



Quantitative Prüfungsdatenanalysen als Beitrag zur Lehrentwicklung – Potenziale und Herausforderungen

FALK SCHEIDIG & MONIKA HOLMEIER

Zusammenfassung

Der Beitrag diskutiert die Potenziale der Analyse von Prüfungsdaten als Impuls zur individuellen und institutionellen Lehrentwicklung. Dies erfolgt am Beispiel einer quantitativen Studie, in deren Rahmen 37.800 Hochschulprüfungen u. a. hinsichtlich Notenniveau und -streuung, Studierendenmerkmalen sowie unterschiedlichen Prüfungsformen wie Hausarbeit und Klausur analysiert wurden. Im Beitrag wird anhand ausgewählter Befunde herausgearbeitet, wie Prüfungsdatenanalysen im Hochschulkontext für verschiedene Zielgruppen funktionsrelevantes Wissen bereitstellen können. Zugleich werden Herausforderungen und Limitationen der Auseinandersetzung mit Prüfungsdaten erläutert.

Schlüsselwörter: Prüfungen; Prüfungsdaten; Hochschulforschung; Learning Analytics; Lehrentwicklung

Quantitative Analyses of Assessment Data as a Contribution to Higher Education Development – Possibilities and Challenges

Abstract

The article discusses how analyses of assessment data can potentially contribute to individual and institutional development in higher education teaching. It presents a quantitative study in which 37,800 single assessments were analysed regarding the given grades, student characteristics, and different assessment types such as homework and exams. The article uses selected findings to show how analyses of assessment data can provide relevant knowledge for various target groups in higher education institutions. Nonetheless, the challenges and limitations of working with such data are also discussed.

Keywords: Exams; assessment data; higher education research; learning analytics; academic development

1 Zur Bedeutung und Erforschung von Hochschulprüfungen

Obwohl Prüfungen elementarer Bestandteil formaler Bildung sind, blieben sie lange „eine Leerstelle pädagogischer Reflexion“. (Ricken & Reh, 2017, S. 247) Mit Blick auf das Missverhältnis zwischen dem Stellenwert von Prüfungen in der pädagogischen Praxis einerseits und ihrer wissenschaftlichen Erschließung andererseits lässt sich für das Feld der Hochschuldidaktik konstatieren, dass Prüfun-

gen in den letzten beiden Dekaden stärker in den Fokus rückten. So wurden etwa Publikationen zur Konstruktion innovativer Prüfungsformen, zu prüfungsbezogenen Gütekriterien und zum kompetenzorientierten Prüfen vorgelegt (Gerick et al., 2017; Macke et al., 2016; Walzik, 2012). Besondere Geltung erlangte dabei mit dem Constructive Alignment eine Perspektive, die dem Verhältnis von Prüfungen, Lehr-/Lernarrangements und Lernzielen besondere Aufmerksamkeit schenkt (Biggs & Tang, 2011). Diese Überlegungen diffundierten in viele Disziplinen, die – wie etwa im Falle der Medizin (Epstein, 2007; Wass et al., 2001) – mitunter eigene Diskurse zu Prüfungsformen zur Beurteilung fach- respektive professionsspezifischer Kompetenzen pflegen. Daneben gewannen Prüfungen auch hochschulpolitisch an Relevanz, etwa mit Blick auf die überdurchschnittlichen Studienabbruchquoten in MINT-Studiengängen (Heublein et al., 2022; Klöpping et al., 2017), auf die Bedeutung von Noten für den Übertritt von der Bachelor- in die Masterphase nach der Bologna-Reform (Binder, 2019), auf geschlechterbezogene und migrationsbegründete Notenunterschiede (OECD, 2019, S. 243–266; Towfigh et al., 2014, 2018) oder auf die international beklagte „grade inflation“ (Gunn & Kapade, 2018; Mansfield, 2001).

Prüfungen stellen als Kernbestand von Studium und Lehre ein genuin hochschuldidaktisches Themen- und Betätigungsfeld dar (so bereits BAK, 1970/2009). Sie gewinnen auch deshalb an Relevanz für die Reflexion über Hochschullehre, weil ihnen ein bedeutsamer Einfluss auf studentisches Lernen zugeschrieben wird („assessment drives learning“) und sie den Studienalltag mitprägen. Dabei ist das Spektrum an ihnen zugeordneten Funktionen breit (siehe auch Reinmann, 2019, S. 619): Prüfungen sollen u. a. (a) Lernanreize stiften, (b) Studierenden formatives Feedback für ihr weiteres Lernen geben, (c) Lernprozesse und Zielerreichung summativ beurteilen, (d) Lernen und Leistungen u. a. in Relation zu Mitstudierenden einordnen, (e) eine Grundlage für Selektion im Studium und auf dem Arbeitsmarkt schaffen, (f) Auskunft über die Eignung für Studium und Beruf erteilen und Prognosen über künftige Leistungen erlauben, (g) akademische und arbeitsweltliche Anforderungen simulieren und den Rahmen für das Testen von Kompetenzen spannen, (h) Lehrenden und Hochschulen Rückmeldungen zu Qualität, Nutzung und Anspruchsniveau von Lernangeboten zukommen lassen.

Die Bedeutung von (qualitätsvollen) Hochschulprüfungen dürfte in der Hochschuldidaktik unbestritten sein: „Students can, with difficulty, escape from the effects of poor teaching, they cannot (by definition, if they want to graduate) escape the effects of poor assessment.“ (Boud, 1995, S. 35) Ihr Stellenwert lässt sich u. a. daran ablesen, dass die Gestaltung von Prüfungen häufig obligatorisches Element hochschuldidaktischer Weiterbildungsprogramme ist (Scheidig & Klingovsky, 2020). In der Forschung zu Hochschulprüfungen spiegelt sich dies nur partiell wider: Empirische Zugänge zu Hochschulprüfungen beziehen sich vorrangig auf Analysen zu einzelnen Prüfungsverfahren wie Multiple-Choice-Prüfungen (z. B. Ledermüller et al., 2015), auf Aufgabenkulturen, insbesondere in domänenspezifischer Ausprägung (z. B. Figas & Hagel, 2017), auf Gestaltungsprinzipien von Prüfungen und Prüfungseffekte (Bücker et al., 2015) oder auf die vergleichende Analyse von Abschlussnoten (Müller-Benedict & Grözing, 2017; Statistisches Bundesamt, 2016; Wissenschaftsrat, 2012).

Mit dem international emergierenden Feld Learning Analytics gewinnt die Verarbeitung von Prüfungsdaten an Bedeutung, z. B. für Studienerfolgsprognosen (Scheidig, 2021b; Scheidig & Holmeier, 2023). Unter Assessment Analytics (als Teilbereich von Learning Analytics) firmieren seit wenigen Jahren Analysen auf der Ebene einzelner Prüfungen, wobei der Fokus hier weniger auf dem Prüfungsergebnis als vielmehr auf dessen Zustandekommen und Detailstruktur liegt (z. B. interpersoneller Vergleich von Prüfungsverhalten, Erkennen von unredlichem Verhalten; für eine Übersicht: Scheidig & Schweinberger, 2022). Durch die zunehmende Verbreitung digitaler Prüfungen – insbesondere im Kontext der Corona-Pandemie (Bandtel et al., 2021; Persike, 2021) – sind die Möglichkeiten für Prüfungsanalysen dieser Art gestiegen. Von dieser Mikroperspektive abzugrenzen sind studiengangbezogene und -übergreifende Prüfungsdatenanalysen, auf die sich dieser Beitrag konzentriert.

Auf Basis ausgewählter Ergebnisse einer Analyse von 37.800 Einzelprüfungen wird im Weiteren erörtert, welchen Beitrag die aus Prüfungsanalysen gewonnenen Erkenntnisse zur individuellen

und institutionellen Lehrentwicklung zu leisten vermögen. Dabei liegt der Fokus weniger auf den Ergebnissen selbst, sondern auf der Diskussion ihrer hochschuldidaktischen Nutzbarmachung und damit verbundener Limitationen.

2 Methodik

Bei der nachfolgend vorgestellten Prüfungsdatenanalyse einer Schweizer Hochschule mit ca. 3.300 Studierenden primär in Lehramtsstudiengängen waren folgende Fragen leitend: Welche Noten erlangen die Studierenden und wie verteilen sich die Noten in verschiedenen Studiengängen, Studienbereichen und Modulen, bei verschiedenen Formen von Leistungsnachweisen und Abschlussarbeiten sowie in Abhängigkeit von Alter und Geschlecht der Studierenden? Die Auswertungen bezogen sich dabei u. a. auf Unterschiede im semesterübergreifenden Zeitverlauf (Quasi-Längsschnitt), auf Unterschiede zwischen Studienelementen (Lehrveranstaltung, Qualifikationsarbeit, studienintegriertes Praktikum usw.), auf die Häufigkeit einzelner Leistungsnachweisformen sowie auf Bestehensquoten, Notenniveau und die Notenstreuung.

Zu diesem Zweck wurden von der Studienadministration die Prüfungsdaten vom Frühjahrssemester 2013 bis zum Herbstsemester 2016 bereitgestellt. In diesen acht Semestern wurden insgesamt 37.800 Prüfungen in sechs Studiengängen absolviert. Die Datensätze enthielten die Veranstaltungsnummer und Bezeichnung der einzelnen Lehrveranstaltungen, in denen ein Leistungsnachweis absolviert wurde, sowie die Leistungsnachweisnote, das Geschlecht und das Geburtsdatum der Studierenden der jeweiligen Lehrveranstaltung. Noten von Abschlussarbeiten waren ebenfalls Bestandteil des Datensatzes.

Der Datensatz wurde in SPSS aufbereitet und ausgewertet, wobei zunächst aus den Veranstaltungsnummern die wesentlichen Informationen extrahiert und darauf aufbauend neue Variablen zum Semester, zum Studiengang, zum Studienbereich sowie zur Modulgruppe generiert wurden. Zudem ließ sich aus der Veranstaltungsbezeichnung u. a. ableiten, ob es sich um eine reguläre Lehrveranstaltung oder um eine Wiederholungsprüfung handelt. Die neu generierten Informationen wurden den Lehrveranstaltungen als neue Variablen zugeordnet. Aus dem Geburtsdatum der Studierenden wurde deren Alter zum letztmöglichen Zeitpunkt der Notenvergabe berechnet (im Frühjahrssemester: jeweils 30. Juni; im Herbstsemester: jeweils 31. Januar). Da den Datensätzen nicht zu entnehmen war, zu welchem Datum der Leistungsnachweis absolviert wurde, handelt es sich beim Alter nicht zwangsläufig um das exakte Alter; die Abweichungen betragen aber maximal ein Lebensjahr.

Neben den Variablen, die sich aus den bereits vorhandenen Daten ableiten ließen, wurden zudem Informationen zu den Verfahren der Leistungsnachweise gesammelt. Dazu wurden die einzelnen Lehrveranstaltungen in den elektronischen Verzeichnissen der entsprechenden Semester gesucht, die Informationen zum Leistungsnachweis extrahiert und zusammengefasst. Eine erste Zusammenstellung ergab dabei 98 verschiedene Verfahren von Leistungsnachweisen. Diese Verfahren wurden im Wege induktiver Kategorienbildung geclustert und Haupt- sowie Subkategorien zugeordnet (siehe Tabelle 1, ähnlich auch Reinmann, 2019, S. 615 f.).

Tabelle 1: Kategorisierung der Leistungsnachweisverfahren

Kategorien von Leistungsnachweisen	Subkategorien
Klausur	keine
schriftliche Ausarbeitung	Haus-/Seminararbeit, Lerntagebuch/Portfolio, Unterrichtseinheit, Forschungsplan
mündlicher Leistungsnachweis	mündliche Prüfung, Kolloquium, Referat
praktischer Leistungsnachweis	keine
kombinierte Leistungsnachweisverfahren	u. a. schriftliche Ausarbeitung und mündlicher Leistungsnachweis, schriftliche Ausarbeitung und praktischer Leistungsnachweis, schriftliche Ausarbeitung und Klausur, mündlicher und praktischer Leistungsnachweis

Bei 17.3 % der im Datensatz enthaltenen Leistungsnachweise konnte keine Zuordnung zu einem der in Tabelle 1 genannten Leistungsnachweisverfahren vorgenommen werden – etwa, weil die Veranstaltung nicht im Veranstaltungsverzeichnis aufzufinden war, weil keine Angaben zum Leistungsnachweis gemacht wurden oder weil mehrere Leistungsnachweisverfahren genannt wurden, sodass eine eindeutige Rubrizierung des Verfahrens nicht möglich war. Bei der Kodierung wurde nach der Zuordnung der verschiedenen Leistungsnachweisverfahren zudem zusätzlich eine Unterscheidung zwischen Team- und Einzelprüfungen (z. B. Referat allein versus in einer Gruppe) sowie zwischen testbasierten und produkt-/projektbasierten Leistungsnachweisen (z. B. Multiple-Choice-Klausur versus Hausarbeit) vorgenommen. Hintergrund war die Hypothese, dass die Sozial- und Darbietungsformen Einfluss auf die Prüfungsleistung und deren Bewertung nehmen.

Nachfolgend werden nur signifikante Ergebnisse ($p < .05$) berichtet. Zur Bestimmung der Signifikanz von Unterschieden wurden in SPSS t-Tests für unabhängige Stichproben sowie einfaktorielle Varianzanalysen gerechnet. Auch wenn es sich bei Noten nicht um eine metrisch skalierte Variable handelt, wurden dennoch parametrische Verfahren angewandt, um die Darstellung und Interpretation der Ergebnisse zu erleichtern (Sedlmeier & Renkewitz, 2013, S. 63).

3 Ausgewählte Ergebnisse

Der zuvor beschriebene Datensatz ermöglichte zahlreiche Auswertungen, um hochschulisch relevante Fragestellungen¹ zu beantworten; im Weiteren werden jedoch lediglich ausgewählte Befunde mit Blick auf verschiedene Leistungsnachweisverfahren vorgestellt.

Es zeigen sich deutliche Unterschiede zwischen den Prüfungsverfahren hinsichtlich der Notenmittelwerte (siehe Tabelle 2). Bei Referaten werden im Vergleich zu anderen Leistungsnachweisverfahren weit überdurchschnittliche Notenmittelwerte erzielt. Demgegenüber bringen textbasierte Prüfungen (Klausur, Haus-/Seminararbeit) und, wenngleich weniger deutlich, mündliche Prüfungen vergleichsweise geringe Notenmittelwerte hervor, aber – gemessen an der größeren Streuung der Noten – ein hohes Potenzial zur Differenzierung zwischen den Studierenden. Eine mögliche Erklärung wäre, dass Studierende bei Referaten die Leistungserbringung besser vorbereiten und steuern können als etwa bei (eher standardisierten) Klausuren oder mündlichen Tests sowie dass sich Referate aufgrund der individuellen Gestaltung und des flüchtigen Darbietungscharakters eher einer kriterialen Beurteilung entziehen. Zudem zentrieren sich Referate meist auf ein vergleichs-

¹ Z. B.: Welche Leistungsnachweisverfahren kommen wann und wie häufig im Studium vor? Unterscheiden sich die Studiengänge hinsichtlich der studentischen Leistungen respektive ihrer Benotung? Was sind – gemessen an der Durchfallquote und am Notendurchschnitt – die anspruchsvollsten Lehrveranstaltungen eines Studiengangs? Lassen sich deutliche Effekte bezüglich eines Zusammenhangs von Noten und Geschlecht oder Alter in Studiengängen oder einzelnen Studienbereichen nachweisen? Wie hoch ist insgesamt das Notenniveau und lässt sich im Zeitverlauf eine Zunahme guter Noten („grade inflation“) konstatieren? Unterscheiden sich die Noten zwischen Studienmodellen oder -phasen, z. B. Bachelor- versus Masterstudiengänge?

weise eng abgegrenztes Thema und nicht, wie bei anderen Prüfungsformen, auf mehrere Themen. Somit können sich die Studierenden sehr viel fokussierter vorbereiten und womöglich ihre persönlichen Interessen bei der Themenwahl oder -entfaltung zur Geltung bringen.

Tabelle 2: Notendurchschnitte ausgewählter Leistungsnachweisverfahren (Schweizer Notenskala: 6 = ausgezeichnet, 4 = genügend, 3.5 bis 1 = nicht bestanden)

Leistungsnachweis	N	M	SD
Referat	615	5.44	0.56
mündliche Prüfung	6276	5.03	0.79
Haus-/Seminararbeit	6969	4.92	1.03
Klausur	9125	4.89	0.82

Eine Erklärung für die vergleichsweise niedrigen Notenmittelwerte bei textbasierten Prüfungen liefert möglicherweise der Umfang des Kontakts zwischen Studierenden und Prüfenden. Sowohl bei den schriftlichen Ausarbeitungen als auch bei den Klausuren stehen die Akteure – von Ausnahmen abgesehen – in der Prüfungssituation in einem eher unpersönlichen Verhältnis. Zu vermuten wäre, dass diese Distanz eine mögliche Zurückhaltung in der Vergabe unterdurchschnittlicher Noten abbaut, die sich bei persönlichem Kontakt z. B. aufgrund von Rollenkonflikten, Konfliktvermeidungsstrategien oder Subjektivitätseinflüssen ergeben kann. Gleichzeitig gewährt die Distanz weniger Möglichkeiten der leistungsbezogenen Verständigung und Einordnung, z. B. durch direktes Nachfragen, durch den Dialog über den Prüfungsinhalt oder durch das Ermitteln individueller Stärken und Schwächen in Bezug auf den Prüfungsinhalt. Eine weitere Erklärung könnte sein, dass dort, wo die Bewertung – im Gegensatz zu mündlichen und praktischen Verfahren – vollständig nachvollzogen und gegebenenfalls wiederholt werden kann (da die Prüfungsleistung in schriftlicher Form vorliegt), die Notenmittelwerte deswegen unterdurchschnittlich sind, weil den Prüfenden bei diesen Verfahren mehr Zeit für die Notengebung zur Verfügung steht und die zu bewertende Leistung differenzierter erfasst werden kann. Zudem kann die Vergabe schlechter Noten auf Basis vollständig dokumentierter Prüfungsergebnisse besser nachvollziehbar begründet werden als bei Verfahren, in denen die Notenvergabe und -kommunikation unmittelbar nach der Leistungserbringung erfolgt (mit Implikationen etwa für die Rechtssicherheit).

Der strukturelle Unterschied zwischen Tests einerseits, bei denen Studierende zu einem definierten Zeitpunkt unter Anwesenheit der prüfenden Person(en) eine Leistung erbringen und die Vorarbeit nicht Gegenstand der Bewertung ist (z. B. Klausur, mündliche Prüfung, praktische Prüfung), und produkt-/projektorientierten Prüfungen andererseits, bei denen Studierende über einen längeren Zeitraum selbststrukturiert ein Produkt erstellen oder ein Projekt bearbeiten (z. B. Haus-/Seminararbeit, Gestaltung eines Objekts), schlägt sich nicht mit einheitlicher Tendenz in den Noten nieder. Weder lassen sich durchweg signifikante Notenunterschiede nachweisen, noch sind die wenigen gefundenen Unterschiede in ihrer Richtung gleich. Die Vermutung bestätigt sich nicht, dass produkt-/projektorientierte Prüfungen aufgrund des relativen Freiheitsgrades der Erstellung u. a. in Bezug auf Zeitinvestition, Themenwahl/-vertiefung, Darbietungsform sowie eigenverantwortliche Arbeitsweise bessere Noten befördern im Vergleich zu fremdgesteuerten Tests (bei denen Studierende die normierten Anforderungen nur eingeschränkt antizipieren können und die Gefahr besteht, beim Lernen unbeabsichtigt Schwerpunkte auf weniger prüfungsrelevante Inhalte zu legen). Gründe hierfür könnten in der stärkeren Gewichtung der sozialen Bezugsnorm bei produkt-/projektorientierten Prüfungen liegen aufgrund eines womöglich weniger engen kriterialen Erwartungsbildes dieser offeneren Leistungsnachweise, der höheren Anspruchshaltung der Prüfenden oder auch der Einflussnahme primär nicht-inhaltlicher Kompetenzbereiche (Selbstorganisation, Ausdruck und Rechtschreibung usw.).

Einschränkend ist anzumerken, dass die Rangfolge der Leistungsnachweisverfahren gemessen am Notenmittelwert zwischen den sechs betrachteten Studiengängen nicht einheitlich ist. Es variiert je nach Studiengang, bei welchem Verfahren die höchsten bzw. die niedrigsten Notenmittelwerte nachweisbar sind. Dies könnte zum einen in den jeweiligen Studierendengruppen begründet sein, denen verschiedene Leistungsnachweisverfahren unterschiedlich entgegenkommen (z. B. weil Studierende in Masterstudiengängen im Gegensatz zu Bachelorstudierenden aufgrund der längeren akademischen Sozialisation mit einigen hochschulischen Leistungsnachweisverfahren vertrauter sind). Die ungleiche Reihenfolge kann zum anderen auch damit erklärt werden, dass die Ausgestaltung der Verfahren in den jeweiligen Studiengängen unterschiedlich erfolgt. So können die Prüfungen zu unterschiedlichen Zeitpunkten im Semester und Curriculum, in unterschiedlichem Umfang oder in unterschiedlicher Häufigkeit stattfinden, Letzteres verbunden mit prüfungsbezogenen Übungseffekten.

Neben den Ergebnissen, die sich auf die verschiedenen Leistungsnachweisverfahren beziehen, lieferte die Prüfungsdatenanalyse weitere Befunde, die eine Berücksichtigung verdienen: Der Notenmittelwert von Teamprüfungen ($M = 5.18$, $SD = 0.63$) liegt höher als bei Einzelprüfungen ($M = 5.00$, $SD = 0.78$). Der Notendurchschnitt von Bachelorarbeiten ($M = 5.24$, $SD = 0.60$) und Masterarbeiten ($M = 5.18$, $SD = 0.88$) rangiert jeweils über dem Notendurchschnitt des verwandten Leistungsnachweisverfahrens Haus-/Seminararbeit ($M = 4.92$, $SD = 1.03$). Der Notendurchschnitt von Wiederholungsprüfungen ($M = 4.43$, $SD = 1.02$) liegt unter dem Notendurchschnitt von Erstprüfungen ($M = 5.04$, $SD = 0.81$); d. h., Studierende steigern sich zwar tendenziell in Zweitprüfungen, erreichen aber durchschnittlich auch im zweiten Versuch nicht das Niveau der Gesamtgruppe. Dieses Muster durchzieht sich durch alle Auswertungskategorien (Semester, Studiengang, Prüfungsform).

4 Potenziale zur Lehrentwicklung

Die zuvor skizzierte Prüfungsdatenanalyse kann als Impulsgeber für die Lehrentwicklung fungieren, weil sie relevante (und zuweilen brisante) Erkenntnisse für verschiedene Akteure im hochschulischen Mehrebenensystem generiert, vom Rektorat und hochschulischen Gremien über dezentrale Leitungsstrukturen (Fakultät, Department, Institut) und die Studiengangsleitung und -koordination bis hin zu den Lehrenden und Studierenden, des Weiteren auch für hochschuldidaktische Arbeitsstellen, Stabseinheiten wie die Studienberatung und das Qualitätsmanagement sowie – insbesondere mit Blick auf Selektivität und Notenniveau – für außerhochschulische Stakeholder in Politik und Praxis. Der Hochschul- und hochschuldidaktischen Forschung, die sich bislang vorrangig auf einzelne Prüfungsverfahren oder Abschlussnoten konzentriert, stellen derlei Auswertungen ebenfalls neue Erkenntnisse in Aussicht. Daneben besitzen sie im Hinblick auf fachspezifische Lehr-, Lern- und Prüfungskulturen und domänenspezifische Wissenserwerbsprozesse durchaus Relevanz für intradisziplinäre Verständigungsprozesse über das Studium im eigenen Fach. Für Studierende und Lehrende liefern Prüfungsdatenanalysen u. a. Referenzwerte, um die eigene Leistung respektive Prüfung im Gesamtensemble des Studiengangs einzuordnen. Sie vermag so auch Vermutungen und Behauptungen, etwa hinsichtlich vermeintlich anforderungsarmer Module oder der mitunter berüchtigten Selektionswirkung einzelner Prüfungen, zu verifizieren oder falsifizieren. Ebenso besteht das Potenzial, Probleme überhaupt zu identifizieren, die heterogenitäts- und diskriminierungssensitive Aufmerksamkeit einfordern, z. B. starke Notenunterschiede in Abhängigkeit von Geschlecht, Vorbildung, sozialer und kultureller Herkunft.

Die genannten Beispiele zeigen, dass Resultate von Prüfungsdatenanalysen Anlass für vertiefende quantitative und qualitative Analysen geben, etwa zur Erhellung der zuvor referierten Effekte einzelner Leistungsnachweisverfahren oder für prädiktive Analysen zur Reduktion von Prüfungsmisserfolgen und Studienabbrüchen. Prüfungsdatenanalysen können normative Klärungen anstoßen, beispielsweise zur Anwendung oder Ablehnung von Bezugsnormen, zur Angemessenheit von

Leistungsanforderungen und zur anvisierten Selektivität, oder auch konkrete Beschlüsse fundieren, z. B. die Bevorzugung bestimmter Prüfungsformate oder die Formulierung von Prüfungsstandards.

Die quantitative Auswertung von Prüfungsdaten liefert für Hochschulen neben steuerungsrelevantem Wissen insbesondere auch hochschuldidaktische Impulse: Einzelne Resultate, z. B. überdurchschnittliche Notenmittelwerte bei Referaten oder Teamprüfungen, können die prüfungsbezogene Reflexion, Kommunikation und Entwicklungsarbeit anregen: Warum werden in einem bestimmten Setting deutlich bessere Noten erlangt und entsprechen die Notendurchschnitte tatsächlich dem subjektiven Leistungseindruck der Dozierenden? Der datengestützte Austausch über Prüfungen fördert die Entwicklung von „assessment literacy“ (Wollersheim & Pengel, 2016) von Dozierenden und die bewusste(re) Wahl und Ausgestaltung von Leistungsnachweisverfahren sowie die Formulierung von Beurteilungskriterien. Ein solcher Austausch kann einerseits aktiv durch explizite Thematisierung von ausgewählten oder bereichsspezifisch vertieften Ergebnissen einer Prüfungsdatenanalyse im Rahmen von Gremiensitzungen, bei Teamsitzungen von Professuren oder in Austauschformaten unter und zwischen Lehrenden und Studierenden (Studiengangskonferenz, Fachschaftstreffen usw.) erfolgen. Mindestens ebenso bedeutsam und wirkungsvoll erscheint ein Referenzieren in Kontexten, in denen Prüfungen oder damit verbundene Aspekte (z. B. Dropout, Prüfungsangst, Studierbarkeit) Gegenstand sind; hierzu zählen u. a.

- Prozesse der Prüfungskonzeption, -planung, -durchführung, -auswertung, -überarbeitung,
- Prozesse der Betreuung und Beratung von Studierenden allgemein sowie im Kontext konkreter Prüfungsvorbereitung, -nachbesprechung und Lernberatung,
- die Modul- und Studiengangsentwicklung,
- Projekte zur Weiterentwicklung des Prüfungswesens, z. B. die Implementation von E-Assessments,
- das Studiengangsmonitoring und -reporting, Evaluationen und das Qualitätsmanagement,
- hochschuldidaktische Weiterbildungs- und Beratungsangebote, in denen Prüfungskonzeption, -durchführung, -beurteilung und Studierendenbetreuung eine Zuwendung erfahren,
- die interne und externe Kommunikation gegenüber Studierenden, Studieninteressierten, der Bildungspolitik und -verwaltung sowie Akteuren in korrespondierenden Praxis-/Arbeitsfeldern mit Blick auf Anforderungen und Leistungen im Studium.

Jeder Kontext verlangt dabei eine je spezifische Form der Thematisierung von Prüfungsdatenauswertungen für eine adressaten- und kontextadäquate Ansprache, Verständigung und Aufmerksamkeitslenkung, so z. B. die Komprimierung vielschichtiger Ergebnisse zu einzelnen Thesen, Leitfragen oder Kennzahlen. Dies erweist sich angesichts der nachfolgend beschriebenen Interpretationsbedürftigkeit von Prüfungsdaten als keineswegs trivial.

5 Herausforderungen

Wesentlich für die Arbeit mit Prüfungsdaten sind gemeinhin die Ergebnisse der analysierten Prüfungen, entweder binär mit pass/fail kodiert oder als Wertzuweisung im Notensystem, die den Ausgangspunkt vieler (obschon nicht aller) Auswertungen bilden. Idealtypisch basieren Noten auf einer objektiven, reliablen und validen Prüfung und beschreiben einzig die Leistung der Studierenden in einem bestimmten Leistungsnachweisverfahren; leistungsferne Merkmale sollten demgemäß keinen Einfluss nehmen. Inwieweit in einem Datensatz vorhandene Noten diesem Anspruch gerecht werden, kann ohne standardisiertes Leistungsmaß nicht geprüft werden. Die Vielzahl möglicher Einflüsse auf das Zustandekommen von Noten und Notendifferenz verlangt jedoch generell eine kritische Würdigung von Prüfungsdaten: Einwirkende Faktoren sind u. a. die Prüfungsform und -gestaltung (digital versus analog, Umfang, Dauer, Schwierigkeitsgrad, erlaubte Hilfsmittel, Passung zu Inhalt und didaktischem Design der zugehörigen Lehrveranstaltung, Zeitpunkt im Semester und Studium), die Beurteilung der Prüfungsleistungen (Bezugsnorm, Bewertungskriterien, Gütekrite-

rien, Verzerrungseffekte), studierendenspezifische Faktoren (Prüfungsvorbereitung, Vorbildung, Leistungsvermögen, Motivation, Angst) sowie situative Faktoren (Uhrzeit, Wochentag, räumliche Umgebung, aktuelle Ereignisse, Zwischenfälle). Ebenso spiegeln sich in Noten die Lehr- und Betreuungsleistungen von Dozierenden sowie Bedingungen des Lehrens und Lernens innerhalb der Hochschule und Studiengänge.

Eine weitere – je nach Studiengang unterschiedlich ausgeprägte – Problematik erwächst aus der zeitlichen Dimensionierung der Curricula: Der Zeitverlauf des Studiums führt zu lehrveranstaltungsübergreifenden Kompetenzentwicklungs- und Wissenserwerbsprozessen, prüfungsbezogenen Übungseffekten, variierenden Ansprüchen zwischen den Lehrveranstaltungen und Studienphasen und schließlich zu wechselnden Gruppenzusammensetzungen in Lehrveranstaltungen und Prüfungen (mit wahrscheinlichen Referenzgruppeneffekten), die das Zustandekommen studentischer Leistungen und die Notengebung beeinflussen, aber in der quer- und längsschnittlichen Betrachtung nur schwer abbildbar sind. Eine numerisch isolierte, diese Aspekte nicht beachtende Analyse von Noten geht mit Informationsdefiziten einher und befördert die Notwendigkeit weiterführender Analysen, interpretativer Zugänge und diskursiver Erschließung von Befunden – nah bei den für die betreffenden Prüfungen zuständigen Organisationseinheiten und Personen und gegebenenfalls unter Einbezug studentischer Perspektiven auf ebenjene Prüfungen.

Das Erfordernis und die grundlegende Herausforderung der Interpretation von Befunden einer Prüfungsdatenanalyse soll nachfolgend an zwei Beispielen illustriert werden, bei denen sich jeweils konträre Deutungsmuster ergeben:

Beispiel I: Der zuvor referierte Befund, dass bei Teamprüfungen höhere Notendurchschnitte als in Einzelprüfungen erzielt werden, kann (A) entweder das intendierte Resultat der didaktischen Anlage sein, dass sich Studierende gegenseitig unterstützen, motivieren, ergänzen, leistungsschwächere Studierende von leistungsstärkeren Studierenden Erklärungen und Hilfestellungen erhalten und leistungsstärkere Studierende ihr Wissen dadurch vertiefen und festigen, indem sie für leistungsschwächere Studierende den Lernstoff aufbereiten, oder (B) als Beleg für nicht sanktioniertes „soziales Faulenzen“ in Teams gelten. Darüber hinaus kann unterstellt werden, dass Teamprüfungen nicht zu differenzierender Notengebung tendieren, weil beispielsweise Leistungsunterschiede schwerer erkennbar sind oder weil leistungsschwächere Studierende in leistungsheterogenen Gruppen bevorteilt werden (bei schwachen Gruppenleistungen schlechte Noten auch jenen Studierenden mit hohem Leistungsvermögen zu geben, wird vermutlich als schwerwiegender angesehen, als eher leistungsschwachen Studierenden in Teams mit guten Studierenden eine gute Note zu geben). Zugleich könnte auch unterstellt werden, dass bei Teamprüfungen das Effizienzkriterium (eine große Prüfungslast bewältigen) zulasten des Bemühens um differenzierte Beurteilung einzelner studentischer Leistungen geht.

Beispiel II: Dass bei Bachelor- und Masterarbeiten (zu den Rahmenbedingungen siehe Scheidig, 2021a) der Notendurchschnitt über dem allgemeinen Notenniveau sowie deutlich über dem Notenniveau des verwandten Leistungsnachweisverfahrens Haus-/Seminararbeit rangiert, kann (A) als Resultat besonders intensiver und individueller Betreuung (im Vergleich zu lehrveranstaltungsgebundenen Leistungsnachweisen), des Leistungszuwachses im Studium (aus dem am Ende geschöpft werden kann), eines Übungseffekts (durch das vorgelagerte Verfassen mehrerer Haus- und Seminararbeiten) oder des hervorgehobenen Gewichts dieser Leistungsnachweisform (hohe ECTS-Punktzahl, langer Zeitraum der Erstellung, zeit- und formatbedingt singuläre Bearbeitungstiefe, summative Leistungsdemonstration am Ende des Studiums, persönlicher und akademischer symbolischer Status, Ausweis auf dem Abschlusszeugnis) interpretiert werden. Es kann aber auch (B) als Indiz dienen für Rollenkonflikte der Betreuungspersonen (infolge der oftmals intensiven, vertrauensbasierten Begleitung und der dozierendenseitigen Mitverantwortung bei Themenwahl, -konkretisierung und -bearbeitung) oder für einen Milde-Effekt sprechen (wegen der besonderen Bedeutung der Note u. a. für den Eintritt in ein Masterstudium, den Arbeitsmarkt oder die Zulassung zur Promotion und insbesondere der Folgeschwere eines Nicht-Bestehens).

Die beiden Beispiele legen zum einen offen, dass und warum direkte Ableitungen aus Prüfungsdatenanalysen erschwert werden und Bedarf daran besteht, weiterführende Informationen zum Zustandekommen der konkret betrachteten Noten (auf der Ebene der Notenvergabe) zu gewinnen und in die Interpretation einzubeziehen: Begründen sich beispielsweise bessere Noten bei Teamprüfungen tatsächlich im Effekt gegenseitiger studentischer Unterstützung, sind diese aufgrund ihrer leistungssteigernden Wirkung wohl zu fördern, findet hier aber eine nicht leistungsbegründete Verbesserung schlechterer Studierendenleistungen statt, so ist berechtigterweise eine Umgestaltung dieses Prüfungstyps oder gar ein Verzicht angezeigt. Zum anderen deuten die Beispiele bereits das hochschuldidaktische Sensibilisierungspotenzial an: Dozierende, die Teamprüfungen durchführen oder Abschlussarbeiten betreuen, erhalten einen datenbasierten Stimulus, um ihre Beurteilungen und letztlich die Notenvergabe einer kritischen Selbstbeobachtung zu unterziehen, die idealiter in Optimierungen mündet (z. B. im Hinblick auf das Erwartungsbild, die Bewertungskriterien, Formen kooperativen Prüfens).

6 Fazit

Prüfungsdatenanalysen eröffnen interessante Möglichkeiten der Informationsgewinnung, z. B. längsschnittliche Analysen über größere Zeiträume, um etwa Veränderungen im Kontext der Corona-Pandemie oder Effekte einer gewandelten Studierendenzusammensetzung zu analysieren. Als hochschuldidaktisches Erkenntnisinstrument sind sie aber voraussetzungsreich: Aus einer Prozessperspektive gilt es zunächst, Erkenntnisinteressen zu klären, um relevante Daten und Datenquellen zu identifizieren. Ein zielgerichteter Datenbezug erfordert den Zugang zu sensiblen Daten unter Wahrung datenschutzrechtlicher Anforderungen und verlangt oftmals auch eine Vermittlungsleistung zwischen mehreren hochschulischen Akteuren und Organisationseinheiten mit verschiedenem Grundauftrag und unterschiedlichen Handlungslogiken (z. B. Studienadministration, Studiengangsentwicklung, Studiendekanat, Studiengangsleitung). Unter Umständen sind mehrere Datenquellen zu konsultieren und Kontextdaten einzubeziehen, wie etwa Informationen zu konkreten Prüfungsformen, die Modulhandbüchern, Prüfungsordnungen oder Vorlesungsverzeichnissen zu entnehmen sind. Die Verarbeitung der Daten setzt in der Regel umfangreiche Data-Preprocessing-Schritte voraus, da Daten hinsichtlich ihrer Genauigkeit, Vollständigkeit und Konsistenz geprüft und notwendige Korrekturen vorgenommen werden müssen. Der Umgang mit den Daten erfordert neben statistischen Kenntnissen ein Bewusstsein für ethische Implikationen der Datenanalyse (z. B. negative Folgen für einzelne Dozierende oder Studierende antizipieren) und ist der Risikominimierung verpflichtet (insbesondere mit Blick auf Datenschutz und Datensicherheit). Frameworks für die Implementierung von Learning Analytics an Hochschulen können eine Orientierung für die Erschließung von Prüfungsdaten bieten (Scheidig & Holmeier, 2021), wobei konkrete Umsetzungsschritte insbesondere von den lokalen Akteurskonstellationen und Zugangswegen zu Prüfungsdaten abhängen, die je nach Hochschulstruktur und Analyseebene variieren. Prüfungsdatenanalysen können bei umfangreichen Vorhaben immense zeitliche Ressourcen binden, die zum möglichen Ertrag der Analysen ins Verhältnis gesetzt werden müssen. Eine dauerhafte und erleichterte Nutzung von Prüfungsdatenanalysen kann durch die technische Umsetzung automatisierter Echtzeitanalysen erzielt werden, die sich via Dashboards abrufen lassen und nutzerseitig keine speziellen Statistikkenntnisse erfordern. Solche Lösungen sind jedoch mit hohem Initialaufwand verbunden sowie mit Aktualisierungsbedarf, sobald sich relevante Veränderungen bei der gekoppelten Verwaltungssoftware oder Studiengangsstrukturen ergeben.

Jenseits technischer, statistischer und rechtlicher Hürden besteht eine zentrale Herausforderung darin, mit Prüfungsdatenanalysen nützliche Erkenntnisse zu generieren und diese hochschulintern wirksam zu disseminieren. Verschiedene Anspruchsgruppen, für die prüfungsbezogene Erkenntnisse potenziell relevant sind, sollten frühzeitig eingebunden werden. Idealiter generieren Prüfungsdatenanalysen gezielt Antworten auf Fragen, die vorab in einem partizipativen Prozess er-

mittelt oder gemeinsam entwickelt wurden. Das eigentliche Potenzial von Prüfungsdatenanalysen entfaltet sich erst mit deren Nutzung, etwa wenn sie zur Reflexion von Prüfungspraktiken, zu Weiterentwicklungen im Prüfungswesen oder zu Optimierungen konkreter Prüfungen führen. Das Formulieren ebenen- bzw. adressatenspezifischer Thesen oder pointierter Reflexionsfragen, die detaillierte Berichte oder Präsentationen mit statistischen Auswertungen zu Prüfungen flankieren, kann die Beachtung und Nutzung der Ergebnisse von Prüfungsdatenanalysen fördern. Auch das Verweisen auf Ergebnisse aus Prüfungsdatenanalysen in verschiedenen Hochschulkontexten und -gremien oder das Angebot für vertiefende Analysen auf Studiengangs-, Studienbereichs- oder Modulebene können die Nutzungswahrscheinlichkeit für eine entwicklungsförderliche Auseinandersetzung mit den Ergebnissen erhöhen. Der direkte Dialog mit der Abnehmerseite mag sich nicht nur auf die Bereitschaft zur Nutzung bereitgestellter Informationen förderlich auswirken, sondern kann auch einem gemeinsamen Verständnis der Befunde zuarbeiten, da z. B. Studiengangsleitungen und Dozierende erfahrungsgesättigte Interpretationsangebote einbringen können.

Gleichwohl zeigen die hier berichteten Befunde, dass sich mitunter breite Interpretationskorridore öffnen und damit das Ableiten konkreter Maßnahmen erschwert wird. Zur Vermeidung voreiliger Schlussfolgerungen muss den komplexen Zusammenhängen des Zustandekommens von Noten besondere Beachtung geschenkt werden. Eine diskursiv angelegte Ergebnisinterpretation kann zur multiperspektivischen Entschlüsselung von Prüfungseffekten auf der mikrodidaktischen Ebene des Lehrens und Lernens, der spezifischen Leistungssituation sowie der konkreten Prüfungsgestaltung und -beurteilung beitragen. Diese Anstrengungen sind insofern gerechtfertigt und lohnend, als Prüfungsdatenanalysen relevante Erkenntnisse für verschiedene Akteure im hochschulischen Mehr Ebenensystem generieren und ihnen das Potenzial innewohnt, neue Zugänge und Inhalte zur hochschuldidaktischen Thematisierung von Prüfungskultur und -qualität bereitzustellen, wie sie in diesem Beitrag skizziert wurden.

Literatur

- Bandtel, M., Baume, M., Brinkmann, E., Bedenlier, S., Budde, J., Eugster, B., Ghoneim, A., Halbherr, T., Persike, M., Rampelt, F., Reinmann, G., Sari, Z. & Schulz, A. (2021). *Digitale Prüfungen in der Hochschule. Whitepaper einer Community Working Group aus Deutschland, Österreich und der Schweiz*. Hochschulforum Digitalisierung. Abrufbar unter: https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_Whitepaper_Digitale_Pruefungen_Hochschule.pdf (23.05.2024).
- Biggs, J. B. & Tang, C. (2011). *Teaching for Quality Learning at University: What the Student Does*. McGraw-Hill.
- Binder, D. (2019). Wer beginnt (k)ein Masterstudium? Determinanten der Übertrittspläne von Bachelorstudierenden. In A. Pausits, R. Aichinger & M. Unger (Hrsg.), *Quo vadis, Hochschule? Studienreihe Hochschulforschung Österreich* (S. 225–246). Waxmann.
- Boud, D. (1995). Assessment and learning: Contradictory or complementary? In P. Knight (Hrsg.), *Assessment for Learning in Higher Education* (S. 35–48). Kogan Page.
- Bücker, S., Deimling, M., Durduman, J., Holzhäuser, J., Schnieders, S., Tietze, M., Sayeed, S. & Schneider, M. (2015). Prüfung. In M. Schneider & M. Mustafić (Hrsg.), *Gute Hochschullehre: Eine evidenzbasierte Orientierungshilfe. Wie man Vorlesungen, Seminare und Projekte effektiv gestaltet* (S. 119–152). Springer.
- Bundesassistentenkonferenz (BAK) (1970/2009). *Forschendes Lernen – Wissenschaftliches Prüfen*. Universitäts-VerlagWebler.
- Epstein, R. M. (2007). Assessment in Medical Education. *New England Journal of Medicine*, 356(4), 387–396.
- Figas, P. & Hagel, G. (2017). A Task is not a Task. Empirical Results about the Quality of Instructional Tasks in Higher Education. In IEEE Computer Society (Hrsg.), *Proceedings of 2017 IEEE Global Engineering Education Conference* (S. 679–682). IEEE. <https://doi.org/10.1109/EDUCON.2017.7942920>
- Gerick, J., Sommer, A. & Zimmermann, G. (Hrsg.) (2017). *Kompetent Prüfungen gestalten: 53 Prüfungsformate für die Hochschullehre*. Waxmann. <https://doi.org/10.36198/9783838548401>
- Gunn, A. & Kapade, P. (2018). The university grade inflation debate is going global. *University World News*, 25. Mai 2018. Abrufbar unter: <https://www.universityworldnews.com/post.php?story=20180523095429859> (23.05.2024).

- Heublein, U., Hutzsch, C. & Schmelzer, R. (2022). *Die Entwicklung der Studienabbruchquoten in Deutschland*. DZHW Brief 05/2022. Abrufbar unter: https://doi.org/10.34878/2022.05.dzhw_brief (23.05.2024).
- Klöppling, S., Scherfer, M., Gokus, S., Dachsberger, S., Krieg, A., Wolter, A., Bruder, R., Ressel, W. & Umbach, E. (2017). *Studienabbruch in den Ingenieurwissenschaften. Empirische Analyse und Best Practices zum Studienerfolg*. Acatech. Abrufbar unter: https://www.acatech.de/wp-content/uploads/2018/03/acatech_STUDIE_Studienabbruch_Web-1.pdf (23.05.2024).
- Ledermüller, K., Nettekoven, M. & Weiler, M. (2015). Automatisierte Analyse von Multiple-Choice Prüfungen. In O. Vettori, G. Salmhofer, L. Mitterauer & K. Ledermüller (Hrsg.), *Eine Frage der Wirksamkeit? Qualitätsmanagement als Impulsgeber für Veränderungen an Hochschulen* (S. 173–189). UniversitätsVerlag-Webler.
- Macke, G., Hanke, U., Viehmann-Schweizer, P. & Raether, W. (2016). *Kompetenzorientierte Hochschuldidaktik: Lehren – vortragen – prüfen – beraten*. Beltz.
- Mansfield, H. C. (2001). Grade Inflation: It's Time to Face the Facts. *Chronicle of Higher Education*, 47(30), B24.
- Müller-Benedict, V. & Grözinger, G. (Hrsg.) (2017). *Noten an Deutschlands Hochschulen. Analysen zur Vergleichbarkeit von Examensnoten 1960 bis 2013*. Springer VS.
- OECD (Hrsg.) (2019). *Bildung auf einen Blick. OECD-Indikatoren 2019*. wbv Media.
- Persike, M. (2021). Digitales Prüfen: Didaktik, Umsetzung und Evidenz für die neue Prüfungsnormalität an Hochschulen. In I. Neiske, J. Osthusenrich, N. Schaper, U. Trier & N. Vöing (Hrsg.), *Hochschule auf Abstand: Ein multiperspektivischer Zugang zur digitalen Lehre* (S. 327–353). transcript.
- Reinmann, G. (2019). Forschendes Lernen prüfen: Hochschuldidaktische Gedanken zu einer Theorie des Prüfens. *Zeitschrift für Pädagogik*, 65(4), 608–626.
- Ricken, N. & Reh, S. (2017). Prüfungen – Systematische Perspektiven der Geschichte einer pädagogischen Praxis. *Zeitschrift für Pädagogik*, 63(3), 247–258.
- Scheidig, F. (2021a). Abschlussarbeiten an der Pädagogischen Hochschule FHNW. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 39(1), 31–37.
- Scheidig, F. (2021b). Von datengestützter Optimierung und gläsernen Studierenden: Beispiele, Potenziale und Probleme von Learning Analytics. In A. Thielsch, C. Bade & L. Mitterauer (Hrsg.), *Ursprünge hinterfragen – Vielfalt ergründen – Praxis einordnen. ReGeneration Hochschullehre* (S. 185–196). wbv Media.
- Scheidig, F. & Klingovsky, U. (2020). Hochschuldidaktik als Grenzfall wissenschaftlicher Weiterbildung. Eine Programmanalyse hochschuldidaktischer Angebote in der Schweiz. In W. Jütte, M. Kondratjuk & M. Schulze (Hrsg.), *Hochschulweiterbildung als Forschungsfeld. Kritische Bestandsaufnahme und Perspektiven* (S. 297–318). wbv Media.
- Scheidig, F. & Holmeier, M. (2021). Learning Analytics aus institutioneller Perspektive: Ein Orientierungsrahmen für die hochschulische Datennutzung. In Hochschulforum Digitalisierung (Hrsg.), *Digitalisierung in Studium und Lehre gemeinsam gestalten. Innovative Formate, Strategien und Netzwerke* (S. 215–231). Springer VS. https://doi.org/10.1007/978-3-658-32849-8_13
- Scheidig, F. & Holmeier, M. (2023). Prognose von Studienabbrüchen angehender Lehrkräfte mit Maschinellem Lernen: Potentiale der Nutzung von Studierenden- und Studienverlaufsdaten aus dem Campus-Management-System. *Unterrichtswissenschaft*, 51(3), 489–509. <https://doi.org/10.1007/s42010-023-00182-1>
- Scheidig, F. & Schweinberger, K. (2022). Assessment Analytics – Daten digitaler Prüfungen auswerten. In B. Berendt, A. Fleischmann, G. Salmhofer, N. Schaper, B. Szczyrba, M. Wiemer & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre. Ergänzungsband Nr. 108* (H 3.9). DUZ.
- Sedlmeier, P. & Renkewitz, F. (2013). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler*. 2. Aufl. Pearson.
- Statistisches Bundesamt (Hrsg.) (2016). *Prüfungen an Hochschulen 2015*. Fachserie 11, Reihe 4.2. Statistisches Bundesamt.
- Towfigh, E. V., Traxler, C. & Glöckner, A. (2014). Zur Benotung in der Examensvorbereitung und im ersten Examen: Eine empirische Analyse. *Zeitschrift für Didaktik der Rechtswissenschaft*, 1(1), 8–27.
- Towfigh, E. V., 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>
- Walzik, S. (2012). *Kompetenzorientiert prüfen. Leistungsbewertung an der Hochschule in Theorie und Praxis*. Buchdr.
- Wass, V., Van der Vleuten, C., Shatzer, J. & Jones, R. (2001). Assessment of clinical competence. *Lancet*, 357(9260), 945–949.

- Wissenschaftsrat (Hrsg.) (2012). *Prüfungsnoten an Hochschulen im Prüfungsjahr 2010. Arbeitsbericht mit einem Wissenschaftspolitischen Kommentar des Wissenschaftsrates*. Abrufbar unter: https://www.wissenschaftsrat.de/download/archiv/2627-12.pdf?__blob=publicationFile&v=4 (23.05.2024).
- Wollersheim, H.-W. & Pengel, N. (2016). Von der Kunst des Prüfens – Assessment Literacy. *HDS.Journal*, 2016(2), 14–32.

Autor:innen

Prof. Dr. Falk Scheidig. Ruhr-Universität Bochum, Institut für Erziehungswissenschaft, Bochum, Deutschland; Orcid-ID: 0009-0000-7837-2574; E-Mail: falk.scheidig@rub.de

Dr. Monika Holmeier. Pädagogische Hochschule FHNW, Stabsstelle Lehr- und Curriculumsentwicklung, Windisch, Schweiz; E-Mail: monika.holmeier@fhnw.ch



Zitiervorschlag: Scheidig, F. & Holmeier, M. (2024). Quantitative Prüfungsdatenanalysen als Beitrag zur Lehrentwicklung – Potenziale und Herausforderungen. *die hochschullehre*, Jahrgang 10/2024. DOI: 10.3278/HSL2435W. Online unter: wbv.de/die-hochschullehre



die hochschullehre

Interdisziplinäre Zeitschrift für Studium und Lehre

Die Open-Access-Zeitschrift **die hochschullehre** ist ein wissenschaftliches Forum für Lehren und Lernen an Hochschulen.

Zielgruppe sind Forscherinnen und Forscher sowie Praktikerinnen und Praktiker in Hochschuldidaktik, Hochschulentwicklung und in angrenzenden Feldern, wie auch Lehrende, die an Forschung zu ihrer eigenen Lehre interessiert sind.

Themenschwerpunkte

- Lehr- und Lernumwelt für die Lernprozesse Studierender
- Lehren und Lernen
- Studienstrukturen
- Hochschulentwicklung und Hochschuldidaktik
- Verhältnis von Hochschullehre und ihrer gesellschaftlichen Funktion
- Fragen der Hochschule als Institution
- Fachkulturen
- Mediendidaktische Themen

wbv.de/die-hochschullehre



Alle Beiträge von **die hochschullehre** erscheinen im Open Access!