ORIGINALIA › PEER REVIEW

Der Zusammenhang zwischen dem Geschlecht von Studierenden und Prüfenden sowie der Prüfungsformate auf den Studienerfolg im Lehramtsstudium Sport

Andre Magner, Malte Jetzke, Till Utesch

Korrespondierender Autor
Andre Magner
Institut für Sportwissenschaft
Arbeitsbereich Bildung und Unterricht im Sport
Universität Münster

Schlüsselwörter
Studienerfolg, Sportstudierende, Lehramt, Geschlechterdifferenzen, Prüfungsformat, Curriculum

Keywords
Academic success, sports students, teacher education, gender differences, mode of assessment, curricula

Zitieren Sie diesen Beitrag wie folgt:
Magner, A., Jetzke, M. & Utesch, T. (2023). Der Zusammenhang zwischen dem Geschlecht von Studierenden und Prüfenden sowie der Prüfungsformate auf den Studienerfolg im Lehramtsstudium Sport. *Zeitschrift für Studium und Lehre in der Sportwissenschaft, 6*(2), 5-20.

Zusammenfassung

Studienerfolg ist sowohl für Hochschulen als auch für Studierende erklärtes Ziel des tertiären Bildungsbereichs. Geschlechterunterschiede zugunsten der Frauen sind dabei in den meisten Bildungseinrichtungen seit längerer Zeit evident. Auch im sportwissenschaftlichen Studium erzielen Studentinnen bessere Noten in Abschlüssen, obwohl sich die Intelligenz zwischen den Geschlechtern nicht unterscheidet. Eine Analyse im Lehramtsstudium Sport und auf Prüfungsebene inkl. Prüfungsformate steht noch aus und wird in diesem Beitrag adressiert. Grundlage der Studie ist eine Vollerhebung der Prüfungsleistungen von Lehramtsstudierenden mit dem Fach Sport (n = 2.868) mit n = 130 Prüfenden über einen Zeitraum von zehn Jahren. Die Archivdaten eines sportwissenschaftlichen Instituts weisen hierbei eine binäre Geschlechterunterscheidung (51 % Männer) auf. In der vorliegenden Studie wurden gekreuzt-klassifizierte lineare gemischte Modelle aufgestellt und geschätzte Randmittel berechnet, um Geschlechterunterschiede vor dem Hintergrund unterschiedlicher Prüfungsformate zu untersuchen. Einen besonderen Fokus nahmen hier die für das sportwissenschaftliche Studium typischen fachpraktischen Prüfungen ein. Zudem wurde das Geschlecht der Prüfenden als Prädiktor einbezogen und für die Hochschulzugangsberechtigung kontrolliert. Die Ergebnisse zeigen, dass Sportstudentinnen insgesamt signifikant besser abschneiden als ihre Kommilitonen. Zudem unterscheiden sich die Noten je nach Prüfungsformat: Frauen erreichen in schriftlichen Klausuren und fachpraktischen Prüfungen signifikant bessere Ergebnisse als Sportstudenten, die in Abschlussarbeiten signifikant besser abschneiden, während sich in mündlichen Prüfungen keine Geschlechterunterschiede zeigen. In fachpraktischen Prüfungen des Studiums schneiden Frauen in sechs von acht Studieninhalten mit bis zu 0,33 Notenpunkten besser ab, mit Ausnahme von Spielsportarten, in denen die Männer um 0,07 bessere Noten erhalten. Weiterhin zeigt sich, dass Prüferinnen an Studentinnen bessere Noten vergeben als an Studenten. Die Ergebnisse werden vor dem Hintergrund der Prädiktoren von Studienerfolg diskutiert, um daraus hochschuldidaktische Konsequenzen für das Lehramtsstudium Sport abzuleiten. Zusammengefasst wird eine datenbasierte Grundlage vorgelegt, um eine Diskussion über die Passung von Berufsanforderungen und Curricula sowie der Prüfungskriterien und -formate anzuregen.

The influence of students' and examiners' gender on academic success in teacher training in physical education

Abstract: Academic success is a declared goal of both universities and students in tertiary education. Gender differences in favor of women have been evident for some time in most educational institutions. Female students also achieve better grades in sports science studies, although there is no difference in intelligence between the sexes. An analysis at the level of sports teacher education and examinations, including modes of assessments, is still pending and will be addressed in this article. The basis of the study is a complete survey of the examination performance of student teachers of physical education (n = 2.868) with n = 130 examiners over a ten-year period. The ten-year archive data from a sports science institute show a binary gender distinction (51 % male). In the present study, cross-classified linear mixed models and marginal means were used to investigate gender differences in the context of different modes of assessment. Particular attention was paid to the practical examinations typical of sports science programmes. In addition, the gender of the examiner was included as a predictor and university entrance qualification was controlled for. The results show that, overall, female students perform significantly better than their male counterparts. There are differences according to the modes of assessment: women perform significantly better in written exams and subject-specific practical exams than male sports students, who perform significantly better in final theses, while there are no gender differences in oral exams. In practical exams, women perform up to 0.33 grade points better in six out of eight subjects, with the exception of game sports, where men perform 0.07 grade points better. Female examiners were also found to give better marks to female students than to male students. The results are discussed against the background of the predictors of academic success, in order to derive consequences of university didactics for teacher training in physical education. In summary, a data-based foundation is presented to stimulate discussion on the fit between professional requirements and curricula as well as criteria and modes of assessment.

1. EINLEITUNG

Wer als Junge oder Mann[1] in eine Bildungsinstitution eintritt, hat qua Geschlecht bereits seit einigen Dekaden schlechtere Chancen auf eine erfolgreiche Laufbahn. Weltweit sind schlechtere (Schul-) Leistungen von Jungen, insbesondere im Lesen und Schreiben, größere Lernschwierigkeiten, verspätete Einschulungen, schlechtere Übergangsempfehlungen und häufigere Schulabbrüche ebenso wie geringere Motivation und größere soziale Auffälligkeiten mindestens seit den PISA-Studien bekannt (Berlin-Institut für Bevölkerung und Entwicklung, 2015; Hadjar et al., 2014; OECD, 2019; Reilly et al., 2019). Auch in der tertiären Bildung ist zunehmend eine Schieflage zu beobachten. Verschiedene Studien berichten, dass Frauen fast durchgehend bessere Noten und mehr Abschlüsse als Männer erzielen und dabei weniger Zeit benötigen (Kuh et al., 2006; Milun et al., 2016; Röwert et al., 2017; van den Berg & Hofman, 2005). 25- bis 34-jährige Frauen erreichen in allen OECD-Ländern mit höherer Wahrscheinlichkeit (52 %) einen Abschluss im Tertiärbereich als gleichaltrige Männer (39 %) - eine Differenz, die im vergangenen Jahrzehnt in vielen Ländern noch zugenommen hat (OECD, 2021). In ihrer Meta-Analyse von Studienergebnissen aus einer Zeitspanne von fast 100 Jahren identifizieren Voyer und Voyer (2014) den robusten Effekt, dass Frauen von der Grundschule bis zur Universität in allen Bildungsinstitutionen bessere Noten erreichen als Männer, obwohl bzgl. der Intelligenz keine Unterschiede zwischen Männern und Frauen bestehen (Hattie, 2009).

Vor dem Hintergrund der öffentlichen Aufmerksamkeit für Gleichstellungsfragen ist es überraschend, dass dieses Ungleichgewicht in der Hochschulbildung selten bildungspolitisch und öffentlich diskutiert wird. Nach Artikel 3, Absatz 2 des Grundgesetzes der Bundesrepublik Deutschland sind Männer und Frauen gleichberechtigt, ein Grundsatz, dem der Staat und damit auch (Hoch-) Schulen folgen sollten und demnach bestehende Nachteile und Hindernisse zu beseitigen sind. Dies betrifft auch nachgewiesene Geschlechterunterschiede in verschiedenen Curricula, Unterrichtsformen oder Prüfungsformaten, welche strukturelle Benachteiligungen für Männer (Reilley et al., 2019) oder Frauen (Towfigh et al., 2014, 2018) vermuten lassen. Ein häufig angeführtes Argument, Männer seien auf dem Arbeitsmarkt aber bevorteilt, muss hier mit Blick auf die Chancengleichheit im Bildungssystem entkräftet werden: Verschiedene Benachteiligungen lassen sich nicht gegeneinander aufrechnen, weil das zu einer noch stärkeren Verfestigung unterschiedlicher Lebenschancen führen würde (Berlin-Institut für Bevölkerung und Entwicklung, 2015).

Sport wird häufig als letzte männliche Domäne bezeichnet (Eliot, 2010). Während Frauen in vielen Bereichen des öffentlichen und privaten Lebens ihre oft historisch gewachsenen Nachteile aufholen, scheint der Sport nach wie vor männlich bestimmt zu sein: Die Maxime vieler Sportvereine sind an wettbewerbsorientierten eher maskulinen Vorlieben ausgerichtet und auch im Sportunterricht orientieren sich Lehrkräfte zumeist an den Interessen von Jungen, beispielsweise um ihren antizipierten Widerstand zu vermeiden (Mutz & Burrmann, 2014). Auch im Studienfach Sportwissenschaft scheint diese Dominanz auf den ersten Blick vorhanden zu sein: Zwischen 2008 und 2020 absolvierten mehr Männer als Frauen Prüfungen in sportwissenschaftlichen Studiengängen, bei den Lehramtsprüfungen im Fach Sport war dieses Verhältnis aber ausgeglichen (Statistisches Bundesamt, 2021). Die für die berufliche (Weiter-)Beschäftigung hoch relevante Abschlussnote zeigt hier - parallel zu anderen Fächern -, dass die Lehramtsstudentinnen das Fach Sport deutlich besser abschließen als ihre männlichen Kommilitonen. Diese nationalen Befunde werden durch internationale Studien bestätigt (Huntley et al., 2017; Sheard, 2009).

Während Erfolgskriterien auf Ebene des Studienabschlusses über zentrale Melderegister vergleichsweise leicht zu organisieren und damit zu vergleichen sind, gibt es nur spärliche Befunde, die die Notengebung über den Verlauf des Studiums untersuchen. In der empirischen Hochschulforschung wird hier zumeist die Studieneingangsphase betrachtet, die als valider Prädiktor für den Studienabschluss gilt (Trautwein & Bosse, 2017). Allerdings erscheint besonders der Blick auf die gesamte Bandbreite der Prüfungen und Prüfungsformate über den Studienverlauf lohnend, zumal heutzutage in modularen Bachelor- und Masterstudiengängen die Abschlussnoten meist aus der Kumulation der Einzelprüfungen berechnet werden – schlechte Leistungen über den Studienverlauf tragen seit den Bologna-Reformen entsprechend zu einem schlechten Studienabschluss bei. Untersuchungen, die den Studienerfolg im Lehramtsstudium Sport über längere Zeit und auf einzelne Teil-

[1] Im Beitrag wird von einem binären Geschlechterverhältnis ausgegangen. Wenngleich gesellschaftliche Entwicklungen mittlerweile auch vielfältige Geschlechteridentitäten berücksichtigen, ist in der Bildungspolitik und -forschung nach wie vor eine stark dichotome Ausrichtung vorzufinden (z. B. Rüdiger et al., 2021). Auf den Sport trifft dies u.a. vor dem Hintergrund geschlechterdifferenzierter Bewertungsmaßstäbe in besonderem Maße zu. Diese Einschränkung wird im abschließenden Kapitel diskutiert und eingeordnet.

CC BY 4.0 DE

prüfungen unter Berücksichtigung möglicher Geschlechterdifferenzen beleuchten, gibt es nach unserem Kenntnisstand bis heute nicht. Hier stellt sich die offene Frage, ob sich in Betrachtung der Einzelprüfungen des sportwissenschaftlichen Lehramtsstudiums geschlechtsbedingte Differenzen in der Bewertung zeigen und, falls zutreffend, ob es Studieninhalte und/oder Prüfungsformate gibt, für die das in besonderer Weise gilt. Im Studiengang Sportwissenschaft werden die klassischen Prüfungsformate (schriftliche Klausuren, mündliche Prüfungen, Abschlussarbeiten) um eine sportmotorische Dimension erweitert: fachpraktische Prüfungen. Im vorliegenden Beitrag werden Noten aus Einzelprüfungen Lehramtsstudierender mit dem Fach Sport vom Wintersemester 2011/12 bis zum Sommersemester 2021 analysiert, um eventuelle Geschlechterdifferenzen in verschiedenen Prüfungsformaten zu identifizieren.

2. THEORETISCHER RAHMEN UND KONZEPTUALISIERUNG

Ein erhöhtes Forschungsinteresse am Studienerfolg ist spätestens seit den Bologna-Reformen Anfang der 2000er Jahre zu verzeichnen (Sarcletti, 2021). Dabei wird eine genaue, forschungstheoretische Annäherung besonders durch die vielfältigen Definitionen von und entsprechend vielfältigen konzeptionellen, methodischen oder empirischen Zugängen zu Studienerfolg erschwert (York et al., 2015). Diese reichen von offensichtlichen Beschreibungen wie dem „erfolgreichen Abschluss eines Studiums“ (Fries, 2007, S. 145) als individuelles Kriterium des Erfolgs für Studierende über weitere und gut messbare Kriterien wie Drop-Out-Quoten, Studiendauer, Abschluss- und Zwischennoten als Erfolg eines Studiengangs (Heinze, 2018) bis hin zu subjektiven Determinanten von Studienerfolg wie Lernstrategien, Studienzufriedenheit, Motivation, Verhalten der Lehrenden (Schneider & Preckel, 2017; Westermann et al., 1996) sowie sozialer und akademischer Integration in das System Hochschule (Tinto, 1975). Auch post-universitäre Kriterien wie Beschäftigungsfähigkeit oder Persönlichkeitsentwicklung werden ebenso untersucht wie die institutionellen Rahmenbedingungen selbst (Grunschel & Dresel, 2021). Dies führt zu integrativen Modellen des Studienerfolgs, die sowohl produkt- als auch prozesshafte Facetten abbilden (Heinze, 2018; Kuh et al., 2006).

Mittlerweile sind neben Fachkulturen auch kommissions-, studiengang-, abschluss- oder bundeslandspezifische Einflussgrößen bekannt (Grözinger, 2015; Hu, 2005; Tsarouha, 2017). Dennoch sind gute Noten nach wie vor nachvollziehbare Primärziele von Lehramtsstudierenden: Sowohl Plätze im Vorbereitungsdienst als auch Einstellungen an Schulen werden häufig nach der Studienabschlussnote vergeben (MSB NRW, 2021; Rauin, 2007). Auch für die weitere universitäre Laufbahn ist die Abschlussnote entscheidend, sodass der Wissenschaftsrat selbst Studienerfolg als „Studienabschluss mit guter Note oder gute[n] Zwischenprüfungsnoten“ (Wissenschaftsrat, 2004, S. 87) definiert. Hochschulnoten werden zudem als objektiv und reliabel eingeschätzt: In vorliegenden Meta-Analysen (Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000) wurden selbst vergebene, von Gleichaltrigen vergebene sowie von Lehrenden vergebene Noten verglichen. Die hohen Korrelationen zeigen, dass alle drei Gruppen mehrheitlich ähnliche Noten vergeben.

Allerdings steht die Notengebung an Hochschulen seit längerem in der Kritik. So hat nicht nur der Wissenschaftsrat (2012) bereits moniert, dass Prüfungsleistungen an Hochschulen zu selten an transparenten und kriterialen Benotungssystemen orientiert sind. Im Lehramtsstudium Sport existieren an einigen Stellen - nämlich in bestimmten fachpraktischen Prüfungen - solche kriterialen Benotungssysteme. Diese Kriterien orientieren sich meist an den anatomischen und physiologischen Unterschieden zwischen Männern und Frauen. Männer müssen für gleiche Noten laut diesen Norm- und Notentabellen absolut betrachtet in bestimmten Sportarten z. T. deutlich bessere Leistungen erbringen als Frauen (Stiller & Kahlert, 2021). Diese geschlechterdifferenzierten Bewertungsmaßstäbe sind besonders in Sportarten wie Leichtathletik oder Schwimmen wahrnehmbar, bei denen Schnellkraft- und Ausdauerleistungen gemessen werden. Frauen können hier durch anatomische und physiologische Voraussetzungen nicht die gleichen Resultate erzielen wie Männer (Hottenrott, 2015). Dass es auch innerhalb der Geschlechtergruppen anatomische und physiologische Unterschiede gibt, wird in diesen Kriterien und darüber hinaus nicht explizit berücksichtigt. Beispielsweise werden für einen 2m großen Mann und einen 1,60m großen Mann dieselben Leistungsstandards im Hochsprung, im Kugelstoßen, im Bodenturnen oder beim Tanz angelegt. Die starke Wirkmacht sportpraktischer Studieninhalte und Prüfungen im Studium für das Lehramtsfach Sport ist dabei seit längerem thematisiert worden (Klinge, 2002). Um schließlich hochschuldidaktische und -politische Entscheidungen zur Maximierung des Studienerfolgs beider Geschlechter zu stützen, ist das Ziel dieses Beitrags, detaillierte Analysen zu Teilnoten über verschiedene Prüfungsformate und den Verlauf des Studiums vorzulegen (vgl. Abb. 1).

3. FORSCHUNGSSTAND

Die Rolle des Geschlechts beim Studienerfolg ist vielfach untersucht worden (Richardson et al., 2012; Voyer & Voyer, 2014). Einen wichtigen Einflussfaktor stellen Prüfungsformate dar. Studien aus der Rechtswissenschaft berichten, dass Männer in mündlichen Prüfungen - trotz schlechterer Eingangsvoraussetzungen - signifikant bessere Ergebnisse erzielen als Frauen (Towfigh et al., 2014, 2018). Die Autoren geben als Grund eine höhere Selbstsicherheit und größere Aktivität von Männern in Prüfungsgesprächen an. Breda und Hillion (2016) zeigten in Analysen zu mündlichen Prüfungen mit angehenden Lehrkräften, dass Geschlechterdifferenzen zugunsten der zahlenmäßig unterrepräsentierten Gruppe auftreten, z. B. Frauen in MINT-Fächern. In schriftlichen Prüfungen sind Männer bereits früh im Nachteil, wobei die Unterschiede sogar im Laufe der Zeit größer werden (Reilly et al., 2019). Woodfield et al. (2005) und Smith (2004) erklären Ursachen und konnten zeigen, dass Studentinnen sich angemessener auf Klausuren vorbereiten und daher mit höherer Wahrscheinlichkeit eine bessere Note erzielen. Die Meta-Analyse von Voyer und Voyer (2014) gibt Hinweise darauf, dass diese Ergebnisse auch auf Abschlussarbeiten übertragbar sein könnten. Lediglich in Multiple-Choice Klausuren scheinen Männer besser abzuschneiden als Frauen, wenn diese so konfiguriert sind, dass falsche Antworten nicht bestraft werden (Espinosa & Gardeazabal, 2020). Studien, die den Einfluss des Prüfungsformats auf Prüfungsleistungen im Lehramtsstudium Sport aus Geschlechterperspektive untersuchen, liegen nach unserer Kenntnis bislang nicht vor.

Auch für sportwissenschaftliche Studiengänge liegen Ergebnisse aus Studien vor, die das Ziel verfolgten, Determinanten von Studienerfolg zu identifizieren. Kuśnierz et al. (2020) zufolge sind Frauen signifikant motivierter und entsprechend erfolgreicher im Hinblick auf die akademische Leistung als Männer. Zudem scheinen

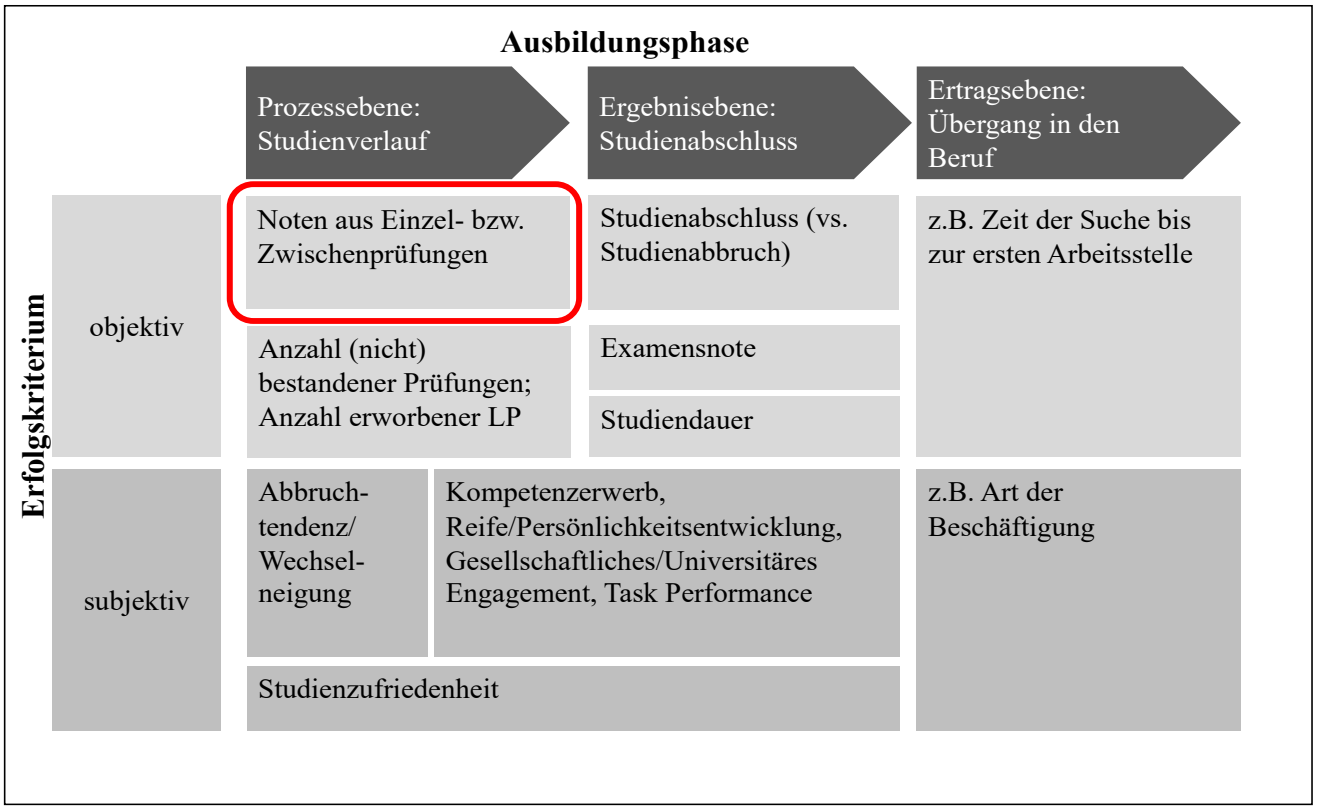

Abb. 1: *Konzeptualisierung des Studienerfolgs (vgl. Heinze, 2018, S. 42)*

schulische Eingangsvoraussetzungen prädiktiv für den Studienerfolg im Fach Sport zu sein, wodurch Frauen bessere Noten erzielen (Şahin et al., 2018). Auch andere Forschungsgruppen untersuchten das Studien- und Lernverhalten von Sportstudierenden und stellten neben besseren Noten der Frauen fest, dass diese offenbar ihre Lernstrategie effektiver dem jeweiligen Prüfungsformat anpassen (Peters et al., 2007).[2] Es ist daher nicht überraschend, dass Frauen in sportwissenschaftlichen Studiengängen neben besseren Noten auch höhere Abschlusswahrscheinlichkeiten attestiert werden (Huntley et al., 2017). Auch bei Troche et al. (2010) erwies sich die Schulabschlussnote als valider Prädiktor für den Studienerfolg; darüber hinaus waren die Leistungen der Männer in fachpraktischen Veranstaltungen für Studiengänge der Sekundarstufe II signifikant besser als die der Frauen. Diese Studie stellt als eine der wenigen die sportpraktischen Studieninhalte des Studiums explizit heraus, auch wenn diese nicht weiter differenziert wurden.

Das Lehramtsstudium Sport beinhaltet neben fachwissenschaftlichen und fachdidaktischen Inhalten bis zu 40 % fachpraktische Studieninhalte und Prüfungen (dvs, 2020). Der Studienerfolg ist damit maßgeblich auch durch sportmotorische Fähigkeiten bestimmt. In der Prüfungskultur der sportwissenschaftlichen Studiengänge sind geschlechterdifferenzierte Bewertungsmaßstäbe - wie auch im Leistungssport - gebräuchlich. Besonders in Sportarten wie Leichtathletik oder Schwimmen werden u. a. Schnellkraft- und Ausdauerleistungen gemessen, die aufgrund physiologischer und anatomischer Geschlechterunterschiede von Männern und Frauen nicht gleichermaßen erreicht werden können. Hier führen insbesondere kardiopulmonale, zelluläre sowie hormonelle und metabole Geschlechterdifferenzen zu Unterschieden in der sportlichen Leistungsfähigkeit (Hottenrott, 2015). Empirische Untersuchungen im Hinblick auf die Überprüfung motorischer Fähigkeiten aus einer Geschlechterperspektive im Sportstudium sind zwar rar, jedoch anatomisch und physiologisch evident. In der Schulsportforschung wird zudem vermutet, dass Sportlehrkräfte häufig unbewusst das Leistungsniveau der Jungen als Bezugsgröße heranziehen, wodurch sich in geschlechterhomogenen Gruppen Nachteile für Mädchen ergeben (Mutz & Burrmann, 2014). Hierdurch wird die Notengebung beeinflusst, obwohl die Fitness insgesamt im Kindesalter als stabil und gleichwertig gilt (Utesch et al., 2018; Niessner et al., 2020). Da Jungen im schulischen Sportunterricht dadurch - in allen Jahrgangsstufen (Gerlach et al., 2006) - bessere Noten als Mädchen erzielen, ist theoretisch zu erwarten, dass diese Unterschiede auch in der (koedukativen) Hochschulbildung Bestand haben und Männer auch im Lehramtsstudium Sport in fachpraktischen Prüfungen die Frauen bezogen auf ihre Noten übertreffen könnten.

Neben individuellen Faktoren auf Prüflingsebene könnten auch weitere Faktoren wie das Geschlecht der Prüfenden einen Einfluss auf Prüfungsergebnisse haben. Insgesamt scheinen Prüferinnen an beide Geschlechter bessere Noten zu vergeben als ihre männlichen Kollegen (Jewell & McPherson, 2012; Wiskin et al., 2004). Towfigh et al. (2018) berichten in Bezug auf mündliche Prüfungen, dass Studentinnen bessere Noten erreichen, wenn auch eine Frau Teil der Prüfungskommission ist. Bessere Abschlusschancen für Studentinnen bei Dozentinnen sind ebenfalls bestätigt worden (Butler & Christensen, 2003). Im besonderen Feld des Sports mit männlich konnotierten Praxen sind zumindest im Sportunterricht für alle Altersklassen bessere Noten der Jungen insbesondere bei männlichen Sportlehrern (Hoven, 2017) evident. Es kann zusammenfassend auch eine Verzerrung auf Hochschulebene vermutet werden.

[2] Dass das Prüfungsformat einen Einfluss auf die Lernstrategien hat, konnte bereits nachgewiesen werden (Lindner et al., 2015). So führen Multiple-Choice-Prüfungen zu oberflächlichem Lernen, während offene Prüfungsformate eher tiefergehendes Lernen fördern. Geschlechterunterschiede in der Bearbeitung von Multiple-Choice-Prüfungen werden von Riener und Wagner (2017) diskutiert.

CC BY 4.0 DE

4. ZIEL DER STUDIE

Ziel der Studie ist die Untersuchung des bisher noch zu wenig betrachteten Zusammenhangs von Geschlecht und Studienerfolg im Lehramtsstudium Sport. Während schulischer Sportunterricht sowie Vereinssport stark männlich dominiert sind, erzielen Frauen in sportwissenschaftlicher Hochschulbildung national und international deutlich bessere (End-)Resultate als Männer. Unklar ist, wie sich diese Ergebnisse auf der Prozessebene des Studienverlaufs manifestieren. Für Lehramtsstudierende stellt eine gute Abschlussnote als Kumulation von Einzelnoten im Hinblick auf ihre Einstellungschancen in den Lehrberuf ein besonders wichtiges Ziel dar. Operationalisiert wird der Studienerfolg in dieser Studie daher parallel zur internationalen Literatur anhand von Noten in Einzelprüfungen (schriftliche Klausuren, mündliche Prüfungen, Abschlussarbeiten, fachpraktische Prüfungen) unter der statistischen Kontrolle der Note für die Hochschulzugangsberechtigung (HZB), die individuell schon einen wichtigen Prädiktor für Studienerfolg darstellt.

Auf Basis bisher vorliegender Forschungsbefunde konnte gezeigt werden, dass empirische Zusammenhänge zwischen Geschlecht und Studienerfolg zum Nachteil der Männer, auch in der sportwissenschaftlichen Hochschulbildung, bestehen. Wir vermuten einen Zusammenhang auch in der vorliegenden Stichprobe im Lehramtsstudium Sport über alle Teilnoten hinweg zu finden (H1). In der Literatur außerhalb des Sports zeigt sich dieser Effekt aber differenziert über verschiedene *Prüfungsformate*: Männliche Studierende erzielen in schriftlichen Prüfungen schlechtere Leistungen als weibliche, während sie in mündlichen Prüfungen bessere Noten erhalten als ihre Kommilitoninnen. Analog vermuten wir, dass Männer (H1a) insgesamt, (H1b) in schriftlichen Klausuren und (H1c) in Abschlussarbeiten schlechter abschneiden und dass Männer (H1d) in mündlichen Prüfungen und (H1e) in fachpraktischen Prüfungen besser abschneiden als Frauen (Reilly et al., 2019; Şahin et al., 2018; Huntley et al., 2017; Sheard, 2009). Zudem explorieren wir, ob sich die Noten in bestimmten fachpraktischen Studieninhalten hinsichtlich des Geschlechts unterscheiden.

Zweitens wird der Einfluss des *Geschlechts der Prüfenden* auf die Noten von Männern und Frauen untersucht (H2). (Hoch-)Schulische Forschungen zeigen, auch im sportwissenschaftlichen Kontext, dass Prüferinnen insgesamt bessere Noten vergeben als Prüfer (Jewell & McPherson, 2012; Wiskin et al., 2004). Laut Towfigh et al. (2018) geben Prüferinnen Frauen bessere Noten als Männern. Für Prüfer ist uns kein solcher Effekt bekannt. Daher untersuchen wir den Einfluss des Geschlechts der Prüfenden, also ob Prüfende Männern und Frauen unterschiedliche Noten geben.

5. METHODE

Die vorliegende Studie basiert auf einer Vollerhebung der Prüfungsdaten des Instituts für Sportwissenschaft der Universität Münster vom Wintersemester 2011/12 bis zum Sommersemester 2021. Mit einem positiven Votum der Ethikkommission des Fachbereichs wurde die Datenerhebung und -auswertung von der zuständigen Universitätsverwaltung sowie dem Institutsdirektor des Instituts für Sportwissenschaft unter strengen datenschutzrechtlichen Auflagen genehmigt (BERA, 2018). Es wurden anonymisierte soziodemografische Daten (Alter, Geschlecht, Nationalität, Zweitfach) sowie Prüfungsdaten (Prüfungsnoten, Fachsemester, Geschlecht der Prüfenden) erhoben.

Stichprobe

Gegenstand der Untersuchung waren 30.508 Prüfungsleistungen der Lehramtsstudierenden[3] mit Studienfach Sport der Universität Münster vom Wintersemester 2011/12 bis Sommersemester 2021 (vgl. Tab. 1). Von den insgesamt *n* = 2.868 Studierenden waren 1.461 männlich (51 %), wobei das Geschlecht in den Archiveinträgen des Prüfungsamts nur binäre Einträge enthielt. Das Durchschnittsalter bei Studienbeginn betrug 20,6 Jahre (*SD* = 2,66) und die Studierenden benötigten im Durchschnitt 6.68 Semester (*SD* = 1.63) zum Abschluss des Bachelors und 4,06 Semester (*SD* = 1,08) zum Abschluss des Masters. Insgesamt vergaben *n* = 130 Prüfende (60 % männlich, nur binär erhoben) die Noten.

Operationalisierung

Die Prüfungsleistungen werden in Teilnoten ausgewertet. Die Noten liegen von 1,0 (sehr gut) bis 5,0 (mangelhaft) vor. Die Teilnoten können dabei in 0,1-Schritten angegeben sein (1,0; 1,1; 1,2; 1,3 usw.). Diese Teilnoten wurden in die vier gemäß den Studienordnungen festgelegten Prüfungsformate eingeteilt: schriftliche Klausur, mündliche Prüfung, fachpraktische Prüfung und Abschlussarbeit. In schriftlichen Klausuren wurden fachwissenschaftliche Inhalte (darunter Sportpädagogik und -geschichte, Sportsoziologie, Sportpsychologie, Bewegungswissenschaft, Trainingswissenschaft, Sportmedizin) sowie fachdidaktische Inhalte (Sportdidaktik) geprüft. Die Klausuren zu fachwissenschaftlichen Inhalten sind dabei im Studienverlauf des Bachelorstudiums verortet. Mündliche Prüfungen können in einer fachwissenschaftlichen Disziplin absolviert werden und finden im Masterstudium statt. Fachpraktische Prüfungen werden im Bachelor absolviert und bestehen aus sportmotorischen und theoretischen Teilprüfungen. Die Gewichtungen werden in den jeweiligen Modulkonferenzen festgelegt (WWU, 2018). Die Prüfungen fokussieren dabei die Inhalte Spielen (*Spielsportarten*), Laufen, Springen, Werfen (*Leichtathletik*), Bewegen an und mit Geräten (*Turnen*), Bewegen im Wasser (*Schwimmen*), Gestalten, Tanzen, Darstellen (*Gymnastik/Tanz*) sowie sportübergreifende Kompetenzen (*Fitness, Natursport* und *Trendsport*).[4] Basis für die Studieninhalte sind die ländergemeinsamen Anforderungen der Lehrerbildung (KMK, 2019) sowie das von der Deutschen Vereinigung für Sportwissenschaft, der Arbeitsgemeinschaft

[3] Präziser formuliert werden hier die lehramtsbezogenen Studiengänge untersucht. Die Studiengänge Grundschule, Haupt-, Real-, Sekundar- und Gesamtschulen (HRSGe) sowie Berufskolleg sind explizite Lehramtsstudiengänge, während der 2-Fach-Bachelor (Gymnasium) polyvalent ausgerichtet ist. Der Großteil der Studierenden im 2-Fach-Bachelor strebt aber anschließend den Master of Education an, welcher dann wieder lehramtsbezogen ist. Auch weicht die Terminologie der einzelnen Lehramtsstudiengänge der Hochschule mitunter von den Vorgaben der KMK (2019) ab, die bspw. das *Lehramt am Berufskolleg* als *Lehramt an Beruflichen Schulen* bezeichnet.

[4] Die lt. Studienordnungen offiziellen Bezeichnungen der fachpraktischen Veranstaltungen an der Universität Münster wurden jeweils kursiv gesetzt und werden im weiteren Verlauf der Studie auch so genutzt. Diese Veranstaltungen werden grundsätzlich von Sportlehramtsstudierenden aller Studiengänge gemeinsam besucht; eine Ausnahme bilden hier Studierende des Grundschullehramts, die z. T. gesonderte Schwerpunkte erhalten (z. B. Kinderleichtathletik). Da die für diese Studie relevanten Prüfungsformate jedoch für alle Studienordnungen gleich sind, werden die fachpraktischen Studieninhalte zusammengefasst. Diese korrespondieren mit den von der KMK vorgegebenen fachpraktischen Studieninhalten, welche dort unter der Überschrift “Bewegungskompetenz und sportliches Können” (2019, S. 63) zusammengefasst werden. Es sei angemerkt, dass die Bezeichnungen der fachpraktischen Veranstaltungen dieser Hochschule auf eine sehr starke Sportartenorientierung hindeuten.

Tab. 1: Statistische Kennwerte der Stichprobe

	Männlich, *n*= 15.769	**Weiblich, *n*= 14.739**	***p***	**Effektgröße**
Metrische Variablen	*M (SA)*	*M (SA)*		
Prüfungsnote	2,75 (1,02)	2,45 (0,87)	< 0,001	0,320[a]
Fachsemester	3,88 (2,27)	3,44 (1,85)	< 0,001	0,214[a]
HZB	2,50 (0,55)	2,13 (0,52)	< 0,001	0,692[a]
Alter bei Studienbeginn	21,30 (3,02)	20,25 (2,26)	< 0,001	0,397[a]
Alter bei Prüfung	23,13 (3,07)	21,85 (2,35)	< 0,001	0,469[a]
Kategorielle Variablen	*n* (%)	*n* (%)		
Angestrebter Abschluss				
Bachelor	14.918 (94,6 %)	13.732 (93,2 %)	< 0,001	0,030[b]
Master	851 (5,4 %)	1.007 (6,8 %)		
Angestrebte Schulform				
Gymnasium	11.613 (73,6 %)	8.389 (56,9 %)		
Berufskolleg	1.264 (8,0 %)	1.135 (7,7 %)	< 0,001	0,281[b]
Grundschule	667 (4,2 %)	3.425 (23,2 %)		
HRSGe	2.225 (14,1 %)	1.790 (12,1 %)		
Prüfende				
Männlich	8.790 (63,7 %)	7.627 (57,8 %)	< 0,001	0,060[b]
Weiblich	5.003 (36,3 %)	5.560 (42,2 %)		
Prüfungsformat				
Schriftliche Klausur	6.693 (44,7 %)	6.262 (44,3 %)		
Abschlussarbeit	598 (4,0 %)	660 (4,7 %)		
Mündliche Prüfung	272 (1,8 %)	264 (1,9 %)	0,042	0,017[b]
Fachpraktische Prüfung	7.404 (49,5 %)	6.941 (49,1 %)		
Inhalt Fachpraxis				
Fitness	379 (5,1 %)	401 (5,8 %)		
Gymnastik/Tanz	1.081 (14,6 %)	1.088 (15,7 %)		
Leichtathletik	1.101 (14,9 %)	1.040 (15,0 %)		
Natursport	158 (2,1 %)	129 (1,9 %)		
Schwimmen	1.044 (14,1 %)	1.020 (14,7 %)	< 0,001	0,069[b]
Spielsportarten	2.321 (31,3 %)	2.135 (30,8 %)		
Trendsport	343 (4,6 %)	157 (2,3 %)		
Turnen	977 (13,2 %)	971 (14,0 %)		

[a]Cohen's *d*; [b]Cramer's *V*

Sportpsychologie, dem Fakultätentag Sportwissenschaft sowie dem Deutschen Sportlehrerverband verabschiedete Kerncurriculum Sportwissenschaft (dvs et al., 2017). Abschlussarbeiten im Lehramtsstudium Sport sind für die betrachteten Studiengänge fakultativ und können auch im Zweitfach oder in den Bildungswissenschaften geschrieben werden. In dieser Studie wurden nur Abschlussarbeiten im Fach Sportwissenschaft betrachtet. Tabelle 2 gibt einen Überblick über die Anteile von Fachwissenschaft, -praxis und -didaktik am sportwissenschaftlichen Lehramtsstudium.

Statistische Überlegungen und Auswertung

Die Datenanalysen wurden mit R Studio (Version 1.1.463; R Core Team, 2018) und vornehmlich den Paketen *lme4* (Bates et al., 2015) für Spezifizierung der Mehrebenenmodelle sowie *emmeans* (Lenth, 2021) für die Berechnung der Randmittel durchgeführt. Das Signifikanzniveau wurde auf $\alpha = 0{,}05$ festgelegt. Die Effektgrößen in den standardisierten Maßen werden nach Funder und Ozer (2019) interpretiert, sodass ein Effekt $0{,}05 < |ß| < 0{,}2$ als klein ($0{,}1 < d < 0{,}4$), ein Effekt $0{,}2 \leq |ß| < 0{,}3$ ($0{,}4 \leq d < 0{,}63$) als mittel und ein Effekt $|ß| \geq 0{,}3$ ($d \geq 0{,}63$) als groß interpretiert wird. Alle berechneten Modelle werden als offener Code und reproduzierbare Ergebnisse als R Markdowm Datei im

CC BY 4.0 DE

Tab. 2: Übersicht der Bestandteile des sportwissenschaftlichen Lehramtsstudiums an der Universität Münster am Beispiel der Schulform Gymnasium/Gesamtschule

	Leistungspunkte		Anteil der Gesamtnote in %	
	Bachelor	Master	Bachelor	Master
Fachwissenschaft	43	12	55	45
Fachpraxis	24	0	35	0
Fachdidaktik	8	13	10	55
Gesamt	75	25	100	100

Anmerkungen. Die Aufteilung weicht zwischen den einzelnen Schulformen leicht ab.

Supplement zur Verfügung gestellt (Utesch & Magner, 2023). Aus datenschutzrechtlichen Gründen können die Daten nicht offen hochgeladen werden, bei Fragen bitten wir um Kontaktaufnahme mit dem Erstautor.

Zunächst wurden demografische Merkmale zwischen den beiden Geschlechtern deskriptiv dargestellt und Unterschiede untersucht. Hierfür wurden t-Tests für unabhängige Stichproben für die HZB, das Alter der Studierenden bei Studienbeginn und Prüfung, die Note und das Fachsemester bei Prüfung sowie Chi²-Tests für die angestrebte Schulform, das Geschlecht der Prüfenden, das Prüfungsformat und die Studieninhalte der Fachpraxis berechnet. Falls bei Variablen keine signifikanten Unterschiede vorlagen, wurden Tests auf Annahme der Gleichheit (tost) vorgenommen.

Um die Varianzverteilung in der geschachtelten Datenstruktur einzubeziehen, wurden zunächst Intraklassenkorrelationen (ICC) bei nicht-konditionalen Modellen (Nullmodelle) für Einzelprüfungen geschachtelt in Studierenden, Einzelprüfungen geschachtelt in Zweitfächern und Einzelprüfungen geschachtelt in Prüfenden berechnet. Eine substantielle Variation liegt bei mehr als 5 % Varianz vor, die durch einen Cluster erklärt wird. Diese Variation der Noten ergab sich innerhalb und zwischen Studierenden (ICC = 0,185) sowie innerhalb und zwischen Prüfenden (ICC = 0,135) jedoch nicht innerhalb und zwischen Zweitfächern (ICC = 0,035). Zur Überprüfung der Hypothesen 1a-e sowie 2 und der explorativen Fragestellungen, wurden somit kreuzweise klassifizierte lineare gemischte Modelle (*cross-classified linear mixed models*) mit der abhängigen Variable Note mit Cross-Random-Effekten berechnet, da sie eine gleichzeitige Schätzung der Varianz zwischen den Studierenden und zwischen den Prüfenden ermöglichen (Judd et al., 2012). Daher wurden die Level-Faktoren Prüflinge und Prüfende mit jeweils Random Intercepts berücksichtigt, um beide Quellen der Varianz zu nutzen. Dies bedeutet, dass bei der Spezifikation der Modelle die Realität so abgebildet wird, dass jedem Studierenden eine unterschiedliche Durchschnittsnote (Intercept) erlaubt wird und dass es jedem Prüfenden erlaubt ist, eine unterschiedliche Durchschnittsnote zu vergeben. Die Modelle berücksichtigen hierbei Schätzungen der Zufallsvarianzen (gesamt σ, τ(Studierende), τ(Prüfende)) und den ICC des Modells. Das marginale R^2 gibt dabei die Aufklärung der fixen Effekte und das konditionale R^2 die Aufklärung der fixen und zufälligen Effekte an. Diese Analyse ermöglicht es, die Hypothesen 1 und 2 so sparsam wie möglich in einem einzigen Modell auszuwerten. Hierfür wurde der Datensatz zunächst in ein langes Format mit 30.508 Zeilen (1 Zeile pro vergebener Note) gebracht. Das finale Modell umfasste folgende Prädiktoren: das Geschlecht des Studierenden (dichotom), das Geschlecht des Prüfenden (dichotom), das Prüfungsformat (4-stufig diskret: schriftliche Klausur, mündliche Prüfung, fachpraktische Prüfung, Abschlussarbeit), die Kontrollvariablen HZB (kontinuierlich) sowie den angestrebten Abschluss (dichotom: Bachelor bzw. Master). Zudem wurden die Interaktionsterme *Geschlecht_Studierende x Geschlecht_Prüfende* sowie *Geschlecht_Studierende x Prüfungsformat* eingefügt. Es wurden Kontraste für das Geschlecht der Studierenden gebildet und geschätzte Randmittelwerte nach Kontrolle durch die Kovariaten des Modells (Estimated Marginal Means) berechnet. Dies bedeutet, dass Mittelwertsunterschiede unter Kontrolle aller anderen Variablen berichtet werden können, ähnlich wie sie bei varianzanalytischem Vorgehen ohne Berücksichtigung der Mehrebenenstruktur berichtet werden.

Zur Beantwortung der Hypothesen wurden das Regressionsgewicht *B* in der Metrik der gegebenen Noten (zwischen 1 (*sehr gut*) und 5 (*mangelhaft*), inkl. Zwischennoten) sowie das standardisierte Regressionsgewicht ß interpretiert. Aus der Perspektive praktischer Bedeutsamkeit werden Notenunterschiede unter 0,05 Notenpunkte dabei als marginale Unterschiede betrachtet. Notenunterschiede bis 0,15 werden als mittlere Unterschiede betrachtet und größere Notenunterschiede als große Unterschiede. Bei der Berechnung von Kontrasten und der geschätzten Randmittel der Noten von Männern und Frauen werden Mittelwertsunterschiede in der Metrik von Noten (*B*) sowie der Teststatistik (*z*) berichtet. Zudem werden die marginalen Mittelwerte und Standardfehler berichtet. Da es sich um eine Differenz handelt, ist zu erwähnen, dass positive Werte (*B*) hierbei bedeuten, dass Frauen bessere Noten aufweisen als Männer.

Zur Beantwortung der Hypothese 1a wird der Haupteffekt Geschlecht interpretiert, wobei gemäß der Hypothese ein negativer Wert im Regressionsgewicht bedeuten würde, dass Frauen bessere Noten aufweisen als Männer. Deskriptiv werden die marginalen Randmittelwerte angegeben. Zur Beantwortung der Hypothesen 1b-e werden zunächst die Interaktionseffekte *Geschlecht x Prüfungsformat [jeweiliges Prüfungsformat]* interpretiert sowie anschließend die geschätzten Randmittel der Noten von Männern und Frauen in Bezug auf den jeweiligen Kontrast Prüfungsform (H1b) schriftliche Klausur, (H1c) Abschlussarbeit, (H1d) mündliche Prüfung, (H1e) fachpraktische Prüfung. Hierfür werden die Mittelwertsunterschiede (*B*) sowie die standardisierten Effekte (*z*)

berichtet. Positive Werte im Mittelwertunterschied (*B*) bedeuten, dass Frauen bessere Noten aufweisen als Männer. Post-hoc wird zudem analysiert, wie sich die Prüfungsformen untereinander unterscheiden.

Zur Beantwortung der Hypothese 2 wird der Haupteffekt Geschlecht_Prüfende interpretiert. Ein negativer Wert bedeutet, dass Frauen bessere Noten vergeben. Zur Beantwortung der Frage, ob es einen doppelten Geschlechtereffekt bei der Notengebung gibt, wird der Interaktionseffekt *Geschlecht_Studierende* x *Geschlecht_Prüfende* und die zugehörigen geschätzten Randmittel interpretiert.

Darüber hinaus untersuchen wir, ob es Unterschiede in den Noten von Männern und Frauen zwischen fachpraktischen Inhalten gibt. Hierfür wird der Datensatz nach fachpraktischen Prüfungen gefiltert und bei der Modellierung wird die Variable *Prüfungsformat* ersetzt durch die Variable *Fachpraxis* (diskrete Variable, enthält alle unterschiedlichen fachpraktischen Prüfungen, z. B. Schwimmen, Fußball, Turnen, usw.). Das oben beschriebene Vorgehen der Modellierung wird anschließend wiederholt.

6. ERGEBNISSE

Zunächst werden die deskriptiven Variablen bzgl. Geschlechterunterschieden untersucht: Hier zeigt sich, dass in allen Variablen signifikante Unterschiede vorliegen (vgl. Tab. 1): Frauen starten ihr Studium mit einer besseren HZB, sind bei Studienbeginn jünger und streben häufiger die Schulform Grundschule und seltener das Gymnasium an. Im Studienverlauf erreichen sie bessere Einzelnoten und absolvieren die Teilprüfungen in einem früheren Fachsemester.

Zur inferenzstatistischen Überprüfung der Hypothesen wurde ein gekreuzt klassifiziertes lineares gemischtes Modell mit der abhängigen Variable Note und den Prädiktoren Geschlecht der Studierenden, Geschlecht der Prüfenden, Prüfungsformat und deren Interaktion berechnet (s. Tab. 3, σ = 0,56, τ(Studierende) = 0,09, τ(Prüfende) = 0,09, ICC = 0,25, marginales R^2 = 0,13, konditionales R^2 = 0,348). Der große Unterschied zwischen marginalem und konditionalem R^2 zeigt, dass die zufälligen Effekte das Modell stark verbessern. Darüber hinaus wurde für die HZB und auch für den angestrebten Abschluss (Bachelor/Master) kontrolliert.

Die Effekte der Kontrollvariablen der HZB (*B* = 0,41, ß = 0,25, *p* < 0,001) und des Studiengangs (*B* = -0,16, ß = -0,17, *p* < 0,001) sind signifikant: Studierende mit besseren (vs. schlechteren) HZBs (mittlere Effektgröße) und Studierende im Masterstudium (vs. Bachelorstudium) erreichen bessere Noten (kleine bis mittlere Effektgröße).

In Bezug auf Hypothese 1a ergab der Haupteffekt des Geschlechts (*B* = -0,17, ß = -0,18, *p* < 0,001) einen signifikanten kleinen Effekt zugunsten der Frauen. Frauen erreichen über alle Prüfungen hinweg bessere Noten als Männer. Die geschätzten Randmittel des Haupteffekts zeigen deskriptiv einen kontrollierten Mittelwertunterschied von 0,05 Notenpunkten. In Hypothese 1b-e wurde das Prüfungsformat untersucht. Die signifikanten Interaktionseffekte liegen dabei zwischen 0,11 < ß < 0,35 (*p* < 0,001) und zeigen kleine bis mittlere Effektgrößen: In schriftlichen Klausuren (H1b) zeigt sich ein signifikanter Unterschied von *B* = 0,22 Notenpunkten (*z* = 10,32, *p* < 0,001), wobei Männer (*M* = 2,69) schlechtere Noten aufweisen als Frauen (*M* = 2,47). In Abschlussarbeiten (H1c) schneiden Männer (*M* = 1,92) signifikant besser ab als Frauen (*M* = 2,03) mit einem Unterschied von *B* = -0,11 Notenpunkten (*z* = 2,10, *p* = 0,036). In mündlichen Prüfungen (H1d) zeigen sich zwischen Männern (*M* = 1,98) und Frauen (*M* = 2,00) keine signifikanten Unterschiede (*z* = -0,34, *p* = 0,731). In fachpraktischen Prüfungen (H1e) schneiden Männer (*M* = 2,27) signifikant schlechter ab als Frauen (*M* = 2,15) mit einem Unterschied von *B* = -0,11 Notenpunkten (*z* = 5,79, *p* < 0,001).

Bei der Auswertung der Hypothese 2, dass Prüferinnen insgesamt bessere Noten vergeben, zeigt sich ein signifikanter Haupteffekt mit kleiner Effektgröße für das Geschlecht des Prüfenden (*B* = 0,15, ß = 0,16, *p* = 0,026), wobei in unserer Stichprobe Prüferinnen insgesamt schlechtere Noten vergeben als Prüfer. Weiterhin wird der Zusammenhang mit dem Geschlecht der Prüfenden auf die Benotung von Männern bzw. Frauen untersucht, wobei ein signifikanter Interaktionseffekt mit kleiner Effektgröße vorliegt (*B* = 0,10, ß = 0,10, *p* = 0,001). In der geschätzten Randmittelanalyse zeigt sich, dass Prüferinnen Studenten (*M* = 2,29) im Gegensatz zu Studentinnen (*M* = 2,19) um *B* = 0,1 Notenpunkte schlechter bewerten (*z* = 3,13, *p* = 0,001) während Prüfer Studenten (*M* = 2,14) im Gegensatz zu Studentinnen (*M* = 2,14) nicht unterschiedlich bewerten (*B* = 0,0, *z* = 0,09, *p* = 0,929).

Post-hoc Analysen zeigen zudem, dass schriftliche Prüfungen insgesamt zu signifikant und mit einer sehr großen Effektgröße deutlich schlechteren Noten führen als mündliche Prüfungen (*B* = -0,72,ß = -0,76, *p* < 0,001), als Abschlussarbeiten (*B* = -0,77, ß = -0,82, *p* < 0,001) und als fachpraktische Prüfungen (*B* = -0,42, ß = -0,45, *p* < 0,001).

Im zweiten Schritt wird die Notengebung in den fachpraktischen Kursen separat nach Inhalten untersucht (s. Tab. 3, σ = 0,44, τ(Studierende) = 0,08, τ(Prüfende) = 0,06, ICC = 0,25, marginales R^2 = 0,164, konditionales R^2 = 0,373). Der große Unterschied zwischen marginalem und konditionalem R^2 zeigt, dass die zufälligen Effekte das Modell stark verbessern. Die Effekte der Kontrollvariablen der HZBs (*B* = 0,36, ß = 0,24, *p* < 0,001) sind hier deskriptiv vergleichbar mit denen über alle Prüfungen.

In fachpraktischen Prüfungen zeigt sich ebenfalls ein Haupteffekt in Bezug auf das Geschlecht (*B* = -0,15, ß = -0,18, *p* < 0,001) mit einem signifikanten kleinen Effekt zugunsten der Frauen. Die geschätzten Randmittel des Haupteffekts zeigen deskriptiv einen kontrollierten Mittelwertunterschied von 0,15 Notenpunkten (*z* = 6,53, *p* < 0,001). Zudem zeigt sich, dass das Geschlecht der Prüfenden keinen signifikanten Haupteffekt (*B* = -0,12, ß = -0,14, *p* < 0,091), aber einen signifikanten Interaktionseffekt mit kleiner Effektgröße zeigt (*B* = -0,10, ß = -0,12, *p* < 0,004). In den Randmitteln zeigt sich, dass Prüferinnen Studenten (*M* = 2,45) im Gegensatz zu Studentinnen (*M* = 2,24) um *B* = 0,20 Notenpunkte schlechter bewerten (*z* = 6,77, *p* = 0,001) während Prüfer Studenten (*M* = 2,33) im Gegensatz zu Studentinnen (*M* = 2,23) um 0,1 Notenpunkte schlechter bewerten (*B* = 0,10, *z* = 3,492, *p* = 0,005). In den fachpraktischen Prüfungen zeigen sich folgende Mittelwertunterschiede, wobei positive Werte (*B*) bessere Noten für Frauen bedeuten (vgl. Tab. 4): Fitness (*B* = 0,20, *z* = 3,539,*p* = 0,004), Gymnastik/Tanz (*B* = 0,33, *z* = 8,604, *p* = 0,001), Leichtathletik (*B* = 0,16, *z* = 4,247, *p* = 0,001), Natursport (*B* = 0,01, *z* = 0,128, *p* = 0,898), Schwimmen (*B* = 0,11, *z* = 2,954, *p* = 0,003), Spielsportarten (*B* = -0,07, *z* = -2,612, *p* = 0,009), Trendsport (*B* = 0,18, *z* = 2,341, *p* = 0,019), (*B* = 0,20, *z* = 3,539, *p* = 0,004) und Turnen (*B* = 0,28, *z* = 7,716, *p* = 0,001). Zusammengefasst gibt es im Natursport keine bedeutsamen Unterschiede, den kleinsten Unterschied danach zugunsten der Männer in Spielsportarten (0,07 Punkte) und in den weiteren sechs Inhalten bedeutsame Unterschiede zwischen 0,1 und 0,33 Notenpunkten zugunsten der Frauen.

CC BY 4.0 DE

Tab. 3: Gekreuzt-klassifiziertes lineares gemischtes Modell für Noten in verschiedenen Prüfungsformaten

Prädikatoren	*B*	ß	*KI* (95 %)	*ß KI* (95 %)	*p*
(Konstante)	1,73	0,12	1,62 – 1,85	0,02 – 0,21	< 0,001
Geschlecht [w]	-0,17	-0,18	-0,21 – -0,13	-0,23 – -0,14	< 0,001
Geschlecht Prüfende [w]	0,15	0,16	0,02 – 0,28	0,02 – 0,29	0,026
Prüfungsformat [Abschlussarbeit]	-0,77	-0,82	-0,84 – -0,69	-0,89 – -0,74	< 0,001
Prüfungsformat [Mündliche Prüfung]	-0,72	-0,76	-0,84 – -0,59	-0,90 – -0,63	< 0,001
Prüfungsformat [Fachpraktische Prüfung]	-0,42	-0,45	-0,47 – -0,38	-0,50 – -0,40	< 0,001
Hochschulzugangsberechtigung	0,41	0,25	0,38 – 0,44	0,23 – 0,27	< 0,001
Angestrebter Abschluss [Master]	-0,16	-0,17	-0,22 – -0,10	-0,23 – -0,10	< 0,001
Geschlecht [w] x Geschlecht Prüfende [w]	-0,10	-0,10	-0,14 – -0,05	-0,15 – -0,06	< 0,001
Geschlecht [w] x Prüfungsformat [Abschlussarbeit]	0,33	0,35	0,23 – 0,43	0,24 – 0,45	< 0,001
Geschlecht [w] x Prüfungsformat [Mündliche Prüfung]	0,25	0,27	0,09 – 0,41	0,10 – 0,43	0,002
Geschlecht [w] x Prüfungsformat [Fachpraktische Prüfung]	0,11	0,11	0,06 – 0,15	0,07 – 0,16	< 0,001
Random Effects					
σ^2	0,56				
τ_{00} Studierende	0,09				
τ_{00} Prüfende	0,09				
ICC	0,25				
Marginal R^2 / Conditional R^2	0,130 / 0,348				

Anmerkungen. Bezugskategorien im vorliegenden Modell sind Studenten und schriftliche Klausuren im Bachelor und Prüfer. Die HZB wurde nicht zentriert. KI beschreibt das Konfidenzintervall. Insgesamt wurden n = 23.460 Prüfungen modelliert.

7. DISKUSSION

Ziel dieser Studie war es herauszufinden, ob über Prüfungsformate und über den Verlauf des gesamten Lehramtsstudiums Sport Geschlechterunterschiede in Prüfungsleistungen identifiziert werden können. Zudem wurde die Notengebung in fachpraktischen Veranstaltungen der sportwissenschaftlichen Hochschulbildung untersucht. Gegenstand der Analysen waren alle Prüfungsleistungen im Lehramtsstudium Sport einer deutschen Hochschule über einen Zeitraum von zehn Jahren mit insgesamt 2.868 Studierenden und 130 Prüfenden. Zu den Prüfungsformaten gehörten schriftliche Prüfungen, mündliche Prüfungen, Abschlussarbeiten und die insbesondere für das Fach Sport charakteristischen fachpraktischen Prüfungen. Grundsätzlich war und ist unsere Prämisse, dass - unter Kontrolle der HZB, die den Studienerfolg maßgeblich prädiziert - Männer und Frauen gleichermaßen gut für das Lehramtsstudium Sport geeignet sind.

Diese Studie kommt dem Desiderat nach, Noten auf Teilprüfungen und Prüfungsformate differenziert und langfristig über den Verlauf des Studiums zu betrachten. Unsere Ergebnisse korrespondieren insgesamt mit Ergebnissen, die in internationalen Studien berichtet werden: Für das Fach Sport, in dem auch sportmotorische Leistungen erbracht werden müssen, sind bessere Abschlussnoten für Frauen evident (Huntley et al., 2017; Peters et al., 2007; Sheard, 2009; Troche et al., 2010). Es wurde im Einklang zur Literatur (Voyer & Voyer, 2014) festgestellt, dass Frauen in den zehn Jahren insgesamt besser abschnitten als Männer. Männer erzielen über die Prüfungsformate hinweg deutlich schlechtere bis leicht bessere Noten. Frauen erzielen in schriftlichen Klausuren sowie fachpraktischen Prüfungen bessere Noten als Männer, während in Bezug auf mündliche Prüfungen keine Unterschiede festgestellt wurden und in Abschlussarbeiten Geschlechterunterschiede zugunsten der Männer vorliegen. In sechs von acht fachpraktischen Prüfungen erhalten Frauen signifikant bessere Noten als Männer, mit besonders großen Effekten bei kompositorischen Studieninhalten. Spielsportarten sind der einzige Inhaltsbereich, in dem Männer marginal (unter 0,07 Notenpunkte) bessere Noten erzielen als Frauen.

Tab. 4: Gekreuzt-klassifiziertes lineares gemischtes Modell für Noten in der Fachpraxis

Prädikatoren	***B***	**ß**	***KI* (95 %)**	***ß KI* (95 %)**	***p***
(Konstante)	1,21	-0,41	1,05 – 1,37	-0,58 – -0,25	< 0,001
Geschlecht [w]	-0,15	-0,18	-0,26 – -0,04	-0,32 – -0,04	0,010
Geschlecht Prüfende [w]	0,12	0,14	-0,02 – 0,26	-0,02 – 0,31	0,091
Studieninhalt [Gymnastik/Tanz]	0,30	0,36	0,15 – 0,45	0,18 – 0,53	< 0,001
Studieninhalt [Leichtathletik]	0,46	0,55	0,37 – 0,56	0,44 – 0,67	< 0,001
Studieninhalt [Natursport]	0,09	0,11	-0,09 – 0,27	-0,10 – 0,32	0,318
Studieninhalt [Schwimmen]	0,67	0,80	0,50 – 0,84	0,59 – 1,01	< 0,001
Studieninhalt [Spielsportarten]	0,06	0,07	-0,08 – 0,19	-0,10 – 0,23	0,429
Studieninhalt [Trendsport]	-0,00	-0,00	-0,16 – 0,16	-0,19 – 0,19	0,995
Studieninhalt [Turnen]	0,63	0,75	0,48 – 0,78	0,57 – 0,93	< 0,001
Hochschulzugangsberechtigung	0,36	0,24	0,33 – 0,39	0,22 – 0,26	< 0,001
Geschlecht [w] x Geschlecht Prüfende [w]	-0,10	-0,12	-0,17 – -0,03	-0,21 – -0,04	0,004
Geschlecht [w] x Studieninhalt [Gymnastik/Tanz]	-0,13	-0,16	-0,26 – 0,00	-0,31 – 0,00	0,051
Geschlecht [w] x Studieninhalt [Leichtathletik]	0,04	0,05	-0,09 – 0,17	-0,10 – 0,20	0,548
Geschlecht [w] x Studieninhalt [Natursport]	0,19	0,23	-0,02 – 0,40	-0,03 – 0,48	0,080
Geschlecht [w] x Studieninhalt [Schwimmen]	0,09	0,11	-0,04 – 0,22	-0,05 – 0,26	0,172
Geschlecht [w] x Studieninhalt [Spielsportarten]	0,27	0,33	0,16 – 0,39	0,19 – 0,47	< 0,001
Geschlecht [w] x Studieninhalt [Trendsport]	0,02	0,02	-0,16 – 0,21	-0,20 – 0,25	0,825
Geschlecht [w] x Studieninhalt [Turnen]	-0,09	-0,10	-0,21 – 0,04	-0,25 – 0,05	0,190
Random Effects					
σ^2	0,44				
τ_{00} Studierende	0,08				
τ_{00} Prüfende	0,06				
ICC	0,25				
Marginal R^2 / Conditional R^2	0,164 / 0,373				

Anmerkungen. Bezugskategorien im vorliegenden Modell sind Studenten in der Sportart Fitness im Bachelor und Prüfer. Die HZB wurde nicht zentriert. KI beschreibt das Konfidenzintervall. Insgesamt wurden n = 12.180 Prüfungen modelliert.

CC BY 4.0 DE

Tab. 5: Beschreibung der Randmittel in Noten fachpraktischer Prüfungen

		Männer - Frauen			Männer		Frauen	
		B	*z*	*p*	M	SF	M	SF
Studieninhalt	Fitness	0,20	3,54	< 0,001	2,11	0,06	1,91	0,06
	Gymnastik/Tanz	0,33	8,60	< 0,001	2,41	0,05	2,08	0,05
	Leichtathletik	0,16	4,25	< 0,001	2,57	0,06	2,41	0,06
	Natursport	-0,01	0,13	0,898	2,20	0,07	2,19	0,08
	Schwimmen	0,11	2,95	0,003	2,78	0,07	2,67	0,07
	Spielsportarten	-0,07	-2,61	0,009	2,17	0,04	2,24	0,04
	Trendsport	0,18	2,34	0,019	2,11	0,05	1,93	0,07
	Turnen	0,29	7,72	< 0,001	2,74	0,05	2,45	0,05

Anmerkungen. B steht für die Differenz der Mittelwerte zwischen Männern und Frauen. Positive Werte bedeuten, dass Frauen bessere Noten aufweisen. Die Mittelwerte (M) und die Standardfehler (SF) resultieren aus den geschätzten Randmitteln des gerechneten Modells, d. h. sie sind unter Kontrolle der Hochschulzugangsberechtigung zu interpretieren und wurde nicht einfach aus der Stichprobe als Mittelwerte berechnet. z steht für den Z-Test der Differenz.

Praktische Relevanz der Ergebnisse

Es konnte gezeigt werden, dass das Geschlecht der Prüfenden einen substanziellen Einfluss auf die Notenvergabe bei Männern und Frauen hat, wobei Prüferinnen Studenten schlechtere Noten gaben als Studentinnen, während bei Prüfern keine Geschlechterunterschiede konstatiert werden konnten. Diese Ergebnisse basieren auf einer sehr großen Datenbasis, die eine Vollerhebung über zehn Jahre darstellt. Sie wurden unter Kontrolle der geschachtelten Datenstruktur für Prüfende und für Studierende sowie unter Kontrolle der Note der HZB, dem angestrebten Studienabschluss (Bachelor/Master) sowie den Prüfungsformaten erzielt und werden daher von den Autoren dieser Studie als sehr robust eingeordnet. Die Interpretation der hier beschriebenen Effekte und Effektgrößen sollte aus unserer Sicht vor dem Hintergrund der praktischen Relevanz vorgenommen werden. Dabei weisen Funder und Ozer (2019) darauf hin, dass auch Effekte kleiner bis mittlerer Größe neben langfristig kumulierenden Wirkweisen insbesondere auch praktisch kurzfristige Relevanz besitzen. Durchgängig konnten in dieser Studie Effektgrößen im statistisch kleinen bis mittleren Bereich und praktisch im bedeutsamen Bereich festgestellt werden. Da die gefundenen Geschlechterunterschiede in Noten der Studierenden dieser Größe entsprechen, sind neben langfristig kumulierenden Effekten auch Effekte auf unmittelbarer individueller Ebene anzunehmen. Einzelnoten schlagen sich additiv in der Abschlussnote des Studiums nieder, sodass sich bereits bei den kleineren Effekten eine Benachteiligung bei Einzelnoten insgesamt und über die Zeit summiert (Funder & Ozer, 2019).

Die Überlegungen darüber, wann und auf welche Weise die Effekte bei der Notengebung in ihrer Stärke oder ihren Folgen kumulieren werden, halten wir für besonders wichtig. Diese Kumulation findet einerseits auf individueller Ebene im Laufe des Studiums statt, wenn die Notengebung in einem einzelnen Anlass stattfindet, im zeitlichen Verlauf eines Studiums, wenn sich bspw. im Studienverlauf des Bachelors viele Einzelnoten zu einer Abschlussnote gewichtet mitteln und andererseits auf struktureller Ebene, wenn sich Notentendenzen über Gruppen hinweg manifestieren. Darüber hinaus können sich Sozialisationseffekte bilden, wenn beispielsweise Praktika oder Stellen als studentische Hilfskräfte über (Teil-)Noten vergeben werden. Werden die hier gefundenen Effektgrößen mit der Literatur (z. B. in der pädagogischen Lernforschung; Hattie, 2009) verglichen, stellen wir fest, dass sie sich in ähnlicher bis größerer Höhe bewegen.

Das Ergebnis, dass Frauen besser abschneiden, lässt sich zunächst durch existierende Befunde auf der Prozessebene erklären, etwa dass Studentinnen sich angemessener auf Klausuren vorbereiten (Woodfield et al., 2005; Smith, 2004), oft motivierter (Kuśnierz et al., 2020) und engagierter (Sheard, 2009) sind oder ihre Lernstrategien besser anpassen (Peters et al., 2007). Obwohl Prüferinnen jedoch in der internationalen Literatur eher bessere Noten vergeben (Jewell & McPherson, 2012; Wiskin et al., 2004), vergaben sie in der vorliegenden Stichprobe schlechtere Noten. Zudem zeigte sich, dass sie Studenten gegenüber Studentinnen schlechtere Noten gaben, was sich nicht durch oben genannte Erklärungsversuche begründen lässt, da Prüfer diese Tendenz nicht aufwiesen. Dieser Effekt findet sich parallel in der Literatur, da hochschulische Notenvorteile für Frauen bei Prüferinnen bereits in anderen Fächern belegt wurden (Towfigh et al., 2018). Allerdings ist es auf der Grundlage der vorliegenden Archivdaten nicht möglich, Prozesse zu erforschen (und es war auch nicht das Ziel dieser Arbeit), die die unterschiedlichen Benotungstendenzen von Prüferinnen und Prüfern erklären würden. Die empirische Schulforschung liefert hier zumindest Belege dafür, dass Frauen im Bildungskontext besser angepasste lernrelevante Merkmale und Verhaltensweisen zeigen und dadurch bei gleichen Kompetenzen bessere Noten als Männer erhalten (Hannover & Kessels, 2011). Eine Erklärung dieses Befundes auf Hochschulebene bleibt ein spannendes Desiderat für zukünftige Forschungsarbeiten.

Bedeutung von Prüfungsformaten im Lehramtsstudium Sport

Dass Studenten schließlich in fast allen Prüfungsformaten der fachwissenschaftlichen und -didaktischen Inhalte im Lehramtsstudium Sport schlechtere Noten als ihre Kommilitoninnen erzielen, lässt sich in der Richtung als Muster beschreiben und ist hinsichtlich Geschlechtergerechtigkeit insgesamt ein alarmierendes Ergebnis. Die von Mutz und Burrmann (2014) für den Sportunterricht festgestellten systematischen Nachteile für Mädchen kehren sich im Lehramtsstudium Sport in systematische Nachteile für Männer um. Befunde, wonach Männer im Sportstudium oft nicht die richtige Lernstrategie wählen, existieren (Peters et al., 2007). Nach dem *Constructive-Alignment-Konzept* von Biggs und Tang (2011) müssen Lehr-Lernaktivitäten und Prüfungen passgenau aufeinander abgestimmt sein. Wenn Männer aber über einen Zeitraum von zehn Jahren schlechter abschneiden, weil für sie beispielsweise Teile bestehender Curricula, Bewertungskriterien oder häufig genutzte Prüfungsformate weniger Passung aufweisen, müssen auch Institute und Lehrende ihre Methoden und Anforderungsprofile grundsätzlich hinterfragen (Schneider & Preckel, 2017). An dieser Stelle könnten auch innovative Prüfungsformate installiert (Gerick et al., 2018) oder den Studierenden sogar – oft gewünschte – Wahlmöglichkeiten zwischen verschiedenen Prüfungsformaten eingeräumt werden (Bülow-Schramm, 2013). Neben den gängigen summativen Prüfungen in modularen Studiengängen bieten besonders formative Rückmeldeformate große Chancen zur Steigerung studentischen Lernens sowie der Lernmotivation, von denen speziell Männer profitieren könnten (Hernández, 2012). In den Noten von Abschlussarbeiten wurden sogar leichte Geschlechterunterschiede zugunsten der Männer gefunden, was ebenfalls für formative Rückmeldungen spricht - nicht selten sind Bachelor- und Masterarbeiten durch ein starkes Eigeninteresse der Studierenden sowie eine engmaschige Betreuung seitens der Hochschule geprägt.

Geschlechterunterschiede in fachpraktischen Prüfungen

Eine besondere Perspektive wird den fachpraktischen Prüfungen zuteil, da Noten hier teilweise auf der Grundlage von normierten Leistungs- bzw. Notentabellen (z. B. in der Leichtathletik oder im Schwimmen) vergeben werden und Männer absolut gesehen bessere Leistungen erbringen müssen als Frauen (Stiller & Kahlert, 2021). In beiden Sportarten werden z. T. nach Geschlecht differenzierte Bewertungsmaßstäbe angelegt, um anatomische und physiologische Unterschiede auszugleichen. In diesen Sportarten sollten durch die Normtabellen objektiv gleiche Noten resultieren, was sich empirisch nicht zeigt. In der fachpraktischen Prüfung Leichtathletik schneiden Frauen etwa um 0,15 Notenpunkte besser ab als Männer, im Schwimmen um etwa 0,11 Notenpunkte. Diese Befunde zeigen, dass die vermeintlich objektiven Bewertungsmaßstäbe Frauen bevorzugen bzw. Männer benachteiligen könnten. Naheliegend ist, die Norm- und Notentabellen auf den Prüfstand zu stellen und ggf. Anpassungen in Richtung Geschlechtergerechtigkeit sowohl bzgl. der körperlichen Voraussetzungen zwischen als auch innerhalb der Geschlechter vorzunehmen.

Allerdings zeigen sich noch größere Notenunterschiede in den kompositorischen Studieninhalten Gymnastik/Tanz (0,33) und Turnen (0,28) zugunsten der Frauen. Dieses Ergebnis legt nahe, dass einige Sportarten ggf. immer noch geschlechtstypisch konnotiert und stärker mit weiblichen Körper- und Bewegungsidealen verbunden sein könnten (Frohn & Süßenbach, 2012), an denen gezeigte Leistungen evaluiert werden. Eine objektive und ergebnisoffene Prüfung der curricularen Inhalte sowie der angelegten Bewertungskriterien bzw. deren Auslegungen in Bezug auf die Berufsanforderungen und unter Berücksichtigung von Gleichstellungsaspekten ist hier aus unserer Sicht angezeigt. Es gibt zwar grundsätzliche Hinweise darauf, dass Männer möglicherweise größere Schwierigkeiten mit den theoretischen Anteilen fachpraktischer Prüfungen haben könnten (Kuśnierz et al., 2020; Sheard, 2009), diese sollten sich jedoch aus unserer Perspektive auf alle fachpraktischen Prüfungen gleichermaßen auswirken. Vorab war die Hypothese aus der Literatur abgeleitet worden, dass Männer in der Fachpraxis besser abschneiden müssten, da mehr Männer durch deutlich höhere Aktivität und Mitgliedschaften in prüfungsrelevanten Spielsportarten häufiger einschlägig vereinssportlich sozialisiert werden – mit Ausnahme von Hockey in der Altersklasse 15-18 Jahre und Volleyball insgesamt (DOSB, 2021). Dennoch zeigen sich hier die meisten Unterschiede zugunsten besserer Noten von Frauen.

Prüfungskultur im Lehramtsstudium Sport

Unsere Ergebnisse zeigen Unterschiede der Prüfungen bzw. Prüfungsformate im Lehramtsstudiengang Sport. Es gibt keinen sachlichen Grund dafür, dass schriftliche Prüfungen (durchschnittliche Note für Männer 2,69) und mündliche Prüfungen (durchschnittliche Note für Männer 1,98) oder fachpraktische Prüfungen (durchschnittliche Note für Männer 2,27) so deutlich unterschiedliche Mittelwerte aufweisen sollten oder dass es zwischen der am besten bewerteten fachpraktischen Prüfung (Fitness für Frauen durchschnittlich 1,91) und der am schlechtesten bewerteten fachpraktischen Prüfung (Schwimmen für Männer durchschnittlich 2,78) so einen großen Unterschied geben sollte. Vielmehr sollte sich ähnliche Verteilungen zeigen. Man könnte die Fragestellung ableiten, ob alle Prüfungen gleichermaßen kompetenzorientiert gestaltet sind und vergleichbare Anforderungen aufweisen. Prüfende sollten die Prüfungsinhalte und -kriterien wissenschaftlichen Prüfungen unterziehen und in Bezug auf die Teilprüfungen untersuchen, indem sie beispielsweise Trennschärfen einzelner Aufgaben (z. B. für die Geschlechter) berechnen, wie es in der Testtheorie für Leistungstests bei wissenschaftlichen Testverfahren vorgesehen wäre. Hierfür müssten entsprechend zeitliche Ressourcen zur Verfügung gestellt werden. Nach unserer Kenntnis ist das Prüfen jeder Person erlaubt, die einen Masterabschluss erworben hat und eine passende Stelle antritt. Von universitärer Seite werden zwar hochschuldidaktische Fortbildungsmöglichkeiten angeboten (Fendler & Gläser-Zikuda, 2013), diese sind jedoch nicht verpflichtend, sodass einheitliche Bewertungsexpertise und -kriterien in unserer Stichprobe und darüber hinaus nicht durchgängig zu vermuten sind, obwohl sich Noten aus einer Interaktion zwischen Prüfenden, gezeigten Leistungen und Prüfkriterien ergeben.

Die vielfältigen Inhalte im sportwissenschaftlichen Lehramtsstudium sprechen nicht dafür, dass diese in nur vier verschiedenen Prüfungsformaten adäquat examiniert werden können, wobei die schriftliche Klausur neben fachpraktischen Prüfungen mit Abstand das meistgenutzte Format darstellt und für fast die Hälfte der Prüfungen eingesetzt wird. Dass Männer in schriftlichen Klausuren schlechter abschneiden als Frauen, ist gut belegt (Reilly et al., 2019). Hier braucht es Prüfungsformate bzw. -kriterien, welche kein Geschlecht systematisch benachteiligen. Dass in der vorliegenden Studie in mündlichen Prüfungen keine und in Abschlussarbeiten nur sehr schwache Geschlechterunterschiede (zugunsten der Männer) gefunden wurden, spricht zumindest dafür, sowohl vermehrt wissenschaftliche Hausarbeiten als auch mündliche Prüfungen anstelle schriftlicher Klausuren einzusetzen. Hinderlich sind hier „too many long-held beliefs and standard operating practices" (Kuh et al., 2006, S. 106)

CC BY 4.0 DE

der Universitäten sowie zeitliche Ressourcen der Prüfenden und der Institute, welche die Wahl der Prüfungsformate eher beeinflussen als inhaltliche Überlegungen. Es ist zu vermuten, dass wissenschaftlich geprüfte Prüfungen mit deutlich höheren Kosten aufgrund des Mehraufwands für Prüfungen einhergehen würden. Dennoch sind hier die vom Wissenschaftsrat (2012) verlangten transparenten und kriterialen Bewertungsmaßstäbe zu verfolgen, um valide Dokumentations-, Selektions- und Rückmeldefunktion von Prüfungsnoten zu gewährleisten.

Kuśnierz et al. (2020) gehen einen Schritt weiter und schlagen sogar geschlechtsspezifische Studienordnungen für die Sportwissenschaft vor, welche neben Inhalten auch verschiedene Prüfungsformate abbilden könnten. Im Hinblick auf eine diversitätssensible und gleichstellungsorientierte Entwicklung der Studien- und Prüfungskultur ist dieser Vorschlag aber kritisch zu sehen, da er Geschlechterdifferenzen verstärken und Stereotypisierungen sogar betonen könnte. Aktuell werden eher individuelle, kompetenzorientierte und eigenverantwortliche Prüfungssituationen diskutiert, welche sich an wissenschaftlichen Erkenntnissen orientieren (Wissenschaftsrat, 2022). Die Hochschule hat daher in der Prüfungskonzeption sicherzustellen, dass das Prüfungsformat mit den Lehr-/Lernmethoden und -zielen, Kompetenzen und Lernergebnissen übereinstimmt (Hochschulrektorenkonferenz, 2015), damit eine gerechte Prüfung gewährleistet wird.

Limitationen und zukünftige Forschungen

Im Hinblick auf die vorgelegten Ergebnisse sind einige Limitationen zu beachten: Obwohl eine Vollerhebung aus zehn Jahren vorliegt, basieren die vorgelegten Daten ausschließlich auf Prüfungsleistungen an einer einzelnen Hochschule. Obwohl alle erbrachten Prüfungsleistungen einbezogen wurden, können leider keine Teilprüfungsnoten (z. B. in der Fachpraxis, die sich aus einer praktischen und theoretischen Note zusammensetzt) genutzt werden. Die untersuchten Prüfungsformate sind am untersuchten Standort durch Prüfungsordnungen vorgegeben und limitiert. Sie sollten an Standorten, an denen diese häufig eingesetzt werden, künftig weiter differenziert untersucht werden. Hinweise, dass Männer z. B. in Multiple-Choice Klausuren bessere Ergebnisse erzielen, existieren bereits (Espinosa & Gardeazabal, 2020). Die vorliegende Studie hat den Anspruch, vorsichtige Tendenzen darzulegen, die in weiteren Forschungsarbeiten an anderen Standorten untersucht werden sollten. Da die Studieninhalte jedoch grundsätzlich durch Empfehlungen von Institutionen wie der dvs oder der KMK gerahmt sind, könnten die Ergebnisse auch auf andere Standorte des sportwissenschaftlichen Lehramtsstudiums übertragbar sein, wenngleich durch die föderalen Strukturen keine bundesweiten Standards - auch im Hinblick auf Prüfungsformate - zu gewährleisten sind. Dies wird u. a. an der - zumindest laut den Prüfungsordnungen und Modulhandbüchern - eher starken Sportartenorientierung in der Fachpraxis an der untersuchten Hochschule deutlich.

Die schlechteren Resultate der Männer müssen nicht zuletzt vor dem Hintergrund der Studienwahlmotivation betrachtet werden. Frauen beginnen ihren Bildungsweg im Tertiärbereich, auch in dieser Studie, bereits oft mit einer deutlich besseren HZB. Diese allein gilt mit hohen Korrelationen (ca. r = .40) bereits als ein wichtiger und valider Prädiktor für Studienerfolg (Trapmann et al., 2007). Deshalb wurde für diesen Faktor in allen inferenzstatistischen Analysen kontrolliert und die hier berichteten Ergebnisse lassen sich für ‘gleiche Eingangsvoraussetzungen = gleiche HZB‘ interpretieren. Es sind für alle Geschlechter frühzeitige Unterstützungsmöglichkeiten im Studienverlauf auf struktureller und persönlicher Ebene wünschenswert (Blanz, 2014; Kuh et al., 2006). Zusätzlich sind weitere Untersuchungen der komplexen studentischen Eingangsvoraussetzungen, insbesondere im Hinblick auf Lernstrategien (Schiefele et al., 2003), nötig. Inwieweit allerdings gute Schülerinnen auch zu erfolgreicheren Sportwissenschaftlerinnen und Sportlehrerinnen werden, kann diese Studie nicht beantworten.

Einschränkend erwähnen möchten wir für die vorliegende Studie ebenfalls, dass die Kategorie Geschlecht als isoliertes dichotomes biologisches Geschlecht betrachtet wird, wie es in den Prüfungsämtern der Hochschule vorliegt. Da nicht-binäre Geschlechtsidentitäten nicht in den Archivdaten vorhanden waren, konnten diese nicht berücksichtigt werden. Weitere mögliche Faktoren oder Kategorisierungsvariablen wie z. B. die ethnische Herkunft oder ein Zuwanderungshintergrund waren explizit kein Thema dieser Arbeit und deren Untersuchung ist auf der vorliegenden Datenbasis auch nicht möglich. Auch die Berücksichtigung weiterer wichtiger Faktoren wie das soziale Umfeld der Studierenden und weitere aus der Literatur herleitbare Faktoren, welche die Notengebung beeinflussen, kann diese Archivstudie nicht leisten (Tsarouha, 2017; Grözinger, 2015). Solche Analysen müssen weiterhin als Forschungsdesiderat bezeichnet werden. Schließlich sind sport(-lehramts)spezifische Untersuchungen wünschenswert, um z. B. Studienzufriedenheit (Westermann et al., 1996) oder den Einfluss der Noten auf die Leistung im Beruf (Roth et al., 1996) aus einer Geschlechterperspektive zu beleuchten. Die Ergebnisse können Ansätze liefern, um den Studienerfolg aller Geschlechter zu optimieren.

LITERATUR

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

BERA. (2018). *Ethical guidelines for educational research* (4. Aufl.). British Educational Research Association.

Berlin-Institut für Bevölkerung und Entwicklung. (2015). *Schwach im Abschluss: Warum Jungen in der Bildung hinter Mädchen zurückfallen - und was dagegen zu tun wäre*. Eigenverlag.

Biggs, J. B. & Tang, C. S. (2011). *Teaching for quality learning at university: What the student does* (4th edition). McGraw-Hill Society for Research into Higher Education & Open University Press.

Blanz, M. (2014). How do study satisfaction and academic performance interrelate? An investigation with students of Social Work programs. *European Journal of Social Work*, *17*(2), 281–292. https://doi.org/10.1080/13691457.2013.784190

Breda, T. & Hillion, M. (2016). Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science*, *353*(6298), 474–478. https://doi.org/10.1126/science.aaf4372

Bülow-Schramm, M. (2013, 15. November). *Das Projekt USuS - Studienverläufe und Studienerfolg: Ergebnisse und Folgerungen [Vortrag]*. Campus Innovation, Hamburg.

Butler, D. M. & Christensen, R. (2003). Mixing and Matching: The Effect on Student Performance of Teaching Assistants of the Same Gender. *PS: Political Science & Politics*, *36*(4), 781–786. https://doi.org/10.1017/S1049096503003135

DOSB. (2021). *Bestandserhebung 2021*. Frankfurt am Main. Deutscher Olympischer Sportbund e. V.

Deutsche Vereinigung für Sportwissenschaft. (2020). *Online-Studienführer Sportwissenschaft der dvs.* Abruf am 10.05.2023 unter www.dvs-sportstudium.de

Deutsche Vereinigung für Sportwissenschaft, Fakultätentag Sportwissenschaft, Arbeitsgemeinschaft Sportpsychologie & Deutscher Sportlehrerverband. (2017). *Kerncurriculum: Ein-Fach-Bachelor Sportwissenschaft.*

Eliot, L. (2010). *Wie verschieden sind sie? Die Gehirnentwicklung bei Mädchen und Jungen*. Berlin Verlag.

Espinosa, M. P. & Gardeazabal, J. (2020). The Gender-bias Effect of Test Scoring and Framing: A Concern for Personnel Selection and College Admission. *The B.E. Journal of Economic Analysis & Policy*, *20*(3), Artikel 20190316. https://doi.org/10.1515/bejeap-2019-0316

Falchikov, N. & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research*, *59*(4), 395–430. https://doi.org/10.3102/00346543059004395

Falchikov, N. & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, *70*(3), 287–322. https://doi.org/10.3102/00346543070003287

Fendler, J. & Gläser-Zikuda, M. (2013). Lehrerfahrung und der „Shift from teaching to learning". *Zeitschrift für Hochschulentwicklung*, *8*(3), 15–28. https://doi.org/10.3217/zfhe-8-03/03

Fries, M. (2007). *Eignungsfeststellungsverfahren und Studienerfolg: Können Eignungskriterien den Studienerfolg prognostizieren?* Steinmeier.

Frohn, J. & Süßenbach, J. (2012). Geschlechtersensibler Schulsport: Den unterschiedlichen Bedürfnissen von Mädchen und Jungen im Sport mit Genderkompetenz begegnen. *Sportpädagogik*, *36*(6), 2–7.

Funder, D. C. & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gerick, J., Sommer, A. & Zimmermann, G. (2018). *Kompetent Prüfungen gestalten: 53 Prüfungsformate für die Hochschullehre*. Waxmann. https://doi.org/10.36198/9783838548401

Gerlach, E., Kussin, U., Brandl-Bredenbeck, H. P. & Brettschneider, W.-D. (2006). Der Sportunterricht aus Schülerperspektive. In Deutscher Sportbund (Hrsg.), *DSB-SPRINT-Studie: Eine Untersuchung des Schulsports in Deutschland* (S. 115-152). Meyer & Meyer.

Grözinger, G. (2015). Einflüsse auf die Notengebung an deutschen Hochschulen. Eine Analyse am Beispiel der Wirtschaftswissenschaften. *Die Hochschule: Journal für Wissenschaft und Bildung*, *24*(2), 94–114. https://doi.org/10.25656/01:16217

Grunschel, C. & Dresel, M. (2021). Studienerfolg und Studienabbruch: Gruppenspezifische Untersuchungen ihrer Bedingungen - Einführung in das Themenheft. *ZeHf – Zeitschrift für empirische Hochschulforschung*, *4*(1), 5–12. https://doi.org/10.3224/zehf.v4i1.02

Hadjar, A. Krolak-Schwerdt, S., Priem, K., & Glock, S. (2014). Gender and educational achievement. *Educational Research*, *56*(2), 117–125. https://doi.org/10.1080/00131881.2014.898908

Hannover, B. & Kessels, U. (2011). Sind Jungen die neuen Bildungsverlierer? Empirische Evidenz für Geschlechterdisparitäten zuungunsten von Jungen und Erklärungsansätze. *Zeitschrift für Pädagogische Psychologie, 25*(2), 89–103. https://doi.org/10.1024/1010-0652/a000039

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Educational research.* Routledge Taylor & Francis Group.

Heinze, D. (2018). *Die Bedeutung der Volition für den Studienerfolg*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-19403-1

Hernández, R. (2012). Does continuous assessment in higher education support student learning? *Higher Education*, *64*(4), 489–502. https://doi.org/10.1007/s10734-012-9506-7

Hochschulrektorenkonferenz. (2015). *Lernergebnisse praktisch formulieren* (Nexus - Impulse für die Praxis Nr. 2). Bonn.

Hottenrott, K. (2015). Die sportliche Frau: Leistungsphysiologische Unterschiede zum Mann - Relevanz für die ärztliche Beratung. *Gynäkologie*(3), 6–10.

Hoven, S. (2017). *Geschlechtergerechtigkeit im koedukativen Sportunterricht: Eine empirische Analyse zur Genderkompetenz von Sportlehrkräften in der gymnasialen Sekundarstufe I* [Dissertation]. Deutsche Sporthochschule Köln, Köln.

Hu, S. (2005). *Beyond Grade Inflation: Grading Problems in Higher Education* (6. Aufl.). *ASHE Higher Education Report: Bd. 30.*

Huntley, T., Whitehead, A., Cullinane, D., Nixon, S. & Huntley, E. (2017). Investigating relationships between and within entry pathways on a sport related programme and the degree outcome obtained. *Cogent Education*, *4*(1), 1–13. https://doi.org/10.1080/2331186X.2016.1276665

CC BY 4.0 DE

Jewell, R. T. & McPherson, M. A. (2012). Instructor-Specific Grade Inflation: Incentives, Gender, and Ethnicity. *Social Science Quarterly*, *93*(1), 95–109. https://doi.org/10.1111/j.1540-6237.2011.00827.x

Judd, C. M., Westfall, J. & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54–69. https://doi.org/10.1037/a0028347

Klinge, A. (2002). Was bildet eigentlich in der Sportlehrer(aus-)bildung? In P. Elflein, P. Gieß-Stüber, R. Laging & W.-D. Miethling (Hrsg.), *Qualitative Ansätze und Biographieforschung in der Bewegungs- und Sportpädagogik* (1. Aufl., S. 153-158). Afra-Verl.

KMK. (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung: Beschluss der Kultusministerkonferenz vom 16.10.2008 i. d. F. vom 16.05.2019.* Kulturministerkonferenz (KMK).

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K. & Hayek, J. C. (2006). *What Matters to Student Success: A Review of the Literature: Commissioned Report for the National Symposium on Postsecondary Student Success: Spearheading a Dialog on Student Success*. National Postsecondary Education Cooperative.

Kuśnierz, C., Rogowska, A. M. & Pavlova, I. (2020). Examining Gender Differences, Personality Traits, Academic Performance, and Motivation in Ukrainian and Polish Students of Physical Education: A Cross-Cultural Study. *International journal of environmental research and public health*, *17*(16), 1–21. https://doi.org/10.3390/ijerph17165729

Lenth, R. (2021). *Emmeans: Estimated Marginal Means, aka Least-Squares Means*. Abruf am 10.05.2023 unter https://CRAN.R-project.org/ package=emmeans

Lindner, M. A., Strobel, B., & Köller, O. (2015). Multiple-Choice-Prüfungen an Hochschulen? Ein Literaturüberblick und Plädoyer für mehr praxisorientierte Forschung. Zeitschrift für Pädagogische Psychologie, 29(3-4), 133–149. https://doi.org/10.1024/1010-0652/a000156

Milun, T., Mardesic, J. & Kovac, G. (2016). Do gender and age impact student success? *Contemporary issues in economy & technology*, *2*, 198–208.

MSB NRW. (2021). *Ordnung des Vorbereitungsdienstes und der Staatsprüfung für Lehrämter an Schulen: OVP.* Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. Abruf am 10.05.2023 unter https://recht.nrw.de/lmi/owa/br_text_anzeigen?v_id=10000000000000000681

Mutz, M. & Burrmann, U. (2014). Sind Mädchen im koedukativen Sportunterricht systematisch benachteiligt? Neue Befunde zu einer alten Debatte. *Sportwissenschaft*, *44*(3), 171–181. https://doi.org/10.1007/s12662-014-0328-x

Niessner, C., Utesch, T., Oriwol, D., Hanssen-Doose, A., Schmidt, S. C. E., Woll, A., Bös, K. & Worth, A. (2020). Representative Percentile Curves of Physical Fitness From Early Childhood to Early Adulthood: The MoMo Study. *Frontiers in Public Health, 8*, 458. https://doi.org/10.3389/fpubh.2020.00458

OECD. (2019). *PISA 2018 Results: Where all students can succeed (Volume II)*. OECD. https://doi.org/10.1787/b5fd1b8f-en

OECD. (2021). *Bildung auf einen Blick 2021: OECD-Indikatoren*. Paris. Bundesministerium für Bildung und Forschung.

Peters, D., Jones, G. & Peters, J. (2007). Approaches to Studying, Academic Achievement and Autonomy, in Higher Education Sports Students. *The Journal of Hospitality Leisure Sport and Tourism*, *6*(2), 16–28.

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Abruf am 10.05.2023 unter https://www.R-project.org/

Rauin, U. (2007). Im Studium wenig engagiert – im Beruf schnell überfordert: Studierverhalten und Karrieren im Lehrerberuf – Kann man Risiken schon im Studium prognostizieren? *Forschung Frankfurt*, *25*(3), 60–64.

Reilly, D., Neumann, D. L. & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, *74*(4), 445–458. https://doi.org/10.1037/amp0000356

Richardson, M., Abraham, C. & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353–387. https://doi.org/10.1037/a0026838

Riener, G. & Wagner, V. (2017). Shying away from demanding tasks? Experimental evidence on gender differences in answering multiple-choice questions. *Economics of Education Review, 59*, 43–62. https://doi.org/10.1016/j.econedurev.2017.06.005

Roth, P. L., BeVier, C. A., Switzer III, F. S. & Schippmann, J. S. (1996). Meta-Analyzing the Relationship Between Grades and Job Performance. *Journal of Applied Psychology*, *81*(5), 548–556. https://doi.org/10.1037/0021-9010.81.5.548

Röwert, R., Lah, W., Dahms, K., Berthold, C. & Stuckrad, T. von. (2017). *Diversität und Studienerfolg: Studienrelevante Heterogenitätsmerkmale an Universitäten und Fachhochschulen und ihr Einfluss auf den Studienerfolg - eine quantitative Untersuchung*. Gütersloh. CHE Centrum für Hochschulentwicklung.

Rüdiger, C., Jansen, M. & Rjosk, C. (2021). „Paul ist nicht so gut in Deutsch": Geschlechtsdifferenzielle Benotung im Fach Deutsch - eine Sekundäranalyse der Daten des IQB-Bildungstrends 2015. *Psychologie in Erziehung und Unterricht*, *68*, 1–17. https://doi.org/10.2378/peu2021.art08d

Şahin, E., Çekin, R. & Yazıcılar Özçelik, İ. (2018). Predictors of Academic Achievement among Physical Education and Sports Undergraduate Students. *Sports*, *6*(1). https://doi.org/10.3390/sports6010008

Sarcletti, A. (2021). *Studienerfolg und Studienabbruch: Ein Überblick über die Dimensionen des Studienerfolgs und die theoretischen Grundlagen*. TUDpress.

Schiefele, U., Streblow, L., Ermgassen, U. & Moschner, B. (2003). Lernmotivation und Lernstrategien als Bedingungen der Studienleistung: Ergebnisse einer Längsschnittstudie. *Zeitschrift für Pädagogische Psychologie*, *17*(3/4), 185–198. https://doi.org/10.1024//1010-0652.17.3.185

Schneider, M. & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Sheard, M. (2009). Hardiness commitment, gender, and age differentiate university academic performance. *The British journal of educational psychology*, *79*(1), 189–204. https://doi.org/10.1348/000709908X304406

Smith, F. (2004). "It's not all about grades": accounting for gendered degree results in geography at Brunel University. *Journal of Geography in Higher Education*, *28*(2), 167–178. https://doi.org/10.1080/0309826042000242422

Statistisches Bundesamt. (2021). *Prüfungen an Hochschulen 2020* (Bildung und Kultur Nr. 4.2). Wiesbaden. Statistisches Bundesamt (Destatis).

Stiller, T. & Kahlert, D. (2021). Die hochschulische Lehramtsausbildung im Fach Sport – kritische Reflexionen und Alternativen. *Zeitschrift für Studium und Lehre in der Sportwissenschaft*, *4*(1), 14–20. https://doi.org/10.25847/zsls.2020.032

Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, *45*(1), 89–125.

Towfigh, E., Traxler, C. & Glöckner, A. (2014). Zur Benotung in der Examensvorbereitung und im ersten Examen. *Zeitschrift für Didaktik der Rechtswissenschaft*, *1*(1), 8–27. https://doi.org/10.5771/2196-7261-2014-1-8

Towfigh, E., Traxler, C. & Glöckner, A. (2018). Geschlechts- und Herkunftseffekte bei der Benotung juristischer Staatsprüfungen. *Zeitschrift für Didaktik der Rechtswissenschaft*, *5*(2), 115–142. https://doi.org/10.5771/2196-7261-2018-2-115

Trapmann, S., Hell, B. & Weigand, S. (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

Trautwein, C. & Bosse, E. (2017). The first year in higher education—critical requirements from the student perspective. *Higher Education*, *73*(3), 371–387. https://doi.org/10.1007/s10734-016-0098-5

Troche, S. J., Anderman, C. P. & Rammsayer, T. H. (2010). Die Vorhersage des Studienerfolgs in sportwissenschaftlichen Studiengängen durch sportmotorische Eignungsprüfungen und Schulleistung. *Spectrum für Sportwissenschaften*, *22*(2), 64–79.

Tsarouha, E. (2017). Typologie der Einflussgrößen auf die Notengebung. In V. Müller-Benedict & G. Grözinger (Hrsg.), *Noten an Deutschlands Hochschulen* (S. 117-169). Springer Fachmedien Wiesbaden.

Utesch, T., & Magner, A. (2023). Online Supplement to Der Zusammenhang zwischen dem Geschlecht von Studierenden und Prüfenden sowie der Prüfungsformate auf den Studienerfolg im Lehramtsstudium Sport. Abruf am 10.05.2023 unter osf.io/jv324

Utesch, T., Zinner, J. & Büsch, D. (2018). Stabilität der physischen Fitness im Kindesalter. *German Journal of Exercise and Sport Research*, *48*(3), 404-414. https://doi.org/10.1007/s12662-018-0500-9

van den Berg, M. N. & Hofman, W. H. A. (2005). Student Success in University Education: A Multi-measurement Study of the Impact of Student and Faculty Factors on Study Progress. *Higher Education*, *50*(3), 413–446. https://doi.org/10.1007/s10734-004-6361-1

Voyer, D. & Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620

Westermann, R., Heise, E., Spies, K. & Trautwein, U. (1996). Identifikation und Erfassung von Komponenten der Studienzufriedenheit. *Psychologie in Erziehung und Unterricht*, *43*, 1–22.

Wiskin, C. M. D., Allan, T. F. & Skelton, J. R. (2004). Gender as a variable in the assessment of final year degree-level communication skills. *Medical education*, *38*(2), 129–137. https://doi.org/10.1111/j.1365-2923.2004.01746.x

Wissenschaftsrat. (2004). *Empfehlungen zur Reform des Hochschulzugangs*. Berlin.

Wissenschaftsrat. (2012). *Prüfungsnoten an Hochschulen im Prüfungsjahr 2010: Arbeitsbericht mit einem Wissenschaftspolitischen Kommentar des Wissenschaftsrates* (Drs. 2627-12). Hamburg.

Wissenschaftsrat. (2022). *Empfehlungen für eine zukunftsfähige Ausgestaltung von Studium und Lehre*. https://doi.org/10.57674/q1f4-g978

Woodfield, R., Earl-Novell, S. & Solomon, L. (2005). Gender and mode of assessment at university: should we assume female students are better suited to coursework and males to unseen examinations? *Assessment & Evaluation in Higher Education*, *30*(1), 35–50. https://doi.org/10.1080/0260293042003243887

Westfälische Wilhelms-Universität Münster. (2018). *Prüfungsordnung für das Fach Sport zur Rahmenordnung für die Bachelorprüfungen innerhalb des Zwei-Fach-Modells*. Münster. Westfälische Wilhelms-Universität Münster. Abruf am 10.05.2023 unter https://www.uni-muenster.de/imperia/md/content/wwu/ab_uni/ab2018/ausgabe11/beitrag02.pdf

York, T. T., Gibson, C. & Rankin, S. (2015). Defining and Measuring Academic Success. *Practical Assessment, Research, and Evaluation*, *20*(5), 1–20. https://doi.org/10.7275/hz5x-tx03

CC BY 4.0 DE