien und Methoden in den Sozialwissenschaften* (S. 131–143). Wiesbaden: Westdeutscher Verlag. https://doi.org/10.1007/978-3-322-80451-8_10
- Correll, S. J. (2001). Gender and the career choice process. The role of biased self-assessments. *American Journal of Sociology*, 106(6), 1691–1730. <https://doi.org/10.1086/321299>
- Correll, S. J. (2004). Constraints into preferences. Gender, status, and emerging career aspirations. *American Sociological Review*, 69(1), 93–113. <https://doi.org/10.1177/000312240406900106>
- Deaton, A. & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, 210, 2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>
- DeGEval – Gesellschaft für Evaluation e.V. (2016). *Standards für Evaluation*. Erste Revision 2016. Mainz: DeGEval – Gesellschaft für Evaluation e.V.
- Diekmann, A. (2012). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Heidelberg: Springer.
- Duflo, E., Glennerster, R. & Kremer, M. (2007). *Using randomization in development economics research*. A toolkit. Discussion Paper Nr. 6059. Centre for Economic Policy Research, London.
- Ertl, B., Luttenberger, S. & Paechter, M. (2014). Stereotype als Einflussfaktoren auf die Motivation und die Einschätzung der eigenen Fähigkeiten bei Studentinnen in MINT-Fächern. *Gruppendynamik und Organisationsberatung*, 45(4), 419–440. <https://doi.org/10.1007/s11612-014-0261-3>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Hiller, S. (2014). *Ergebnisbericht zur wissenschaftlichen Evaluation von EnterTechnik*. Stuttgart: Dialogik gGmbH.

- Hurrelmann, K., Bauer, U., Grundmann, M. & Walper, S. (2015). *Handbuch Sozialisationsforschung*. Weinheim: Beltz.
- Jones, M. G., Howe, A. & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education*, 84(2), 180–192. [https://doi.org/10.1002/\(sici\)1098-237x\(200003\)84:2<180::aid-sce3>3.0.co;2-x](https://doi.org/10.1002/(sici)1098-237x(200003)84:2<180::aid-sce3>3.0.co;2-x)
- Kenkmann, T. (2005). 5. *Brandenburgische Sommer-Universität für Schülerinnen in Naturwissenschaft und Technik*. Abschlussbericht.
- Kessels, U. & Hannover, B. (2006). Zum Einfluss des Image von mathematisch-naturwissenschaftlichen Schulfächern auf die schulische Interessenentwicklung. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule*. Abschlussbericht des DFG-Schwerpunktprogramms (S. 350–369). Münster: Waxmann.
- Kessels, U. (2015). Bridging the gap by enhancing the fit: How stereotypes about STEM clash with stereotypes about girls. *International Journal of Gender, Science and Technology*, 7(2), 280–296.
- Kim, M. & Hunter, J. E. (1993). Relationships among attitudes, behavioral intentions, and behavior. A meta-analysis of past research, part 2. *Communication Research*, 20, 331–364. <https://doi.org/10.1177/009365093020003001>
- Kray, J. (2019). *Entwicklungspsychologie. Ein Überblick für Psychologiestudierende und -interessierte*. Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-57761-5>
- Lent, R. W., Brown, S. D. & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice and performance. *Journal of Vocational Behavior*, 45, 79–122. <https://doi.org/10.1006/jvbe.1994.1027>
- Löther, A. & Girlich, J. (2011). Frauen in MINT-Fächern: Bilanzierung der Aktivitäten im hochschulischen Bereich. *Materialien der Gemeinsamen Wissenschaftskonferenz. Heft 21*. Bonn.
- Matthes, S. (2019). *Warum werden Berufe nicht gewählt? Die Relevanz von Attraktions- und Aversionsfaktoren in der Berufsfindung*, Institut für Berufsbildung, Bonn.
- Mokhonko, S., Nickolaus, R. & Windaus, A. (2014). Förderung von Mädchen in Naturwissenschaften. Schülerlabore und ihre Effekte. *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 143–159. <https://doi.org/10.1007/s40573-014-0016-2>
- Mokhonko, S. (2016). *Nachwuchsförderung im MINT-Bereich. Aktuelle Entwicklungen, Fördermaßnahmen und ihre Effekte*. Stuttgart: Franz Steiner.
- Müller, C. E. (2016). Wirkungsbewertung mittels retrospektiver Pretests: Eine kritische Würdigung. *Zeitschrift für Evaluation*, 15(2), 221–239.
- Müller, C. & Albrecht, M. (2016). The future of impact evaluation is rigorous and theory-driven. In R. Stockmann & W. Meyer (Hrsg.), *The future of evaluation. Global trends, new challenges, shared perspectives* (S. 283–293). New York: Palgrave Macmillan. https://doi.org/10.1057/9781137376374_21
- Nationales MINT Forum (2018). *Orientierungsrahmen für wirkungsvolle Arbeit außerschulischer MINT-Initiativen*. Ein praktischer Leitfaden zur Selbstanalyse. Berlin. Verfügbar unter: https://www.nationalesmintforum.de/fileadmin/medienablage/content/publikationen_und_empfehlungen/publikationen/2018/NMF_Leitfaden_Web.pdf
- Nickolaus, R. & Mokhonko, S. (Hrsg.) (2016). *In fünf Schritten zum zielführenden Evaluationsdesign. Eine Handreichung für Bildungsinitiativen im MINT-Bereich* (acatech Materialien). München: acatech – Deutsche Akademie der Technikwissenschaften. Verfügbar unter: <https://www.acatech.de/publikation/in-fuenf-schritten-zum-zielfuehrenden-evaluationsdesign-eine-handreichung-fuer-bildungsinitiativen-im-mint-bereich/>
- Obergruber, N. (2018). *Microeconomic analysis of individual and institutional determinants of education and occupational choice*. Inaugural-Dissertation zur Erlangung des Grades Doctor oeco-

- nomiae publicae (Dr. oec. publ.) an der Ludwig-Maximilians-Universität München.
<https://doi.org/10.5282/edoc.22892>
- Patton, M. (1997). *Utilization-focused evaluation: The new century text*. Thousand Oaks: Sage.
- Pöllmann-Heller, K. & Rudolph, C. (2019). Frauen in das MINT-Studium. Ambivalenzen und Potenziale von hochschulischen Förderprozessen. In C. Driesen & A. Ittel (Hrsg.), *Der Übergang in die Hochschule. Strategien, Organisationsstrukturen und Best Practices an deutschen Hochschulen* (S. 91–102). Münster: Waxmann.
- Prenzel, M., Lankes, E. & Minsel, B. (2000). Interessenentwicklung in Kindergarten und Grundschule. Die ersten Jahre. In U. Schiefele & K.-P. Wild (Hrsg.), *Interesse und Lernmotivation. Untersuchungen zu Entwicklung, Förderung und Motivation* (S. 11–30). Münster: Waxmann.
- Randall, D. M. & Wolff, J. A. (1994). The time interval in the intention-behaviour relationship. Meta-analysis. *British Journal of Social Psychology*, 33, 405–418.
<https://doi.org/10.1111/j.2044-8309.1994.tb01037.x>
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report*, 2(5), 36–38.
- Shapiro, J. R. & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, 66, 175–183.
<https://doi.org/10.1007/s11199-011-0051-0>
- Sheu, H., Lent, R. W., Brown, S. D., Miller, M. J., Hennessy, K. D. & Duffy, R. D. (2010). Testing the choice model of social cognitive career theory across Holland themes: A meta-analytic path analysis. *Journal of Vocational Behavior*, 76(2), 252–264.
<https://doi.org/10.1016/j.jvb.2009.10.015>
- Siegler, R., Eisenberg, N., De Loache, J. & Saffran, J. (2016). *Entwicklungspsychologie im Kindes- und Jugendalter*. Heidelberg: Springer. https://doi.org/10.1007/978-3-662-47028-2_16
- Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28.
<https://doi.org/10.1006/jesp.1998.1373>
- Solga, H. & Pfahl, L. (2009). Doing Gender im technisch-naturwissenschaftlichen Bereich. In J. Milberg (Hrsg.), *Förderung des Nachwuchses in Technik und Naturwissenschaft* (S. 155–218). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-01123-8_4
- Statistisches Bundesamt (2019). *DESTATIS – Bildung, Forschung, Kultur: Studierende in MINT-Fächern*. Verfügbar unter:
<https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/studierende-mint-faechern.html>
- Stöger, H., Duan, X., Schirmer, S., Greindl, T. & Ziegler, A. (2013). The effectiveness of a one-year online mentoring program for girls in STEM. *Computers & Education*, 69, 408–418.
<https://doi.org/10.1016/j.compedu.2013.07.032>
- Taskinen, P. (2010). *Naturwissenschaften als zukünftiges Berufsfeld für Schülerinnen und Schüler mit hoher naturwissenschaftlicher und mathematischer Kompetenz: Eine Untersuchung von Bedingungen für Berufserwartungen* (Dissertation). Christian-Albrechts-Universität zu Kiel, Kiel.
- Taskinen, P. H., Schütte, K. & Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career. The role of school factors, individual interest and science self-concept. *Educational Research and Evaluation*, 19(8), 717–733.
<https://doi.org/10.1080/13803611.2013.853620>
- Weinhardt, F. (2017). Ursache für Frauenmangel in MINT-Berufen? Mädchen unterschätzen schon in der fünften Klasse ihre Fähigkeiten in Mathematik. *DIW Wochenbericht*, 45/2017.
- Wentzel, W. (2008). „Ich will das und das ist mein Weg!“ *Junge Frauen auf dem Weg zum Technikberuf. Qualitative Interviews mit ehemaligen Girls' Day-Teilnehmerinnen in Ausbildung und Studium*. Bielefeld: Kompetenzzentrum Technik – Diversity – Chancengleichheit.

- Widmer, T. (1996). *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*. Bern: Haupt.
- W. K. Kellogg Foundation (2004). *Logic Model Development Guide. Using Logic Models to bring together Planning, Evaluation and Action*. Michigan. Verfügbar unter:
<https://www.bttop.org/sites/default/files/public/W.K.%20Kellogg%20LogicModel.pdf>
- Wooldridge, J. (2013). *Introductory econometrics: A modern approach*. Mason: Nelson Education.
- Zeisberg, I. (2013). Light up your life. Einfluss geschlechtsspezifischer Maßnahmen in außerschulischen Lernorten auf MINT-Interesse und Berufswahl. *Wissenschaftliche Schriften der WWU Münster*, Reihe VI, 12.

Kontakt:

Stefanie Brenning, M.A. · Prof. Dr. Elke Wolf
Hochschule München
Lothstraße 64
80335 München
E-Mail: s_brenning@gmx.de
E-Mail: elke.wolf@hm.edu

Test- und notenbasierte Erfolgsprognose im wirtschaftswissenschaftlichen Masterstudium: Befunde aus dem Nationalen Bildungspanel

Johannes Schult

Zusammenfassung: Fachspezifische Studierfähigkeitstests haben sich auch im deutschsprachigen Raum als valide Prädiktoren des Bachelorstudien Erfolgs und somit als hilfreiche standardisierte Auswahlinstrumente erwiesen. Anhand der Daten des Nationalen Bildungspanels (NEPS) wurde untersucht, ob sich ein vergleichbares Muster auch für das konsekutive Masterstudium zeigt. Die Analysestichprobe umfasste 118 Studierende der Wirtschaftswissenschaften, die zum Ende ihres Bachelorstudiums einen fachspezifischen Kompetenztest bearbeiteten und ein Masterstudium aufnahmen. Der Test zeigte eine niedrige Validität ($r = -.23$) bezüglich der Masternoten. Es zeigte sich dabei keine inkrementelle Validität über die Bachelornote hinaus. Bei gleichem Testscore waren Frauen im Schnitt um 0.17 Notenpunkte besser als Männer. Die Ergebnisse sprechen gegen den Einsatz des wissenslastigen Fähigkeitstests zur Studierendenauswahl im Master. Gelingensfaktoren für eine erfolgreiche Studierendenauswahl werden im Kontext von Prädiktorenkombinationen diskutiert.

Schlüsselwörter: Studierfähigkeitstest, Validität, Wirtschaftswissenschaften, Studienerfolg, Geschlecht, differenzielle Prognose

Test-based and grade-based prediction of success in Economics Masters: Findings from the National Educational Panel Study

Abstract: Subject-specific competence tests have been proven to be valid predictors of academic success in Bachelor studies; such tests are therefore useful standardized tools in student selection procedures. Using data from the National Educational Panel Study (NEPS), we investigated whether the pattern is similar for academic success in subsequent Master studies. The analysis sample consisted of 118 economics students who took a subject-specific competence test at the end of their Bachelor degree and began a subsequent Master. The test showed a weak validity ($r = -.23$) regarding the Master grade point average (GPA). It had no incremental validity over the Bachelor GPA. Given equal test scores, women outperformed men by 0.17 grade points on average. The results suggest that the knowledge-based test is not suited for Master student selection. Success factors for student selection procedures are discussed in the context of combining multiple predictors.

Keywords: admission test, validity, economics, academic performance, sex, differential prediction

1 Einleitung

Die Vergabe von Studienplätzen bewegt sich im Spannungsfeld zwischen Bestenauslese und Chancenfairness. Zulassungsverfahren betreffen auch Masterstudiengänge. Denn weder in der Konzeption noch in der Praxis gibt es so viele Masterstudienplätze, wie es (interessierte) Bachelorabsolventinnen und -absolventen gibt (Kultusministerkonferenz, 2011). Die Instrumente, die für die Studierendenauswahl eingesetzt werden, sollen dabei in erster Linie prognostisch valide sein, also den zukünftigen Studienerfolg möglichst gut vorhersagen. Aber auch Kriterien wie die Kosten und die Logistik des Auswahlprozesses und die Anfälligkeit für Verfälschungen müssen berücksichtigt werden. Während die Vorhersage des Studienerfolgs im Bachelor bereits vielfach untersucht wurde (vgl. Schult, Hofmann & Stegt, 2019), gibt es zur prognostischen Validität von Auswahlkriterien für das konsekutive Masterstudium kaum Forschungsbefunde aus dem deutschen Hochschulbetrieb. Entsprechend finden sich für das Masterstudium auch keine gesicherten Befunde zu Gruppenunterschieden bei der Leistungsprognose, die sich auf Personenmerkmale wie beispielsweise das Geschlecht beziehen. In der vorliegenden Studie wird deshalb die prognostische Validität eines wirtschaftswissenschaftlichen Kompetenztests hinsichtlich des Studienerfolgs im Masterstudium untersucht. Im Fokus stehen dabei die operative Validität¹ der Testleistung, die inkrementelle Validität des Tests über bisherige Noten hinaus sowie die geschlechtsspezifische differenzielle Prognose.

1.1 Studienerfolgsprognose

Die prognostische Validität eines Auswahlinstruments gibt an, wie gut es den späteren Studienerfolg vorhersagen kann. Der Studienerfolg umfasst viele, häufig miteinander verbundene Aspekte. Dazu gehören ein erfolgreicher Abschluss, subjektive Studienzufriedenheit, berufliche Chancen und damit verbunden auch möglichst gute Noten (Stemler, 2012). Studiennoten haben eine besondere Bedeutung in der Validitätsforschung, da sie mit den anderen genannten Erfolgsindikatoren positiv zusammenhängen und bereits während des laufenden Studiums vergeben werden und entsprechend erfasst werden können (Hell, Trappmann & Schuler, 2008; Strahan & Credé, 2015). Zudem fließen im Masterstudium die Noten ab dem ersten Semester in den Abschluss ein. Dadurch ergibt sich ein unmittelbarer Indikator des Lernfortschritts und Wissenserwerbs im Studium.

Bachelorabschlussnoten werden häufig als (zentrales) Auswahlkriterium für zulassungsbeschränkte Studiengänge herangezogen. Noten, die von verschiedenen Institutionen vergeben wurden, sind allerdings nur bedingt vergleichbar.² Die Notenvergabepraxis wie auch die abgeprüften Inhalte variieren von Hochschule zu Hochschule (Gaens, 2018). Es gibt auch Benotungsunterschiede zwischen Universitäten und Fachhochschulen (Weimar, Schauburger, Borowski & Prinz, 2017). Da etwa 40% der Studierenden für das Masterstudium die Hoch-

-
- 1 Die operative Validität ist der für selektionsbedingte Variabilitätseinschränkungen und Kriteriumsunreliabilität korrigierte Validitätskoeffizient.
 - 2 In diesem Sinne entschied das Bundesverfassungsgericht (BVG), dass die Hochschulen Studienplätze „nicht allein und auch nicht ganz überwiegend nach dem Kriterium der Abiturnoten“ vergeben dürfen (Bundesverfassungsgericht, 2017, Abs. 209).

schule wechseln (Bericht der Bundesregierung, 2015), ist die notenbasierte Masterstudienplatzvergabe entsprechend problematisch. Aus konzeptueller Sicht sind Noten zudem eine Bewertung erbrachter Bildungsleistungen und keine Abschätzung des erwarteten zukünftigen Studiums- und Berufserfolgs (Schuler & Schult, 2018). Diese Problematik tritt im Zuge der Bologna-Reform deutlich zutage. Als standardisiertes Auswahlinstrument, dessen Ergebnisse ein hohes Maß an Vergleichbarkeit bieten, werden an manchen Hochschulen deshalb fachspezifische Studierfähigkeitstests eingesetzt. Im Jahr 2005 wurden in Deutschland bei 3,3% der zulassungsbeschränkten Studiengänge fachspezifische Studierfähigkeitstests zur Auswahl eingesetzt (überwiegend in Kombination mit anderen Auswahlkriterien wie Abiturnoten oder Auswahlgesprächen). Dabei waren in allen 19 betrachteten wirtschaftswissenschaftlichen Studiengängen allgemeine oder fachspezifische Tests Bestandteil der Auswahlverfahren (Heine, Briedis, Didi, Haase & Trost, 2006). Aufgrund der Bologna-Reform haben Fähigkeitstests seither weiter an Bedeutung gewonnen (Kouwenaar, 2015).

Studierfähigkeitstests zielen „auf die Operationalisierung der für ein erfolgreiches Studium wesentlichen kognitiven Fähigkeiten“ ab (Hell, Trapmann & Schuler, 2007, S. 252). Während manche Tests vor allem fachbezogenes schlussfolgerndes Denken erfassen, gibt es auch Verfahren, die studienrelevantes Vorwissen abfragen.³ Studierfähigkeitstests kommen insbesondere in stark nachgefragten Fächern wie Medizin (Schwabbe et al., 2018), Psychologie (Janke & Dickhäuser, 2018) und Wirtschaftswissenschaften (Weimar et al., 2017) zum Einsatz. Da üblicherweise weniger Masterplätze als Bachelorplätze vorhanden sind, verschärft sich die Selektionsproblematik bei der Masterzulassung. Das Verhältnis der Bewerbungen zu vorhandenen Plätzen variiert zwischen Studienfächern, Hochschulen und Jahrgängen. Für die Hochschulen kommt es bei der Studierendenauswahl sowohl darauf an, die Zahl der später im Studium erfolgreichen Personen zu maximieren, als auch die Zahl der später im Studium nicht erfolgreichen Personen zu minimieren. Eine Übersicht über die Kosten-Nutzen-Abwägung bei der Auswahlquote (Bewerbungen pro Studienplatz) einerseits und bei der Erfolgserwartung bei einer zufälligen Auswahl (Basiserfolgsrate ohne spezifische Zulassungskriterien) andererseits findet sich bei Hell et al. (2008). Eine zentrale Größe bei den Abwägungen ist dabei die prognostische Validität im Sinne der Korrelation von Auswahlkriterium und Studienerfolg. Zumindest bei hochselektiven Studiengängen gilt: Ein Auswahlverfahren ist umso hilfreicher, desto höher seine Validität ist.

1.2 Empirische Befunde zur Vorhersagekraft von Bachelornoten und Testleistungen

Zur prognostischen Validität von Bachelorabschlussnoten gibt es im deutschsprachigen Raum erstaunlich wenige empirische Untersuchungen. Bei Berner Psychologie-Studierenden korrelierten die Bachelor- und die Masterabschlussnoten hoch mit $r = .58$ ($n = 256$; Troche, Mosimann & Rammsayer, 2014); bei Züricher Informatik-Studierenden betrug die Korrelation sogar $r = .65$ ($n = 181$; Zimmermann, Brodersen, Heinemann & Buhmann, 2015). In beiden Studien gab es keine Zulassungsbeschränkungen für das Masterstudium.

3 Es gibt weitere valide nichtkognitive Prädiktoren (z.B. Gewissenhaftigkeit; vgl. Schuler & Schult, 2018). Auswahlverfahren – und entsprechend auch der vorliegende Beitrag – nutzen aber überwiegend kognitive Prädiktoren, weil deren Messung weniger anfällig gegenüber bewussten Verfälschungen ist.

Obwohl in den deutschen Hochschulverwaltungen eigentlich eine Fülle von Registerdaten zur Bestimmung der Notenvalidität existiert, fehlen in Studien zur Mastererfolgsprognose die Bachelornoten als Prädiktoren (z.B. Madani, Melzer & Müller, 2013). Eine US-amerikanische Meta-Analyse schätzt die operative Validität⁴ von Bachelornoten deutlich niedriger auf $\rho = .35$ ($k = 28$, $n = 5,609$; Kuncel, Credé & Thomas, 2007). Für $n = 77$ Studierende der Wirtschaftswissenschaften an einer niederländischen Universität betrug die Korrelation ebenfalls „nur“ $r = .30$ (Schwager, Hülshager, Bridgeman & Lang, 2015).

Einen Überblick über die prognostische Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum bieten zwei Metaanalysen, die für wirtschaftswissenschaftliche Studiengänge Validitäten von $\rho = .33$ ($k = 7$, $n = 441$; Hell et al., 2007) und für neuere Studien ab 2005 $\rho = .38$ ($k = 10$, $n = 2,678$; Schult et al., 2019) fanden. Für die Erfolgsprognose im Masterstudium zeigen sich in internationalen Studien etwas höhere Werte (Kuncel & Hezlett, 2007), z.B. für den Graduate Management Admission Test (GMAT) $\rho = .47$ ($k = 29$, $n = 5,201$; Kuncel et al., 2007). Im deutschsprachigen Raum findet sich bislang nur eine Studie für den Masterbereich (Stegt & Bergholz, 2018), obgleich im Rahmen der Bologna-Reform viele Hochschulen Studierfähigkeitstests zur Masterstudierendenauswahl einsetzen (vgl. Borowski, Schauburger & Weimar 2018). Im Rahmen des Masterzulassungsverfahrens bearbeiteten $n = 841$ Studierende der Betriebs- und Volkswirtschaftslehre (aus zwei verschiedenen Universitäten) den fachspezifischen Studierfähigkeitstest TM-WISO (Test für Masterstudiengänge in Wirtschafts- und Sozialwissenschaften)⁵. Die Korrelation zwischen Testleistung und Masterabschlussnote betrug $r = -.37$ (Stegt & Bergholz, 2018).⁶ In der vorliegenden Studie wird nun untersucht, ob sich diese Befunde auf einen anderen wirtschaftswissenschaftlichen Fähigkeitstest und Studierende von mehreren verschiedenen Hochschulen übertragen lassen. Angesichts der oben diskutierten Notenproblematik (vgl. Gaens, 2018) und der bereits gängigen Auswahlpraxis scheint es zudem angezeigt, die inkrementelle Validität der Testleistung über die Bachelorabschlussnote hinaus zu betrachten.

1.3 Testfairness bei der Vorhersage des Studienerfolgs

Ein wichtiger Aspekt der Testfairness im Auswahlkontext ist die unverzerrte Erfolgsvorhersage (d.h. die Abwesenheit einer differenziellen Prognose; vgl. AERA, APA & NCME, 2014). Eine Vorhersage wird dann als fair angesehen, wenn für verschiedene Personengruppen dasselbe (lineare Regressions-)Modell für die Prognose verwendet werden kann. Eine Verzerrung liegt hingegen dann vor, wenn Personen aus verschiedenen Gruppen bei gleicher Testleistung unterschiedlich gut im Studium abschneiden (vgl. Fischer, Schult & Hell, 2013). Im Fokus stehen häufig Geschlechtsunterschiede, da im Sinne der Bildungsge-

4 Die operative Validität (korrigiert für selektionsbedingte Variabilitätseinschränkungen und Kriteriumsunreliabilität) wird nachfolgend stets mit ρ bezeichnet zur klaren Abgrenzung vom unkorrigierten Korrelationskoeffizient r .

5 Der Test bestand aus den Teilen „Planen in Studium und Beruf“, „Texte analysieren“, „Wirtschaftliche Zusammenhänge formalisieren“ und „Wirtschaftsgrafiken interpretieren“ mit jeweils 18 gewerteten Multiple-Choice-Aufgaben. Die Testungen hatten eine Bearbeitungszeit von 3:50 Stunden und fanden computerbasiert unter Aufsicht statt. Die Gesamtreliabilität betrug $r = .87$ (Stegt & Bergholz, 2018).

6 Das negative Vorzeichen des Validitätskoeffizienten ergibt sich daraus, dass die Notenskala in Deutschland so gepolt ist, dass niedrige Zahlen gute Leistungen und hohe Zahlen schlechte Leistungen angeben.

rechtigkeit kein Geschlecht beim Hochschulzugang sowie bei der Beurteilung der Studienleistung benachteiligt werden soll. Denn bereits im Studium können Unterschiede auftreten, die sich im Berufsleben weiterentwickeln, so dass Frauen für die gleiche Arbeit oft weniger Lohn erhalten als Männer (Anger & Schmidt, 2010; Schult, 2012). Für den Studienfeldbezogenen Beratungstest (SFBT) Wirtschaftswissenschaften wurde bei der Vorhersage der Noten im Bachelor-Studium nach einem Jahr die Leistung der Frauen unterschätzt, konkret erhielten Frauen im Studium durchschnittlich 0.12 Notenpunkte mehr als Männer mit der gleichen Testleistung ($n = 356$; Fischer, Schult & Hell, 2015). Die metaanalytischen Befunde zum US-amerikanischen GRE (Graduate Record Examination) zeigen hingegen keine differenzielle Prognose: Bei gleicher GRE-Leistung schneiden Frauen im Studium lediglich um durchschnittlich 0.03 Standardabweichungen besser ab ($k = 13$, $n = 2,589$, $p = .96$; Fischer et al., 2013). Inwiefern sich im deutschsprachigen Raum Geschlechtsunterschiede bei der Vorhersage des Masterstudien Erfolgs durch Testleistungen zeigen, ist bislang unklar und gehört ebenfalls zu den Forschungsfragen, die nachfolgend anhand des Nationalen Bildungspanels (NEPS) untersucht werden.

1.4 Fragestellungen

Forschungsfrage 1 lautet: Wie valide ist der fachspezifische Kompetenztest bezüglich der Noten im Masterstudium? Forschungsfrage 2 behandelt die inkrementelle Validität des Testverfahrens über die Bachelorabschlussnote hinaus. Außerdem wird die geschlechtsspezifische Fairness der Vorhersage exploriert (Forschungsfrage 3).

2 Methode

2.1 Stichprobe und Studiendesign

Diese Arbeit nutzt Daten des NEPS (Startkohorte 5: Studierende⁷). Diese Startkohorte bestand aus $N = 18,252$ Studienanfängerinnen und -anfängern, die im Wintersemester 2010/11 in Deutschland ihr Studium aufnahmen (vgl. Blossfeld, Roßbach & von Maurice, 2011; Steinwede & Aust, 2012). Die vorliegende Studie fokussiert auf wirtschaftswissenschaftliche Studiengänge, da aufgrund des hohen Auslastungsgrads in diesem Feld oftmals fachspezifische Studierfähigkeitstests eingesetzt werden. Untersucht wird deshalb die Teilstichprobe der Studierenden, die (a) 2014 den wirtschaftswissenschaftlichen Kompetenztest bearbeitet haben⁸, (b) zum Testzeitpunkt ein wirtschaftswissenschaftliches Bachelorstudium verfolgten oder bereits abgeschlossen hatten und (c) anschließend ein Masterstudium begannen. Die fi-

7 Die Daten des NEPS (beschrieben unter <https://dx.doi.org/10.5157/NEPS:SC5:12.0.0>) wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird NEPS vom Leibniz-Institut für Bildungsverläufe e.V. (LIfBi) an der Otto-Friedrich-Universität Bamberg in Kooperation mit einem deutschlandweiten Netzwerk weitergeführt.

8 Es konnten aufgrund begrenzter Ressourcen und der komplexen bundesweiten Erhebungslogistik nicht alle 1701 potenzielle Fälle getestet werden. Es wurden durch die Ziehung von 15 regionalen Clustern 601 Personen für die Kompetenztestung ausgewählt. Es konnten 342 Testungen realisiert werden, von denen 338 auswertbare Ergebnisse lieferten (Lauterbach, 2015).

nale Analysestichprobe umfasste $n = 118$ Personen (Geburtsjahr: $M = 1989$, $SD = 2.2$). Davon gaben 97 an, ihren Master an einer Universität zu machen.⁹ Neun Personen hatten einen Migrationshintergrund.¹⁰ Ausländische Studierende sind in der Stichprobe deutlich unterrepräsentiert (1% im Vergleich zu 14% in der Erstsemesterpopulation¹¹). Der Frauenanteil in der Stichprobe liegt mit 52% dagegen nahe am Populationswert von 51%). Eine Subgruppenanalyse wäre gerade im Fall von Hochschulwechsellern eine Möglichkeit, um Notengebungsunterschiede näher zu beleuchten (vgl. Gaens, 2018; Weimar et al., 2017); in der vorliegenden Analysestichprobe umfasst die Teilstichprobe jedoch deutlich unter 30 Fälle, weshalb die Analysen sich auf die Gesamtstichprobe beschränken. Dieses Vorgehen liefert eine Validitätsschätzung im Sinne einer Studierendenauswahl, bei der Bachelornoten von anderen Hochschulen gleich gewertet werden wie Bachelornoten der eigenen Hochschule.

Als Prädiktorvariablen wurden die Kompetenztestleistung und die Bachelorabschlussnote verwendet. Demografische Angaben wie Geschlecht und Alter wurden beim Paneleintritt erfragt. Die erreichten Durchschnitts- und Abschlussnoten im Studium wurden laufend in weiteren Befragungswellen erhoben. Die Kompetenztestung fand im ersten Quartal 2014 statt. Knapp die Hälfte der Studierenden in der Analysestichprobe hatte in diesem Jahr auch ihren Bachelorabschluss erworben (Bachelorabschluss 2013: 45%, 2014: 43%, 2015: 8%, 2016: 3%).

Als Erfolgskriterium wurde die Durchschnittsnote im Masterstudium verwendet. Lag (noch) keine Abschlussnote vor ($n = 51$), wurde die zuletzt berichtete Durchschnittsnote im laufenden Masterstudium verwendet. Die Erhebung der Variablen im Zeitverlauf wird in Abbildung 1 schematisch illustriert. Die durchschnittliche Zeit zwischen dem Kompetenztest und der Kriteriumsleistung betrug im Schnitt $M = 2.3$ Jahre ($SD = 0.7$ Jahre). Die unterschiedlichen Zeitabstände ergeben sich daraus, dass zwar alle Personen im gleichen Zeitraum getestet wurden (und gleichzeitig ihr Bachelorstudium aufnahmen), jedoch nicht alle zur gleichen Zeit ihre Abschlüsse machten.

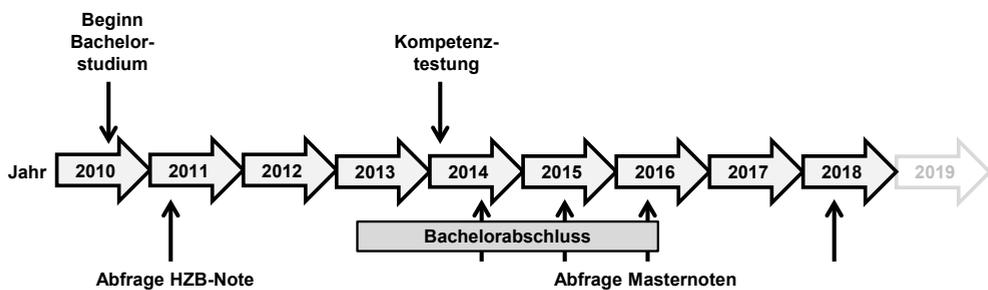


Abbildung 1: Zeitschiene der verwendeten Erhebungszeitpunkte im Studienverlauf (HZB = Hochschulzugangsberechtigung)

- 9 Bedingt durch das Stichprobendesign sind in den Wirtschaftswissenschaften Studierende an Universitäten überrepräsentiert (vgl. Hochschulartanteile bei Kieser & Küpper, 2001).
- 10 Ein Migrationshintergrund lag vor, wenn die Zielperson im Ausland geboren wurde, keine deutsche Staatsbürgerschaft besaß oder eine nichtdeutsche Muttersprache hatte.
- 11 Erstsemesterpopulation = Studienanfängerinnen und -anfänger in Betriebswirtschaftslehre im ersten Hochschulsesemester im Wintersemester 2010/11 (Statistisches Bundesamt, 2011, Tabelle ZUS-08).

2.2 Instrumente

Der wirtschaftswissenschaftliche Kompetenztest basiert auf übersetzten und adaptierten Aufgaben des mexikanischen EGEL (Exámenes Generales de Egreso de la Licenciatura; Centro Nacional de Evaluación para la Educación Superior, 2011) und des US-amerikanischen TUCE (Test of Understanding in College Economics; Walstad, 2007). Die Tests erfassen für das Studium relevante Lerninhalte. Im Rahmen von Vorstudien und Expertenratings wurden 36 Aufgaben ausgewählt, die die Bereiche Marketing, Organisation, Finanzierung, Rechnungswesen, Mikroökonomie und Makroökonomie mit jeweils sechs Multiple-Choice-Aufgaben abdecken. Im direkten Vergleich mit dem TM-WISO (vgl. Stegt & Bergholz, 2018) zielt der eingesetzte Kompetenztest stärker auf wirtschaftswissenschaftliche Kenntnisse aus dem Erststudium ab als auf vorwissensfreie, fachrelevante Fähigkeiten. Die papierbasierten Einzeltestungen wurden durch eine Interviewerin bzw. einen Interviewer administriert. Die Bearbeitungszeit betrug im Schnitt 42 Minuten ($SD = 6.9$ Minuten). Nach Ausschluss von zwei Items wegen *differential item functioning* (DIF) wurden die verbleibenden 34 Aufgaben Rasch-skaliert. Für die Schätzungen der Kompetenz der Studierenden wurden gewichtete Likelihood-Schätzungen (*weighted likelihood estimates*, WLE) als Personenparameter verwendet ($M = 0$, $SD = 0.88$, $n = 338$). Die WLE-Reliabilität lag bei .77 (für Details zur Entwicklung, Durchführung und Skalierung des Tests s. Lauterbach, 2015).

2.3 Auswertungsplan

Zur Bestimmung der Validität der einzelnen Prädiktorvariablen hinsichtlich der Studienleistung im Master wurden bivariate Korrelationskoeffizienten berechnet (Forschungsfrage 1). Die operative Validität wurde durch die Korrelation nach der Korrektur für die Unreliabilität des Kriteriums (nicht aber des Prädiktors) sowie für *restriction of range* geschätzt. Als Kriteriumsreliabilität der Masternoten wurde $r = .93$ angenommen (Beatty, Walmsley, Sackett, Kuncel & Koch, 2015).

Die inkrementelle Validität wurde mit einer Reihe von linearen Regressionen zur Vorhersage der Masterstudienleistung bestimmt. Im ersten Schritt wurde die Bachelorabschlussnote als Prädiktor verwendet. Anschließend wurden Bachelorabschlussnote und Testleistung als Prädiktoren verwendet und der Zuwachs an erklärter Varianz (ΔR^2) auf Signifikanz getestet (Forschungsfrage 2). Die Prädiktorvariablen gingen zentriert in die Regressionen ein. Ferner wurde auch die inkrementelle Varianz der Testleistung über die Prädiktion durch die (beim Paneleintritt erfragte) Hochschulzugangsberechtigungsnote (HZB) sowie über die Prädiktion durch HZB- und Bachelornote hinaus berechnet.

Die differenzielle Prognose wurde mit einer moderierten multiplen Regression getestet (Cleary-Modell, vgl. Cleary, 1968; Fischer et al., 2013). Neben der Testleistung als Prädiktor wurde das Geschlecht (männlich = 0, weiblich = 1) sowie der dazugehörige Interaktionsterm mit aufgenommen (Forschungsfrage 3). Die Testleistung wurde auch für diese Regression zentriert, um die Interpretation des Intercepts als Gruppenunterschied bei mittlerem Leistungsniveau zu ermöglichen. Der Interaktionsterm und der Gruppenfaktor geben in solchen Modellen die gruppenspezifische Verzerrung bei der Studienerfolgsprognose an.

Für das Datenmanagement und die Auswertung wurde Stata 13.1 verwendet.¹² Als Signifikanzniveau wurde $\alpha = .05$ gewählt.

3 Ergebnisse

Die deskriptive Statistik der interessierenden Variablen wird zusammen mit den dazugehörigen Korrelationen in Tabelle 1 aufgelistet. Die erste Forschungsfrage betrifft die Korrelation zwischen Testleistung und Masternotenschnitt. Diese betrug $r = -.23$ ($p < .001$). Das negative Vorzeichen des Validitätskoeffizienten ergibt sich dadurch, dass höhere Testwerte stärkere Leistungen anzeigen, während bei den Noten niedrigere Werte ein besseres Abschneiden angeben. Der Effekt bleibt auch nach der Korrektur für *restriction of range*¹³ und für die Kriteriumsunreliabilität klein mit einer operativen Validität von $\rho = -.24$.

Bezüglich der zweiten Forschungsfrage gab es keine signifikante inkrementelle Validität des wirtschaftswissenschaftlichen Kompetenztests über die Bachelorabschlussnote hinaus; der Validitätszuwachs betrug praktisch null ($\Delta R^2 < .01$, $F(1, 113) = 0.00$, $p = .96$).¹⁴

Tabelle 1: Bivariate Korrelationen und deskriptive Statistiken der Variablen

	2.	3.	4.	5.	<i>n</i>	<i>M</i>	<i>SD</i>
1. HZB-Note	.55*	-.40*	.49*	-.08	108	2.17	0.58
2. Bachelornote		-.37*	.55*	-.28*	116	2.10	0.48
3. Testleistung (WLE)			-.23*	-.22*	118	0.38	0.92
4. Masternote				-.15	118	1.85	0.43
5. Geschlecht (weiblich)					118	0.52	0.50

* $p < .05$, HZB = Hochschulzugangsberechtigung, WLE = weighted likelihood estimation.

Forschungsfrage 3 adressiert geschlechtsspezifische Verzerrungen bei der differenziellen Prognose. Die Regressionsgeraden zur Vorhersage der Masternoten sind in Abbildung 2 getrennt für Frauen und Männer dargestellt. Die beiden geschlechtsspezifischen Steigungen unterschieden sich nicht signifikant ($p = .11$; vgl. Tabelle 2). Der signifikante Intercept-Unterschied hingegen zeigt, dass Frauen mit einer durchschnittlichen Testleistung im Schnitt um $\Delta b = -0.17$ Punkte bessere Masternoten hatten als Männer mit derselben (durchschnittlichen) Testleistung.

12 Die verwendeten Auswertungsbefehle und die Logdatei der Auswertung können unter <https://osf.io/3dejj/> eingesehen werden.

13 Tatsächlich führte die Korrektur für *restriction of range* zu einer leichten Validitätsverringerng, da die Streuung der Testwerte in der Gesamttestung ($SD = 0.88$) etwas kleiner ausfiel als in der Analytestichprobe ($SD = 0.92$).

14 Die Testleistung zeigte auch weder über die HZB-Note hinaus noch über die gemeinsam berücksichtigte Bachelor- und HZB-Note hinaus inkrementelle Validität (jeweils $\Delta R^2 < .01$, $p > .48$).

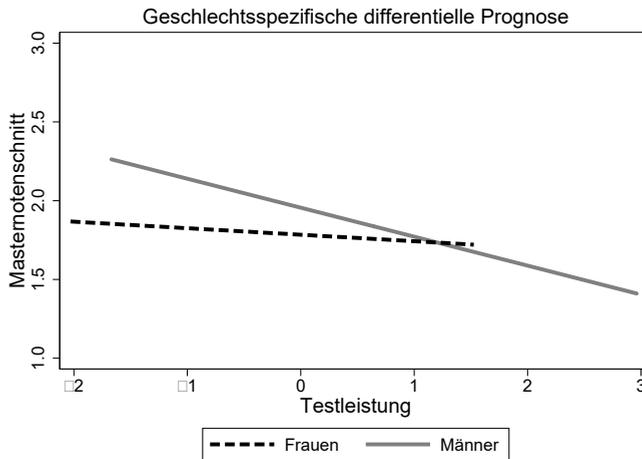


Abbildung 2: Geschlechtsspezifische Regressionsgeraden zur Prognose der Masternoten durch die Kompetenztestleistung. Die Geraden sind für die Spannweite der jeweiligen Testleistungen eingezeichnet.

Tabelle 2: Moderierte multiple Regression zur Vorhersage der Masternote gemäß Cleary (1968) mit Test auf gleiche Steigungen (Interaktionsterm) und gleiche Intercepts (Geschlechtsterm)

Prädiktor	b (SE)
Intercept	1.95* (0.06)
Testleistung ^{zcn}	-0.18* (0.09)
Geschlecht (männlich = 0, weiblich = 1)	-0.17* (0.08)
Testleistung ^{zcn} × Geschlecht	0.14 (0.09)

* $p < .05$, $n = 118$, ^{zcn} = Variable wurde zentriert, $R^2 = .11^*$

4 Diskussion

Anhand der Studierenden-Kohorte des NEPS wurde untersucht, welche prognostische Validität ein wirtschaftswissenschaftlicher Kompetenztest hinsichtlich des Erfolgs im Masterstudium aufweist. Die Testleistung, die kurz vor bzw. nach Ende des Bachelorstudiums erhoben wurde, korrelierte niedrig mit den Noten im Masterstudium. Der Test ist somit weniger valide als wirtschaftswissenschaftliche Studierfähigkeitstests im Bachelor- oder Grundstudium (vgl. Hell et al., 2007; Schult et al., 2019). Aufgrund der fortgeschrittenen Studierfahrung im Master sowie der möglichen Selbstselektion der Studierenden erklärt der Kompetenztest vermutlich nicht mehr Studienleistungsvarianz.

Eine inkrementelle Validität der Kompetenztestleistung über die Bachelor-Noten hinaus konnte nicht gefunden werden. Die Bachelor-Noten allein erklärten 31% der Masterleistungsvarianz ($R^2 = (.5548)^2 = .3078$). Die Interkorrelationen zwischen den Prädiktorva-

riablen deuten darauf hin, dass überlappende kognitive Fähigkeiten von den Noten und dem Test erfasst werden. Der eingesetzte Test war nicht nur weniger valide, sondern auch kürzer als übliche Studierfähigkeitstests (vgl. z.B. 72 gewertete TM-WISO-Items bei Stegt & Bergholz, 2018). Weiterhin erfasst der eingesetzte Kompetenztest stärker Kenntnisse als kognitive Fähigkeiten. Die hohe Korrelation zwischen den Bachelornoten und den Testwerten deutet auf eine Nähe der Testinhalte zum Bachelorlernstoff hin. Die höheren Validitäten, die sich für den TM-WISO finden (Stegt & Bergholz, 2018), stammen möglicherweise auch daher, dass dieses Testverfahren weniger bisherige Kenntnisse und wesentlich stärker fachrelevante kognitive Fähigkeiten erfasst.

Frauen schnitten bei gleicher Testleistung im Studium etwas besser ab als Männer. Der Geschlechtsunterschied bei der Vorhersage von 0.17 Notenpunkten liegt knapp unter dem metaanalytischen Unterschied von 0.24 Notenpunkten (Fischer et al., 2013). Die Ursache für die Unterschätzung der Studienleistung von Frauen dürfte nicht bei einzelnen Aufgaben liegen, da der hier verwendete Test bereits bei der Erstellung auf Geschlechts-DIF untersucht wurde, wobei sich keine Verzerrungen zeigten (Lauterbach, 2015). Für zukünftige Studien größerer Kohorten wäre die Fairness in Abhängigkeit des Migrationshintergrunds zu untersuchen, da sich hier in den Wirtschaftswissenschaften teils deutliche Testleistungsunterschiede zeigen (Happ, Förster & Beck, 2018).

Aus Sicht der besten Vorhersage fällt die Beurteilung der Ergebnisse nicht schwer. Die Bachelorabschlussnote liefert bereits eine gute Vorhersage des Studienerfolgs im Master, ungeachtet der möglicherweise eingeschränkten Vergleichbarkeit der Abschlüsse bei einem Wechsel der Hochschule zwischen Bachelor und Master. Die Kompetenztestleistung hingegen bietet im vorliegenden Fall keinen Mehrwert bezüglich der Erfolgsprognose. Da die Bachelorabschlussnoten einfach erfassbar sind, während ein Auswahltest logistisch und ökonomisch aufwändig ist, unterstützen unsere Befunde die bisherige Zulassungspraxis an deutschen Hochschulen, bei der Noten aus dem Erststudium eine zentrale Rolle spielen und Zulassungstests nur marginal vertreten sind. Es gibt jedoch trotzdem zumindest eine mögliche Relevanz der gefundenen Testvalidität: Die Benotungsunterschiede beim Abitur haben dazu geführt, dass Abiturnoten in Deutschland aktuell nicht mehr allein als Kriterium für die Bachelorstudierendenauswahl verwendet werden dürfen. Sollten Hochschulen (oder gar zukünftige Hochschulgesetze) die Notengebung im Bachelor zu wenig vergleichbar finden, braucht man alternative (nicht zusätzliche) Zulassungskriterien. Dann sind Studierfähigkeitstests selbst ohne inkrementelle Validität eine Option. Freilich ist der hier untersuchte Test in seiner jetzigen Form und Länge keine zufriedenstellende Lösung. Darüber hinaus muss berücksichtigt werden, dass es möglicherweise eine Benachteiligung weiblicher Kandidatinnen gibt und dass sozio-ökonomisch schwache Personen durch (hohe) Kosten für die Teilnahme an Zulassungsverfahren von vorneherein vor einem weiterführenden Studium abgeschreckt werden. Die Verwendung von Aufgaben mit offenem Aufgabenformat kann der geschlechtsspezifischen differenziellen Prognose möglicherweise entgegenwirken (Bridgeman & Lewis, 1994). Allgemein müssen die Testfairness-Richtlinien beachtet werden (AERA, APA & NCME, 2014), damit das Testverfahren für alle Teilnehmenden transparent abläuft und alle Personengruppen gleichermaßen Vorbereitungsmöglichkeiten bekommen. Die Testteilnahme im NEPS weicht hier von echten Auswahlverfahren durch die Freiwilligkeit und den Testzweck (Forschung statt Selektion) ab.

4.1 Limitationen

Obwohl das NEPS eine umfangreich angelegte Panelerhebung ist, reicht der Umfang der Analysestichprobe leider nicht für weiterführende Analysen bezüglich Hochschulart, Hochschulstandort, Migrationshintergrund usw. aus. Als Erfolgskriterium wurden ausschließlich Noten betrachtet, da weitere Dimensionen des Studienerfolgs nur für kleine Fallzahlen vorliegen. So braucht es bei Langzeitstudierenden noch weitere Jahre, bis der Studienabschluss (oder gegebenenfalls auch der Studienabbruch) eindeutig festgestellt werden kann. Die Studienzufriedenheit wurde nur bis zur achten Welle erfasst, in der viele Personen in der Analysestichprobe ihr Masterstudium noch nicht oder gerade erst begonnen hatten.

Studierfähigkeitstests werden nur in Ausnahmefällen als alleiniges Auswahlkriterium herangezogen. Die im Rahmen der dritten Forschungsfrage explorierten Vorhersagemodelle mit mehreren Prädiktoren berücksichtigen dies. Die Relevanz weiterer Variablen wie zum Beispiel intrinsisches Interesse oder Leistungsmotivation (Fischer et al., 2015; Richardson, Abraham & Bond, 2012) wurde jedoch nicht näher betrachtet. Wenn diese Aspekte in zukünftigen Studien aufgegriffen werden, sollte neben der prognostischen Validität jedenfalls die Möglichkeit der bewussten Verfälschung im Bewerbungskontext kritisch reflektiert werden. Interessentests können beispielsweise im Rahmen von beratenden Self-Assessments ein sinnvoller Bestandteil von Auswahlverfahren sein (Hell, 2009).

Die vorliegende Studie beschränkt sich auf eine Fachrichtung und dort auf einen Kompetenztest. Sie hat somit eher praxeologischen und exemplarischen Charakter. Eine Übertragung des Befundmusters auf andere Studiengänge ist teilweise plausibel, da eine fachspezifische Fähigkeitsmessung, die auf einer Anforderungsanalyse für den jeweiligen Studiengang basiert, im Regelfall valide Prognosen erlaubt. Andererseits gibt es in den meisten Fächern (mit der Ausnahme von Maschinenbau) eine homogenere Verteilung der Bachelornoten im guten bis sehr guten Bereich (Wissenschaftsrat, 2012, S. 55). Im Fall von Noteninflation ist eine niedrigere Validität der Bachelornoten und damit einhergehend eine größere inkrementelle Validität von Studierfähigkeitstests zu erwarten. Wenn sich US-amerikanische Befunde übertragen lassen, dürfte die prognostische Validität in anderen Fächern tendenziell höher ausfallen (vgl. Kuncel & Hezlett, 2007).

Die vorliegende Studie unterliegt den Einschränkungen, die sich bei der Analyse von Sekundärdaten ergeben. Zusätzliche Variablen lassen sich nicht nachträglich erheben. Personen verlassen trotz entsprechender Bemühungen die Panelstudie, was freilich auch eine Herausforderung für längsschnittliche Primärstudien ist. Zudem wurde im konkreten Fall lediglich ein Studierfähigkeitstest eingesetzt. Dem steht der Vorteil gegenüber, dass die Daten über einen langen Zeitraum hinweg an einer bundesweiten Stichprobe erhoben wurden und dadurch die Bildungslaufbahn der Befragten umfangreich abgebildet wird. Das NEPS ist selbstverständlich nicht die einzige geeignete Datenquelle für Validitätsstudien. Die administrativen Register von Hochschulen enthalten genügend Daten, um mit größeren Fallzahlen für eine breitere Auswahl von Studienfächern die Validität der eingesetzten Auswahlverfahren zu evaluieren. Derartige Forschungsarbeiten sucht man jedoch vergeblich. Vermutlich liegt das zum einen an den datenschutzrechtlichen Bestimmungen bezüglich personenbezogener Daten; zum anderen scheint es Vorbehalte gegenüber solchen Untersuchungen zu geben, da die Hochschule im Fall von Nullbefunden unter Rechtfertigungsdruck geraten würde und das Auswahlverfahren anpassen sollte.

4.2 Fazit

Da es zur Validität von Auswahlverfahren für Masterstudiengänge im deutschsprachigen Raum bislang kaum empirische Untersuchungen gibt, liefert die vorliegende Studie wichtige erste Erkenntnisse für Hochschulen, die fachspezifische Kompetenztests einsetzen bzw. einsetzen wollen. Der wirtschaftswissenschaftliche Kompetenztest, der im NEPS eingesetzt wird (Lauterbach, 2015), besitzt prognostische Validität hinsichtlich der Studienleistung im Master. Die Vorhersagekraft fällt niedriger aus als die mittlere Validität für fachspezifische Studierfähigkeitstests in den Wirtschaftswissenschaften (vgl. Schult et al., 2019). Anhand der NEPS-Daten konnte also gezeigt werden, dass der untersuchte Test fachlicher Kompetenzen in konsekutiven Masterstudiengängen zwar eine valide Erfolgsprognose liefert, jedoch in der aktuellen Form als (alleiniges) Instrument für die Studierendenauswahl nicht geeignet ist.

Literatur

- American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME] (2014). *Standards for educational and psychological testing* (5th ed.). Washington, DC: American Educational Research Association.
- Anger, C. & Schmidt, J. (2010). Gender Pay Gap: Gesamtwirtschaftliche Evidenz und regionale Unterschiede. *IW-Trends*, 37, 1–15.
- Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R. & Koch, A. J. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice*, 34(4), 31–40.
<https://doi.org/10.1111/emip.12096>
- Blossfeld, H.-P., Roßbach, H.-G. & von Maurice, J. (Hrsg.) (2011). Education as a lifelong process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft: Sonderheft 14*.
- Bericht der Bundesregierung (2015). *Umsetzung des Bologna-Prozesses 2012–2015 in Deutschland*. Verfügbar unter:
https://www.bmbf.de/files/Bericht_der_Bundesregierung_zur_Umsetzung_des_Bologna-Prozesses_2012-2015.pdf
- Borowski, S.; Schauburger, M. & Weimar, D. (2018). Zugangs- und Zulassungsvoraussetzungen an deutschen Hochschulen: Eine deskriptive Bestandsaufnahme. *Hochschulmanagement*, 13(2), 41–48.
- Bridgeman, B. & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement*, 31, 37–50.
<https://doi.org/10.1111/j.1745-3984.1994.tb00433.x>
- Bundesverfassungsgericht (2017). Urteil des Ersten Senats vom 19. Dezember 2017 – 1 BvL 3/14 – Rn. (1–253). Verfügbar unter: http://www.bverfg.de/e/ls20171219_1bv1000314.html
- Centro Nacional de Evaluación para la Educación Superior (2011). *Exámenes Generales para el Egreso de la Licenciatura*. Mexico City: DF.
- Cleary, T. A. (1968). Prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115–124. <https://doi.org/10.1111/j.1745-3984.1968.tb00613.x>
- Fischer, F. T., Schult, J. & Hell, B. (2013). Sex-specific differential prediction of college admission tests: A meta-analysis. *Journal of Educational Psychology*, 105, 478–488.
<https://doi.org/10.1037/a0031956>
- Fischer, F., Schult, J. & Hell, B. (2015). Unterschätzung der Studienleistungen von Frauen durch Studierfähigkeitstests: Erklärbar durch Persönlichkeitseigenschaften? *Diagnostica*, 61, 34–46.
<https://doi.org/10.1026/0012-1924/a000120>

- Gaens, T. (2018). *Der Einfluss leistungskonformer und leistungsexterner Prüfungsbedingungen auf die Notengebung an deutschen Hochschulen: Eine empirische Untersuchung der langfristigen Entwicklung von Examensnoten*. Unveröffentlichte Dissertation, Europa-Universität Flensburg. Verfügbar unter: <https://nbn-resolving.org/urn:nbn:de:01168-ssoar-56104-3>
- Happ, R., Förster, M. & Beck, K. (2018). Eingangsvoraussetzungen von Studierenden der Wirtschaftswissenschaften mit und ohne Migrationshintergrund. *Zeitschrift für empirische Hochschulforschung*, 2(1), 6–22. <https://doi.org/10.3224/zehf.v2i1.01>
- Heine, C., Briedis, K., Didi, H.-J., Haase, K. & Trost, G. (2006). *Bestandsaufnahme von Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern: Eine Bestandsaufnahme* (HIS Kurzinformation A 3/2006). Hannover: Hochschul-Informations-System.
- Hell, B., Trapmann, S. & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. *Empirische Pädagogik*, 21, 251–270.
- Hell, B., Trapmann, S. & Schuler, H. (2008). Synopse der Hohenheimer Metaanalysen zur Prognostizierbarkeit des Studienerfolgs und Implikationen für die Auswahl- und Beratungspraxis. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 43–54). Göttingen: Hogrefe.
- Hell, B. (2009). Selbsttests zur Studienorientierung: Nützliche Vielfalt oder unnützer Wildwuchs? In G. Rudinger & K. Hörsch (Hrsg.), *Self-Assessment an Hochschulen: Von der Studienfachwahl zur Profilbildung* (S. 9–20). Bonn: Bonn University Press.
- Janke, S. & Dickhäuser, O. (2018). Zur prognostischen Güte von Zulassungskriterien im Psychologiestudium für Studienerfolgsindikatoren. *Psychologische Rundschau*, 69, 160–168. <https://doi.org/10.1026/0033-3042/a000383>
- Kieser, A. & Küpper, H.-U. (2001). Ein Konzept zur Einführung von Bachelor- und Master-Studiengängen an wirtschaftswissenschaftlichen Fakultäten. *Beiträge zur Hochschulforschung*, 23(4), 6–26.
- Kouwenaar, K. (2015). Changing paradigms: Towards competency-assessment in admission to master's programmes in Europe: A Review. *Tuning Journal for Higher Education*, 3, 99–135. [https://doi.org/10.18543/tjhe-3\(1\)-2015pp99-135](https://doi.org/10.18543/tjhe-3(1)-2015pp99-135)
- Kultusministerkonferenz (2011). *Situation im Masterbereich und statistische Erfassung von Masterstudienplätzen: Eine Diskussionsgrundlage*. Beschluss der Kultusministerkonferenz vom 29.04.2011.
- Kuncel, N. R., Credé, M. & Thomas, L. L. (2007). A meta-analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and undergraduate grade point average (UGPA) for graduate student academic performance. *Academy of Management Learning & Education*, 6, 51–68. <https://doi.org/10.5465/amle.2007.24401702>
- Kuncel, N. R. & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315, 1080–1081. <https://doi.org/10.1126/science.1136618>
- Lauterbach, O. (2015). *Erfassung wirtschaftswissenschaftlicher Fachkompetenzen von Studierenden in Startkohorte 5 des Nationalen Bildungspanels – Technischer Bericht* (Aktualisierung NEPS Working Paper No. 51). Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.
- Madani, R., Melzer, B. & Müller, M. (2013). *Prognose des Studienerfolgs im MBA-Studium an der Universität Potsdam* (Working Paper). Servicestelle zur Förderung lebenslangen Lernens an der Universität Potsdam.
- Richardson, M., Abraham, C. & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138, 353–387. <https://doi.org/10.1037/a0026838>
- Schuler, H. & Schult, J. (2018). Prädiktoren von Studien- und Berufserfolg. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5. Aufl.) (S. 645–653). Weinheim: Beltz.

- Schult, J. (2012). Prädiktoren des Berufserfolgs von Hochschulabsolventen: Befunde aus dem Sozio-Ökonomischen Panel. *Wirtschaftspsychologie* 14(4), 82–91.
- Schult, J., Hofmann, A. & Stegt, S. J. (2019). Leisten fachspezifische Studierfähigkeitstests im deutschsprachigen Raum eine valide Studienerfolgsprognose? Ein metaanalytisches Update. *Zeitschrift für Entwicklungs- und Pädagogische Psychologie*, 51, 16–30. <https://doi.org/10.1026/0049-8637/a000204>
- Schwager, I. T., Hülshager, U. R., Bridgeman, B. & Lang, J. W. (2015). Graduate Student Selection: Graduate record examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European University. *International Journal of Selection and Assessment*, 23, 71–79. <https://doi.org/10.1111/ijsa.12096>
- Schwibbe, A., Lackamp, J., Knorr, M., Hissbach, J., Kadmon, M. & Hampe, W. (2018). Medizinstudierendenauswahl in Deutschland. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 61, 178–186. <https://doi.org/10.1007/s00103-017-2670-2>
- Statistisches Bundesamt (2011). Deutsche und ausländische Studierende im Wintersemester 2010/2011 in den 20 am stärksten besetzten Studienfächern. Fachserie 11, Reihe 4.1, WS 2010/2011. Verfügbar unter: https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00005831
- Stegt, S. J. & Bergholz, L. (2018). Vorhersage des Studienerfolgs in konsekutiven Masterstudiengängen mithilfe eines kognitiven Eignungstests. *Hochschulmanagement*, 13(4), 101–107.
- Steinwede, J. & Aust, F. (2012). *NEPS Startkohorte 5 – CATI-Haupterhebung Herbst 2010 B52 (Methodenbericht)*. Bonn: infas.
- Stemler, S. E. (2012). What should university admissions tests predict? *Educational Psychologist*, 47, 5–17. <https://doi.org/10.1080/00461520.2011.611444>
- Strahan, S. & Credé, M. (2015). Satisfaction with college: Re-examining its structure and its relationships with the intent to remain in college and academic performance. *Journal of College Student Retention: Research, Theory & Practice*, 16(4), 537–561. <https://doi.org/10.2190/CS.16.4.d>
- Troche, S., Mosimann, M. & Rammsayer, T. (2014). Die Vorhersage des Studienerfolgs im Masterstudiengang Psychologie durch Schul- und Bachelorstudienleistungen. *Beiträge zur Hochschulforschung*, 36(1), 30–45.
- Walstad, W. B. (2007). *Test of understanding in college economics*. New York: National Council on Economic Education.
- Weimar, D., Schauburger, M., Borowski, S. & Prinz, J. (2017). Selektionseffekte von Zulassungstests an Universitäten. *Betriebswirtschaftliche Forschung und Praxis*, 69, 694–716.
- Wissenschaftsrat (2012). *Prüfungsnoten an Hochschulen im Prüfungsjahr 2010 Arbeitsbericht mit einem Wissenschaftspolitischen Kommentar des Wissenschaftsrates* (Drucksache 2627-12). Hamburg: Wissenschaftsrat.
- Zimmermann, J., Brodersen, K. H., Heinemann, H. R. & Buhmann, J. M. (2015). A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance. *Journal of Educational Data Mining*, 7, 151–176.

Kontakt:

Dr. Johannes Schult
Institut für Bildungsanalysen Baden-Württemberg
Heilbronner Straße 172
70191 Stuttgart
E-Mail: jutze@jutze.com

Akzeptanz und Nutzung psychologischer Verfahren zur Auswahl von Promovierenden

Max Knaut, Matthias Ziegler

Zusammenfassung: Die Auswahl von Promovierenden ist für Wissenschaftsorganisationen von hoher Relevanz. Dennoch existieren bisher kaum Studien zu diesen Prozessen. Die Arbeit adressiert diese Lücke, indem sie Akzeptanzurteile für Auswahlverfahren bei $N = 170$ potenziell Bewerbenden erfragte. Frei geführte Auswahlgespräche, berufliche Erfahrungen, Arbeitsproben und strukturierte Auswahlgespräche erhielten die höchsten Akzeptanzwerte. Persönlichkeitseigenschaften der Befragten und vorher erbrachte Leistungen in den Verfahren hatten keinen signifikanten Einfluss auf die Akzeptanzurteile. Welche Verfahren eingesetzt werden, hatte Einfluss auf eine mögliche Bewerbungsabsicht. Zusätzlich wurde bei $N = 164$ Promovierenden erhoben, wie diese tatsächlich ausgewählt wurden. Promotionsstellen wurden am häufigsten durch frei geführte Auswahlgespräche und persönliche Kontakte besetzt. Es werden Implikationen für die Gestaltung von Auswahlverfahren von Promovierenden mit Blick auf Akzeptanz, Validität und Fairness diskutiert.

Schlüsselwörter: Akzeptanz von Auswahlverfahren, Promovierende, Wissenschaftsorganisationen, Validität, Fairness

Acceptance and application of psychological selection procedures for doctoral candidates

Abstract: Selection processes for doctoral candidates are of great relevance for scientific organizations. However, there is only a limited number of studies on this subject. This study addresses this gap by asking $N = 170$ potential applicants for acceptance ratings of different selection methods. Unstructured interviews, professional experience, work samples and structured interviews had the highest ratings in acceptance. Participants' personality traits and previous performance in the selection methods had no impact on acceptance ratings. The selection method used had an impact on the decision whether or not someone applied for a doctoral position. In addition, $N = 164$ doctoral candidates were asked how doctoral candidates were actually selected. Unstructured interviews and personal contacts were most often used to fill doctoral positions. Implications for doctoral candidates' selection processes are discussed with respect to acceptance, validity, and fairness.

Keywords: acceptance of selection processes, doctoral candidates, scientific organizations, validity, fairness

1 Einleitung

Universitäten und andere Wissenschaftsorganisationen befinden sich in immer größer werdenden Wettbewerben. Zum einen müssen sie ihre auf Empirie basierenden Forschungsergebnisse gegen emotional aufgeladene Mythen, Irrglauben und Missinformationen verteidigen, damit deren Verbreiterinnen und Verbreiter nicht die Deutungshoheit für Teile der Gesellschaft gewinnen können. Gleichzeitig geht es für Wissenschaftsorganisationen darum, im internationalen und nationalen Vergleich in Forschung und Lehre zu den Besten zu gehören. Um für beide Wettkämpfe gut aufgestellt zu sein, ringen Wissenschaftsorganisationen um Ressourcen und vor allem um die besten Köpfe. Mit den besten Forscherinnen und Forschern in den eigenen Reihen können Wissenschaftsorganisationen erfolgreich Erkenntnisse generieren und diese effektiv verbreiten und so gegen Missinformationen vorgehen. Ebenfalls können sie mit exzellentem wissenschaftlichem Output die Reputation weiter steigern, Gelder akquirieren und attraktiv für weitere Top-Wissenschaftlerinnen und Top-Wissenschaftler bleiben. Ähnliches gilt auch für die Qualität der Lehre. Lehrende sollten zum Profil und zur Lehrkultur einer Hochschule passen und die Weiterentwicklung des Curriculums vorantreiben (Wissenschaftsrat, 2017). So sind Hochschulen im Wettbewerb um kompetente Studierende gut aufgestellt und können mit guter, forschungsnaher Lehre für eine hochwertige wissenschaftliche Qualifizierung der Studierenden sorgen. Um die besten Köpfe zu gewinnen und für die genannten Herausforderungen gut aufgestellt zu sein, ist die faire und professionelle Auswahl von Wissenschaftlerinnen und Wissenschaftlern ein wichtiger Baustein (Braun, Hentschel, Peus & Frey, 2015). Dieser Auswahlprozess betrifft alle Positionen innerhalb von Wissenschaftsorganisationen und beginnt mit der Auswahl von Promovierenden. Die Auswahl auf dieser Karriereebene ist von großer Bedeutung, da die Promotion den Einstieg in die Wissenschaft darstellt. Für viele ist sie der Beginn von Wissenschaftskarrieren oder, um mit den Worten von Abele-Brehm und Bühner (2016) die Relevanz der Personalauswahl in der Wissenschaft zu untermauern: „Die Entwicklung einer Wissenschaft steht und fällt mit ihren Akteuren“ (S. 250).

Trotz dieser Bedeutsamkeit für Wissenschaftsorganisationen existieren kaum Studien über die Auswahlverfahren von Promovierenden. Es gibt keine empirischen Belege darüber, mit welchen Verfahren Promovierende für ihre Stellen und Stipendien ausgesucht werden. Auch fehlt es an Befunden, welche Auswahlverfahren für Promotionsstellen von potenziell Bewerbenden akzeptiert werden. Daher adressieren wir diese Forschungslücken mit dieser Arbeit.

2 Herausforderungen bei der Auswahl von Promovierenden

Im Jahr 2017 legten über 200,000 Studierende in Deutschland eine universitäre Abschlussprüfung ab, die es ihnen erlaubt, eine Promotion zu beginnen (Statistisches Bundesamt, 2017). Hinzu kommen Studierende aus dem Ausland, welche die formalen Kriterien für eine Promotion erfüllen. Wissenschaftsorganisationen, die Promotionsstellen oder Stipendien zu vergeben haben, sehen sich mit mindestens zwei Herausforderungen bei der Auswahl von Promovierenden konfrontiert:

- Wissenschaftsorganisationen sollten von leistungsstarken Absolventinnen und Absolventen als attraktiver Arbeitgeber wahrgenommen werden. Die Konkurrenz für Wissenschaftsorganisationen ist dabei groß. Innerhalb der Wissenschaft befinden sie sich im Wettkampf mit nationalen und internationalen, mit privaten und öffentlichen Universitäten. Außerhalb der Wissenschaft müssen sie mit Unternehmen aus der Privatwirtschaft konkurrieren. Diese sind häufig in der Lage, mehr zu bezahlen als die an Tarifverträge gebundenen Wissenschaftsorganisationen (Fabian & Briedis, 2009; Falk, 2010). Dadurch, dass Absolventinnen und Absolventen entscheiden, wo sie sich bewerben oder wo sie sich weiterqualifizieren möchten, selektieren sie selbst. Dieser Selbstselektionsprozess der Bewerbenden setzt ein, bevor die Organisationen mit ihrer Auswahl beginnen können (Ryan & Ployhardt, 2000). Die erste Herausforderung ist es also, als Wissenschaftsorganisation attraktiv für Absolventinnen und Absolventen zu sein. Hierfür haben Wissenschaftsorganisationen nur ein kleines Zeitfenster, da es Quereinstiege in die Wissenschaft kaum gibt: Über 80% der Promovierenden beginnen ihre Promotion innerhalb von 18 Monaten nach dem Ende des Studiums (Haus et al., 2012).
- Im nächsten Schritt gilt es für Wissenschaftsorganisationen, unter allen Bewerbenden die am besten geeigneten auszuwählen und dabei den Auswahlprozess fair durchzuführen. Dafür sollten Auswahlverfahren so gestaltet werden, dass die Prognosekraft möglichst hoch und Chancengleichheit gegeben ist. Dies bedeutet, es sollten diejenigen Bewerbenden ausgesucht werden, die mit größter Wahrscheinlichkeit erfolgreich promovieren. Dafür bedarf es den Einsatz kriteriumsvalider Auswahlverfahren. Zusätzlich sollten alle Bewerbenden gleichbehandelt und -bewertet werden. Nur so kann gewährleistet werden, dass die besten Köpfe ausgewählt werden, ohne dabei Gruppen oder einzelne Bewerbende gewollt oder ungewollt zu bevorzugen oder zu diskriminieren. Die zweite Herausforderung ist demnach die sachgemäße Anwendung wissenschaftlich fundierter Methoden (Diagnostik und Testkuratorium, 2018; DIN, 2016). So kann die Wahrscheinlichkeit für Fehlbesetzungen minimiert werden, und zugleich sinkt die Wahrscheinlichkeit, die Potenziale geeigneter Kandidatinnen und Kandidaten zu übersehen.

2.1 Herausforderung 1: Attraktiv für Absolventinnen und Absolventen durch die Akzeptanz von Auswahlverfahren für Promotionsstellen

Aus der Forschung wissen wir, dass die Sicht der Bewerbenden auf Auswahlverfahren in der Personalauswahl verschiedene und weitreichende Folgen für Unternehmen und Organisationen haben kann: Beispielsweise hat die Akzeptanz von Auswahlverfahren einen Einfluss auf das Bewerbungsverhalten. Sie wirkt sich positiv auf die Absicht aus, ein Stellenangebot anzunehmen, und hilft dabei, Bewerbende nicht an die Konkurrenz zu verlieren (Chapman, Uggerslev, Carroll, Piasentin & Jones, 2005). Die Akzeptanz von Auswahlverfahren beeinflusst darüber hinaus das Erleben und Verhalten im Beruf: Positive Akzeptanzurteile stehen unter anderem im Zusammenhang mit späterer Leistung im Job, geringeren Kündigungsabsichten und dem Wohlbefinden von Mitarbeiterinnen und Mitarbeitern (Hausknecht, Day & Thomas, 2004). Auch das Image der Organisation profitiert von positiven Akzeptanzurteilen: Arbeitgeber werden attraktiver beurteilt und die Bereitschaft, an-

deren den Arbeitgeber zu empfehlen, steigt (Hausknecht et al., 2004). Dabei erhalten Arbeitsproben und Auswahlgespräche die besten Akzeptanzurteile (Anderson, Salgado & Hülsheger, 2010; Hausknecht et al., 2004). Für die Gestaltung von Auswahlverfahren und für den Umgang mit den Bewerbenden kann es wichtig sein, inwieweit die Akzeptanzurteile auf Charakteristika der Bewerbenden und nicht nur auf die Merkmale der Verfahren zurückzuführen sind. Hier gibt es keine einheitlichen Befunde. Es existieren mehrere Studien, die jedoch zum Teil widersprüchliche Zusammenhänge zwischen Akzeptanzurteilen und Persönlichkeitseigenschaften berichten (z.B. Merkulova, Melchers, Kleinmann, Annen & Tresch, 2014; Oostrom, Born, Serlie & Van Der Molen, 2010; Truxillo, Bauer, Campion & Paronto, 2006).

Zusätzlich zu den genannten Konsequenzen sind Akzeptanzurteile bei der Entscheidung, mit welchen Verfahren Bewerbende selektiert werden, von Bedeutung. Um Bewerbende nicht abzuschrecken und die genannten Folgen von negativen Akzeptanzurteilen zu vermeiden, werden in der Praxis oftmals Verfahren eingesetzt, bei denen davon ausgegangen wird, dass diese von Bewerbenden akzeptiert werden (Schuler, Hell, Trapmann, Schaar & Boramir, 2007; König, Klehe, Berchthold & Kleinmann, 2010). Sind empirische Daten zu Akzeptanzurteilen nicht bekannt oder es gibt keine, werden diese von den verantwortlichen Personen vermutet. Derartige Spekulationen sind jedoch oftmals falsch (Beermann, Kersting, Stegt & Zimmerhofer, 2013; Kersting, 2008). Das bedeutet, in der Praxis werden vermutlich Verfahren eingesetzt, bei denen fälschlicherweise davon ausgegangen wird, dass diese von Bewerbenden akzeptiert werden.

Die genannten Gründe und Befunde sprechen sehr dafür, die Perspektive der Bewerbenden bei der Gestaltung von Auswahlverfahren für Wissenschaftlerinnen und Wissenschaftler zu berücksichtigen. Die beschriebenen Befunde gelten über verschiedene Branchen und Arbeitsfelder hinweg. Wissenschaftsorganisationen und deren Arbeitsgruppen könnten sich dies zunutze machen und ihre Attraktivität durch den Einsatz von akzeptierten Verfahren steigern (Hausknecht et al., 2004). Die Forschung zu Akzeptanzurteilen hat sich in Wissenschaftskontexten mit der Auswahl von Studierenden zu Studienbeginn (u.a. Hell & Schuler, 2005; Stegt, Didi, Zimmerhofer & Seegers, 2018), der Auswahl von Studierenden für weiterführende Masterstudiengänge (Herde, Stegt & Preckel, 2016) und mit Berufungsverfahren für Professorinnen und Professoren (Abele-Brehm & Bühner, 2016) beschäftigt. Die Ergebnisse zu den Akzeptanzurteilen bei der Auswahl von Studierenden werden im Folgenden zusammengefasst.

Bei der Auswahl für Studienplätze erhalten studienfachspezifische Fähigkeitstests, Abschlussnoten, frei geführte Auswahlgespräche, berufliche Erfahrung, strukturierte Auswahlgespräche und Motivationsschreiben von Studierenden die höchsten Akzeptanzurteile, während Arbeitsproben, Persönlichkeitstests, Wartezeit, Intelligenztests und Abiturnoten die niedrigsten Akzeptanzurteile erhalten (Stegt et al., 2018). Darüber hinaus fallen die Urteile positiver aus, wenn Studierende ihre bisherigen Leistungen in dem Verfahren positiv bewerten (Herde et al., 2016). Dies entspricht den Befunden zum Prozess der Akzeptanzbildung in Unternehmen. Bewerbende akzeptieren die Verfahren, bei denen sie einen Bezug zur späteren Tätigkeit erkennen können, die sie für augenscheinlich und prognostisch valide halten und bei denen sie ihre bisherigen Leistungen in diesem Verfahren positiv einschätzen (Anderson et al., 2010; Chan & Schmitt, 2004; Hausknecht et al., 2004). Die Be-

funde unterscheiden sich dabei nicht zwischen den Studiengängen. Auch in den Fächern, in denen entsprechendes diagnostisches Wissen vermittelt wird, kommt es zu den beschriebenen Akzeptanzurteilen (Stegt et al., 2018).

Des Weiteren wurde der Zusammenhang zwischen den Persönlichkeitseigenschaften der *Big Five* und Akzeptanzurteilen bei der Studierendenauswahl untersucht. Die Befunde sind uneindeutig und die wenigen gefundenen Zusammenhänge schwach (Herde et al., 2016; Stegt et al., 2018).

In einer Befragung zu Berufungsverfahren von Professorinnen und Professoren war die Akzeptanz von Arbeitsproben und strukturierten Auswahlgesprächen hoch, während Assessment Center weniger akzeptiert wurden (Abele-Brehm & Bühner, 2016).

Empirische Studien zu Akzeptanzurteilen für die Auswahl von Promovierenden liegen bisher nicht vor. Die vorliegende Studie möchte diese untersuchen. Damit sollen die bisherige Forschungslücke geschlossen und Indizien gesammelt werden, ob Ergebnisse aus der Forschung an Berufstätigen bzw. Studierenden generalisiert werden können. Wir erwarten eine weitgehende Replikation der Befunde aus den Studien zur Auswahl für Studienplätze (Herde et al., 2016; Stegt et al., 2018), da diese Urteile sich auf die Auswahl von Studierenden beziehen und in einem ähnlichen Kontext gebildet werden. Konkret bedeutet dies, dass wir erwarten, dass frei geführte Auswahlgespräche, strukturierte Auswahlgespräche, Arbeitsproben, Empfehlungsschreiben und berufliche Erfahrung höhere Akzeptanzwerte erhalten als Abschlussnoten, Intelligenztests, Motivationsschreiben, Persönlichkeitstests und persönliche Beziehungen. Auch gehen wir davon aus, dass die selbst eingeschätzten früheren Leistungen in einem Verfahren positiv mit den Akzeptanzwerten in dem entsprechenden Verfahren korrelieren.

Die Zusammenhänge zwischen den Persönlichkeitseigenschaften der *Big Five* und Akzeptanzurteilen sind uneindeutig (Merkulova et al., 2014; Oostrom et al., 2010; Truxillo et al., 2006). Bei Studierenden sind die berichteten Zusammenhänge schwach. Bei der Studie mit der größten Stichprobe beträgt die höchste Korrelation $r = .14$ zwischen der Bewertung von Auswahlgesprächen und Offenheit für Neues (Stegt et al., 2018). Aufgrund dieser Befundlage explorieren wir, ob es Zusammenhänge zwischen den Persönlichkeitseigenschaften der *Big Five* und Akzeptanzurteilen für Auswahlverfahren bei Promovierenden gibt. Dies hätte praktische Implikationen. Wenn Bewerbende mit bestimmten Persönlichkeitseigenschaften beispielsweise bestimmte Auswahlverfahren weniger akzeptieren sollten, könnten diese besser bei der Auswahl von Verfahren von den Wissenschaftsorganisationen berücksichtigt werden.

Bei der Personalauswahl in Unternehmen haben die Akzeptanzurteile einen Einfluss auf die Entscheidung, ob sich Bewerbende dem Auswahlverfahren aussetzen möchten und sich überhaupt bewerben (Hausknecht et al., 2004). Gehen potenziell Bewerbende davon aus, dass in einem Einstellungsverfahren Instrumente eingesetzt werden, die sie beispielsweise als nicht augenscheinlich bewerten, sinkt die Wahrscheinlichkeit, dass diese sich bewerben. Sollte dieser Befund auch für Promotionsstellen gelten, wäre dies von hoher praktischer Relevanz. Es empfiehlt sich grundsätzlich, im Vorfeld transparent darzustellen, welche Auswahlverfahren zum Einsatz kommen. Sofern dies geschieht, könnte der Einsatz von akzeptierten Verfahren zu einer Steigerung der Anzahl an Bewerbenden beitragen. Entsprechend wollen wir Hinweise finden, ob die eingesetzten Auswahlverfahren zur Vergabe einer Promotionsstelle die Entscheidung von Studierenden, sich zu bewerben, beeinflussen.

2.2 Herausforderung 2: Der Einsatz wissenschaftlicher Verfahren für die Auswahl von Promovierenden

Für die Entscheidung, mit welchen Verfahren selektiert werden soll, sollten neben Akzeptanzurteilen vor allem die sogenannten Hauptgütekriterien (Objektivität, Reliabilität und Validität) sowie die Fairness berücksichtigt werden (Diagnostik und Testkuratorium, 2018; DIN, 2016). Dabei gilt es unter anderem Verfahren zu nutzen, die kriteriumsvalide und fair sind, um eine hohe Prognosekraft und Chancengleichheit zu gewährleisten. Dies entspricht auch der Forderung des Wissenschaftsrats (2014), in Deutschland ein „faires, qualitätsorientiertes und wettbewerbsorientiertes Wissenschaftssystem“ zu installieren (S. 6). Aus Meta-Analysen in Unternehmen über verschiedene Berufe ist bekannt, dass Arbeitsproben, Intelligenztests und strukturierte Interviews zu den validesten Verfahren gehören und Berufserfolg am besten vorhersagen (Schmidt & Hunter, 1998). Diese Verfahren erfüllen bei sachgemäßer Anwendung auch das Kriterium der Fairness und sorgen für Chancengleichheit unter den Bewerbenden (Diagnostik und Testkuratorium, 2018; DIN, 2016). Sie sind somit eine notwendige Bedingung, um die besten Köpfe auswählen zu können. Aus der Literatur zur Studierendenauswahl ist bekannt, dass Schulnoten und fachspezifische Studierfähigkeitstests zu den validesten Verfahren gehören, um Studienerfolg zu prognostizieren (Hell, Trapmann & Schuler, 2007; Schult, Hofmann & Stegt, 2019; Trapmann, Hell, Weigand & Schuler, 2007). Für Auswahlgespräche wurden schlechtere Validitätswerte berichtet, wobei freie Gespräche hier noch schlechter abschnitten als strukturierte Gespräche (Hell, Trapmann & Schuler, 2008).

Zur Vorhersage von Promotionserfolg durch verschiedene Auswahlverfahren gibt es bisher kaum systematische Studien. Die Meta-Analyse von Kuncel, Hezlett und Ones (2001) liefert Hinweise, dass standardisierte Tests und Abschlussnoten aus dem Studium in verschiedenen Disziplinen prädiktiv valide für verschiedene Erfolgskriterien sind. Andere Verfahren wurden jedoch nicht untersucht.

Es gibt Hinweise, dass nur eine Minderheit der Promovierenden über eine Bewerbung und ein formelles Auswahlverfahren zur Promotionsstelle kommt (Berning & Falk, 2006). Über die Verwendungshäufigkeit eignungsdiagnostischer Verfahren in Wissenschaftsorganisationen, die Promotionsstellen besetzen oder Promotionsstipendien vergeben, gibt es bisher keine empirisch gesicherten Erkenntnisse. Belastbare Zahlen gibt es hier lediglich aus der Personalauswahl in Unternehmen (Benit & Soellner, 2013; Schuler et al., 2007; Ziegler, Danay & Maaß, 2012). In der Studie von Schuler et al. (2007) wurde auch erfasst, mit welchen Verfahren Unternehmen Trainees auswählen. Die Karrierestufe von Trainees ist mit Promovierenden vergleichbar. Tabelle 1 zeigt den Einsatz in Unternehmen zur Auswahl von Trainees und die Prognosekraft in Form der Kriteriumsvalidität verschiedener Verfahren für Berufserfolg sowie das Vorliegen von Fairness bei sachgemäßer Anwendung.

Tabelle 1: Einsatz, Kriteriumsvalidität und Fairness von Auswahlverfahren in Unternehmen

Verfahren	Einsatz Personalauswahl in Unternehmen ^a	Kriteriumsvalidität ^b	Fairness ^c
Strukturiertes Auswahlgespräch	72 %	.51	x
Assessment Center	50 %	.37	x
Frei geführtes Auswahlgespräch	21 %	.38	
Arbeitsprobe	13 %	.54	x
Persönlichkeitstest	6 %	.31 (Gewissenhaftigkeit)	x
Empfehlungsschreiben	5 %	.26	x
Intelligenztest	4 %	.51	x
Persönliche Kontakte	k. A.	k. A.	
Berufliche Erfahrung / Praktika	k. A.	.18	x
Motivations-/Begründungsschreiben	k. A.	k. A.	x
Abschlussnote	k. A.	k. A.	x

Anmerkungen: ^a Angaben über den Einsatz zur Auswahl von Trainees aus Schuler et al. (2007); ^b Angaben zu den Kriteriumsvaliditäten aus Schmidt & Hunter (1998); ^c bei sachgemäßer Anwendung nach Diagnostik und Testkuratorium (2018). k. A. = keine Angabe.

Die vorliegende Studie soll Hinweise darauf liefern, mit welchen Verfahren Promovierende ausgewählt werden. Die Befunde können Anhaltspunkte liefern, ob Wissenschaftsorganisationen das Potenzial von Auswahlverfahren nutzen, um die besten Köpfe für sich zu gewinnen. Entsprechend prüfen wir, ob die Nutzung psychologischer Verfahren in der Promovierenden-Auswahl in Wissenschaftsorganisationen der in deutschen Unternehmen für Trainees entspricht.

3 Methode

3.1 Stichprobe für Herausforderung 1

Die Daten wurden zwischen Januar und Oktober 2018 mittels einer Online-Umfrage gewonnen. Für die Fragestellungen, die sich auf Herausforderung 1 beziehen, wurden Studierende befragt. An diesen potenziellen Bewerbenden sollten sich Auswahlverfahren von Promotionsstellen orientieren, da sich die meisten mit Abschluss des Studiums für oder gegen eine Promotion entscheiden (Hauss et al., 2016). Auch Studierende mit dem (vorläufigen) Abschlussziel Bachelor wurden befragt, da diese zu einem späteren Zeitpunkt potenzielle Bewerbende sein können oder unter bestimmten Bedingungen mit einem Bachelorabschluss eine Promotion beginnen können (HRK Hochschulrektorenkonferenz, 2007).

Hierfür wurden alle Studierendenvertretungen an deutschen Universitäten angeschrieben, deren Kontakte öffentlich waren. Insgesamt wurden 1390 Vertretungen angeschrieben, aus denen sich eine Stichprobe von 170 Studierenden von 41 verschiedenen Universitäten ergab. Die TU Dresden war mit 10% der Teilnehmenden am stärksten vertreten, gefolgt von der FU Berlin (8.8%) und der Westfälische Wilhelms-Universität Münster (6.5%). Um

das Studienfach zu erfragen, wurde das Schlüsselverzeichnis der Studenten- und Prüfungsstatistik des Statistischen Bundesamtes genutzt. Für die Fächergruppen Geisteswissenschaften, Rechts-, Wirtschafts- und Sozialwissenschaften, Mathematik und Naturwissenschaften wurden auch einzelne Studienbereiche (wie beispielsweise Biologie oder Chemie) zur Auswahl gegeben, um ein differenziertes Bild zu erhalten, als es mit den Fächergruppen möglich gewesen wäre. Die Teilnehmenden gaben am häufigsten an, Psychologie (16.5%), Medizin (10.6%) oder Wirtschaftswissenschaften (11.2%) zu studieren. Im Durchschnitt waren die Studierenden 24.6 Jahre ($SD = 4.05$) alt. 35.9% strebten einen Bachelorabschluss, 47.6% einen Masterabschluss und 16.5% ein Staatsexamen an. Unter den Befragten gaben 40% an, auf jeden Fall promovieren zu wollen; 35.3% waren noch unentschieden, und 24.7% planten nach dem Studium definitiv etwas anderes zu machen (Tabelle 2). Den höchsten Erfahrungswert hatte das frei geführte Auswahlgespräch. 69.4% gaben an, mit diesem Verfahren bereits Erfahrung gesammelt zu haben. Lediglich 4.7% gaben an, schon mindestens einmal einen Persönlichkeitstest in einem Auswahlverfahren durchgeführt zu haben. Dies war der niedrigste Erfahrungswert unter den Verfahren.

Tabelle 2: Promotionsvorhaben nach Abschlüssen (in Prozent)

	ich möchte auf jeden Fall promovieren	bei mir ist die Entscheidung noch nicht gefallen	ich möchte nicht promovieren
Bachelor	18.0	39.3	42.6
Master	42.7	21.0	35.8
Staatsexamen	78.6	3.6	17.9

3.2 Akzeptanzmessung

Analog zu den Erhebungen von Hell und Schuler (2005), Herde et al. (2016) und Stegt et al. (2018) wurden die Teilnehmenden gefragt, wie sie einzelne Verfahren zur Auswahl von Promovierenden auf einer vierstufigen Rating-Skala von sehr schlecht (1) bis sehr gut (4) beurteilen. Für ein geteiltes Verständnis der einzelnen Verfahren wurden diese teilweise mit Beschreibungen gezeigt (Tabelle 3). Die Verfahren wurden in randomisierter Reihenfolge dargeboten.

Tabelle 3: Kurzbeschreibungen der Auswahlverfahren

Verfahren	Beschreibung
Abschlussnote Studium	
Arbeitsprobe	Verfahren, die Stichproben des zukünftigen Arbeitsverhaltens von Promovierenden darstellen; z.B. Präsentation, Rollenspiel
Assessment Center	Kombination mehrerer unterschiedlicher eignungsdiagnostischer Verfahren unter Einsatz mehrerer Beurteiler/innen
Berufliche Erfahrung / Praktika	
Empfehlungsschreiben	
Frei geführtes Auswahlgespräch	
Intelligenztest	Kognitiver Leistungstest zur Erfassung allgemeiner Intelligenz
Motivations-/Begründungsschreiben	Ein vom Promotionsbewerbenden verfasster Aufsatz zu motivationalen Hintergründen bzw. zu einem vorgegebenen Thema
Persönliche Kontakte	Persönliche Beziehungen vor dem Auswahlverfahren (zu potenziellen Vorgesetzten, Kolleginnen/Kollegen oder Entscheiderinnen/ Entscheidern). Dazu gehören auch Beziehungen über Dritte
Persönlichkeitstest	Fragebogen zur Beschreibung der Ausprägung von Persönlichkeitseigenschaften
Strukturiertes Auswahlgespräch	Einheitliche, vorbestimmte Abfolge von Fragen oder zumindest Themenbereichen

Zusätzlich wurden die Teilnehmenden gebeten, die einzelnen Verfahren gemäß ihrer Akzeptanz in eine Rangreihe zu bringen. Für jedes Verfahren wurde erhoben, ob die Teilnehmenden Erfahrung mit dem Verfahren gesammelt haben. Wurde dies bejaht, mussten die Teilnehmenden ihre bisherige Leistung in dem entsprechenden Verfahren auf einer vierstufigen Rating-Skala (1 = sehr schlecht, 4 = sehr gut) einschätzen. Darüber hinaus wurde eine Liste mit allen Verfahren vorgelegt. Aus dieser sollten die Studierenden die Verfahren nennen, bei denen sie davon ausgehen, dass sie bei der Auswahl von Promovierenden derzeit genutzt werden. Hier waren Mehrfachnennungen möglich.

3.3 Persönlichkeitsmessung

Die Persönlichkeitseigenschaften der *Big Five* wurden mit der deutschen Kurzversion des Big-Five-Inventory (BFI-K, Form S; Rammstedt & John, 2005) erhoben. Bei diesem Fragebogen mussten die Teilnehmenden 21 Aussagen über sich selber auf einer fünfstufigen Rating-Skala von „sehr unzutreffend“ (1) bis „sehr zutreffend“ (5) einschätzen. Die internen Konsistenzen (Cronbachs Alpha) lagen zwischen .63 (Gewissenhaftigkeit) und .82 (Neurotizismus).

3.4 Weitere Variablen

Zusätzlich wurden Alter, Geschlecht, Studienfächer, Universität, Abschluss und Fachsemester erhoben. Es wurde die Frage gestellt, ob die Studierenden planen zu promovieren und ob die eingesetzten Verfahren bei der Entscheidung, sich auf eine Promotionsstelle zu bewerben, eine Rolle spielen.

3.5 Stichprobe für Herausforderung 2

Um die Häufigkeit der Verfahren zu ermitteln, wurden von allen deutschen Universitäten mit Promotionsrecht jeweils zehn zufällig ausgewählte Promovierende angeschrieben mit der Bitte, an einer kurzen Online-Umfrage teilzunehmen. Insgesamt 164 Promovierende von 46 verschiedenen Universitäten nahmen teil. Der häufigste genannte Fachbereich, in dem die Teilnehmenden promovierten, waren Sprach- und Kulturwissenschaften (12.8%), gefolgt von den Wirtschaftswissenschaften (10.7%) und den Rechtswissenschaften (8%). 16.6% gaben an, an der Eberhard-Karls-Universität Tübingen zu promovieren, 16% an der Universität Ulm und 10.7% an der Humboldt-Universität zu Berlin.

3.6 Messung der eingesetzten Verfahren

Den Promovierenden wurde eine Liste mit möglichen Verfahren und kurzen Erklärungen zu den Verfahren (Tabelle 3) vorgelegt, aus der sie die Verfahren nennen sollten, mit denen sie für ihre Beschäftigung ausgewählt wurden. In die Berechnungen gingen ausschließlich die Promovierenden ein, die eine Form der Beschäftigung an der Universität (Haushaltsstelle oder Drittmittelstelle) hatten oder ein Promotionsstipendium bezogen. Freie Promotionen oder an Anstellungsverhältnisse in der Wirtschaft gekoppelte Promotionen wurden ausgenommen, da für diese keine Auswahlverfahren für Promotionsstellen im engeren Sinne durchgeführt wurden.

3.7 Weitere Variablen

Die Promovierenden wurden nach ihrem Geschlecht, Alter, Promotionsfach bzw. -fächern, dem Datum des Beginns der Promotion, Universität und Form der Beschäftigung gefragt.

4 Ergebnisse Herausforderung 1

Ein Methodenvergleich bei Herde et al. (2016) ergab, dass die Ergebnisse mit Ratingskalen und Rangreihen sehr ähnlich sind. Dies gilt auch für die vorliegende Studie. Aufgrund der hohen Vergleichbarkeit berichten wir aus ökonomischen Gründen ausschließlich die Ergebnisse der Ratingskala.

4.1 Akzeptanzurteile

Abbildung 1 zeigt die Mittelwerte der Bewertungen und die dazugehörigen 95% Konfidenzintervalle für alle Studierenden und für diejenigen, die sich vorstellen können, zu promovieren. In der Gruppe aller Studierenden werden frei geführte Auswahlgespräche, berufliche Erfahrungen, Arbeitsproben und strukturierte Auswahlgespräche signifikant besser bewertet als Assessment Center, Empfehlungsschreiben, Abschlussnoten, Persönlichkeitstests, persönliche Kontakte und Intelligenztests, da sich die Konfidenzintervalle nicht überlappen. Motivations- und Begründungsschreiben stehen zwischen diesen beiden Gruppen von Verfahren, da sich das Konfidenzintervall sowohl mit Verfahren aus der ersten als auch mit Verfahren aus der zweiten Gruppe überlappt.

In der Gruppe derjenigen, die sich vorstellen können zu promovieren, ergibt sich ein ähnliches Bild. Intelligenztests werden hier zusätzlich signifikant am wenigsten akzeptiert. Das Konfidenzintervall überlappt sich mit keinem Konfidenzintervall eines anderen Verfahrens¹.

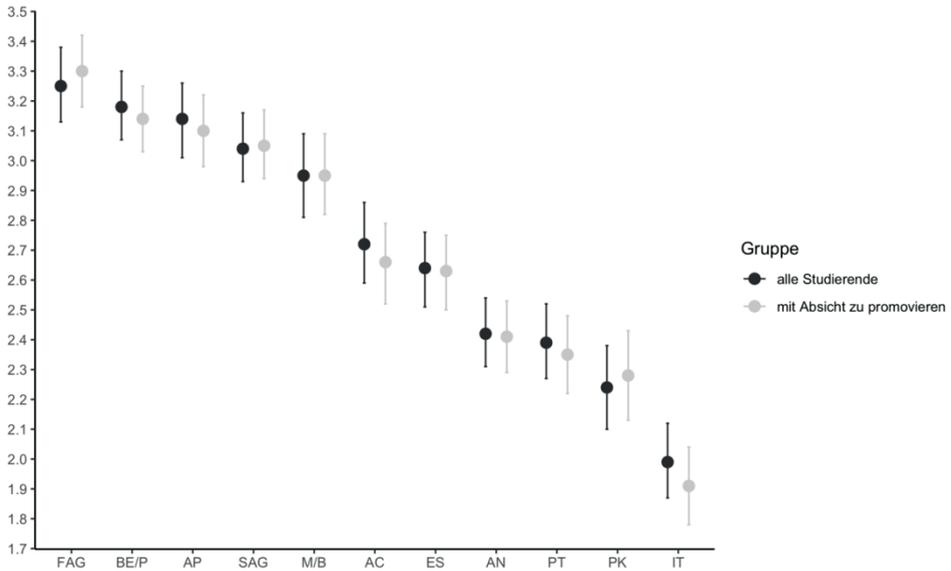


Abbildung 1: Barplots der Mittelwerte der Akzeptanzbewertungen und die dazugehörigen 95% Konfidenzintervalle

Die Stichprobe bilden 170 Studierende, die Verfahren zur Auswahl von Promovierenden auf einer vierstufigen Rating-Skala (1 = sehr schlecht, 4 = sehr gut) bewerteten. FAG = Frei geführtes Auswahlgespräch; BE/P = Berufliche Erfahrung/Praktika; AP = Arbeitsprobe; SAG = Strukturiertes Auswahlgespräch; M/B = Motivations-/Begründungsschreiben; AC = Assessment Center; ES = Empfehlungsschreiben; AN = Abschlussnote; PT = Persönlichkeitstest; PK = Persönliche Kontakte; IT = Intelligenztest

Tabellen 4 und 5 zeigen die dazugehörigen Mittelwerte, Standardabweichungen und die 95% Konfidenzintervalle.

¹ Auch die differenzierten Betrachtungen nach Studienfach, Geschlecht und Semesterzahl führten zu vergleichbaren Ergebnissen in den Akzeptanzurteilen. Es machte ebenfalls keinen Unterschied für die Akzeptanzurteile, ob die Studierenden Vorerfahrung in den einzelnen Verfahren hatten oder nicht.

Tabelle 4: Akzeptanzurteile aller Studierenden sortiert nach absteigender Akzeptanz ($N = 170$)

Verfahren	<i>M</i>	<i>SD</i>	95% KI
Frei geführtes Auswahlgespräch	3.25	0.77	3.13 – 3.38
Berufliche Erfahrung / Praktika	3.18	0.72	3.07 – 3.30
Arbeitsprobe	3.14	0.77	3.01 – 3.26
Strukturiertes Auswahlgespräch	3.04	0.73	2.93 – 3.16
Motivations-/Begründungsschreiben	2.95	0.88	2.81 – 3.09
Assessment Center	2.72	0.83	2.59 – 2.86
Empfehlungsschreiben	2.64	0.79	2.51 – 2.76
Abschlussnote	2.42	0.73	2.31 – 2.54
Persönlichkeitstest	2.39	0.78	2.27 – 2.52
Persönliche Kontakte	2.24	0.90	2.10 – 2.38
Intelligenztest	1.99	0.80	1.87 – 2.12

Tabelle 5: Akzeptanzurteile derjenigen, die promovieren wollen oder deren Entscheidung noch nicht gefallen ist, sortiert nach absteigender Akzeptanz ($N = 128$)

Verfahren	<i>M</i>	<i>SD</i>	95% KI
Frei geführtes Auswahlgespräch	3.30	0.75	3.18 – 3.42
Berufliche Erfahrung / Praktika	3.14	0.72	3.03 – 3.25
Arbeitsprobe	3.10	0.76	2.98 – 3.22
Strukturiertes Auswahlgespräch	3.05	0.73	2.94 – 3.17
Motivations-/Begründungsschreiben	2.95	0.87	2.82 – 3.09
Assessment Center	2.66	0.84	2.52 – 2.79
Empfehlungsschreiben	2.63	0.81	2.50 – 2.75
Abschlussnote	2.41	0.76	2.29 – 2.53
Persönlichkeitstest	2.35	0.82	2.22 – 2.48
Persönliche Kontakte	2.28	0.93	2.13 – 2.43
Intelligenztest	1.91	0.82	1.78 – 2.04

4.2 Selbsteingeschätzte Leistungen

In diese Berechnung gingen nur die Probandinnen und Probanden ein, die bereits Erfahrung mit dem jeweiligen Instrument gesammelt hatten und so ihre Leistung einschätzen konnten. Der Zusammenhang zwischen der selbst eingeschätzten bisherigen Leistung in einem Verfahren und den Akzeptanzurteilen wurde mittels Pearson-Korrelation nach Bonferroni-Korrektur für multiple Testungen bestimmt. Lediglich beim Motivations- und Begründungsschreiben gab es einen signifikanten, positiven Zusammenhang ($r = .49$). Studierende, die ihre Leistung im Schreiben eines Motivations- und Begründungsschreibens positiv einschätzten, akzeptierten diese stärker.

4.3 Persönlichkeitsfaktoren

Nach Bonferroni-Korrektur gab es nur einen statistisch signifikanten Zusammenhang zwischen den Persönlichkeitseigenschaften und den Akzeptanzurteilen. Studierende, die sich als offener für Neues beschrieben, akzeptierten Motivations-/Begründungsschreiben stärker ($r = .27$).

4.4 Einfluss der angenommenen Verfahren auf die Entscheidung, sich zu bewerben

Von denjenigen, die sicher promovieren wollen, gaben 48.5% an, dass die eingesetzten Verfahren zur Auswahl von Promovierenden ihre Entscheidung, sich zu bewerben, maßgeblich beeinflussen würden. Von denjenigen, die sich bezüglich einer Promotion noch unsicher waren, waren es 58.3% (s. Tabelle 6).

Tabelle 6: Hat das Verfahren einen Einfluss auf die Bewerbungsentscheidung?

	Ich möchte auf jeden Fall promovieren	Bei mir ist die Entscheidung noch nicht gefallen
ja	48.5%	58.3%
nein	45.6%	30.0%
keine Angabe	5.9%	11.7%

4.5 Ergebnisse Herausforderung 2: Eingesetzte Verfahren zur Auswahl von Promovierenden

Das meistverwendete Verfahren war das frei geführte Auswahlgespräch (57.2%), gefolgt von persönlichen Beziehungen und Kontakten (55.6%), die zu den Entscheiderinnen und Entscheidern schon vor dem Verfahren bestanden. Am wenigsten wurden Assessment Center (1.6%), Intelligenztests (0.5%) und Persönlichkeitstests (0.5%) eingesetzt (Tabelle 7)². Dies ist ein anderes Bild als bei der Personalauswahl in Unternehmen. Diese suchten ihre Trainees vorwiegend mit strukturierten Auswahlgesprächen und Assessment Centern aus und nutzten diese beiden Verfahren häufiger als Wissenschaftsorganisationen bei der Auswahl ihrer Promovierenden. Arbeitsproben, Intelligenztests und Persönlichkeitstests wurden ähnlich selten eingesetzt, während das frei geführte Auswahlgespräch deutlich häufiger in Wissenschaftsorganisationen als in Unternehmen zum Einsatz kam.

2 Auch die differenzierten Betrachtungen nach Promotionsfach, Geschlecht, Beschäftigungsart und der Zeit seit der Anmeldung führten zu vergleichbaren Häufigkeiten der verwendeten Verfahren.

Tabelle 7: Einsatz der Verfahren bei der Auswahl von Promovierenden

Verfahren	Einsatz Promovierendenauswahl	Einsatz Personalauswahl ^a
Frei geführtes Auswahlgespräch	57.2%	21%
Persönliche Kontakte	55.6%	k. A.
Abschlussnote	43.9%	k. A.
Motivations-/Begründungsschreiben	35.3%	k. A.
Berufliche Erfahrung / Praktika	30.0%	k. A.
Empfehlungsschreiben	23.0%	5%
Strukturiertes Auswahlgespräch	15.0%	72%
Arbeitsprobe	13.4%	13%
Assessment Center	1.6%	50%
Intelligenztest	0.5%	4%
Persönlichkeitstest	0.5%	6%

Anmerkungen: ^a Angaben über den Einsatz zur Auswahl von Trainees aus Schuler et al. (2007); k. A. = keine Angabe.

5 Diskussion

Die vorliegende Arbeit beschäftigte sich mit zwei Herausforderungen, mit denen sich Wissenschaftsorganisationen konfrontiert sehen, wenn sie Promotionsstellen oder Stipendien für Promotionen vergeben. Zum einen wurde untersucht, wie elf verschiedene Verfahren für die Auswahl von Promovierenden von potenziell Bewerbenden akzeptiert werden. Zum anderen untersuchte die Studie, mit welchen Verfahren Promovierende derzeit ausgesucht werden. Bisher war wenig dazu bekannt.

5.1 Herausforderung 1

Frei geführte Auswahlgespräche, berufliche Erfahrungen bzw. Praktika, Arbeitsproben und strukturierte Auswahlgespräche wurden für die Personalauswahl von Promotionsstellen am besten akzeptiert. Dies entspricht weitgehend den gängigen Befunden zur Akzeptanz der Personalauswahl in Unternehmen (Anderson et al., 2010; Hausknecht et al., 2004) und bei der Auswahl von Studierenden zum Studium (Stegt et al., 2018). Bei den Berufungsverfahren von Professorinnen und Professoren hatten Arbeitsproben und strukturierte Auswahlgespräche die höchsten Akzeptanzwerte erhalten (Abele-Brehm & Bühner, 2016). Auch bei der Auswahl von Promovierenden wurden diejenigen Verfahren akzeptiert, die von Laien und nicht diagnostisch ausgebildeten Personen als augenscheinlich und prognostisch valide bewertet werden. Eine Ausnahme waren Arbeitsproben. Diese erhielten bei der Auswahl von Studierenden geringe bis mittlere Akzeptanzurteile, während sie bei der Personalauswahl in Unternehmen, bei der Auswahl von Promovierenden und auch bei Berufungsverfahren von Professorinnen und Professoren hohe positive Urteile erhielten. Die hohen positiven Akzeptanzurteile werden damit erklärt, dass Arbeitsproben einen hohen Bezug zur Zieltätigkeit aufweisen (Hausknecht et al., 2004). Bei der Studierendenauswahl ist es jedoch schwierig, genaue Vorstellungen darüber zu haben, wie eine Arbeitsprobe gestaltet

sein müsste, um die Zieltätigkeit „Studieren“ zu repräsentieren (Herde et al., 2016). Ähnlich wie bei der Personalauswahl in Unternehmen oder bei der Berufung von Professorinnen und Professoren könnten beim Promovieren Vorstellungen über die Zieltätigkeit konkreter sein und somit eine bessere Vorstellung über repräsentative Arbeitsproben (beispielsweise wissenschaftliche Artikel zusammenfassen und präsentieren) vorhanden sein. Dies könnte ein Grund für die unterschiedlichen Akzeptanzurteile von Arbeitsproben sein.

Unterschiede unter den Bewerbenden spielten kaum eine Rolle für die verschiedenen Akzeptanzurteile. Selbsteingeschätzte bisherige Leistungen in den Auswahlverfahren hatten wenig Einfluss auf die Bildung von Akzeptanzurteilen bei der Auswahl von Promovierenden. Dies entsprach nicht den Befunden bei der Personalauswahl in Unternehmen und bei der Auswahl von Studierenden. Die bisherigen Erfahrungen bei Auswahlverfahren von Studierenden stammen vermutlich aus Kontexten (z.B. Studierenden-Jobs) und unterlagen Bedingungen (z.B. Teilzeit, befristet für die Semesterferien oder Studienplätze), die nicht vergleichbar sind mit einer Promotionsstelle. Möglicherweise spielte die selbsteingeschätzte bisherige Leistung aufgrund dieser Diskrepanzen keine Rolle. Promotionsstellen gelten als Berufseinstieg und sind somit ein neuer Kontext, bei denen bisherige Erfahrungen und Leistungen in Auswahlverfahren gegebenenfalls nicht weiterhelfen können.

Die Ausprägungen der Persönlichkeitseigenschaften hatten ebenfalls kaum einen Einfluss auf die Bildung von Akzeptanzurteilen. Ein Grund für den Mangel an empirischen Zusammenhängen könnte die hohe Abstraktionsebene der Messung sein. Es ist zu vermuten, dass es auf der Ebene enger umschriebener Persönlichkeitsfacetten durchaus Zusammenhänge geben kann (Ziegler, Knogler & Bühner, 2009; Kretzschmar, Spengler, Schubert, Steinmayr & Ziegler, 2018).

Die Befunde dieser Studie scheinen sowohl für verschiedene Ausprägungen von Persönlichkeitseigenschaften generalisierbar zu sein, wenn diese so breit erfasst werden wie mit dem BFI-K (Rammstedt & John, 2005), als auch unabhängig von bisherigen Leistungen in den einzelnen Verfahren zu sein.

5.2 Herausforderung 2

Der Einsatz von Verfahren bei der Auswahl von Promovierenden unterschied sich von den Ergebnissen aus Unternehmen. Auswahlprozesse in Unternehmen scheinen umfangreicher zu sein. Trainees werden häufiger mit verschiedenen Verfahren ausgesucht. Eine mögliche Erklärung hierfür können die knappen Ressourcen in den Wissenschaftsorganisationen sein. In vielen Unternehmen stehen für Auswahlverfahren mehr Ressourcen zur Verfügung. Während in Unternehmen ganze Stellen bis zu Abteilungen mit der Konzeption und Durchführung von Auswahlverfahren beschäftigt sein können, sind es bei der Auswahl von Promovierenden meist Personen, die dies zusätzlich zu ihren Aufgaben in Lehre und Forschung machen. Aufwändige Verfahren sind daher aufgrund geringer Ressourcen schwerer umsetzbar.

Die Daten deuten darauf hin, dass viele Promotionsstellen über persönliche Kontakte kombiniert mit frei geführten Auswahlgesprächen vergeben wurden. Diese Zahlen entsprechen Promovierendenbefragungen, in denen über 70% angaben, kein formelles Auswahlverfahren durchlaufen zu haben (Berning & Falk, 2006). Die Urteilsbildung bei frei geführten Auswahlgesprächen und bei persönlichen Kontakten erfolgt nicht nach vorab festgelegten Regeln. Dies wird als klinische Urteilsbildung beschrieben (siehe Schmidt-Atzert,

Krumm & Kersting, 2018). Sie ist bei der Vorhersage von Berufs- und Ausbildungserfolg dem regelgeleiteten Vorgehen, der sogenannten mechanischen oder statistischen Urteilsbildung, unterlegen (Kuncel, Klieger, Connelly & Ones; 2013). Darüber hinaus haben freie Auswahlgespräche eine geringere Prognosekraft (Schmidt & Hunter, 1998; Tabelle 8) und sind anfällig für Verzerrungen durch Beurteilerfehler wie beispielsweise den Halo-Effekt (Posthuma, Morgeson & Campion, 2002). Es ist davon auszugehen, dass nicht alle Personen nach denselben Regeln und Kriterien bewertet werden.

Promovierende mit frei geführten Auswahlgesprächen und über persönliche Kontakte auszuwählen, mag eine zeiteffiziente Methode sein, da derartige Auswahlverfahren kaum vorbereitet werden müssen. Sie haben jedoch potenziell eine geringere Prognosekraft und entsprechen nicht der Forderung nach Chancengleichheit (Wissenschaftsrat, 2014). Die Tatsache, dass anzunehmen ist, dass Promovierende vorher im Rahmen studentischer Hilfskrafttätigkeiten oder von Abschlussarbeiten längere Arbeitsproben abgegeben haben, ist zwar der Prognosekraft zuträglich, mindert jedoch vermeintlich die Objektivität der Auswahlmethode und dürfte vor allem aus der Fairnessperspektive problematisch sein. Schließlich werden studentische Stellen und Abschlussarbeiten vermutlich noch unsystematischer vergeben. Die Daten sprechen dafür, dass Wissenschaftsorganisationen nicht unter allen möglichen Personen auswählen, sondern lediglich unter denen, die sie bereits kennen. Die Gefahr, gute Kandidatinnen und Kandidaten zu übersehen, ist mit derart gestalteten Auswahlverfahren groß.

5.3 Limitationen

Die Stichproben sind mit jeweils über 20 verschiedenen Fächern zwar sehr heterogen, allerdings gibt es bei der Fächerverteilung Abweichungen zur Grundgesamtheit aller Studierenden und Promovierenden in Deutschland. Psychologie ist mit 16.5% bei den Studierenden überrepräsentiert, bei 2.8% in der Grundgesamtheit. Informatik ist mit 1.8% (4% in der Grundgesamtheit) unterrepräsentiert (Statistisches Bundesamt, 2017). Auch waren Studentinnen mit 71.8% überrepräsentiert bei 48.5% in der Grundgesamtheit (Statistisches Bundesamt, 2017). Die verschiedenen Fächer und das Geschlecht hatten zwar keinen Einfluss auf die Akzeptanzurteile oder das eingesetzte Verfahren, die Generalisierbarkeit der Befunde ist dennoch in ausgewogeneren Stichproben zu prüfen. Es ist wahrscheinlich, dass die Befunde auch in einer repräsentativeren Stichprobe gelten, da die Bildung von Akzeptanzurteilen auch über verschiedene Gruppen (Studierende, Angestellte) vergleichbar ist. Diese Vermutung und die Generalisierbarkeit der Befunde von Herausforderung 2 müsste jedoch empirisch getestet werden.

Darüber hinaus könnte die Generalisierbarkeit der Ergebnisse auch eingeschränkt sein, da die Analysen auf geringen Stichprobengrößen basieren, die mit einer geringen Rücklaufquote zusammenhängen könnten. Hier kann es aufgrund des Themas „Auswahlverfahren für Promovierende“ zu Selbstselektionseffekten gekommen sein. Auch konnten aufgrund der zu geringen Zahl an Teilnehmenden keine Mehrebenenanalysen durchgeführt werden. Dies wäre bei zukünftigen Erhebungen mit mehr Teilnehmenden zu begrüßen.

Auch wäre es wertvoll, in weiteren Untersuchungen die Akzeptanzurteile mit mehr als einem Item zu erfragen, um ein differenziertes Bild über verschiedene Aspekte der Akzeptanzurteile zu gewinnen (Kersting, 2008; Bauer et al., 2001).

Für Herausforderung 2 wurden Promovierende befragt, mit welchen Verfahren sie für ihre Stellen ausgesucht wurden. Dies ist nur eine Quelle, die Erinnerungsfehlern unterliegen könnte. Hier könnten Personen befragt werden, die für die Durchführung der Verfahren verantwortlich sind, um die Ergebnisse zu bestätigen. Dies würde zu einer besseren Vergleichbarkeit mit den Daten aus Unternehmen führen, da für diese Verantwortliche der Unternehmen befragt wurden.

5.4 Implikationen für die Praxis

Wissenschaftsorganisationen, die Promotionsstellen zu besetzen haben, sollten ihre Verfahren an mindestens zwei Kriterien ausrichten. Zum einen sollten die Verfahren akzeptiert werden, um Bewerbende nicht zu verschrecken. Über 50% der Studierenden gaben an, dass das eingesetzte Verfahren einen Einfluss darauf habe, ob sie sich bewerben würden (Tabelle 6). Zum anderen sollten die Verfahren wissenschaftlichen Kriterien genügen, um Chancengleichheit und Prognosekraft zu gewährleisten. Solange es keine Metaanalysen zur Prognosekraft für Erfolgsmaße bei Promovierenden gibt, sollte auf bisherige Arbeiten zur Kriteriumsvalidität für Berufserfolg aus Unternehmen zur Orientierung zurückgegriffen werden (z.B. Schmidt & Hunter, 1998). Tabelle 8 integriert die Ergebnisse der vorliegenden Studie und von bisherigen Arbeiten. Für die elf Verfahren werden Akzeptanz und der derzeitige Einsatz aus der vorliegenden Studie dargestellt. Ergänzt wird Tabelle 8 durch Werte der Kriteriumsvalidität aus der Meta-Analyse von Schmidt und Hunter (1998). Für die Bewertung der Fairness bei sachgemäßer Anwendung wurde sich an den Empfehlungen des Diagnostik- und Testkuratoriums (2018) orientiert.

Tabelle 8: Akzeptanz, Validität und Einsatz der Verfahren.

Verfahren	akzeptiert ^a	eingesetzt ^b	valide ^c	fair ^d
Abschlussnote Studium				x
Arbeitsprobe	x		x	x
Assessment Center				x
Berufliche Erfahrung / Praktika	x			x
Empfehlungsschreiben				x
Frei geführtes Auswahlgespräch	x	x		
Intelligenztest			x	x
Motivations-/Begründungsschreiben	x		k. A.	x
Persönliche Kontakte		x	k. A.	
Persönlichkeitstest				x
Strukturiertes Auswahlgespräch	x		x	x

Anmerkungen: ^a signifikant besser akzeptiert als die übrigen Verfahren nach Tabelle 4; ^b mindestens 50% der Promovierenden wurden damit ausgesucht nach Tabelle 7; ^c Kriteriumsvalidität mindestens .5 nach Schmidt und Hunter (1998); ^d bei sachgemäßer Anwendung nach Diagnostik und Testkuratorium (2018); k. A. = keine Angabe.

Es gibt kein Verfahren, das häufig genutzt und gleichzeitig akzeptiert, fair und valide ist.

Strukturierte Auswahlgespräche und Arbeitsproben werden akzeptiert, sind fair und gehören zu den validesten Verfahren. Diese Befunde sollten Entscheiderinnen und Ent-

scheider sich zunutze machen und diese Verfahren bei der Auswahl von Promovierenden anwenden. Ein angemessener Einsatz dieser Verfahren würde die Prognosekraft und Chancengleichheit erhöhen. Gleichzeitig könnte der Einsatz akzeptierter Verfahren die Grundquote erhöhen. Studierende gaben an, sich eher auf Stellen zu bewerben, bei denen akzeptierte Verfahren eingesetzt werden (Tabelle 6).

Aufgrund der Ergebnisse zu den selbsteingeschätzten bisherigen Leistungen und Persönlichkeitseigenschaften liefert diese Studie keinen Anhaltspunkt, Verfahren nach Charakteristika der Bewerbenden individuell auszurichten.

Wichtig ist, dass bei der Konzeption, Durchführung und Evaluation dieser Verfahren diagnostisch geschulte Personen involviert werden. Nur so ist zu gewährleisten, dass die Potenziale dieser Verfahren auch genutzt werden (Diagnostik und Testkuratorium, 2018).

5.5 Fazit

Die vorliegende Studie beinhaltet wertvolle Erkenntnisse zur Auswahl von Promovierenden. Sie liefert wichtige Indizien, um bestehende Forschungslücken zu schließen. Wissenschaftsorganisationen können diese nutzen, um ihre Auswahlverfahren zu verbessern. Der vermehrte Einsatz von akzeptierten, validen und fairen Auswahlverfahren kann ein wichtiger Baustein sein, um die besten Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler für sich zu gewinnen und nicht an andere Wissenschaftsorganisationen oder die Wirtschaft zu verlieren. Die Verwendung von strukturierten Interviews oder Arbeitsproben scheint eine sinnvolle Alternative zum bisherigen Vorgehen zu sein. Mit einem solchen Vorgehen steigt potenziell auch die Chance, die besten Köpfe in den eigenen Reihen zu haben. Mit diesen sind Universitäten und andere Organisationen des Wissenschaftsbetriebs gut aufgestellt, um exzellenten wissenschaftlichen Output zu generieren und diesen erfolgreich gegenüber Missinformationen zu behaupten und zu verbreiten.

Literatur

- Abele-Brehm, A. & Bühner, M. (2016). Wer soll die Professur bekommen? Eine Untersuchung zur Bewertung von Auswahlkriterien in Berufungsverfahren der Psychologie. *Psychologische Rundschau*, 67(4), 250–261. <https://doi.org/10.1026/0033-3042/a000335>
- Anderson, N., Salgado, J. F. & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18(3), 291–304. <https://doi.org/10.1111/j.1468-2389.2010.00512.x>
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P. & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology*, 54(2), 387–419. <https://doi.org/10.1111/j.1744-6570.2001.tb00097>
- Beermann, D., Kersting, M., Stegt, S. & Zimmerhofer, A. (2013). Vorurteile und Urteile zur Akzeptanz von Persönlichkeitsfragebogen. *PersonalQuarterly*, 65(4), 41–45.
- Benit, N. & Soellner, R. (2013). Scientist-practitioner gap in Deutschland: Eine empirische Studie am Beispiel psychologischer Testverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, 57(3), 145–153. <https://doi.org/10.1026/0932-4089/a000111>
- Berning, E. & Falk, S. (2006). Promovieren an den Universitäten in Bayern. *Praxis – Modelle – Perspektiven*. München: Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung.

- Braun, S., Hentschel, T., Peus, C. & Frey, D. (2015). Chancengleichheit durch professionelle Personalauswahl in der Wissenschaft. In C. Peus, S. Braun, T. Hentschel & D. Frey (Hrsg.), *Personalauswahl in der Wissenschaft* (S. 29–48). Berlin: Springer.
<https://doi.org/10.1007/978-3-662-48112-7>
- Chan, D. & Schmitt, N. (2004). An agenda for future research on applicant reactions to selection procedures: A construct-oriented approach. *International Journal of Selection and Assessment*, 12(1-2), 9–23. <https://doi.org/10.1111/j.0965-075X.2004.00260.x>
- Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A. & Jones, D. A. (2005). Applicant attraction to organizations and job choice: a meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology*, 90(5), 928–944.
<https://doi.org/10.1037/0021-9010.90.5.928>
- Diagnostik und Testkuratorium. (2018). *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430*. Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-53772-5>
- DIN Deutsches Institut für Normung e.V. (2016). *DIN 33430: Anforderungen an berufsbezogene Eignungsdiagnostik. Ref. Nr. DIN 33430:2016-07*. Berlin: Beuth.
- Fabian, G. & Briedis, K. (2009). *Aufgestiegen und erfolgreich. Ergebnisse der dritten HIS-Absolventenbefragung des Jahrgangs 1997 zehn Jahre nach dem Examen* (HIS: Forum Hochschule 2/2009). Hannover: HIS.
- Falk, S. (2010). Gleicher Lohn bei gleicher Qualifikation? Eine Analyse der Einstiegsgehälter von Absolventinnen und Absolventen der MINT-Fächer. *Beiträge zur Hochschulforschung*, 32(4), 48–71.
- Hausknecht, J. P., Day, D. V. & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57(3), 639–683.
<https://doi.org/10.1111/j.1744-6570.2004.00003.x>
- Hauss, K., Kaulisch, M., Zinnbauer, M., Tesch, J., Fräßdorf, A., Hinze, S. & Hornbostel, S. (2012). *Promovierende im Profil: Wege, Strukturen und Rahmenbedingungen von Promotionen in Deutschland; Ergebnisse aus dem ProFile-Promovierendenpanel* (iFQ-Qorking Paper, 13). Berlin: iFQ – Institut für Forschungsinformation und Qualitätssicherung.
- Hell, B. & Schuler, H. (2005). Verfahren der Studierendenauswahl aus Sicht der Bewerber. *Empirische Pädagogik*, 19(4), 361–376.
- Hell, B., Trapmann, S. & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. *Empirische Pädagogik*, 21(3), 251–270.
- Hell, B., Trapmann, S. & Schuler, H. (2008). Synopse der Hohenheimer Metaanalysen zur Prognostizierbarkeit des Studienerfolgs und Implikationen für die Auswahl- und Beratungspraxis. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 43–54). Göttingen: Hogrefe.
- Herde, C. N., Stegt, S. & Preckel, F. (2016). Auswahlverfahren für Masterstudiengänge aus Sicht von Bachelorstudierenden. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, 60(3), 145–161. <https://doi.org/10.1026/0932-4089/a000216>
- HRK Hochschulrektorenkonferenz. (2007). Ungewöhnliche Wege zur Promotion? Rahmenbedingungen und Praxis der Promotion von Fachhochschul- und Bachelorabsolventen. *Beiträge zur Hochschulpolitik* 3/2007. Bonn: HRK.
- Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33(9), 420–433.
- König, C. J., Klehe, U. C., Berchtold, M. & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment*, 18(1), 17–27. <https://doi.org/10.1111/j.1468-2389.2010.00485.x>

- Kretzschmar, A., Spengler, M., Schubert, A.-L., Steinmayr, R. & Ziegler, M. (2018). The relation of personality and intelligence – What can the Brunswik symmetry principle tell us? *Journal of Intelligence*, 6(3), 30–67. <https://doi.org/10.3390/jintelligence6030030>
- Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological Bulletin*, 127(1), 162–181. <https://doi.org/10.1037/0033-2909.127.1.162>
- Kuncel, N. R., Klieger, D. M., Connelly, B. S. & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, 98(6), 1060–1072. <https://doi.org/10.1037/a0034156>
- Merkulova, N., Melchers, K. G., Kleinmann, M., Annen, H. & Tresch, T. S. (2014). Effects of individual differences on applicant perceptions of an operational assessment center. *International Journal of Selection and Assessment*, 22(4), 355–370. <https://doi.org/10.1111/ijsa.12083>
- Oostrom, J. K., Born, M. P., Serlie, A. W. & Van Der Molen, H. T. (2010). Effects of individual differences on the perceived job relatedness of a cognitive ability test and a multimedia situational judgment test. *International Journal of Selection and Assessment*, 18(4), 394–406. <https://doi.org/10.1111/j.1468-2389.2010.00521.x>
- Posthuma, R. A., Morgeson, F. P. & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, 55(1), 1–81. <https://doi.org/10.1111/j.1744-6570.2002.tb00103.x>
- Rammstedt, B. & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K). *Diagnostica*, 51(4), 195–206. <https://doi.org/10.1026/0012-1924.51.4.195>
- Ryan, A. M. & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26(3), 565–606. <https://doi.org/10.1177/014920630002600308>
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274. <https://doi.org/10.1037/0033-2909.124.2.262>
- Schmidt-Atzert, L., Krumm, S. & Kersting, M. (2018). Evaluation der Eignungsbeurteilung. In Diagnostik- und Testkuratorium (Hrsg.), *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430* (S. 189–221). Berlin: Springer. <https://doi.org/10.1007/978-3-662-53772-5>
- Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen: Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, 6(2), 60–70. <https://doi.org/10.1026/1617-6391.6.2.60>
- Schult, J., Hofmann, A. & Stegt, S. J. (2019). Leisten fachspezifische Studierfähigkeitstests im deutschsprachigen Raum eine valide Studienerfolgsprognose? Ein metaanalytisches Update. *Zeitschrift für Entwicklungs- und Pädagogische Psychologie*, 51(1), 16–30. <https://doi.org/10.1026/0049-8637/a000204>
- Statistisches Bundesamt (2017). *Bildung und Kultur. Fachserie 11 Reihe 4.2: Prüfungen an Hochschulen 2017*. Wiesbaden: Statistisches Bundesamt.
- Stegt, S. J., Didi, H.-J., Zimmerhofer, A. & Seegers, P. K. (2018). Akzeptanz von Auswahlverfahren zur Studienplatzvergabe. *Zeitschrift für Hochschulentwicklung*, 13(4), 15–35. <https://doi.org/10.3217/zfhe-13-04/02>
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für pädagogische Psychologie*, 21(1), 11–27. <https://doi.org/10.1024/1010-0652.21.1.11>
- Truxillo, D. M., Bauer, T. N., Campion, M. A. & Paronto, M. E. (2006). A field study of the role of big five personality in applicant perceptions of selection fairness, self, and the hiring organiza-

- tion. *International Journal of Selection and Assessment*, 14(3), 269–277.
<https://doi.org/10.1111/j.1468-2389.2006.00351.x>
- Wissenschaftsrat. (2014). *Empfehlungen zu Karrierezielen und -wegen an Universitäten*. Verfügbar unter: <http://www.wissenschaftsrat.de/download/archiv/4009-14.pdf>
- Wissenschaftsrat. (2017). Strategien für die Hochschullehre. Positionspapier. Verfügbar unter: <https://www.wissenschaftsrat.de/download/archiv/6190-17.pdf>
- Ziegler, M., Danay, E. & Maaß, U. (2012). Überschätzter Nutzen? Soziale Netzwerke bei der Personalauswahl. *Wirtschaftspsychologie aktuell*, 19(3), 9–11.
- Ziegler, M., Knogler, M. & Bühner, M. (2009). Conscientiousness, achievement striving, and intelligence as performance predictors in a sample of German psychology students: Always a linear relationship? *Learning and Individual Differences*, 19(2), 288–292.
<https://doi.org/10.1016/j.lindif.2009.02.001>

Kontakt:

M.Sc. Psych. Max Knaut · Prof. Dr. Matthias Ziegler
Humboldt-Universität zu Berlin
Institut für Psychologie
Psychologische Diagnostik
Rudower Chaussee 18
12489 Berlin
E-Mail: max.knaut@hu-berlin.de
E-Mail: zieglema@hu-berlin.de

Wie gehen Hochschullehrende mit ihren Emotionen in der Lehre um? Eine Experience-sampling-Befragung zum situativen Emotionsausdruck von Freude, Angst und Ärger

Katharina Thies, Robert Kordts-Freudinger

Zusammenfassung: Hochschullehrende sind in der Lehre mit Situationen konfrontiert, die in ihnen Emotionen auslösen und damit verbunden eine Regulation der Emotionen fordern können. Für diese wenden Hochschullehrende häufig reaktionsorientierte Emotionsregulationsstrategien an, die u.a. auf den Emotionsausdruck abzielen. Unbekannt ist, ob und wie häufig Hochschullehrende spezifische Emotionen in Lehrsituationen ausdrücken und welche Zusammenhänge zwischen dem Emotionserleben und dem Emotionsausdruck bestehen. Der vorliegende Beitrag untersucht den Emotionsausdruck von Freude, Angst und Ärger bei $N = 80$ Hochschullehrenden durch eine mehrmalige Experience-sampling-Erhebung im Anschluss an $N = 802$ Lehrsituationen. Die Befunde deuten darauf hin, dass Hochschullehrende Freude überwiegend authentisch ausdrücken, während sie Angst und Ärger tendenziell in der Lehrveranstaltung nicht ausdrücken. Für die spezifischen Emotionen ergeben sich unterschiedliche intra- und interindividuelle Zusammenhänge (d.h. innerhalb und zwischen den Hochschullehrenden). So neigt etwa eine Hochschullehrperson bei einem starken Angststerben in einer Lehrsituation dazu, den Emotionsausdruck zu unterdrücken; eine Tendenz, die sich bei Hochschullehrenden mit einem allgemein starken Angststerben nicht fand. Die Ergebnisse sprechen für eine differenzierte Analyse von intra- und interindividuellen Zusammenhängen emotionsbezogener Vorgänge in Lehrsituationen.

Schlüsselwörter: Hochschullehre, Emotionen, Emotionsausdruck, Emotionsregulation, Experience-Sampling Methode

How do university lecturers handle their emotions during teaching? An experience-sampling study of the situational emotion expression of joy, anxiety and anger

Abstract: In their teaching, university lecturers are confronted with situations that can trigger emotions in them and thus demand a regulation of the emotions. For these situations, they often apply reaction-oriented emotion regulation strategies, which, among other things, target the expression of emotions. It is not known whether and how often university lecturers express specific emotions in teaching sessions and what relations exist between the experience of emotions and the expression of emotions. This study investigates the emotional expression of joy, anxiety, and anger among $N = 80$ university lecturers by using an experience-sampling survey following $N = 802$ lectures. The findings indicate that university lecturers express joy mostly authentically, while they tend to suppress the expression of anxiety and anger. Different intra- and interindividual relations (i.e., within and between

the instructors) were found for the specific emotions. For example, a university lecturer with a strong experience of anxiety in a teaching session tends to suppress the emotional expression; a tendency that was not present among university lecturers with a generally strong experience of anxiety. The results emphasize a differentiated analysis of intra- and interindividual correlations of emotional processes in teaching situations.

Keywords: university teaching, emotions, emotion expression, emotion regulation, experience-sampling method

1 Einleitung

Hochschullehrende bewerten den Umgang mit ihren Emotionen als wesentlichen Teil ihrer beruflichen Praxis (Ogbonna & Harris, 2004). Häufig erleben Hochschullehrende Emotionen in der Lehre (z.B. Kordts-Freudinger, 2017), während dieser sie insbesondere in der unmittelbaren Interaktion mit Studierenden wahrscheinlich und vornehmlich von dem (Lern-)Verhalten der Studierenden ausgelöst werden (z.B. Hagenauer & Volet, 2014a; Postareff & Lindblom-Ylänne, 2011). Es kann somit angenommen werden, dass Lehrsituationen im besonderen Maße einen Umgang mit Emotionen von den Hochschullehrenden fordern.

Für Berufe im Service- wie auch Bildungsbereich wird angenommen, dass die nach außen gezeigten Emotionen nicht immer den tatsächlich erlebten Emotionen entsprechen (Ashfort & Humphrey, 1993). Dies trifft auch auf Hochschullehrende zu, die insbesondere den Ausdruck negativer Emotionen in der Lehre herabregulieren oder gar verbergen (Hagenauer & Volet, 2014b; Kordts-Freudinger, 2017; Mendzheritskaya, Hansen, Scherer & Horz, 2018). Eine Limitierung bisheriger Studien besteht in Bezug auf ihre einmalige, retrospektive Messmethodik, die Aussagen auf individueller Ebene der Hochschullehrpersonen einschränken. Eine weitere Limitierung resultiert aus einer mangelnden Fokussierung spezifischer Emotionen, die Angaben zur Qualität einzelner Emotionen im Lehrgeschehen verwehrt.

Der vorliegende Beitrag untersucht den alltäglichen Emotionsausdruck von Hochschullehrenden für eine Auswahl spezifischer Emotionen. Er geht den Forschungsfragen nach, ob und wie Hochschullehrende ihre in der Lehrveranstaltung erlebten spezifischen Emotionen den Studierenden gegenüber zum Ausdruck bringen und wie der jeweilige Emotionsausdruck mit dem spezifischen Emotionserleben zusammenhängt. Die Datenerhebung folgte einem Event-Sampling, einem Verfahren der Experience-sampling-Methode (Csikszentmihalyi, Larson & Prescott, 1977), mittels dessen das Emotionserleben und der Emotionsausdruck der Hochschullehrenden für die spezifischen Emotionen Freude, Angst und Ärger – die wegen ihrer Häufigkeit in der Lehre sowie wegen ihrer Spezifität ausgewählt wurden – direkt im Anschluss an Lehrsituationen erhoben und anschließend unter Beachtung der genesteten Datenstruktur (Messzeitpunkte der Hochschullehrpersonen; Personenebene der Hochschullehrenden) ausgewertet wurden.

2 Stand der Forschung

2.1 Emotionen von Hochschullehrenden in der Lehre

Emotionen können als episodenhafte Zustände definiert werden, die sich in zusammenhängenden Veränderungen in Komponenten des Organismus als Reaktion auf einen vom Individuum als bedeutsam eingestuftem Reiz äußern (Scherer, 2005, S. 23). Die Komponenten umfassen kognitive Prozesse (wie Einschätzungen, Beurteilungen), Affekt (ein positives oder negatives Empfinden), physiologische Vorgängen (wie zentralnervöse Aktivierung, periphere Erregung), Verhaltenstendenzen (darunter Handlungsvorbereitungen und -reaktionen) und den Ausdruck (etwa Mimik, Gestik, Stimmlage) (Scherer, 2003). Dabei unterscheiden sich spezifische Emotionen, wie etwa Freude, Angst oder Ärger, in ihrer Qualität. Dies meint zum Beispiel für die jeweilige Emotion spezifische Veränderungen in den Komponenten (wie dem erfreuten, ängstlichen oder verärgerten Gesichtsausdruck) (Scherer, 2003), eine typische Emotionsvalenz (positiv, negativ) und einen eindeutigen Zeit- und Objektbezug (etwa retrospektive, gegenwärtige, zukünftige Ergebnisse, Tätigkeiten, Ereignisse) (Pekrun, 2006). Die Intensität umfasst das subjektive Gefühl, also das intensive Wahrnehmen einer spezifischen Emotion, etwa in Abgrenzung zu allgemeinen, zeitlich länger überdauernden und weniger intensiven Gefühlslagen oder Stimmungen (z.B. Rosenberg, 1998).

Über die Emotionen von Hochschullehrenden ist bereits einiges bekannt. Qualitative Befragungen, die Emotionen von Hochschullehrenden offen aus Interviewantworten ermittelt haben, fanden häufig positive Emotionen wie Freude und Zufriedenheit (Hagenauer & Volet, 2014a; Postareff & Lindblom-Ylänne, 2011), sowie auch negative emotionale Äußerungen wie Abneigung, Unlust (Postareff & Lindblom-Ylänne, 2011), Frustration und Ärger (Hagenauer & Volet, 2014a). Quantitative Erhebungen, die *spezifische* Emotionen von Hochschullehrenden geschlossen erfasst haben, fanden am häufigsten die positiven Emotionen Freude, Stolz (Kordts-Freudinger, 2017; Stupnisky, Pekrun & Lichtenfeld, 2014) und Zufriedenheit (Kordts-Freudinger, 2017), und unter den negativen Emotionen Ärger (Kordts-Freudinger, 2017; Stupnisky et al., 2014), Frustration, Langeweile (Kordts-Freudinger, 2017) und Angst (Stupnisky et al., 2014). Hochschullehrende erleben demnach eine Bandbreite an Emotionen in der Lehre, wobei insgesamt positive Emotionen die negativen Emotionen in ihrer Häufigkeit zu überwiegen scheinen (z.B. Hagenauer & Volet, 2014a; Kordts-Freudinger, 2017; Postareff & Lindblom-Ylänne, 2011; Stupnisky et al. 2014).

Wie eingangs beschrieben, stellt nicht nur das Erleben eine Komponente von Emotionen dar, sondern auch der Ausdruck. Befunde über den Emotionsausdruck von Hochschullehrenden werden nachfolgend aufgezeigt.

2.2 Der Emotionsausdruck von Hochschullehrenden in der Lehre

Der Emotionsausdruck hilft Individuen dabei, über das emotionale Erleben zu informieren und Reaktionen im Umfeld hervorzurufen (Horstmann, 2003). Darüber trägt er dazu bei, soziale Beziehungen zu stärken oder die eigene soziale Rolle abzusichern (Fischer & Manstead, 2008). Der Ausdruck einer Emotion erfolgt nicht ungefiltert, d.h. Individuen

treffen Einschätzungen über die (Un-)Angemessenheit eines Emotionsausdrucks in einer Situation (Scherer, 2003), etwa basierend auf implizit erworbenen Darbietungsregeln (so genannten *display rules*; Ekman & Friesen, 1969). Formelle und informelle Normen zum Emotionsausdruck bestehen auch im beruflichen Kontext (Ekman & Friesen, 1969; Hochschild, 1979), was besonders auf serviceorientierte Berufe und auch solche im Bildungsbereich zutrifft (Ashfort & Humphrey, 1993). Auch Hochschullehrende bewerten den Ausdruck von Emotionen in der Lehre. So stuft diese Personengruppe einen offenen, unveränderten Ausdruck ihrer positiven Emotionen in der Lehre als unproblematisch und angemessen ein, wohingegen sie einen offenen Ausdruck negativer Emotionen als unprofessionell oder unangemessen bewertet (z.B. Constantini & Gibbs, 2004; Hagenauer & Volet, 2014b; Ogbonna & Harris, 2004). Weitere Studienergebnisse deuten darauf hin, dass Hochschullehrende ihren Emotionsausdruck in der Lehre kontrollieren, d.h. bewusst steuern (Kordts-Freudinger & Thies, 2018). Dabei werden positive Emotionen von den Hochschullehrenden überwiegend offen und authentisch geäußert (Hagenauer & Volet, 2014b), während negative Emotionen von ihnen im Ausdruck tendenziell reguliert werden (Hagenauer & Volet, 2014b; Kordts-Freudinger, 2017).

Allgemein umfasst die Emotionsregulation die unbewussten oder bewussten Prozesse zur Initiierung, Erhaltung und Modifikation des Auftretens, der Intensität und Dauer sowie des Ausdrucks von Emotionen (z.B. Gross, 2008; Koole, 2009). Dies schließt sowohl ein Herauf- als auch ein Herunterregulieren von Emotionen ein: Für ein hedonistisches Wohlbefinden werden etwa positive Emotionen intensiviert und negative Emotionen gemindert, für instrumentelle Zwecke werden etwa auch positive Emotionen reduziert oder negative Emotionen verstärkt (Gross, 2015). Strategien zur Emotionsregulation setzen an verschiedenen Zeitpunkten im Prozess der Emotionsentstehung an (Gross, 2008, 2015). Während entstehungsorientierte Strategien an der zeitlich früheren Auslösung von Emotionen wirken, wirken reaktionsorientierte Strategien an der zeitlich späteren emotionalen Reaktion. Strategien der erstgenannten Gruppe sind die Situationsauswahl (z.B. Aufsuchen oder Vermeiden von emotionsauslösenden Gegebenheiten), Situationsmodifikation (z.B. Umgestaltung emotionsauslösender Umstände), Aufmerksamkeitslenkung (z.B. Steuerung der Aufmerksamkeit auf emotional weniger bedeutsame Merkmale) und kognitive Veränderung (z.B. Neu-/Umbewertung der emotionsauslösenden Reize) (vgl. Gross, 2008; 2015). Zu den letztgenannten Strategien gehört neben der Veränderung der physiologischen Erregung oder des subjektiven Erlebens die Regulation des Emotionsausdrucks, etwa in der Mimik, mittels Intensivierung oder Unterdrückung der Ausdruckskomponente (vgl. Brandstätter, Schüler, Puca & Lozo, 2013).

Für individuelle Unterschiede in der Emotionsregulation wie auch der Strategieverwendung werden vielfältige Ursprünge diskutiert, darunter die Entwicklung und Sozialisation, Persönlichkeitseigenschaften, emotionale Ziele und Funktionen oder auch Anforderungen der jeweiligen Situation (Thompson, 1994). Während zu Beginn des Lebens die Regulation von Emotionen fremdgesteuert erfolgt (etwa durch beruhigendes Verhalten der Eltern), entsteht im Verlauf der Entwicklung Wissen um Emotionen und die Fähigkeit zur Selbstregulation von Emotionen (Kullik & Petermann, 2012). Emotionsregulation ist zudem verbunden mit den Persönlichkeitseigenschaften, so korreliert Extraversion positiv mit der Regulationsstrategie der kognitiven Neubewertung und negativ mit der Unterdrückung von

Emotionen (Matsumoto, 2006). In der letztgenannten Studie vermittelten Persönlichkeitseigenschaften gar kulturelle Unterschiede in der Emotionsregulation. Weitere Unterschiede in der Emotionsregulation resultieren aus den oben bereits genannten individuellen emotionalen oder sozialen Zielen (wie Wohlbefinden, instrumentelle Zwecke; Gross, 2015), und/oder aus den konkreten Anforderungen und Verhaltensnormen der jeweiligen Situation (wie soziale, berufliche Darbietungsregeln; Ekman & Friesen, 1969).

Wie erste empirische Studien ergaben, regulieren Hochschullehrende ihre Emotionen häufig reaktionsorientiert, indem sie den Ausdruck unterdrücken (Hagenauer & Volet, 2014b; Kordts-Freudinger, 2017) oder gar verbergen (Hagenauer & Volet, 2014b). Gelegentlich regulieren Hochschullehrende den emotionalen Ausdruck auch, indem sie eine andere als die tatsächlich erlebte Emotion äußern (Hagenauer & Volet, 2014b). Als Gründe für die Veränderung des Ausdrucks negativer Emotionen finden sich in Befragungen der Hochschullehrenden vor allem instrumentelle Zwecke, wie etwa der Erhalt der Authentizität oder die Erreichung von Lehrzielen (Hagenauer & Volet, 2014b).

Dass der emotionale Ausdruck von Hochschullehrenden Darbietungsregeln folgt, belegen auch die Studien zu kulturellen Unterschieden des Emotionsausdrucks (Hagenauer, Gläser-Zikuda & Volet, 2016; Mendzheritskaya et al., 2018). Für Hochschullehrende in Deutschland wurde seltener ein stärkerer Ausdruck positiver Emotionen und häufiger ein kontrollierter Ausdruck negativer Emotionen gefunden als für Hochschullehrende in Australien (Hagenauer et al., 2016). Im Vergleich zu Russland zeigen Hochschullehrende in Deutschland unter anderem einen schwächeren Ausdruck negativer Emotionen (Menzheritskaya et al., 2018). Weitgehend ähnliche Muster des Emotionsausdrucks weisen deutsche und russische Hochschullehrende in typischen Lehrsituationen auf: In Lehrveranstaltungen zeigen Hochschullehrende ihre Emotionen offenbar seltener authentisch und unterdrücken diese stärker als in Sprechstunden mit Studierenden (Menzheritskaya et al., 2018).

Weitere Studienergebnisse lassen auf einen Zusammenhang zwischen dem Emotionsausdruck und dem Emotionserleben schließen. So korrelierte die Häufigkeit eines unterdrückten Ausdrucks negativer Emotionen positiv mit der Stärke der negativen Emotionen (Kordts-Freudinger, 2017), und die Häufigkeit eines unkontrollierten Ausdrucks positiver Emotionen positiv mit der Stärke positiver Emotionen (Kordts-Freudinger, 2018). Als Ursache für diesen Zusammenhang kommen verschiedene Faktoren infrage, wie etwa die habituelle Tendenz zum Emotionsausdruck, die Emotionsintensität und -valenz. So zeigte ein Experiment von Gross, John und Richards (2000), dass Personen, die habituell zu einem starken Emotionsausdruck tendieren, ihre Emotionen unabhängig von der Intensität stärker ausdrücken als Personen, die habituell zu einem schwächeren Ausdruck neigen. Bei negativen Emotionen war dieser Effekt nur bei einem sehr intensiven Emotionserleben deutlich, was damit erklärt wurde, dass der Ausdruck negativer Emotionen weitaus stärker durch soziale und berufliche Normen, und damit situativen Faktoren, beeinflusst wird als der Ausdruck positiver Emotionen (Gross et al., 2000). Neben der habituellen Neigung zum Ausdruck von Emotionen scheint es somit wahrscheinlich, dass die Emotionsstärke und -valenz den Emotionsausdruck beeinflussen, wenn auch grundsätzlich andere Beeinflussungsrichtungen – wie die der Emotionsstärke durch den Emotionsausdruck – oder auch die wechselseitige Beeinflussung (z.B. Coles, Larsen, & Lench, 2019) nicht auszuschließen sind.

Zusammengenommen weisen die Befunde daraufhin hin, dass Hochschullehrende insbesondere negative Emotionen in der Lehre unterdrücken und/oder gar verbergen, entsprechend gelernter Darbietungsregeln. Ferner scheint die Art der Emotionsregulation über den Emotionsausdruck mit der Stärke des Emotionserlebens zusammenzuhängen. Weitgehend unerforscht ist hingegen, wie Hochschullehrende *spezifische* Emotionen (versus allgemein positive oder negative Emotionen) in der konkreten Lehrsituation ausdrücken und wie das Erleben der spezifischen Emotionen mit dem gezeigten Emotionsausdruck zusammenhängen.

2.3 Messung von Emotionen und Emotionsregulation

Für die Emotionsmessung werden häufig Verfahren eingesetzt, die Emotionen einmalig und retrospektiv erfassen (vgl. Robinson & Clore, 2002). Diese können mit Erinnerungsfehlern durch nachträgliche Attributionen oder kognitive Beurteilungen der Befragten verbunden sein und darüber zu einer Überschätzung der tatsächlich erlebten Emotionen führen (sogenannter *intensity bias*; Buehler & McFarland, 2001; Robinson & Clore, 2002). Zur Reduzierung von Verzerrungseffekten wird eine mehrmalige Emotionsmessung möglichst direkt zum Zeitpunkt des Emotionserlebens bevorzugt (z.B. Augustine & Larsen, 2012). Diese Herangehensweise wird ebenfalls für die Erfassung der Emotionsregulation empfohlen (Dorn, Spindler & Kullik, 2013), da diese ebenso wie die Emotionen selbst episodisch und situativ erfolgt (Thompson, 1994). Die Experience-sampling-Methode (ESM; Csikszentmihalyi et al., 1977) stellt ein Verfahren für die Messung dieser kurzzeitigen und situativen Zustände dar. Die Methode ermöglicht wiederholte Messungen über einen bestimmten Zeitraum zu festgelegten oder unbestimmten Zeiten (Time-Sampling) oder zu definierten Ereignissen (Event-Sampling). Dabei rücken die Messzeitpunkte möglichst dicht an das Alltagsgeschehen der Befragten. Die ESM gilt insgesamt als eine reliable und valide Methode (Csikszentmihalyi & Larson, 1987), auch in der Anwendung von Einzelitems (Gogol et al., 2014).

Ein weiterer Vorteil der ESM betrifft die Datenstruktur: Die wiederholte Messung führt zu mehreren Messzeitpunkten pro Befragungsperson und ermöglicht dadurch intraindividuelle und interindividuelle Auswertungen (Eid, Gollwitzer & Schmitt, 2015). Intraindividuelle Analysen berücksichtigen die Variation auf Ebene der einzelnen Messzeitpunkte pro Person (Ebene 1: Messzeitpunktebene) und erlauben damit eine Untersuchung von Zusammenhängen, die *innerhalb* von Personen vorliegen; interindividuelle Analysen ziehen die Variation zwischen Personen (Ebene 2: Personenebene) heran und ermöglichen dadurch eine Analyse von Zusammenhängen, die zwischen Personen bestehen (Hamaker, 2011; Murayama et al., 2017).

Bereits mehrfach und mit bedeutsamen Befunden wurde die ESM zur Untersuchung von Emotionen im Bildungskontext eingesetzt: Entsprechende Studien zu alltäglichen Emotionen haben gezeigt, dass die einmalige, retrospektive Erhebung die Emotionen überschätzt (z.B. Bieg, Goetz & Lipnevich, 2014; Götz, Becker et al., 2015; Keller, Chang, Becker, Götz & Frenzel, 2014; Roos et al., 2015). Die in ESM-Studien erhobenen Emotionen zeigten ferner insgesamt hohe Anteile intraindividuelle Variation; d.h. Unterschiede in den Emotionsintensitäten lagen vornehmlich innerhalb der Hochschullehrenden auf der Ebene der Messzeitpunkte, was auf einen starken situativen Bezug hinweist (z.B. Götz, Frenzel,

Stöger & Hall, 2010; Götz, Sticca, Pekrun, Murayama & Elliot, 2015; Thies & Kordts-Freudinger, 2019a, 2019b). Diese Situationssensitivität wurde auch für die Emotionsregulationsstrategie der Unterdrückung aufgezeigt (Catterson, Eldesouky & John, 2017). ESM-Studien ergaben ferner Befunde über intraindividuelle Prozesse, u.a. hinsichtlich kognitiver Vorläufer der Emotionen (z.B. Götz et al., 2010; Götz, Sticca et al., 2015; Thies & Kordts-Freudinger, 2019a, 2019b). Inter- und intraindividuell vergleichende Studien über emotionsbezogene Zusammenhänge verweisen zudem auf die statistische und interpretative Unabhängigkeit der Analyseformen (z.B. Götz et al., 2014); die Richtung der Zusammenhänge muss sich dabei nicht zwangsläufig unterscheiden (z.B. Götz, Sticca et al., 2015). Ebenfalls durch ESM-Studien bestätigt wurde die Neigung von Individuen, positive Emotionen herauf- und negative Emotionen herunterzuregulieren (Deng, Sang, & Luan, 2013). Die Befunde der letztgenannten Studie betonen insbesondere für die Reduktion negativer Emotionen deren Adaptivität für das tägliche Erleben, mit kurzfristigen positiven Effekten auch auf das Erleben positiver Emotionen (vgl. Deng et al., 2013). Im Kontext der Emotionen von Hochschullehrenden wurde die ESM allerdings noch selten eingesetzt (vgl. Thies & Kordts-Freudinger, 2019a), so dass der situative Emotionsausdruck und dessen Veränderung bei Hochschullehrenden noch weitgehend unerforscht ist.

3 Ziel und Forschungsfragen der Studie

Aus dem Forschungsstand geht hervor, dass Hochschullehrende auf ihren Emotionsausdruck in der Lehre einwirken. Dabei scheint eine gewisse Variation im Emotionsausdruck vorzuliegen. So beeinflussen Hochschullehrende den Ausdruck von positiven und negativen Emotionen unterschiedlich (vgl. Hagenauer & Volet, 2014b; Kordts-Freudinger, 2017; Kordts-Freudinger & Thies, 2018), und sie wählen einen unterschiedlichen Emotionsausdruck in verschiedenen Lehrsituationen wie etwa die der Lehrveranstaltung oder der Sprechstunde (vgl. Mendzheritskaya et al., 2018). Gelegentlich neigen Hochschullehrende zu einer Simulation nicht erlebter Emotionen (vgl. Hagenauer & Volet, 2014b; Mendzheritskaya et al., 2018). Weithin ungeklärt ist, ob und wie häufig Hochschullehrende spezifische Emotionen ausdrücken oder diese von ihnen über den Ausdruck simuliert werden. Es ist wahrscheinlich, dass der situative Emotionsausdruck der Hochschullehrenden von den bisherigen Befunden abweicht, die auf einmaligen retrospektiven Befragungen (Hagenauer & Volet, 2014b; Kordts-Freudinger, 2017; Kordts-Freudinger & Thies, 2018) oder experimentellen Settings (Menzheritskaya et al., 2018) basieren und damit eher den habituellen, situationsübergreifenden Emotionsausdruck der Hochschullehrenden abbilden. Ziel der vorliegenden Studie es, den situativen Emotionsausdruck der Hochschullehrenden für eine Auswahl spezifischer Emotionen – Freude, Angst und Ärger – in Lehrveranstaltungen zu untersuchen. Die Auswahl der drei Emotionen erfolgte nach (1) ihrer Häufigkeit, (2) ihrem typischen Emotionsausdruck und (3) ihrem eindeutigen Objektfokus: (1) Freude, Angst und Ärger werden von den Hochschullehrenden häufig in der Lehre erlebt (Kordts-Freudinger, 2017; Stupnisky et al., 2014), so dass für diese Emotionen ein häufiger Emotionsausdruck angenommen werden kann. (2) Als sogenannte Primäraffekte gehen die drei Emotionen jeweils mit einem eindeutigen und weitgehend universellen Emotionsausdruck einher (Ek-

man, 1992; Izard, 1981). Aufgrund dieser Eigenschaft eignen sich die drei Emotionen besonders gut für die Untersuchung des Emotionsausdrucks bzw. dessen Veränderung, welche aus verschiedenen Gründen in Lehrsitungen indiziert sein kann (s. Kapitel 2.2). (3) Die Emotionen unterscheiden sich deutlich in ihrem Objektfokus. Freude gilt als Reaktion auf erwünschte Ereignisse, Angst als Reaktion auf unerwünschte oder unsichere prospektive Ereignisse, Ärger als Reaktion auf durch Eigen- oder Fremdverhalten verursachte unerwünschte Ereignisse (Lazarus, 1991; Ortony, Clore & Collins, 1988). So erleben Hochschullehrende Freude bei Lernerfolgen der Studierenden (z.B. Postareff & Lindblom-Ylänne, 2011), Angst bei vorliegenden Unsicherheiten (z.B. Meanwell & Kleiner, 2014), und Ärger bei unerwünschtem studentischen Verhalten (z.B. Hagenauer & Volet, 2014). Die Lehrveranstaltung wurde als Kontext gewählt, da diese offenbar mit einer starken Veränderung des Emotionsausdrucks verbunden ist (vgl. Mendzheritskaya et al., 2018). Ausgehend von einer Variation des situativen Ausdrucks spezifischer Emotionen wird in einem ersten Schritt basierend auf dem Erleben von Freude, Angst und Ärger untersucht, wie häufig Hochschullehrende einen bestimmten Emotionsausdruck für diese Emotionen anwenden.

Forschungsfrage 1: Wie häufig wenden Hochschullehrende einen verstärkten, authentischen, unterdrückten oder simulierten Emotionsausdruck für die in der Lehrsitung erlebten Emotionen Freude, Angst und Ärger an?

Weiter lässt sich aus dem Forschungsstand ein Zusammenhang zwischen dem Emotionserleben und dem Emotionsausdruck ableiten. So war eine häufige Unterdrückung des Ausdrucks negativer Emotionen mit stärkeren negativen Emotionen korreliert (Kordts-Freudinger, 2017), der Ausdruck positiver Emotionen mit positiven Emotionen (Kordts-Freudinger & Thies, 2018). Verbunden ist dieser Zusammenhang vermutlich mit habituellen Tendenzen, der Emotionsstärke und -valenz (Gross et al., 2000). Die Limitierungen der vorliegenden Befunde bestehen auch hier durch einen mangelnden Fokus auf spezifische Emotionen und durch die ausgewählte Erhebungsmethodik. In einem zweiten Schritt werden daher die Zusammenhänge zwischen dem Erleben spezifischer Emotionen und den situativen Veränderungen des Emotionsausdrucks innerhalb der Hochschullehrpersonen und zwischen den Hochschullehrenden untersucht.

Forschungsfrage 2: Welche Zusammenhänge bestehen zwischen dem Erleben von Freude, Angst und Ärger und dem Ausdruck dieser Emotionen auf (a) der Messzeitpunktebene der Lehrsitungen (innerhalb der Hochschullehrperson) und (b) der Personenebene (zwischen den Hochschullehrenden)?

Die Messzeitpunktebene der konkreten Lehrsitungen adressiert dabei die intraindividuelle Variation innerhalb einer Hochschullehrperson, also das episodenhafte, situative Emotionserleben und den episodenhafte, situativen Emotionsausdruck, welche vermutlich stärker durch die jeweiligen emotionalen und sozialen Zielen der Hochschullehrperson und die konkreten Anforderungen (wie etwa der Angemessenheit eines Emotionsausdrucks) der Situation beeinflusst werden. Die Personenebene erfasst die interindividuelle Variation zwischen Hochschullehrenden, also ihre eher situationsübergreifende Tendenz zu einem Emotionserleben und -ausdruck, welche vermutlich eher auf dispositionelle, habituelle Merkmale (etwa Persönlichkeitseigenschaften) der Hochschullehrenden zurückgeht. Die beiden

Ebenen erfassen somit verschiedene Vorgänge des Emotionserlebens und Emotionsausdrucks, so dass die Zusammenhänge nicht nur emotionspezifisch sondern auch intra- und interindividuell unterschiedlich ausfallen könnten.

4 Methode

4.1 Stichprobe

Achtzig Hochschullehrende von deutschsprachigen Universitäten und Fachhochschulen nahmen im Sommersemester 2018 an der Studie teil. Die Akquise zur Teilnahme erfolgte mittels einer E-Mail-Anfrage, die primär über interne hochschuldidaktische Einrichtungen und über hochschulweite Netzwerke (u.a. www.hdw-nrw.de) versandt wurde. Die Anfrage enthielt eine Bitte um Weiterleitung an potentielle Teilnehmer*innen, so dass die Anzahl der erreichten Hochschullehrenden und eine Rücklaufquote nicht herzuleiten waren. Die Teilnahme wurde nicht incentiviert. Unter den Teilnehmenden waren 59 Frauen, 19 Männer und zwei Personen ohne Angabe des Geschlechts. Der überwiegende Anteil der Teilnehmenden verortete sich in den Human-, Sozial- und Kulturwissenschaften (59%), der kleinere Teil in den Ingenieurs-, Wirtschafts-, Naturwissenschaften, im Gesundheitswesen oder der Architektur (41%). 21% der Teilnehmenden waren Professor*innen oder Junior-Professor*innen, 68% wissenschaftliche Mitarbeitende, 9% Lehrkräfte für besondere Aufgaben und 2% Lehrbeauftragte. 76% der Teilnehmenden waren an Universitäten beschäftigt, 19% an Fachhochschulen und 5% an anderen Hochschulen wie Pädagogische Hochschulen oder Kunsthochschulen. Der Großteil der Befragungsgruppe (66%) war unter oder 40 Jahre alt¹, 43% hatten weniger oder 5 Jahre Lehrerfahrung². 68% der Teilnehmenden nannten eine Vollzeitbeschäftigung.

4.2 Datenerhebung und Instrumente

Die Datenerhebung erfolgte durch eine Paper-Pencil-Befragung mit zwei Fragebogen, entsprechend der ESM: Der erste Fragebogen erfasste einmalig die soziodemographischen Angaben (Geschlecht, Alter, Hochschulform, Personalkategorie, Stellenumfang, Anzahl der Dienstjahre) der Hochschullehrenden. Der zweite Fragebogen wurde den Hochschullehrenden in mehrfacher Ausführung zur Verfügung gestellt und erfasste ihr Emotionserleben und ihren Emotionsausdruck. Einem Event-Sampling folgend, füllten die Hochschullehrenden diesen inhaltlich gleichbleibenden Fragebogen jeweils möglichst direkt im Anschluss an bis zu 15 Lehrsitzen (d.h. in Einzelterminen von Lehrveranstaltungen) aus. Der Fragebogen umfasste folgende Einzelitems (adaptiert von Keller & Becker, 2018):

- 1 Das Alter wurde in Kategorien erfasst: ≤ 30 Jahre: $N = 18$; 31-40 Jahre: $N = 35$; 41-50 Jahre: $N = 14$; 51-60 Jahre: $N = 9$; ≥ 61 Jahre: $N = 4$. Gründe für diese Art der Abfrage: (1) Wahrung der Anonymität einer erwarteten kleinen Stichprobe, (2) Zusammenhänge mit dem Alter waren nicht Forschungsgegenstand der Studie.
- 2 Die Lehrerfahrung wurde aus den oben genannten Gründen ebenfalls in Kategorien erfasst: ≤ 5 Jahre: $N = 34$; 6-15 Jahre: $N = 34$; 16-26 Jahre: $N = 9$; ≥ 27 Jahre: $N = 3$.

Das *Emotionserleben* von Freude, Angst und Ärger wurde mit der Frage „Wie stark haben Sie Freude / Angst / Ärger erlebt?“ auf der Ratingskala „1: überhaupt nicht, 2: etwas, 3: mäßig, 4: stark, 5: sehr stark“ erhoben.

Der *Emotionsausdruck* wurde bei vorliegendem Emotionserleben erhoben. Die Frage lautete „Wenn Sie etwas / mäßig / stark / sehr stark Freude / Angst / Ärger erlebt haben: Haben Sie Ihre Freude / Ihre Angst / Ihren Ärger den Studierenden gegenüber zum Ausdruck gebracht (gezeigt, verbalisiert, o.a.)?“ Erhoben wurde auf der Ratingskala: „Ja, und zwar... 1: viel schwächer, 2: etwas schwächer, 3: genau wie erlebt, 4: etwas stärker, 5: viel stärker ... als sie/er war“. Die Antwortstufen wurden nach der Befragung neu zusammengefasst: „etwas stärker“ und „viel stärker“ zu einem *verstärkten Ausdruck*, „etwas schwächer“ und „viel schwächer“ zu einem *unterdrückten Ausdruck*. Die Antwortstufe „genau wie erlebt“ wurde als *authentischer Ausdruck* kodiert.

Die *Simulation des Emotionsausdrucks* wurde bei ausbleibendem Emotionserleben wie folgt erfragt: „Wenn Sie ‚überhaupt nicht‘ Freude / Angst / Ärger erlebt haben: Haben Sie trotzdem Freude / Angst / Ärger den Studierenden gegenüber zum Ausdruck gebracht (gezeigt, verbalisiert, o.a.)?“ Die Ratingskala war „1: nein, 2: ja, etwas, 3: ja, viel“.

4.3 Datenauswertung

Nach Abschluss der Erhebungsphase wurden insgesamt $N = 802$ mehrmalige Fragebogen ($M = 10.02$ pro Lehrperson) und $N = 80$ einmalige Fragebogen ausgewertet. Als Grundlage für die Beantwortung der Forschungsfragen wurden zunächst die absoluten Häufigkeiten und aggregierten Mittelwerte (d.h. ein zusammengefasster Mittelwert aus zuvor für die einzelnen Hochschullehrenden separat berechneter Mittelwerte) jeweils separat für die Emotionen Freude, Angst und Ärger berechnet.

Für die Beantwortung der ersten Forschungsfrage wurden anschließend die Häufigkeiten des Emotionsausdrucks separat für jede Emotion ermittelt. In Vorbereitung der zweiten Forschungsfrage wurden die Intraklassenkorrelationen für die Emotionen und die jeweiligen Ausdrucksweisen berechnet. Die Intraklassenkorrelation (*intraclass correlation*; ICC) ist ein Maß für den Grad der Abhängigkeit der Messwerte und gibt an, wieviel Prozent der Merkmalsvarianz (hier des Emotionserlebens) durch Unterschiede auf der Personenebene (hier zwischen den Hochschullehrenden) erklärt werden (Eid et al., 2015). Bereits ab einer ICC von 0.1 wird eine Analyse der einzelnen Ebenen (hier: Situations- und Personenebene) empfohlen, um die Abhängigkeit der Messwerte zu berücksichtigen und Fehlinterpretationen der Ergebnisse zu vermeiden (Geiser, 2011). Diese liegen etwa dann vor, wenn Zusammenhänge, die auf Personenebene (Ebene 2 hier: zwischen Hochschullehrenden) bestehen, auf Ebene der Individuen (Ebene 1; hier: innerhalb eines/einer Hochschullehrende/n) interpretiert werden (sogenannte *ecological fallacy*; Robinson, 1950).

Für die Beantwortung der zweiten Forschungsfrage wurden separat für die Emotionen jeweils Within- und Between-Korrelationen zwischen dem Emotionserleben und dem Emotionsausdruck berechnet. Within-Korrelationen beschreiben die Beziehung von Variablen innerhalb einer Messeinheit (hier: Zusammenhänge, die innerhalb einer/eines Hochschullehrende/n bestehen); Between-Korrelationen geben Auskunft über die Beziehung von Variablen zwischen Messeinheiten (hier: Zusammenhänge, die zwischen den Hochschullehrenden vorliegen) (*within-group and between-group correlations*; Snijders & Bosker,

2011). Die Berechnungen der Korrelationen erfolgten mit der Statistiksoftware *Mplus 7.4* (Muthén & Muthén, 1998-2015).

5 Ergebnisse

5.1 Emotionserleben und Emotionsausdruck in Lehrsitungen

Die Hochschullehrenden machten rund 800 Angaben zum Erleben der einzelnen Emotionen Freude ($N = 802$), Angst ($N = 799$) und Ärger ($N = 800$). In 771 Lehrsitungen (96%) erlebten sie nach eigener Aussage Freude, in 316 Lehrsitungen (40%) Angst und in 339 Lehrsitungen (42%) Ärger zwischen *etwas* bis *sehr stark* (s. Abbildung 1). Die aggregierten Mittelwerte für das Emotionserleben betragen für Freude $M = 3.58$ ($SD = 0.61$), für Angst $M = 1.59$ ($SD = 0.48$) und für Ärger $M = 1.80$ ($SD = 0.60$). Hochschullehrende scheinen also in nahezu jeder Lehrsitung Freude zu erleben und diese tendenziell intensiv. Weniger häufig und intensiv scheinen Hochschullehrende in den Lehrsitungen Angst oder Ärger zu erleben.

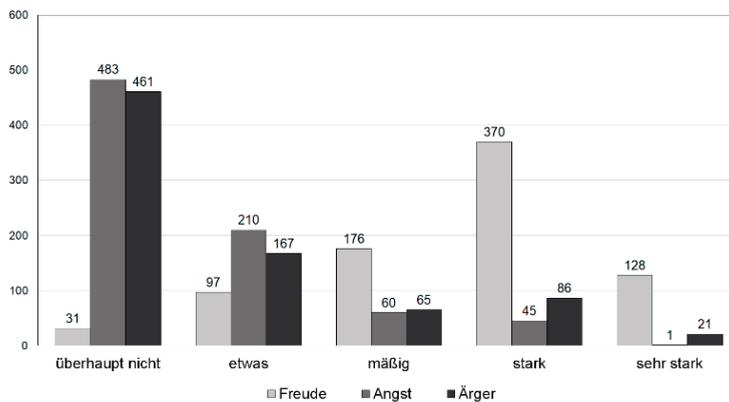


Abbildung 1: Häufigkeiten des Emotionserlebens von Freude ($N = 802$), Angst ($N = 799$) und Ärger ($N = 800$)

Im Hinblick auf die erste Forschungsfrage wurden Häufigkeiten des Emotionsausdrucks für Freude, Angst und Ärger analysiert. Die Daten ergeben, dass die Personen ihre Freude in den meisten Lehrveranstaltungen authentisch, also wie von ihnen erlebt, den Studierenden gegenüber äußerten (vgl. Abbildung 2). In einigen Fällen gaben sie an, ihre Freude stärker als erlebt ausgedrückt zu haben, für weitaus weniger Lehrveranstaltungen berichteten sie einen schwächeren Ausdruck, und in 95 Lehrsitungen haben sie Freude überhaupt nicht ausgedrückt. Die von den Hochschullehrenden erlebte Angst wurde dagegen in den meisten Fällen überhaupt nicht geäußert ($n = 215$); etwa zu gleichen Anteilen gaben sie einen (viel) schwächeren Angstausdruck an. In fünf Lehrsitungen wurde Angst sogar etwas stärker ausgedrückt. Für Ärger verhält es sich ähnlich wie für Angst: in den meisten Lehrsitungen ($n = 128$) brachten die Hochschullehrenden ihren erlebten Ärger überhaupt nicht zum Aus-

druck. Allerdings gaben sie für etwa doppelt so viele Lehrsitzenungen wie für Angst an, ihren Ärger genauso oder stärker als erlebt ausgedrückt zu haben.

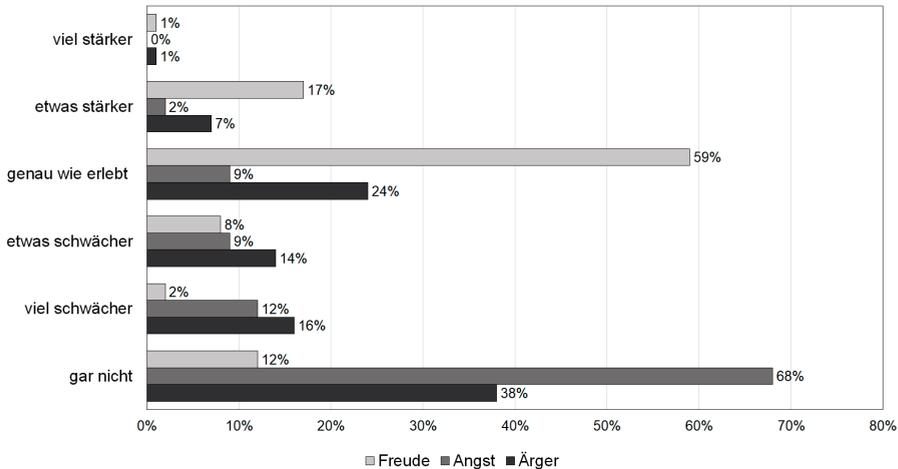


Abbildung 2: Prozentuale Häufigkeiten des Emotionsausdrucks von Freude (N = 771), Angst (N = 316) und Ärger (N = 339)

Für insgesamt 31 Lehrsitzenungen gaben die Hochschullehrenden an, überhaupt keine Freude, für 483 Lehrsitzenungen überhaupt keine Angst und für 461 Lehrsitzenungen überhaupt keinen Ärger erlebt zu haben. Eine Simulation eines Emotionsausdrucks wurde nur für sehr wenige Fälle berichtet ($n = 4$ für Freude, $n = 5$ für Angst und $n = 3$ für Ärger). Eine starke Simulation eines Emotionsausdrucks nannten die Hochschullehrenden lediglich einmal für Freude und einmal für Angst.

5.2 Zusammenhänge auf Lehrsitzenungsebene und auf Personenebene

Die für die Bestimmung der Abhängigkeit der Messwerte ermittelten Intraklassenkorrelationen betragen für das Erleben von Freude $pIC = .29$, von Angst $pIC = .24$ sowie von Ärger $pIC = .22$. Für den Emotionsausdruck variieren diese Indizes zwischen $pIC = .04$ (für den verstärkten Emotionsausdruck von Angst) und $pIC = .25$ (für den verstärkten Ausdruck von Freude). Große Anteile der Varianzen des Emotionserlebens und des Emotionsausdrucks resultieren also aus der Variation der Messwerte *innerhalb* der Hochschullehrenden (Messzeitpunktebene). Eine Berücksichtigung der genesteten Datenstruktur ist somit angezeigt.

Für die Beantwortung der zweiten Forschungsfragen wurden jeweils separat für die Emotionen Freude, Angst und Ärger Within-Korrelationen für die Analyse von Zusammenhängen innerhalb der Hochschullehrenden (Messzeitpunktebene) und Between-Korrelationen für die Untersuchung von Zusammenhängen zwischen den Hochschullehrenden (Personenebene) berechnet. Aufgrund der geringen Fallzahlen für die Emotionssimulation wurde diese von den Analysen ausgenommen. Die Ergebnisse der Within-Korrelationen in Tabelle 1 zeigen einen schwachen negativen Zusammenhang zwischen dem Erle-

ben von Freude und dem unterdrückten Emotionsausdruck. Je stärker eine Hochschullehrperson in einer Lehrsituation Freude erlebt, desto weniger tendiert sie offenbar dazu, ihre Freude im Ausdruck zu unterdrücken. Für das Erleben von Angst ergab sich ein positiver Zusammenhang mit dem unterdrückten Emotionsausdruck. Je stärker eine Hochschullehrperson in einer Lehrsituation Angst erlebt, desto schwächer scheint sie die erlebte Angst den Studierenden gegenüber auszudrücken. Für das Erleben von Ärger fanden sich zwei schwache positive Zusammenhänge mit dem verstärkten und dem authentischen Emotionsausdruck. Je stärker eine Hochschullehrperson in einer Lehrsituation Ärger erlebt, desto mehr wählt sie entweder einen authentischen oder stärkeren Ausdruck für diese Emotion.

Tabelle 1: Zusammenhänge auf Messzeitpunktebene (Within-Korrelationen)

	Freude	Angst	Ärger
Verstärkter Ausdruck	.09	.12	.27*
Authentischer Ausdruck	.08	.12	.12*
Unterdrückter Ausdruck	-.07*	.21*	-.04

Anmerkung: Fallzahlen für Freude sind N = 802 (Clustersize N = 10.03), Angst N = 799 (Clustersize N = 9.98), Ärger N = 800 (Clustersize N = 10) Lehrsituationen; * p < .05.

Die Ergebnisse der Between-Korrelationen in Tabelle 2 geben für das Erleben von Freude der Hochschullehrenden einen positiven Zusammenhang mit dem verstärkten Emotionsausdruck und einen negativen Zusammenhang mit dem unterdrückten Emotionsausdruck. Hochschullehrende, die generell stärker Freude empfinden, drücken ihre Freude offenbar häufiger verstärkt und seltener unterdrückt aus. Für das Erleben von Ärger besteht ein positiver Zusammenhang zu einem verstärkten Emotionsausdruck vor, sodass Hochschullehrende mit einem generell stärkeren Erleben von Ärger diesen tendenziell stärker in ihren Lehrveranstaltungen zum Ausdruck bringen. Für die Emotion Angst ergaben sich keine Zusammenhänge auf der Personenebene.

Tabelle 2: Zusammenhänge auf Personenebene (Between-Korrelationen)

	Freude	Angst	Ärger
Verstärkter Ausdruck	.38**	.19	.67**
Authentischer Ausdruck	.32	-.07	-.18
Unterdrückter Ausdruck	-.48***	.11	-.10

Anmerkung: Die Fallzahlen für Freude, Ärger und Angst betragen jeweils N = 80 Hochschullehrende; ***p < .001, **p < .01, *p < .05.

6 Diskussion

Die vorliegende Studie untersuchte, wie häufig Hochschullehrende einen bestimmten Emotionsausdruck für drei in der Lehre erlebte spezifische Emotionen anwenden und wie dieser mit dem Emotionserleben innerhalb und zwischen Hochschullehrenden verbunden ist. Als

Bedingung für einen Emotionsausdruck wurden zunächst die Häufigkeiten des Emotionserlebens analysiert, mit dem Befund, dass Hochschullehrende in nahezu jeder Lehrsituation Freude erleben, in etwa der Hälfte der Lehrsituationen Angst und Ärger. Die Emotionen können somit als Teil der konkreten Lehrpraxis eingestuft werden und werden von den Hochschullehrenden nicht nur in der einmaligen Rückschau ihrer Lehre wahrgenommen (z.B. Hagenauer & Volet, 2014a; Kordts-Freudinger, 2017; Postareff & Lindblom-Ylänne, 2011; Stupnisky et al., 2014).

Die erste Forschungsfrage untersuchte, wie häufig Hochschullehrende einen schwächeren, authentischen und stärkeren Emotionsausdruck für Freude, Angst und Ärger in Lehrsituationen anwenden. Es ergab sich ein unterschiedlicher Emotionsausdruck für die spezifischen Emotionen. Freude bringen die Hochschullehrenden in Lehrsituationen überwiegend authentisch zum Ausdruck, in nur sehr wenigen Fällen wählen sie einen verstärkten oder unterdrückten Ausdruck für diese Emotion. Dieses Muster ist konform mit den Befunden für positive Emotionen der retrospektiven, einmaligen Studien (Hagenauer & Volet, 2014b; Mendzheritskaya et al., 2018). Mit der vorliegenden ESM-Studie wurde der authentische Ausdruck positiver Emotionen jedoch erstmals im konkreten Fall von Lehrsituationen und für eine spezifische positive Emotion, hier Freude, aufgezeigt. Für Angst und Ärger wurde in der vorliegenden Studie aufgezeigt, dass Hochschullehrende diese Emotionen in den meisten Fällen gar nicht zum Ausdruck bringen, oder sie wird schwächer als tatsächlich erlebt ausgedrückt. Auch dies entspricht weitgehend den Befunden retrospektiver, einmaliger Befragungen (Hagenauer & Volet, 2014b; Mendzheritskaya et al., 2018) und wurde im vorliegenden Beitrag erstmalig in Lehrsituationen und für spezifische negative Emotionen aufgezeigt. Der in der Lehrsituation empfundene Ärger wird hingegen von den Hochschullehrenden bisweilen authentisch oder stärker als tatsächlich erlebt den Studierenden gegenüber zum Ausdruck gebracht. Für die Simulation nicht erlebter Emotionen ergaben sich erneut weitgehend ähnliche Muster zu denen retrospektiver, einmaliger Befragungen (Hagenauer & Volet, 2014b; Mendzheritskaya et al., 2018). Wenn auch methodische und konzeptionelle Unterschiede die Ergebnisse nicht eindeutig vergleichbar machen, so weisen sie darauf hin, dass Hochschullehrende nicht erlebte Emotionen in Lehrsituationen nur äußerst selten simulieren. Insgesamt zeigen die hier vorgestellten Ergebnisse, dass Hochschullehrende im konkreten Fall der Lehrsituation den Ausdruck ihrer erlebten Emotionen beeinflussen. Über die Hintergründe des für die spezifischen Emotionen gewählten Emotionsausdrucks geben die Ergebnisse keine Auskunft, sie können aber möglicherweise Hinweise dafür sein, dass Hochschullehrende ihren Emotionsausdruck vor dem Hintergrund von Darbietungsregeln in der Lehre (vgl. Ashfort & Humphrey, 1993; Ogbonna & Harris, 2004) und/oder ihrer sozial-kommunikativen Funktion (vgl. Fischer & Manstead, 2008; Horstmann, 2003) wählen, wie weiter unten noch diskutiert werden soll. Darüber hinaus beschränken sich die hier vorliegenden Befunde auf Hochschullehrende in Deutschland, so dass Fragen kultureller Unterschiede in diesem Beitrag nicht beantwortet werden können. Kulturelle Unterschiede in der Wahl des Emotionsausdrucks fanden sich in den Studien von Hagenauer und Kolleg*innen (2016) oder Mendzheritskaya und Kolleg*innen (2018) und könnten etwa auch für den situativen Ausdruck spezifischer Emotionen in Lehrveranstaltungen vorliegen.

Die zweite Forschungsfrage adressierte die Zusammenhänge zwischen dem Erleben der spezifischen Emotionen und ihren Emotionsausdruck innerhalb der Hochschullehrpersonen

(auf der Messzeitpunktebene der Lehrsitzen) und zwischen den Hochschullehrenden (auf der Personenebene). Innerhalb der Hochschullehrpersonen fanden sich mehrere Zusammenhänge. Ein stärkeres Erleben von Freude korrelierte negativ mit dem unterdrückten Ausdruck. Nach bisher existierenden Studien erleben Hochschullehrende insbesondere dann Freude, wenn Studierende ein von ihnen erwünschtes Verhalten zeigen und/oder Lernerfolge erzielen (Hagenauer & Volet, 2014a; Postareff & Lindblom-Ylänne, 2011). Möglicherweise drücken Hochschullehrpersonen ihre Freude offen aus, um ihre soziale Beziehung zu den Studierenden zu festigen (vgl. Horstmann, 2003) und/oder Studierende in ihrem Verhalten zu bestärken (vgl. Fischer & Manstead, 2008). Im Fall der Emotion Angst zeigen die Befunde Zusammenhänge zwischen dem stärkeren Erleben von Angst und einem unterdrückten Ausdruck. Dieser Befund kann darauf hinweisen, dass Hochschullehrende den Ausdruck eines starken Erlebens von Angst als für eine Lehrsitzen unangemessen bewerten (vgl. Ashfort & Humphrey, 1993; Ogbonna & Harris, 2004) und ihn daher abschwächen. Wird Angst als eine emotionale Reaktion auf Ereignisse mit unsicherem oder potentiell unerwünschten Ausgang verstanden (Lazarus, 1991; Ortony et al., 1988), und wird berücksichtigt, dass diese Unsicherheiten auch von den Hochschullehrenden als Ursache für Angst identifiziert wurden (z.B. Hagenauer & Volet, 2014a; Meanwell & Kleiner, 2014), könnten Hochschullehrpersonen möglicherweise mit dem Unterdrückung des Emotionsausdrucks beabsichtigen, die Lehr-/Lernsituation nicht durch das Äußern der eigenen Unsicherheiten zu stören und/oder ein sicheres, professionelles Lehrhandeln und -auftreten zu bewahren. Für die Emotion Ärger ergab sich ein auf den ersten Blick überraschender Befund: So war ein stärkeres Erleben von Ärger mit einem authentischen oder verstärkten Emotionsausdruck verbunden. In Lehrveranstaltungen scheint es somit Umstände zu geben, die eine Hochschullehrperson dazu veranlassen, ihrem stark erlebten Ärger in der Lehrsituation Ausdruck zu verleihen. Einen ähnlichen Befund fanden auch Hagenauer und Kolleg*innen (2016) für Hochschullehrende in Deutschland. Danach drücken Hochschullehrende ihren Ärger insbesondere dann aus, wenn sie Studierende auf ein Fehlverhalten hinweisen möchten (Hagenauer et al., 2016). Wenn Ärger als emotionale Reaktion auf eigen- oder fremdverursachte unerwünschte Ereignisse verstanden wird (Lazarus, 1991; Ortony et al., 1988) und studentisches Fehlverhalten eine Quelle für das Erleben von Ärger bei Hochschullehrenden darstellt (z.B. Hagenauer & Volet, 2014a; Postareff & Lindblom-Ylänne, 2011), könnte dieser Befund darauf hinweisen, dass Hochschullehrpersonen den Ausdruck von Ärger situativ für instrumentelle Zwecke anwenden, zum Beispiel um Studierende über ein als unangemessen empfundenen Verhalten zu informieren und bei ihnen eine Reaktionsveränderung hervorzurufen (vgl. Horstmann, 2003).

Zwischen den Hochschullehrenden, also auf Personenebene, fanden sich nur wenige, dafür recht deutliche Zusammenhänge für die spezifischen Emotionen Freude und Ärger. So zeigte sich, dass die zu Freude neigenden Hochschullehrenden in Lehrsitzen dazu tendieren, dieses Erleben im Emotionsausdruck noch zu verstärken und weniger zu unterdrücken. Zudem scheinen zu Ärger neigende Hochschullehrende dazu zu tendieren, diesen generell im Ausdruck zu verstärken. Somit liegen auch personenbezogene oder habituell geprägte Zusammenhänge zwischen dem Emotionserleben und dessen Ausdruck vor, wenn auch diese offenbar nur bei einer Auswahl spezifischer Emotionen bestehen. Beeinflusst werden könnten diese Zusammenhänge etwa durch individuelle Charakteristika der Hochschullehrenden, wie ihrer Einstellung oder ihrer Identifikation mit der Lehre (vgl. Hagenau-

er & Volet, 2014a; Postareff & Lindblom-Ylänne, 2011), oder durch Persönlichkeitsmerkmale (vgl. Revelle & Scherer, 2009; Matsumoto, 2006).

Insgesamt verdeutlichen die Ergebnisse, dass das Erleben und der Ausdruck spezifischer Emotionen bei Hochschullehrenden stark situationssensitiv erfolgen. Damit erscheint eine mehrmalige und situative Erhebung nicht nur für die Erfassung von Emotionen angemessen, sondern auch für den Emotionsausdruck und die -regulation (vgl. Dorn et al., 2013) – auch wenn die deskriptiven Befunde ähnliche Muster zu denen vorheriger Studien aufweisen. Diese Ähnlichkeit kann mit der zeitlichen Verortung der mehrmaligen Befragung zusammenhängen, die zwar zeitlich relativ dicht an einer konkreten Lehrsituation erfolgt ist, aber kurzfristige Erinnerungsfehler und Fehleinschätzungen durch die Hochschullehrenden nicht ausschließt und auch keine Variation im Emotionserleben und -ausdruck innerhalb einer Lehrsituation abbildet. Unabhängig davon zielt der Einsatz von mehrmaligen Befragungen wie der ESM nicht darauf ab, von einmaligen, retrospektiven Erhebungen abweichende Ergebnisse zu generieren (vgl. Goetz, Sticca et al., 2015), sondern der statistischen und interpretativen Unabhängigkeit intra- und interindividueller Phänomene Rechnung zu tragen (vgl. Hamaker, 2011; Murayama et al., 2017). Ein weiterer Erkenntnisgewinn resultiert aus den zum Teil sehr unterschiedlichen intra- und interindividuellen Korrelationen zwischen dem Emotionserleben und dem Emotionsausdruck, über deren Ursache in diesem Beitrag nur Annahmen getroffen werden können. Aus unserer Sicht ist es lohnenswert, die spezifischen Emotionen und den situativen Emotionsausdruck der Hochschullehrenden im Lichte konkreter Lehrereignisse und situativer Merkmale (wie die der emotionalen und instrumentellen Ziele oder die der situativen Einschätzungen zu Verhaltensnormen) auf Ebene der Lehrsituationen und unter Hinzunahme dispositioneller und habitueller Variablen (wie die der Persönlichkeitseigenschaften und der individuellen Charakteristika) auf der Ebene der Personen weiter zu untersuchen.

6.1 Limitierungen

Methodische Eigenschaften der Studien begrenzen ihre Aussagekraft. Die Stichprobe ist nicht repräsentativ für die Hochschullehrenden in Deutschland. Es handelte es sich um eine kleine, selbst-selektierte Stichprobe, die überwiegend aus Hochschullehrenden in den Sozial-, Human- und Kulturwissenschaften bestand, die möglicherweise eine Affinität zum Thema oder zur Methode haben, sowie solchen aus dem wissenschaftlichen Mittelbau. Ein höherer Anteil von Hochschullehrenden aus anderen Disziplinen oder eine Befragung von ausschließlich Professor*innen mit mehr Lehrerfahrung könnte zu anderen Befunden führen. Da eine Verstärkung des Ausdrucks von Angst und Ärger nur sehr selten vorkam, ist die Repräsentativität der Ergebnisse weiter eingeschränkt. Es empfiehlt sich eine Replikation der Befragung an einer größeren Stichprobe oder über einen längeren Befragungszeitraum. Zudem stellen die Korrelationen keine Kausalrichtungen der Zusammenhänge dar. Ein stärkeres Emotionserleben könnte auch aus dem verstärkten Emotionsausdruck resultieren. Zudem zeigte sich für die Emotionen Angst und Ärger ein Erleben von durchschnittlich geringerer oder mittlerer Stärke. Die Hochschullehrenden könnten ihr Emotionserleben noch vor der emotionalen Reaktion mittels entstehungsorientierter Strategien reguliert haben. Außerdem wurden Einzelitems und Selbstberichte eingesetzt. Zukünftige Forschung könnte die Befundlage durch validierte Instrumente oder beobachtende Messmethoden, wie z.B. einer video-basierten Analyse des Emotionsausdrucks, ergänzen.

Literatur

- Ashfort, B. E. & Humphrey, R. H. (1993). Emotional labor in service roles: The influence of identity. *The Academy of Management Review*, 18, 88–115. <https://doi.org/10.2307/258824>
- Augustine, A. A. & Larsen, R. J. (2012). Emotion research. In M. R. Mehl & T. S. Conner (Hrsg.), *Handbook of research methods for studying daily life* (S. 497–510). New York: Guilford.
- Bieg, M., Goetz, T. & Lipnevich, A. A. (2014). What students think they feel differs from what they really feel – Academic self-concept moderates the discrepancy between students' trait and state emotional self-reports. *PLoS One*, 9, e92563. <https://doi.org/10.1371/journal.pone.0092563>
- Brandstätter V., Schüler J., Puca R. & Lozo L. (2013) Emotionsregulation. In V. Brandstätter, J. Schüler, R. Puca & L. Lozo (Hrsg), *Motivation und Emotion* (S. 174–188). Berlin: Springer. https://doi.org/10.1007/978-3-642-30150-6_13
- Buehler, R. & McFarland, C. (2001). Intensity bias in affective forecasting: The role of temporal focus. *Personality and Social Psychology Bulletin*, 21, 1480–1493. <https://doi.org/10.1177/01461672012711009>.
- Catterson, A. D., Eldesouky, L. & John, O. P. (2017). An experience sampling approach to emotion regulation: Situational suppression use and social hierarchy. *Journal of Research in Personality*, 69, 33–43. <https://doi.org/10.1016/j.jrp.2016.04.004>
- Coles, N. A., Larsen, J. T. & Lench, H. C. (2019). A meta-analysis of the facial feedback literature: Effects of facial feedback on emotional experience are small and variable. *Psychological Bulletin*, 145, 610–651. <https://doi.org/10.1037/bul0000194>
- Constantini, P. & Gibbs, P. (2004). Higher education teachers and emotional labour. *The International Journal of Educational Management*, 18(4/5), 243–249. <https://doi.org/10.1108/09513540410538822>
- Csikszentmihalyi, M. & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease*, 175, 526–536. <https://doi.org/10.1097/00005053-198709000-00004>
- Csikszentmihalyi, M., Larson, R. & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, 6, 281–294. <https://doi.org/10.1007/BF02138940>
- Deng, X., Sang, B. & Luan, Z. (2013). Up-and down-regulation of daily emotion: An experience-sampling study of Chinese adolescents' regulatory tendency and effects. *Psychological Reports: Mental & Physical Health*, 113, 552–565. <https://doi.org/10.2466/09.10.PR0.113x22z4>
- Dorn, C., Spindler, G. & Kullik, A. (2013). Erfassung von Emotionsregulationsstrategien – eine Übersicht. *Psychologische Rundschau*, 64, 217–227. <https://doi.org/10.1026/0033-3042/a000176>
- Eid, M., Gollwitzer, M. & Schmitt, M. (2015). *Statistik und Forschungsmethoden* (4. Auflage). Weinheim: Beltz.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169–200. <https://doi.org/10.1080/02699939208411068>
- Ekman, P. & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1, 49–98. <https://doi.org/10.1515/semi.1969.1.1.49>
- Fischer, A. H. & Manstead, A. S. R. (2008). Social functions of emotions. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Hrsg.), *Handbook of emotions* (3rd Ed.) (S. 456–468). New York: Guilford Press.
- Geiser, C. (2011). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung* (2. Aufl.). Wiesbaden: Springer VS.
- Götz, T., Frenzel, A. C., Stöger, H. & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion*, 34, 49–62. <https://doi.org/10.1007/s11031-009-9152-2>

- Götz, T., Haag, L., Lipnevich, A. A., Keller, M. M., Frenzel, A. C. & Collier, A. P. M. (2014). Between-domain relations of students' academic emotions and their judgments of school domain similarity. *Frontiers in Psychology*, 5, 1153. <https://doi.org/10.3389/fpsyg.2014.01153>
- Götz, T., Becker, E. S., Bieg, M., Keller, M. M., Frenzel, A. C. & Hall, N. C. (2015). The glass half empty: how emotional exhaustion affects the state-trait-discrepancy in self-reports of teaching emotions. *Plos ONE*, 10(9), e0137441. <https://doi.org/10.1371/journal.pone.0137441>
- Götz, T., Sticca, F., Pekrun, R., Murayama, K. & Elliot, A. J. (2015). Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach. *Learning and Instruction*, 41, 115–25. <https://doi.org/10.1016/j.learninstruc.015.10.007>
- Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A. & Preckel, F. (2014). 'My questionnaire is too long!' The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, 39, 188–205. <https://doi.org/10.1016/j.cedpsych.2014.04.002>
- Gross, J. J. (2008). Emotion regulation. In M. Lewis, J. M. Haviland, & L. Feldman Barrett (Hrsg.), *Handbook of emotions* (3rd ed.) (S. 497–512). New York: Guilford.
- Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, 26, 1–26. <https://doi.org/10.1080/1047840X.2014.940781>
- Gross, J. J., John, O. P. & Richards, J. M. (2000). The dissociation of emotion expression from emotion experience: A personality perspective. *Personality and Social Psychology Bulletin*, 24, 279–288. <https://doi.org/10.1177/0146167200268006>
- Hagenauer, G. & Volet, S. (2014a) 'I don't think I could, you know, just teach without any emotion': exploring the nature and origin of university teachers' emotions. *Research Papers in Education*, 29, 240–62. <https://doi.org/10.1080/02671522.2012.754929>
- Hagenauer, G. & Volet, S. (2014b). „I dont' hide my feelings, even though I try to“: Insight into teacher educator emotion display. *Australian Educational Researcher*, 41, 261–81. <https://doi.org/10.1007/s13384-013-0129-5>
- Hagenauer, G., Gläser-Zikuda, M. & Volet, S. (2016). University teachers' perceptions of appropriate emotion display and high-quality teacher-student relationship: Similarities and differences across cultural-educational contexts. *Frontline Learning Research*, 4(3) 44–74. <https://doi.org/10.14786/flr.v4i3.236>
- Hamaker, E. L. (2011). Why researchers should think “within-person”, a paradigmatic rationale. In M. R. Mehl & T. S. Conner (Hrsg.), *Handbook of research methods for studying daily life* (S. 43–61). New York: Guilford.
- Hochschild, A. R. (1979). Emotion work, feeling rules, and social structure. *American Journal of Sociology*, 85, 551–75. <https://doi.org/10.1086/227049>
- Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? *Emotion*, 3, 150–66. <https://doi.org/10.1037/1528-3542.3.2.150>
- Izard, C. E. (1981). *Human emotions*. New York: Plenum.
- Keller, M. M. & Becker, E. S. (2018). Erleben und Regulation positiver Emotionen bei Lehrpersonen. In G. Hagenauer & T. Hascher (Hrsg.), *Emotionen und Emotionsregulation in Schule und Hochschule* (S. 165–80). Münster: Waxmann.
- Keller, M. M., Chang, M.-L., Becker, E. S., Götz, T. & Frenzel, A. C. (2014). Teachers' emotional experiences and exhaustion as predictors of emotional labor in the classroom: an experience sampling study. *Frontiers in Psychology*, 5, 1442. <https://doi.org/10.3389/fpsyg.2014.01442>
- Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition and Emotion*, 23, 4–41. <https://doi.org/10.1080/02699930802619031>

- Kordts-Freudinger, R. (2017). Feel, think, teach – Emotional underpinnings of approaches to teaching in higher education. *International Journal of Higher Education*, 6, 217–29. <https://doi.org/10.5430/ijhe.v6n1p217>
- Kordts-Freudinger, R. & Thies, K. (2018). Regulate this! Emotionsregulation und Lehrorientierungen der Hochschullehrenden. In G. Hagenauer & T. Hascher (Hrsg.), *Emotionen und Emotionsregulation in Schule und Hochschule* (S. 211–225). Münster: Waxmann.
- Kullik, A. & Petermann, F. (2012). *Emotionsregulation im Kindesalter*. Göttingen: Hogrefe.
- Lazarus, R.S. (1991). *Emotion and adaptation*. Oxford: Oxford University.
- Matsumoto, D. (2006). Are cultural differences in emotion regulation mediated by personality traits? *Journal of Cross-Cultural Psychology*, 37, 421–437. <https://doi.org/10.1177/0022022106288478>
- Meanwell, E. & Kleiner, S. (2014). The emotional experience of first-time teaching: Reflections from graduate instructors. *Teaching Sociology*, 42, 17–27. <http://doi.org/10.1177/0092055X13508377>
- Mendzheritskaya, J., Hansen, M., Scherer, S. & Horz, H. (2018). “Wann, wie und wem gegenüber darf ich meine Emotionen zeigen?” Regeln der emotionalen Darbietung von Hochschullehrenden in der Interaktion mit Studierenden in unterschiedlichen kulturellen Kontexten. In G. Hagenauer & T. Hascher (Hrsg.), *Emotionen und Emotionsregulation in Schule und Hochschule* (S. 227–241). Münster: Waxmann.
- Murayama, K., Götz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A. & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. *British Journal of Educational Psychology Monograph Series II*, 12, 71–87.
- Muthén, L. K. & Muthén, B. O. (1998–2015). *Mplus User's Guide* (7th Ed.). Los Angeles, CA: Muthén & Muthén.
- Ogbonna, E. & Harris, L. C. (2004). Work intensification and emotional labour among UK university lecturers: An exploratory study. *Organization Studies*, 25, 1185. <https://doi.org/10.1177/0170840604046315>
- Ortony, A., Clore, G. L. & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315–341. <http://doi.org/10.1007/s10648-006-9029-9>
- Postareff, L. & Lindblom-Ylänne S. (2011). Emotions and confidence within teaching in higher education. *Studies in Higher Education*, 36, 799–813. <https://doi.org/10.1080/03075079.2010.483279>
- Revelle, W. & Scherer, K. R. (2009). Personality and emotion. In D. Sander & K. R. Scherer (Hrsg.), *The Oxford companion to emotion and the affective science* (S. 304–306). Oxford: Oxford University.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 347–51. <https://doi.org/10.2307/2087176>
- Robinson, M. D. & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128, 934–60. <https://doi.org/10.1037/0033-2909.128.6.934>
- Roos, A.-L., Bieg, M., Goetz, T., Frenzel, A. C., Taxer, J. & Zeidner, M. (2015). Experiencing more mathematics anxiety than expected? Contrasting trait and state anxiety in high achieving students. *High Ability Studies*, 26, 245–58. <https://doi.org/10.1080/13598139.2015.1095078>
- Rosenberg, E. L. (1998). Level of analysis and the organization of affect. *Review of General Psychology*, 2, 247–270. <https://doi.org/10.1037/1089-2680.2.3.247>
- Scherer, K. R. (2003). Emotion. In W. Stroebe, K. Jonas & M. Hewstone (Hrsg.), *Sozialpsychologie. Eine Einführung* (S. 165–214). Berlin: Springer.

- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44, 695–729. <https://doi.org/10.1177/0539018405058216>
- Snijders, T. & Bosker R. (2011). *Multilevel analysis: an introduction to basic and advanced multi-level modelling* (2nd ed.). London: Sage.
- Stupnisky, R. H., Pekrun, R. & Lichtenfeld, S. (2014). New faculty members' emotions: A mixed method study. *Studies in Higher Education*, 41, 1167–1188. <https://doi.org/10.1080/03075079.2014.968546>
- Thies, K. & Kordts-Freudinger, R. (2019a). German higher education academic staff's positive emotions trough work domains. *International Journal of Educational Research*, 98, 1–12. <https://doi.org/10.1016/j.ijer.2019.08.004>
- Thies, K. & Kordts-Freudinger, R. (2019b). University academics' state emotions and appraisal antecedents: An intraindividual analysis. *Studies in Higher Education*, 44, 1723–33. <https://doi.org/10.1080/03075079.2019.1665311>
- Thompson, R. A. (1994). Emotion regulation: A theme in search of a definition. *Monographs of the Society for Research in Child Development*, 59(2-3), 25–52. <https://doi.org/10.1111/j.1540-5834.1994.tb01276.x>

Kontakt:

Dr. Katharina Thies
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH
Abteilung für Bildungsverläufe und Beschäftigung
Lange Laube 12
30159 Hannover
E-Mail: thies@dzhw.eu

Dr. Robert Kordts-Freudinger
Universität St.Gallen
Institut für Wirtschaftspädagogik
Dufourstrasse 40a
9000 St. Gallen, Schweiz
E-Mail: robert.kordts-freudinger@unisg.ch



Nicole Auferkorte-Michaelis,
Frank Linde (Hrsg.)

Diversität lernen und lehren – ein Hochschulbuch

Diversity Management ist eine Aufgabe, der sich eine Hochschule als Gesamtorganisation stellen muss. Der Band ermöglicht einen umfassenden Blick auf die Arbeit mit Vielfalt an höheren Bildungseinrichtungen. In praktischen Beiträgen teilen Diversity-Expertinnen und -Experten von Hochschulen im In- und Ausland ihr Wissen und ihre Erfahrung zu Diversity-Management in den Kernbereichen Lehre, Forschung, Strategie/Strukturen/Prozesse und Kompetenzentwicklung.

2018 • 411 S. • kart. • 49,90 € (D) • 51,30 € (A)
ISBN 978-3-8474-2046-0 • eISBN 978-3-8474-1067-6



Constance Engelfried,
Pierre L. Ibisch (Hrsg.)

Promovieren an und mit Hochschulen für Angewandte Wissen- schaften

Aktuell existieren in mehreren Bundesländern kontrovers diskutierte Bemühungen, den Hochschulen für Angewandte Wissenschaften bzw. Fachhochschulen das Promotionsrecht zu gewähren. Die Autor_innen reflektieren den aktuellen Diskussionsstand und legen Vorschläge zum konstruktiven Gestalten der zukünftigen Hochschulentwicklung vor.

2016 • 213 S. • kart. • 29,90 € (D) • 30,80 € (A)
ISBN 978-3-8474-0771-3 • eISBN 978-3-8474-0890-1



Jörg Schlömerkemper

Pädagogische Diskurs-Kultur

Über den sensiblen Umgang mit
Widersprüchen in Erziehung und
Bildung

2021 • 196 Seiten • Kart. • 24,90 € (D) • 25,60 € (A)

ISBN 978-3-8474-2461-1 • eISBN 978-3-8474-1599-2

Der Autor regt dazu an, Begriffe und Konzepte der pädagogischen Reflexion kritisch zu durchdenken und nach einem theoretischen Verständnis zu suchen, mit dem Widersprüche und irritierende Ungereimtheiten konstruktiv „aufgehoben“ werden können. An die Stelle der Hoffnung auf absolute Wahrheiten setzt er das Konzept der „Wirksamkeiten“. Wenn deren konkrete Bedeutungen im offenen Diskurs „antinomie-sensibel“ transparent werden, sind Felder der Persönlichkeits-Entwicklung und Formen der Persönlichkeits-Erziehung differenzierter und wirkungsvoller zu bearbeiten.

Dieser Ansatz wird sowohl für allgemeine Erziehungsprozesse wie auch als Entwurf für eine antinomie-sensible Theorie der Schule konkretisiert. Abschließend werden Folgerungen für die Bildungspolitik, die gesellschaftspolitische Rahmung, die Zukunft der Schule und die Aufgaben der Erziehungswissenschaft entwickelt.

www.shop.budrich.de



Ulrich Teichler

Hochschulforschung, was sonst?

Rückblick auf ein Wissenschaftlerleben. Gespräche mit Anna Kosmützky und Christiane Ritterrott

Der Begründer der Hochschulforschung in Deutschland, Ulrich Teichler, blickt auf mehr als fünf Jahrzehnte Hochschulforschung zurück. Wirtschaftswunder, Hochschulexpansion, studentische Bewegung, Experimente und Krisen der 1970er Jahre, Organisationsruhe, Internationalisierung, Ranking- und Management-Kult – all dies sind historische Stationen, die sich in Hochschule und Wissenschaft widerspiegeln. Teichler berichtet als unmittelbar Beteiligter mit Offenheit und Humor und liefert zugleich kluge Analysen.

2020 • 206 S. • kart. • 36,00 € (D) • 37,10 € (A)
ISBN 978-3-96665-023-6 • eISBN 978-3-96665-993-2



Jutta Helm

Die Kindheitspädagogik an deutschen Hochschulen

Eine empirische Studie zur Akademisierung einer pädagogischen Profession

Im Zuge der wachsenden Bedeutung öffentlicher Kindertagesbetreuung und frühkindlicher Bildung kommt der akademischen Ausbildung kindheitspädagogischer Fachkräfte eine Schlüsselposition zu. Die Autorin präsentiert die Ergebnisse einer qualitativen und quantitativen Studie zur Entwicklung und Konzeptualisierung früh- bzw. kindheitspädagogischer Bachelorstudiengänge. Dabei werden die Perspektiven der Studiengangleiter*innen und die Erfahrungen und Einschätzungen der Studierenden untersucht.

2015 • 340 S. • kart. • 39,90 € (D) • 41,10 € (A)
ISBN 978-3-86388-037-8 • eISBN 978-3-86388-235-8



utb.

Nicole Auferkorte-Michaelis
Frank Linde

Diversität in der Hochschullehre – Didaktik für den Lehralltag



Frank Linde
Nicole Auferkorte-Michaelis

Diversität in der Hoch- schullehre – Didaktik für den Lehralltag

Kompetent lehren, Band 13

2021 • 143 Seiten • Kart. • 14,90 € (D) • 15,40 € (A) • utb S

ISBN 978-3-8252-5603-6

Wie können Hochschullehrende der Diversität von Studierenden mit einer bewussten Gestaltung von Lehr-Lern-Interaktionen konstruktiv begegnen? Die Autor*innen unterstützen mit diesem Buch Lehrende bei der Gestaltung ihrer eigenen didaktischen Konzepte für Lehrveranstaltungen. Wohlwissend, dass die Studierenden alle unterschiedlich sind, entstehen Ideen für eine diversitätsgerechte Didaktik.

www.utb-shop.de