

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 Bachelorstudienerfolgs 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 Chancengerechtigkeit. 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 Studienerfolgprognose

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 Bewertungsunterschiede 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 (Schwibbe 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 Notenvvalidität 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 Masterstudienerfolgs 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 Bildungsvläufe 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 Hochschulwechseln 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.

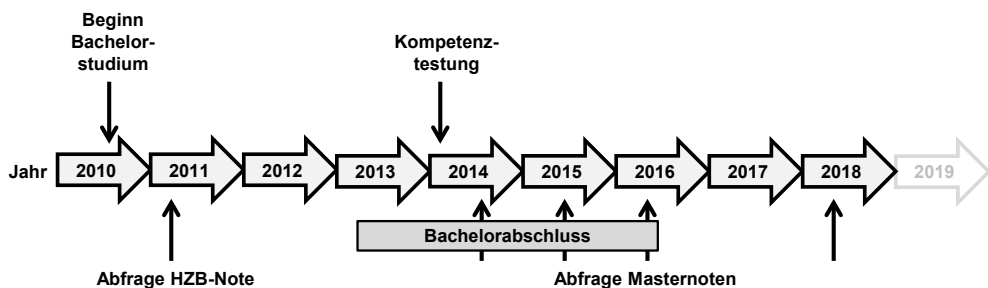


Abbildung 1: Zeitschiene der verwendeten Erhebungszeitpunkte im Studienverlauf (HZB = Hochschulzugangsberechtigung)

- 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/3dcjz/> eingesehen werden.

13 Tatsächlich führte die Korrektur für *restriction of range* zu einer leichten Validitätsverringern, 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 $\Delta R^2 < .01$, $p > .48$).

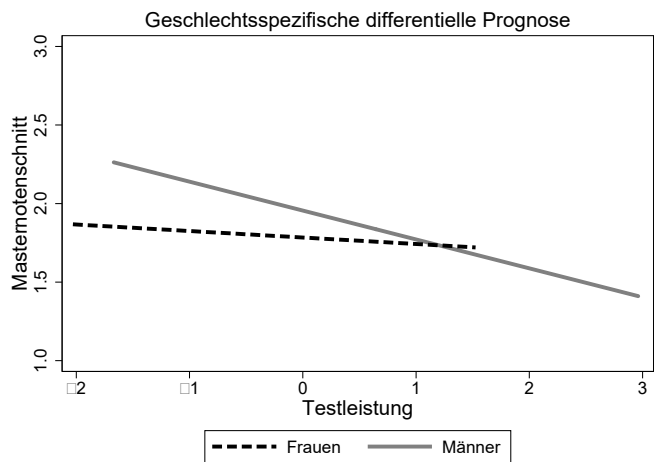


Abbildung 2: Geschlechtsspezifische Regressionsgeraden zur Prognose der Masternoten durch die Kompetenztestleistung. Die Geraden sind für die Spannweite der jeweiligen Testleistungen eingezeichnet.

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 ^{zen}	-0.18* (0.09)
Geschlecht (männlich = 0, weiblich = 1)	-0.17* (0.08)
Testleistung ^{zen} × Geschlecht	0.14 (0.09)

* $p < .05$, $n = 118$, ^{zen} = 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 alleinig 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 vornherein 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_1bvl000314.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:0168-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