

Kölsch, Fabian

Kollaboration und Rechtfertigung. Bewertungskommunikation in der universitären Notenberatung

Zeitschrift für Erziehungswissenschaft 29 (2026) 3, S. 719-736



Quellenangabe/ Reference:

Kölsch, Fabian: Kollaboration und Rechtfertigung. Bewertungskommunikation in der universitären Notenberatung - In: Zeitschrift für Erziehungswissenschaft 29 (2026) 3, S. 719-736 - URN: urn:nbn:de:0111-pedocs-357080 - DOI: 10.25656/01:35708; 10.1007/s11618-026-01377-9

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-357080>

<https://doi.org/10.25656/01:35708>

in Kooperation mit / in cooperation with:



<https://info.oe-deepgreen.de>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Kollaboration und Rechtfertigung

Bewertungskommunikation in der universitären Notenberatung

Fabian Kölsch 

Eingegangen: 17. Juni 2025 / Überarbeitet: 2. Oktober 2025 / Angenommen: 25. November 2025 / Online publiziert: 2. Februar 2026
© The Author(s) 2026

Zusammenfassung Der Aufsatz untersucht das universitäre Bewerten am Beispiel mündlicher Prüfungen in der Geschichtswissenschaft mit dem Ziel, die Kommunikation und Konstellation der universitären Notenberatung konzeptionell wie empirisch zu explizieren. Auf der Basis ethnografischer Feldforschung zeigt er, wie Prüfer:innen kollaborativ an der Produktion, Deliberation und Legitimation universitärer Urteile arbeiten. Im Zentrum steht die Dramaturgie der Notenberatung, die entlang dreier Bezugspunkte analysiert wird: (1) den Bewertungsobjekten, (2) den Bewertungssubjekten sowie (3) dem Bewertungspublikum. Die Notenberatung erweist sich als ein Akt gemeinsamer Selbstvergewisserung, dessen Funktion die Herstellung legitimer Begründungen für kollektiv bindende Notenentscheidungen ist. Mit seinen Analysen leistet der Aufsatz einen bewertungssoziologischen Beitrag zur Erforschung der universitären Bildungspraxis.

Schlüsselwörter Bewertungskonstellationen · Bildungssoziologie · Ethnografie · Mündliche Prüfung · Qualitative Hochschulforschung

✉ Fabian Kölsch

Institut für Soziologie, Johannes Gutenberg-Universität, Jakob Welder-Weg 12, 55128 Mainz, Deutschland
E-Mail: f.koelsch@uni-mainz.de

Collaboration and justification

Evaluative communication in university grading deliberations

Abstract This article examines university grading practices through the case of oral examinations in history, with the aim of conceptually and empirically explicating the modes of communication and the constellation of grade deliberations. Drawing on ethnographic fieldwork, it shows how examiners collaboratively engage in the production, deliberation, and legitimation of academic judgments. The focus lies on the dramaturgy of grade deliberations, analyzed along three reference points: (1) the objects of evaluation, (2) the subjects of evaluation, and (3) the evaluation audience. Grade deliberations emerge as acts of collective self-assurance, whose function is the construction of legitimate reasons for collectively binding grading decisions. With these analyses, the article contributes to the sociology of evaluation by advancing the study of academic practices in higher education.

Keywords Evaluation constellations · Sociology of education · Ethnography · Oral examination · Qualitative higher education research

1 Einleitung

Bewertungen sind ein allgegenwärtiger Bestandteil akademischer Lehre. Neben den flüchtigen Kommentaren zu studentischen Beiträgen in Seminardiskussionen finden sie vor allem am Semesterende statt: etwa in der solitären Korrektur individueller Hausarbeiten, in der quasi-maschinellen Bewertung standardisierter Klausuren sowie in der Beurteilung mündlicher Prüfungen. Im Zentrum dieses Aufsatzes steht die Notenberatung nach mündlichen Prüfungen. Die mündliche Prüfung, die historisch bis zur mittelalterlichen Universität zurückreicht und damit in einer langen Tradition universitärer Gesprächskultur steht (Gindhart und Kundert 2010), ist fest in den Lehrplänen des Hochschulbetriebs verankert. Angesichts der zunehmenden Integration sogenannter ‚Künstlicher Intelligenz‘ in das Bildungssystem erfährt sie – als ein Format, das als weitgehend resistent gegenüber KI gilt – derzeit sogar eine neue Aufwertung. Die interdisziplinäre Hochschulforschung hat sich hingegen erst in jüngerer Zeit mit der Praxis des Prüfens und Bewertens im Allgemeinen und der mündlichen Prüfung im Besonderen beschäftigt (Leonhard et al. 2025; Meer 2019; Tyagunova 2021).

Hier schließt der Beitrag an, indem er die Herstellung, Deliberation und Legitimation akademischer Urteile untersucht. Sein empirischer Fall sind die Situationen der universitären Notenberatung sowie der Notenverkündung nach mündlichen Prüfungen in einer wissenschaftlichen Disziplin, und zwar der Geschichtswissenschaft. Der Beitrag verfolgt sein Anliegen auf der Grundlage ethnografischer Beobachtungen und in Auseinandersetzung mit dem Modell der „Bewertungskonstellationen“ (Meier et al. 2016), mit dem Ziel, zu verstehen, wie die Praxis der Bewertung zur Produktion studentischer Leistungen beiträgt. Diesem Vorhaben liegt ein anerkennungstheoretischer Leistungsbegriff zugrunde, der darauf verweist, dass die studentische Leistung erst durch die Prozesse des Bewertens konstituiert und ratifiziert

wird und dass mit diesen Klassifikationen schließlich auch eine symbolische Auf- oder Abwertung der Studierenden im Feld der Universität verbunden ist (Banerjee et al. 2025, S. 236; Bourdieu 1988, S. 370; Lamont 2012, S. 207).

Der Aufsatz entfaltet seinen Argumentationsgang wie folgt: Nach einem Überblick über den Stand der Forschung (2), werden die konzeptionellen Annahmen der Studie erläutert (3). Anschließend wird ihr methodisches Design vorgestellt (4). Der empirisch-analytische Hauptteil (5) expliziert die Bewertungskonstellation der mündlichen Notenberatung, d.h. die Relation der Studierenden und ihrer Aussagen (als Bewertungsobjekte), der Lehrenden (als Bewertungssubjekte) und des Publikums. Das abschließende Kapitel fasst die Ergebnisse zusammen und diskutiert Grenzen und Anschlussmöglichkeiten der Untersuchung (6).

2 Funktion und Praxis des mündlichen Prüfens

Während testtheoretische Ansätze in der Soziologie und Erziehungswissenschaft mündliche Prüfungen in erster Linie im Hinblick auf ihre begrenzte diagnostische Güte beurteilen (Ingenkamp und Lissmann 2008, S. 139), zielen hochschuldidaktische Studien auf eine Evaluation und Verbesserung des Prüfungsgeschehens ab (Sommer 2018; Wildt und Redecker 2008). Die Allgemeine Pädagogik diskutiert hingegen, wie das Prüfungsgeschehen mit erzieherischem Handeln verknüpft ist und fragt danach, ob die Prüfung reines Organisationserfordernis, ein Selektionsgeschehen oder eine eigenständige Form pädagogischen Handelns ist (Prange und Strobel-Eisele 2006, S. 94f.; Ricken und Reh 2017). Ältere bildungssoziologische Studien befassen sich vor allem mit der gesellschaftlichen Funktion von Prüfungen; diese, so das Argument, objektivieren studentische und schulische Leistungen und dienen damit der Reproduktion einer stratifizierten und an Leistungszertifikaten orientierten Gesellschaft (Kvale 1972; Prahl 1976). In anderen Worten: Die Funktion der Prüfung bestehe darin, soziale Ungleichheiten in gesellschaftlich legitimierte Differenzen zu verwandeln (Bourdieu 2001, S. 46). Wissenssoziologisch lässt sich diese Perspektive bis zu den akademischen Umbrüchen der Post-68er-Generation zurückverfolgen, die einen lebhaften Nährboden für Debatten um „Sinn und Unsinn“ (Prahl 1976) von Hochschulprüfungen bot (Kehm 2001). Argumente dieser prüfungskritischen Kultur sind auch in aktuellen Diskursen präsent (Bülow-Schramm 2008).

Ausgehend von Überlegungen Niklas Luhmanns (2002) lässt sich die Wirkung von Prüfungen auch anders beschreiben. Der systemtheoretische Blick richtet sich nicht auf das Verhältnis von Prüfung und Arbeitsmarkt oder Prüfung und Gesellschaft (wie die o.g. Studien), sondern fragt nach der Funktion von Prüfungen für das Erziehungssystem. Luhmanns Antwortet lautet, dass Zensuren die Ausdifferenzierung des Erziehungssystems entlang der Unterscheidung von besser/schlechter, können/nicht-können, wissen/nicht-wissen ermöglichen und damit überhaupt erst die Beobachtbarkeit von Erziehungserfolgen stiften (Luhmann 2002, S. 63f.). Indem sie ferner Selbst- und Fremdvergleiche ermöglichen, werden Prüfung und Zensur zu zentralen Mitteln der Karrierekonstruktion (Luhmann 2002, S. 71f.). Schließlich lässt sich auch die disziplinierende Wirkung von Prüfungen betonen – nicht nur auf Lernende, sondern auch auf Lehrende: Der Zwang zur Prüfung fungiert, so Luh-

mann und Schorr (1988, S. 257), als „eine Art organisatorisches Korsett“, das die Lehrenden dazu anhält, sich mit den Leistungen der Lernenden auseinanderzusetzen.

Nur wenige Studien haben bislang eine empirische Wende vollzogen und mündliche Prüfungen an Universitäten mikrosoziologisch untersucht (Tyagunova 2021, S. 186). Für den spezifischen Fall projektbezogener Prüfungen im Lehramtsstudium zeigen Herzmann und Liegmann (2020), wie Prüflinge von den Prüfenden dazu angehalten werden, sich als Expert:innen ihrer eigenen Projekte zu inszenieren und wie sie mit dieser Adressierung umgehen. Die Positionierung als Expert:in strukturiert das Prüfungsgespräch insofern, als den Prüflingen ein Wissensvorsprung zugeschrieben wird. Gelingt diese Selbstinszenierung nicht, kann das Gespräch in eine klassische Rollenverteilung zurückfallen: wissende Prüfer:innen auf der einen, sich zu bewährende Prüflinge auf der anderen Seite (Herzmann und Liegmann 2020, S. 741). Leonhard et al. (2025) zeigen wiederum, dass mündliche Prüfungen fachliches Wissen evaluieren und zugleich zu einer Subjektivierung der Studierenden beitragen.

Aus einer konversationsanalytischen Perspektive untersucht Meer (1998, 2019) das mündliche Prüfungsgeschehen und beschreibt es als eine Form institutioneller Kommunikation. Sie zeigt, wie Prüflinge und Prüfer:innen durch Mechanismen des Turn-Takings und der Zuweisung von Rederechten verschiedene diskursive Positionen einnehmen; sie kommt zu dem Schluss, dass die Annahme einer strikten Asymmetrie zwischen Prüfling und Prüfer:in empirisch nicht haltbar ist (Meer 1998, S. 67). Dies wird von Tyagunova (2021) aufgegriffen, die aus einer ethnomethodologischen Perspektive rekonstruiert, wie die Beteiligten im Prüfungsgespräch unterschiedliche Handlungsprobleme bearbeiten. Ein zentrales Problem für die Prüflinge besteht in der Unsicherheit über die Angemessenheit ihrer Antworten. Sie lösen dieses Problem durch interaktive Mechanismen wie das Entschlüsseln der Frageabsicht der Prüfer:innen oder das Vermeiden von dispräferierten Schweigephasen (Tyagunova 2021, S. 216). Diese Gesprächsmechanismen verweisen auf eine grundlegende „Prüfungscompetenz“ der Prüflinge (Tyagunova 2021). In einer Anschlussstudie untersucht Tyagunova (2023) zudem die Artikulationsmuster fachspezifischen Wissens und unterscheidet dabei die Modi der Anwendung, Relationierung und Immanenz. Alle drei stellen spezifische Formen der Wissensproduktion dar, die für die Gesprächsform der Prüfung jeweils angepasst werden.

Den genannten Studien kommt das Verdienst zu, die Vollzugslogik des mündlichen Prüfens für die empirische Forschung zugänglich gemacht zu haben. Auffällig ist aber dennoch, dass in den genannten Studien eine Analyse der Notenberatung fehlt. Dadurch bleibt theoretisch wie empirisch unbestimmt, wie die Konstellation der Notenberatung zur Produktion der Leistung beiträgt. Für den Fall der Schule beschreibt Kalthoff (1996) das Notengespräch der mündlichen Abiturprüfung als Form einer kollektiven Vergewisserung: Er expliziert Mechanismen der Urteilsverhandlung, die vom ‚Spurensuchen‘ über das ‚Vermessen‘ bis hin zur Evaluation der Bildungskarriere des Prüflings reichen. Deutlich wird auch, dass im Fall der Abiturprüfung alle Teilnehmenden (Vorsitzende, Prüfer:innen, Beisitzer:innen und Protokollant:innen) gleichberechtigte Mitspieler:innen bei der Notenfindung sind (Kalthoff 1996, S. 119f.). Ebenfalls für den Fall mündlicher Schulprüfungen haben Solem et al. (2024) die Notenberatung aus konversationsanalytischer Perspektive

untersucht. Demzufolge werden Bewertungsprozesse interaktiv gestaltet und die Beteiligten nähern sich einer *gemeinsamen* Entscheidungsfindung an. Charakteristisch für diese Form der Aushandlung ist, dass erste Notenvorschläge bereits in frühen Phasen des Gesprächs artikuliert werden und die abschließende Notenvergabe kaum von den anfänglichen Einschätzungen abweicht (Solem et al. 2024, S. 1526 f.). Der Aufsatz schließt an die hier skizzierten Perspektiven an und wirft einen eher distanzierten Blick auf die Vollzugslogik des lokalen Geschehens. Über das Schließen der empirischen Lücke hinaus, will er zur konzeptionellen Bestimmung der Situation der Notenberatung beitragen.

3 Bewertungssoziologische Perspektiven

Der Aufsatz schließt an Überlegungen aus dem jungen Forschungsfeld der Soziologie der Bewertung (Nicolae et al. 2019) bzw. den sogenannten Valuation Studies (Lamont 2012) an, die ihren Fokus auf Bewertungsprozesse in unterschiedlichsten Feldern und Funktionssystemen richten. Das rege Interesse an Bewertungen entstammt u. a. der Beobachtung, dass diese in modernen Gesellschaften nicht nur allgegenwärtig, sondern auch überaus wirkmächtig geworden sind: etwa mit Blick auf Universitätsrankings, Google-Bewertungen oder Finanzratings von Staaten. Mit der Leistungsbewertung im Bildungssystem nimmt der Aufsatz hingegen ein vermeintlich unscheinbares, weil alltagsweltlich vertrautes, Bewertungsregime in den Blick.

Eine der grundlegendsten Annahmen der Bewertungssoziologie lautet, dass der Wert einer Sache oder einer Person dieser nicht inhärent gegeben ist, sondern erst durch Prozesse des Prüfens und Bewertens konstituiert wird (Banerjee et al. 2025; Verheyen 2018, S. 74). Mit dieser Perspektive positioniert sich die Bewertungssoziologie gegen einen messtheoretischen Realismus, der von der prinzipiellen Möglichkeit einer neutralen Abbildung – etwa von Leistung, Geschmack oder Intelligenz – ausgeht (Boud et al. 2018, S. 1108 f.). Prüfung und Bewertung geraten in einer solchen realistischen Perspektive ausschließlich als mehr oder weniger valide Messinstrumente in den Blick, wodurch die soziale Dimension des Bewertens auf den Aspekt möglicher Verzerrungen reduziert wird. Dem Aufsatz geht es stattdessen um die Einnahme einer „kontextualisierenden“ Perspektive, die das Bewerten als sozialen Prozess begreift, der in lokale Verhältnisse eingebettet ist (Kalthoff und Dittrich 2016, S. 463). Für den Fall des universitären Bewertens bedeutet dies danach zu fragen, *wie* die Leistung der Studierenden durch die soziomateriellen Praktiken des Bewertens her- und dargestellt wird (Kalthoff 1996; Mummelthey und Rabenstein 2024).

Zur Systematisierung von Bewertungsprozessen schließt der Aufsatz an die Unterscheidung von Werten und Bewerten an. Während mit dem Begriff des Wertens informelle Einschätzungen oder Geschmacksurteile bezeichnet werden, bezieht sich das Bewerten vor allem auf formalisierte Prozeduren, die sich durch institutionalisierte Konventionen oder Kriterien der Beurteilungen auszeichnen. Für den Fall der universitären Bewertung handelt es sich dabei um eine graduelle Unterscheidung, die eine analytische Sortierung von Bewertungsprozessen entlang ihres Härtegrades

bzw. Entscheidungscharakters erlaubt: Wertungen sind erste Einschätzungen, die korrigiert oder infrage gestellt werden können und den Bewertenden einen höheren Spielraum interpretativer Flexibilität ermöglichen. In der Zeitdimension betrachtet handelt es sich bei ihnen um Vor-Urteile, die erst im Verlauf des transsituativen Bewertungsprozesses – von der Notenberatung und Notengebung über die Notenverkündung bis hin zur administrativen Notenverbuchung – sukzessive gehärtet, d. h. in ihrer sozialen Wirkmächtigkeit wie auch ihrer Irreversibilität stabilisiert werden (Kalthoff und Dittrich 2016). Werten und Bewerten sind in diesem Verständnis unterschiedliche Praktiken auf einem Kontinuum der Objektivierung studentischer Leistungen.

Für die vorliegende Untersuchung ist schließlich das Modell der „Bewertungskonstellationen“ (Meier et al. 2016) von Bedeutung, dass in den vergangenen Jahren breit rezipiert und dessen Nutzen für die Analyse von Bewertungen im Bildungssystem bisher kaum herausgearbeitet wurde (siehe jedoch Mummelthey und Rabenstein 2024; Sowada und Peetz 2019). Das Modell soll die Beschreibung und Erklärung der transsituativen Dimension von Bewertungsprozessen ermöglichen (hierzu auch Kalthoff und Kölsch 2025b). Mit ihm geht es den Autor:innen um eine theoretische Öffnung des „methodologischen Situationismus“ (Meier et al. 2016) der Bewertungssoziologie, der das Bewerten ausschließlich auf seine situative Performanz reduziere. Drei zentrale Dimensionen kennzeichnen eine Bewertungskonstellation: (1) Positionen und Relationen, (2) Regeln sowie (3) materiellen Infrastrukturen.

Zu (1): Eine Bewertungskonstellation kann die Positionen der Bewertenden, der Bewerteten und des Publikums umfassen. Diese triadische Konzeption erlaubt nach den Relationen zu fragen, die notwendigerweise im Spiel sind, um Urteile zu produzieren, diese mit Legitimität auszustatten und über konkrete Situationen hinweg zu stabilisieren. Diese modellhafte Vereinfachung, erfordert gleichzeitig eine empirische Spezifizierung. Zum Beispiel suggeriert das Modell eine gewisse Stabilität der Positionen, wodurch unklar bleibt, ob es sich bei den Bewertenden um organisatorische Rollen oder um Sozialfiguren handelt, ob und welchen Unterschied es macht, ob die Bewertungsobjekte Menschen oder Dinge sind oder ob das Publikum als Zuschauer des Bewertungsprozesses oder als ‚Endverbraucher‘ auftritt (Muniesa und Helgesson 2013).

Zu (2): Ferner gehören zu einer Bewertungskonstellationen auch institutionalisierte Regeln, die das Geschehen strukturieren. Dabei ist zwischen formellen und informellen Sachregeln („Was darf Gegenstand der Bewertung werden und was nicht?“), Verfahrensregeln („Wie ist eine Prüfung durchzuführen?“) und Inklusionsregeln („Wer kommt als Prüfling oder Prüfende infrage?“) zu unterscheiden (Meier et al. 2016, S. 317). Viele dieser Regeln bleiben im konkreten Bewertungsprozess implizit und verborgen, können jedoch in Krisenmomenten explikationsbedürftig werden; in solchen Momenten wird erkennbar, dass das Befolgen einer Regel entweder „blind“ erfolgt (Wittgenstein 1984, S. 351) oder interpretative Arbeit verlangt, die immer auch an die Spezifik lokaler Verhältnisse angepasst ist (Boltanski und Thévenot 2014, S. 193). Dies bedeutet auch, dass die Regeln des Bewertens nicht vorausgesetzt werden können, sondern in ihrer institutionellen Verankerung und empirischen Realisierung berücksichtigt werden müssen.

Zu (3): Schließlich sind Bewertungskonstellationen durch materielle Infrastrukturen gestützt. Meier et al. (2016) schließen hierfür an praxis- und materialitätstheoretische Diskussionszweige an, in denen die grundlegende materielle Verfasstheit sozialer Praxis betont wird (Latour 2001). Die Bedeutung von materiellen Trägern ist für die vorliegende Studie vor allem mit Blick auf die Sichtbarmachung und Fixierung von Bewertungsobjekten (etwa das Beschreiben von Papier in Klausuren) sowie auf die Zirkulation von Bewertungsentscheidungen (etwa in Gutachten und Prüfungsprotokollen) relevant. Um das Prüfungsgeschehen für Dritte nachvollziehbar machen zu können, bedarf es vor allem organisatorisch vorgefertigter Materialien, wie Formulare oder Listen, die als Mittler in das Geschehen eingreifen. Die Bewertungspraxis wird demnach als ein Zusammenwirken körperlicher Verhaltensweisen, performativer Sprechakte, materieller Artefakte und Infrastrukturen begriffen – auch über Zeiten und Räume hinweg.

Der Aufsatz begreift diese theoretischen Überlegungen und das Modell der Bewertungskonstellation, dass die Autor:innen selbst als theoretisch offen und voraussetzungsarm beschreiben (Meier et al. 2016, S. 310f.), im Sinne einer sensibilisierenden Heuristik. Hierbei geht es nicht um eine bloße Anwendung eines Modells, sondern um die Spezifizierung seiner Annahmen im Licht der empirischen Beobachtung, d. h. es geht darum, Positionen, Regeln und Infrastrukturen nicht als Determinanten des Bewertens zu begreifen, sondern danach zu fragen, wie diese in der Praxis aufeinander bezogen werden.

4 Bewerten beforschen

Die folgende Analyse stützt sich auf empirisches Datenmaterial, das im Zuge einer ethnografischen Feldforschung in den Jahren 2023 bis 2025 an drei deutschen Universitäten generiert wurde (siehe auch Kalthoff und Kölsch 2025a, 2025b). Im Laufe des Forschungsprozesses wurden Lehrveranstaltungen und Prüfungen teilnehmend beobachtet und protokolliert sowie Interviews mit dem Lehrpersonal und Studierenden geführt und ausgewertet. Der Grad der Teilnahme beschränkte sich im Fall der Prüfungsbeobachtungen auf den stummen Beisitz und die Anfertigung schriftlicher Notizen (Emerson et al. 1995), im Fall der Lehrveranstaltungen auf die beobachtende Begleitung der Vorbereitung und Durchführung von Seminaren, Vorlesungen und Kolloquien. Zudem wurden Dokumente, die im Feld zirkulierten, gesammelt und dokumentenanalytisch ausgewertet (hierzu Prior 2003). Die Wahl des Forschungsansatzes orientierte sich an der Strukturierung des Gegenstandes. Da sich Prüfungen vor allem als Interaktionen unter Anwesenden vollziehen, erfordert ihre Beobachtung die Anwesenheit des/der Forscher:in. Im Gegensatz zur ausschließlichen Nutzung von Tonaufnahmen konnten so auch stumme Praktiken, wie Blicke, Gesten oder das Gekritzelt auf Papier in den Blick genommen werden.

Im Zuge der Feldforschung konnten verschiedene Prüfungsformen wie schriftliche Papierklausuren, studentische Forschungsprojekte oder mündliche Modulabschlussprüfungen in ihrer Durchführung und/oder Bewertung beobachtet werden. Insgesamt wurden sechs mündliche Prüfungen à 20 min an zwei verschiedenen Universitäten beobachtet. Die Dauer der Notenberatung und -verkündung variierte zwi-

schen zwei und fünf Minuten. Besonders die ‚schwierigen‘ Fälle nahmen dabei die meiste Zeit der Prüfer:innen in Anspruch. Die Beobachtungsprotokolle und Interviewtranskripte wurden offen kodiert, kategorisiert und ständig miteinander verglichen (Corbin und Strauss 1998). Die Kodiersitzungen fanden sowohl einzeln als auch in einer Gruppe statt. Während das offene Kodieren den gesamten Datenkorpus in den Blick nahm und der Generierung erster Ideen diente, zielte die Arbeit an den Kategorien auf eine thematische Ordnung bereits gewonnener Codes. Die Kategorienbildung war zugleich durch eine intensivierte Auseinandersetzung mit Konzepten aus der Forschungsliteratur geprägt. So wurden etwa Codes wie ‚Bewertungsdyade‘, ‚Zeugenschaft durch Mitschrift‘ oder ‚Prüfung als Lerngelegenheit‘ aus dem Material entwickelt und durch weitere Beobachtungen präzisiert und geschärft. In Verbindung mit dem Modell der Bewertungskonstellation entstanden aus diesen ursprünglichen Codes, die später zu Kategorien verdichtet wurden, schließlich die drei Bezugspunkte, die im Folgenden jeweils ein eigenes Analysekapitel bilden.

Eine ethnografische Forschungshaltung zeichnet sich jedoch nicht nur durch die schreibende Verfertigung von Beobachtungen, sondern auch durch die Annahme aus, dass der Zugang zum Feld bereits eine erste Möglichkeit zur Analyse bietet (Agar 1980). Mündliche Prüfungen sind in diesem Sinne ein besonderer Forschungsgegenstand, da sie unter Ausschluss der Öffentlichkeit stattfinden. Das heißt, sie kombinieren eine Klausur des Individuums mit Verborgenheit und Diskretion. Die Abgeschlossenheit mündlicher Prüfungen wird auch dadurch deutlich, dass dem Ethnografen einige Male die Teilnahme an mündlichen Prüfungen durch die Prüfenden verwehrt wurde. Die entsprechenden Begründungen lassen sich als Grenzziehungen des Forschungsfeldes interpretieren: Es wurde auf mögliche organisationsinterne Konsequenzen verwiesen (‚Das Prüfungsamt könnte Stress machen‘) oder auf den Schutz der Studierenden (‚Das könnte die Studierenden irritieren‘). Der Hinweis auf die Legitimationsinstanz antizipiert Entscheidungen, die noch nicht getroffen wurden, während der Verweis auf den Schutz der Studierenden das reibungslose Funktionieren der Prüfung als oberstes Gebot markiert. Durch die ethnografische Anfrage scheinen Prüfende zu realisieren, dass sie selbst und ihr Vorgehen Gegenstand einer Beobachtung werden, dass sie also bei ihrem i. d. R. verborgenem Tun gesehen werden und damit der geschützte Raum der Beratung öffentlich wird.

5 Die Konstellation der Notenberatung

5.1 Die Herstellung von Bewertungsobjekten

Anders als schriftliche Hausarbeiten oder Klausuren, die als geschriebene Texte vorliegen und wiederholt begutachtet werden können, handelt es sich bei mündlichen Prüfungen um situative *performances*, die im Moment ihres Vollzugs bereits vergangen sind (Bergmann 1985). Das heißt, dass die Objekte der Bewertung den Bewertenden nicht einfach in fixierter Form gegeben sind. Diesem Flüchtigkeitsproblem begegnen Beisitzer:innen und Prüfer:innen, indem sie das mündliche Gespräch in eine schriftliche Form bringen. Im Sinne Jakobsons (1959, S. 233) handelt es sich um eine „intralingual translation“: „an interpretation of verbal signs by means of

other signs of the same language“. Indem die Beteiligten mündliche Äußerungen in ein schriftliches Notationssystem überführen, transformieren sie diese zugleich. Vergleichbar dem Dolmetschen geschieht dies ad hoc und unter Zeitdruck. Das Prüfungswesen lebt dabei von der Fiktion, dass das Protokoll das Prüfungsgeschehen ‚so wie es war‘ dokumentiert. Ein Beispiel aus einer Beobachtung:

Der Prüfer hat sich während der Prüfung fortlaufend Notizen auf ein weißes DIN-A4-Papier gemacht. Im Vergleich dazu ist das Prüfungsprotokoll des Beisitzers spärlich beschrieben. Er hat nur einzelne Stichwörter aufgeschrieben. Während der Prüfer seine Beurteilung der einzelnen Aussagen ausspricht, haftet sein Blick konstant auf seinen Notizen. Nach jeder Beurteilung setzt er entweder einen Haken oder ein Kreuz mit seinem Stift auf das Papier. Der Protokollant blickt stattdessen kaum auf sein Protokoll und spricht frei.

Die Szene dokumentiert eine doppelte Protokollierung und die Differenz ihres Gebrauchs. Die Prüfungsnotizen dienen der Ergänzung der mündlichen Beurteilung, während das formale Protokoll für die Leistungsbewertung kaum von Nutzen ist; es ist selbstgenügsam und sperrt sich gegen seine Inanspruchnahme als Legitimationsinstanz der Bewertung. Die Mitschrift des Prüfers wiederum erhält im Moment des Abhakens einen listenförmigen Charakter (Brückner und Wolff 2015, S. 344): Entsprechend der Logik einer Inventarliste geht es um die Bestandsaufnahme dessen, was ‚da war‘ und ‚was nicht‘, das heißt zuallererst um eine Sichtbarmachung der Antworten für die Zwecke der Bewertung. Grundlage hierfür ist die „subjektive Semiotik“ (Ehlich 2003, S. 19) der Mitschrift, die Lesbarkeit unter Zeitdruck gewährt. Ein weiteres Beispiel:

Nachdem die Studentin sichtbar gefrustet den Raum verlassen hat, äußert die Prüferin, dass sie sich unsicher ist, ob die vergebene Note (3,3) nicht zu hart gewesen sei. Der Protokollant hält dagegen. Er blickt auf sein Protokoll, liest einzelne Fragen laut vor und kommentiert diese: „Da hatte sie wirklich gar keine Antwort (...), das ist absolutes Basiswissen. Und bei der letzten Frage, hat sie auch nur rumgestottert.“ Der Prüfer daraufhin: „Ja, hast’ recht. Es hilft, das nochmal so vor Augen zu haben.“

Im Kontrast zum vorherigen Beispiel ist es hier der Beisitzer, der mithilfe des Protokolls die infrage gestellte Urteilskraft der Prüferin rehabilitiert und ihr sein ‚Gesicht leiht‘. Bedingung hierfür ist der dokumentarische Charakter einer pragmatisch verstandenen Schrift (Ehlich 1994), die Aussagen über vergangene Performances autorisiert und legitimiert; das Protokoll „bürgt dann im Gegenwert der Schriftzeichen für das Geschehene. Es übernimmt die Funktion eines Wahrheitszeichens für Echtzeit“ (Vismann 2002, S. 137). In diesem Sinne sind Notizen materialisierte Erinnerungen – selektiv erzeugt und mit spezifischem Verwendungszweck (Luhmann 1998, S. 579). Relevant bleibt allein, wie die Prüflinge auf die Fragen reagierten; alles andere – etwa die Schwierigkeit der Fragen oder vergangene Seminareindrücke – bleibt in aller Regel unmarkiert. Ein Protokoll, das für die Zwecke der Organisation ausreichen mag (etwa indem einzelne Stichworte notiert werden), bietet weniger Unterstützung bei der Begründung und Rechtfertigung des Urteils. Umgekehrt erzeugt ein detailliertes Protokoll eher Evidenz bei der Benotung. Für

die Beisitzer:innen bedeutet dies, dass diese hinter der Evidenzfiktion des Protokolls zurücktreten können: Sie sprechen nicht als involvierte Bewerter:innen, sondern im Namen des Dokuments.

Im Gegensatz zu einer realistischen Annahme über die fraglose Gegebenheit des Bewertungsobjekts, zeigt sich hier, wie die Prüfer:innen in dessen Herstellung involviert sind. Indem sie die medientechnische Infrastruktur von organisatorischen Protokollen und Mitschriften für ihre Zwecke bemühen, reagieren sie auf die Flüchtigkeit des Prüfungsgeschehens. Die bewertungssoziologisch interessante Ambivalenz dieser Praxis ist, dass die Mitschriften die Bewertenden dazu auffordern, nicht nur die mündliche Performance, sondern auch die Qualität ihrer eigenen schriftliche Dokumentationsleistung zu bewerten. Die Fiktion des ‚so ist es gewesen‘ hält jedoch die eigene Arbeit am Urteil latent und bestätigt die Eigenständigkeit der Leistung auf Seiten der Studierenden (Verkuyten 2000, S. 469).

5.2 Kollaboratives Urteilen

Auf den ersten Blick legt die formale Rollendifferenzierung von Erst- und Zweitprüfer:innen oder Prüfer:innen und Beisitzer:innen eine asymmetrisch verteilte Entscheidungsmacht nahe, jedoch zeigen erste empirische Analysen, dass die Notenberatung nicht der Logik hierarchischer Weisung, sondern einer Logik gegenseitiger Verständigung folgt (Solem et al. 2024; Syverud und Prøitz 2025). Die Handlungs- und Entscheidungsfähigkeit des Bewertungsobjektes kann demnach nicht einfach theoretisch vorausgesetzt werden, sondern muss einer prozesssensiblen Forschung zum empirischen Problem werden; mit Meier et al. (2016, S. 316) gesprochen, geht es um die Fusion zu einer Bewertungsinstanz. Die kollaborative Dimension des Bewertens zeigt sich schon zu Beginn der Notenberatung, nachdem die Prüflinge den Raum verlassen, haben: Die Bewertenden entspannen sich, sie atmen durch und lachen zusammen, zeigen sich genervt oder äußern sich überschwänglich lobend. Während die Stimmung der Prüfung von taktvoller Höflichkeit geprägt ist, ist jene der Notenberatung ‚aufgelockert‘. Eine Sequenz:

Prüfer: Wie schalldicht sind diese Türen?

Beisitzer: Lachen sollten wir jetzt nicht.

Prüfer: Okay. Aber man merkt, dass der nicht bei mir studiert hat. Der würde sich nicht trauen zu sagen, dass Geschichte organisch ist, mit Fortschritt und Niedergang.

In diesem Auszug werden drei zentrale Aspekte deutlich. Erstens fällt die Distanzierung zum Prüfling auf: Die Aussage, „nicht bei mir studiert“, ist mehr als eine einfache Einschätzung des Prüflings – sie signalisiert, dass der Prüfer sich mit ihm und seinen Aussagen nicht identifiziert. Markiert wird die Differenz zu einem – aus Sicht des Prüfers – schlichten Geschichtsverständnis und damit die Nicht-Passung des Studenten. Ein zweiter Aspekt liegt in der Veränderung der Interaktionsordnung: Vom Gespräch mit dem Prüfling zum Gespräch über ihn, von einer triadischen Konstellation der Prüfung zu einer dyadischen der Beratung. Diese konspirativ anmutende Vergemeinschaftung ermöglicht den Prüfer:innen ihre teilweise unsicheren Urteile ungestört und offen zu formulieren und entlastet von der

Erwartung taktvoller Kommunikation. Ihre Bedingung ist eine räumliche Infrastruktur, die nicht nur die visuelle, sondern auch die auditive Exklusion des Studierenden stützt. Drittens verständigt sich das Prüfungspersonal in einem klar strukturierten Prozess (Anspielung, Thematisierung, Bestätigung) auf eine *gemeinsame* Reaktion: Auf den ersten Blick geht es darum, die Nicht-Hörbarkeit des Gesprächs zu wahren; auf den zweiten Blick führt der Hinweis direkt zu einer ersten Kategorisierung des Prüflings und seiner Leistung. In diesen ersten Momenten der Notenberatung geht es zunächst um eine Vergewisserung über das gemeinsame Gehörte, die sich auf die für die Notenfindung wesentlichen Aspekte beschränkt: Ein Austausch allgemeiner Eindrücke und Wertungen, die sich an Gegensätzen festmachen (sicher/unsicher; vorbereitet/nicht-vorbereitet; wissen/nicht-wissen, etc.). Ein Beispiel aus einem Benotungsgespräch:

- Prüferin: Also ich fand den Einstieg ziemlich unstrukturiert und schwammig.
Also er war auf die Prüfungsform nicht vorbereitet.
- Beisitzer: Im zweiten Teil besser.
- Prüferin: Genau. Da hat er nicht so lang geredet und kam auf den Punkt.
Also ich denke wir reden hier über eine eins Komma ((1,)).
- Beisitzer: Ja. Insgesamt fehlte ein bisschen Struktur. Die Quellen hätte er noch bisschen genauer beschreiben–
- Prüferin: Ja, das war an ein paar Stellen so bisschen ein Problem. Also würde ich sagen eins Komma drei ((1,3)).
- Beisitzer: Unterschreib ich.

Der Beginn des Ausschnitts dokumentiert die Legitimationsbedürftigkeit negativer (Teil-)Bewertungen. Diese gelten nur dann als legitim, wenn sie eindeutig auf die Eigentätigkeit der Prüflinge zurückgeführt werden können (hierzu auch Verkuyten 2000, S. 455). Indem die Prüferin ihre anfängliche Bewertung mit einem Kommentar zur mangelhaften Vorbereitung des Studenten hinsichtlich des Prüfungsablaufs einleitet, deutet sie zugleich an, dass über die inhaltliche Dimension kein Klärungsbedarf besteht. Der Prüfer greift dies anschließend auf und bestätigt damit die stillschweigende Übereinkunft in Bezug auf die inhaltliche Dimension der Prüfung. Ferner drückt sich im Gebrauch der Personalpronomen (*ich* denke, *wir* reden hier über eine eins Komma‘) ein erster Verständigungsversuch aus, der einen Benotungsraum absteckt. Die sich anschließende negative Teilwertung des Beisitzers (*Die Quellen‘*) liefert der Prüferin wiederum Begründungsmaterial für eine Notenofferte, die zum Abschluss ratifiziert wird.

Neben solchen reibungslos ablaufenden Beratungen finden sich im Material auch Formen der Gegenrede, in denen sich die Beisitzer:innen als Für- oder Widersprecher einer aus ihrer Sicht zu milden oder zu strengen Beurteilung positionieren. Diese Stellungnahmen erfolgen in beide Richtungen: Erstprüfer:innen treten mitunter als Fürsprecher der Prüflinge auf und werden dafür von Beisitzer:innen kritisiert – oder umgekehrt. Ein Beispiel:

- Prüferin: Wenn man fragt, sagen Sie etwas über den Autor und dann sagt er, der ist in der Schweiz geboren. Wenn schon solche Sachen nicht passen. Man

muss ihm immerhin zugutekommen lassen, dass er aufgeregt war.

Beisitzer: Ja, ne. Aber das waren die anderen auch.

Prüferin: Ja. Also drei Komma drei ((3,3))?

Hier klären die Akteure die Frage, welche sozialen Merkmale in angemessener Weise als entlastend beziehungsweise belastend in die Bewertung einbezogen werden dürfen. Während ein unvorbereiteter Prüfling in vollem Umfang für seine Leistung verantwortlich gemacht wird (wie im ersten Beispiel gezeigt), kann ein nervöser Prüfling prinzipiell als partiell entlastet gelten. Nervosität wird aus Sicht der Prüferin grundsätzlich als Indiz für eine Entlastung von Verantwortung gewertet, verliert jedoch durch den Verweis des Beisitzers auf das Gleichheitsgebot („die anderen auch“) an Legitimität (Boltanski und Thévenot 2014, S. 192). Der Beisitzer reguliert dadurch das Nähe-Distanz-Verhältnis der Prüferin zum Prüfling und fordert eine Orientierung an der Norm der Leistungsgerechtigkeit ein.

Mit diesen Überlegungen zur Notenberatung lässt sich die Frage nach der Konstruktion von Leistung präzisieren: Die Gespräche folgen weniger der Fiktion einer durch Kriterien gestützten Messung, sondern vielmehr dem Sammeln angemessener wie unangemessener Begründungen für erste Eindrücke und Schätzurteile. Ich schließe hiermit an praxeologisch orientierte Entscheidungstheorien an, die das Verhältnis von Entscheidungsprozess und -ergebnis nicht als Kausalkette, sondern als retroaktive Herstellung von Gründen für schon getroffene Ad-Hoc-Entscheidungen konzipieren (Dahmen 2022; Garfinkel 1967; Schmidt 2022). Bewerten erscheint insofern weniger ein öffentlicher Austausch guter Gründe zu sein, die im Ergebnis zu einem eindeutigen Urteil führen, sondern vielmehr eine Produktion plausibler und legitimer Begründungen für schnelle Notenurteile, die auf Einschätzungen zurückgehen, die sich im Laufe des Prüfungsgeschehens herausbildeten. Ferner zeigt die Analyse entgegen der Vorannahme einer organisatorisch stabilisierten Rollendifferenzierung, dass die Bewertenden gemeinsam an der Produktion von Leistung partizipieren und dabei füreinander je unterschiedliche Positionen einnehmen. Das Bewertungssubjekt, verstanden als eine entscheidungsfähige Einheit, bildet sich somit erst im Vollzug des gemeinsamen Urteilens heraus.

5.3 Die Rechtfertigung der Bewertung und ihr Publikum

In einem weiten Sinne kann das Publikum der Bewertung unter anderem Familien oder Kommiliton:innen der Prüflinge sowie Kolleg:innen der Prüfenden umfassen. Ferner sind in einer pragmatistischen Perspektive (etwa Mead 1973) die Bewertenden und die Bewerteten jeweils füreinander und für sich selbst ein Publikum; deutlich wird dies etwa an Selbstbeobachtungen und Selbstbeschreibungen. In einem engeren Sinne zählt vor allem die Hochschulorganisation zum Publikum, die über Regeln und zirkulierende Dokumente (etwa Prüfungsformulare) auf das Geschehen zugreift. Indem sie die Prüfung infrastrukturell rahmt, kann sie zwar nicht den Prüfungsprozess als solchen beeinflussen, jedoch über die Notenentscheidung und das im Protokoll dokumentierte Wissen den Prozess der Leistungsobjektivierung administrativ vervollständigen. Der Fall der mündlichen Notenberatung weist darü-

ber hinaus noch eine weitere Publikumsrolle auf: die Prüflinge selbst. Ein Beispiel aus einer Notenverkündung:

Prüfer: Herzlichen Glückwunsch sie haben die Prüfung mit Bravour bestanden. Das ist eine glatte eins Komma null ((1,0)). Sie haben sich wirklich auf den Punkt genau vorbereitet und da war im Prinzip alles Wichtige drin. Also wirklich ganz ganz prima und ich bedaure es wirklich sehr, da hab ich mich mit meinem Kollegen verständigt, dass Sie im Seminar nicht mehr sagen.

Für einen kurzen Moment lockert sich der ernste Blick des Prüflings und ein stolzes Lächeln erscheint, das er jedoch ebenso schnell wieder verbirgt, bevor er in einem selbstkritischen Ton antwortet:

Prüfling: Ja, das ist ne schlechte Angewohnheit geworden, das muss ich abstellen.

Prüfer: Ja, Sie haben durchaus mehr zu sagen als viele andere, die sich die ganze Zeit melden. Also machen Sie das. Ja, sehr schön, dann herzlichen Glückwunsch und schönen Tag.

Die Notenverkündung beginnt im untersuchten Sample fallübergreifend mit einer Beglückwünschung, die den rituellen Charakter des Geschehens unterstreicht und das Bestehen als Routineleistung markiert. Im obigen Fall wird die Besonderheit des Bestehens durch den Zusatz „mit Bravour“ hervorgehoben. Die affektive Dimension des professoralen Lobs zeigt sich in der direkten Reaktion des Studierenden, der dieses mit Stolz quittiert. Besonders relevant ist hierbei die Wertung der mangelnden Beteiligung im Seminar, mit der eine in die Zukunft gerichtete Anrufung verbunden ist, derer sich der Prüfling unterwirft, indem er Besserung gelobt. Die abschließende symbolische Anerkennung, die sich über eine Abwertung anderer Teilnehmer:innen vollzieht, inszeniert die universitäre Leistungsordnung als Wettbewerb und den Studierenden als zukünftigen Primus. Solche prospektiven Kommentare sind nicht ungewöhnlich und markieren die Studierenden als Lernende. Dies zeigt sich auch im folgenden Beispiel, das an den bereits bekannten Fall des ‚unstrukturierten‘ Studenten anschließt:

Prüferin: Herzlichen Glückwunsch, Sie haben die Prüfung bestanden, und zwar mit der Note Eins Komma Drei ((1,3)). Man hat gemerkt, Sie haben sich mit den Texten auseinandergesetzt und ein hohes Reflexionsniveau bewiesen. Vielleicht, warum es keine glatte Eins geworden ist. Sehr sperrig war die Frage nach den Quellen. Und tatsächlich, das haben Sie in der Prüfung selber gemerkt, war so ein bisschen Struktur. Das ist auch so Prüfungstaktik. Wenn ich nach den Quellen frage, dann nennen Sie doch einfach erstens, zweitens, drittens, viertens und die eine Quelle ist am wichtigsten.

Nach der obligatorischen Beglückwünschung, der Bekanntgabe der Note und einem kurzen Lob erfolgt ein Rahmenwechsel hin zur Rechtfertigung. Die Eindeutigkeit der Bewertung wird von der Prüferin unter anderem dadurch hervorgehoben, dass der Student seine Fehlleistung noch während der Prüfung selbst erkannt habe –

eine Naturalisierung der Leistung, die von allen Beteiligten anerkannt wird (Verkuyten 2000, S. 462). Die Rechtfertigung der Bewertung vollzieht sich hier über die Einbindung des Studenten. Mit ihrem abschließenden Hinweis auf die ‚Prüfungstaktik‘ wechselt die Prüferin den Rahmen von der Leistungsbeurteilung hin zur pädagogischen Belehrung. Dies zeigt sich auch in anderen Fällen:

Prüfer: Also ich hatte den Eindruck, dass es Ihnen schwergefallen ist, das, was sie gelesen haben auf den Punkt zu bringen. Also das ist etwas, was man auch üben kann. Zum Beispiel habe ich nach Quellen gefragt. Und dann fangen sie mit mathematischen Modellen an und das ist eben nicht das, was man als Historiker unter Quellen versteht.

Die Begründung der Leistung wird hier ebenfalls zur Lerngelegenheit der Studierenden. Auf diese Weise verschiebt sich der Fokus von der abgeschlossenen Prüfung hin zu einer fortgesetzten Bildungsbiografie, in der die einzelne mündliche Prüfung lediglich eine Etappe in einem auf Dauer angelegten Prüfungsparcours darstellt. Verbunden wird dies mit einer Erwartung das Sprachspiel der Geschichtswissenschaft zu beherrschen. Eine klare Botschaft: Mit dieser Leistung ist und wird man nicht Teil der akademischen Gemeinschaft.

Die Fallspezifität des Bewertungspublikums lässt sich folgendermaßen zusammenfassen: Im Gegensatz zu anderen Bewertungskonstellationen wird das studentische Publikum der mündlichen Prüfung direkt im Anschluss mit der Beurteilung konfrontiert. Dies erhöht den Rechtfertigungsaufwand auf Seiten der Bewertenden, die die vermeintliche Eindeutigkeit der Leistungsfeststellung sowohl gegenüber der Organisation (durch Protokolle) als auch gegenüber den Studierenden nachweisen bzw. in Szene setzen müssen. Durch diesen transsituativen Publikumsbezug gewinnt das Sammeln überzeugender Begründungen in der Situation der Notenberatung zusätzliche Relevanz. Ferner transformiert die Notenverkündung auch das Objekt der Bewertung: Während es in der Beratung um die Konstruktion einer gemeinsamen Erinnerung an das Gesagte geht, die zum Objekt der Bewertung wird, vollzieht die Verkündung und Rechtfertigung der Note eine Identifizierung der Prüfungsleistung mit den Prüflingen; zentral für diese individualisierende Zurechnung der studentischen Leistung ist die Invisibilisierung der Bewertungsleistung auf der Seite der Prüfenden. Dass diese nicht immer gelingt, zeigt sich etwa im studentischen Klatsch über unfaire oder besonders milde Prüfer:innen, der wiederum voraussetzt, dass die Prüflinge ihrerseits eine Selbstbewertung vorgenommen haben, von der die offizielle Bewertung abweicht.

6 Fazit und Ausblick

Der Aufsatz untersuchte Prozesse der Notengebung im Kontext mündlicher Hochschulprüfungen. Sein Ziel war es zu verstehen, wie in und durch Notenberatungen die Leistung der Studierenden fabriziert und (schriftlich-mündlich) legitimiert wird. Ausgehend vom Modell der Bewertungskonstellationen wurde danach gefragt,

wie in der Notenberatung Bewertungsobjekte hergestellt, diese durch Bewertungs-subjekte begutachtet und das Bewertungspublikum transsituativ durch Protokolle, Rechtfertigungen und Belehrungen eingebunden wird. Die Notenberatung mobilisiert hierfür verschiedenste Ressourcen: kollaborative und autoritative Sprechakte, räumliche Trennungen, Schriftzeichen und Dokumente.

Aufgrund des explorativen Charakters der Studie und des kleinen Samples bleiben einige Fragen unbeantwortet: So wurden etwa nur jene Notenberatungen beobachtet, in denen die Bewertenden in einem kollegialen Verhältnis zueinanderstanden. Es sind aber auch andere Beziehungskonstellationen denkbar, etwa im Fall sich gegenseitig unbekannter Prüfer:innen (Solem et al. 2024). Für die empirische Vielfalt möglicher Beziehungseffekte gilt es empirisch sensibel zu bleiben – auch und gerade aus einer gerechtigkeits-theoretischen Perspektive, die sich für mögliche Ungleichheitseffekte von Bewertungen interessiert. Des Weiteren hat der Aufsatz mit der Geschichtswissenschaft eine Disziplin beforscht, die sich womöglich von anderen Disziplinen und ihren spezifischen Prüfungskulturen unterscheidet, etwa hinsichtlich ihres akademischen Selbstverständnisses, einer spezifischen Studierendenzusammensetzung oder auch unterschiedlicher (Aus)bildungsziele und damit verbundenen Erwartungen. Diesbezüglich bietet das Modell der Bewertungskonstellation eine Möglichkeit für systematische Fallvergleiche zwischen unterschiedlichen Bewertungsformen und -kulturen – ein Anliegen, dass die Autor:innen explizit teilen (Meier et al. 2016, S. 321).

Der Beitrag hat ferner auch auf Grenzen des bewertungssoziologischen Rahmens aufmerksam gemacht. Zwar erweist sich das Modell der Bewertungskonstellation als hilfreiche Heuristik, um die transsituative Dynamik von Bewertungsprozessen zu erfassen, es bleibt jedoch zu schematisch, wenn es um die Beschreibung und Erklärung konkreter Bewertungssituationen geht. In diesem Zusammenhang weist der Aufsatz auf vier offene Baustellen hin, die sich auch für einen zukünftigen Dialog von Bildungsforschung und Bewertungssoziologie als instruktiv erweisen könnten: Erstens suggeriert das Modell eine gewisse Stabilität der Bewertungsobjekte und blendet damit die lokale Objektivierungsarbeit der Bewertenden aus. Am Fall der mündlichen Prüfung wird deutlich, dass Bewertungsobjekte nicht einfach in vergegenständlichter Form vorliegen, sondern in ihrem Objektstatus aktiv hergestellt und stabilisiert werden müssen – in diesem Fall durch dialogische und schriftliche Verfahren der Erinnerung und Fixierung von Sinn. Wie solche Konstruktionsprozesse bei anderen Prüfungsverfahren – etwa Large-Scale-Assessments – ablaufen, bleibt Gegenstand weiterer empirischer Forschung. Zweitens setzt das Modell die Entscheidungsfähigkeit von Bewertenden als selbstverständliche Tatsache voraus. In diesem Zusammenhang hat der Aufsatz dargestellt, wie die Prüfenden miteinander um eine gemeinsam verantwortbare Entscheidung ringen und sich die Legitimität ihrer Urteile zu- oder absprechen, d. h. dass das Entscheiden im Fall mündlicher Prüfungen konsequenterweise als eine deliberative Praxis begriffen werden muss. Drittens ist mit dem Publikum der Bewertung auch die Frage nach dem Rezipient:innenstatus von Bewertungen aufgerufen. Der Aufsatz hat diesbezüglich auf die Relation von Bewertungssituation und studentischem Publikum aufmerksam gemacht. Das Argument diesbezüglich lautete, dass die Fabrikation belastbarer Rechtfertigungen erst vor dem Hintergrund eines Publikums relevant wird, das von den Bewertenden nicht

als passive Konsument:innen, sondern als aktive Rezipient:innen der Entscheidung angesprochen wird. Subjektivierungstheoretisch gesprochen geht es hierbei um eine Anrufung zur bildungsbedeutsamen Auseinandersetzung mit der ‚eigenen‘ Leistung. Schließlich bleibt viertens unklar – und eine Diskussion dieser Problematik würde den Rahmen des vorliegenden Aufsatzes sprengen – wie sich das Modell der Bewertungskonstellation in Relation zu höherstufig generalisierten Strukturmustern des Bildungs- oder Erziehungssystems verhält, d. h. wie das Verhältnis von transsituativer Prüfungspraxis und ihrer gesellschaftlichen Funktion zu begreifen ist; eine Frage, die nicht zuletzt auch die Relation von empirischer Forschung und Gesellschaftsanalyse tangiert.

Förderung Dem Aufsatz liegt ein Forschungsprojekt („Universitäre Urteile. Die Bewertung studentischer Leistungen in den Ingenieurs- und Geschichtswissenschaften.“ KA 1235/10-1) zugrunde, das durch die DFG gefördert wurde.

Funding Open Access funding enabled and organized by Projekt DEAL.

Interessenkonflikt F. Kölsch gibt an, dass kein Interessenkonflikt besteht.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Agar, M. H. (1980). *The professional stranger. An informal introduction to ethnography*. Academic Press.
- Banerjee, M., Feindt, G., Kalthoff, H., & Wihstutz, B. (2025). Leistung im Vergleich. Objektivierungen trotz Differenzierungskontingenz. In A. Friedrichs, S. Hirschauer, P. Hofmann & G. Schabacher (Hrsg.), *Humandifferenzierung im Vergleich* (S. 236–266). Velbrück.
- Bergmann, J. (1985). Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit. Aufzeichnungen als Daten der interpretativen Soziologie. In W. Bonss & H. Hartmann (Hrsg.), *Entzauberte Wissenschaft* (S. 299–320). Schwartz.
- Boltanski, L., & Thévenot, L. (2014). *Über die Rechtfertigung*. Hamburger Edition.
- Boud, D., Dawson, P., Bearman, M., Bennett, S., Joughin, G., & Molloy, E. (2018). Reframing assessment research: through a practice perspective. *Studies in Higher Education*, 43(7), 1107–1118.
- Bourdieu, P. (1988). *Homo Academicus*. Suhrkamp.
- Bourdieu, P. (2001). Die konservative Schule. In P. Bourdieu (Hrsg.), *Wie die Kultur zu den Bauern kommt* (S. 25–46). VSA-Verlag.
- Brückner, F., & Wolff, S. (2015). Die Listen der Organisation – Und der Blick zwischen die Zeilen. In S. Wilz & V. v. Groddeck (Hrsg.), *Formalität und Informalität in Organisationen* (S. 341–370). Springer VS.
- Bülow-Schramm, M. (2008). Hochschuldidaktische Prüfungskritik revisited unter Bologna-Bedingungen. In S. Dany, B. Szczyrba & J. Wildt (Hrsg.), *Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen* (S. 27–44). Bertelsmann.
- Corbin, J. M., & Strauss, A. L. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory*. SAGE.

- Dahmen, S. (2022). Konturen einer praxistheoretischen Professions- und Entscheidungsforschung. *Empirische Pädagogik*, 36(1), 31–47.
- Ehlich, K. (1994). Funktion und Struktur schriftlicher Kommunikation. In H. Günther & O. Ludwig (Hrsg.), *Schrift und Schriftlichkeit. Writing and Its Use* (S. 18–41). De Gruyter.
- Ehlich, K. (2003). Universitäre Textarten, universitäre Struktur. In K. Ehlich & A. Steets (Hrsg.), *Wissenschaftlich schreiben – lehren und lernen* (S. 13–28). De Gruyter.
- Emerson, R. M., Fretz, R. I., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes*. University of Chicago Press.
- Foucault, M. (2019). *Überwachen und Strafen* (21. Aufl.). Suhrkamp.
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Prentice Hall.
- Gindhart, M., & Kundert, U. (Hrsg.). (2010). *Disputatio 1200–1800. Form, Funktion und Wirkung eines Leitmediums universitärer Wissenskultur*. De Gruyter.
- Herzmann, P., & Liegmann, A. (2020). Mündliche Prüfungen im Kontext des Forschenden Lernens. (Re-)Adressierungen als Inszenierung studentischer Expertise. *Zeitschrift für Pädagogik*, 66(5), 727–745.
- Ingenkamp, K., & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6. Aufl.). Beltz.
- Jakobson, R. (1959). On linguistic aspects of translation. In R. Jakobson (Hrsg.), *On translation* (S. 232–239). Harvard University Press.
- Kalthoff, H. (1996). Das Zensurenpanoptikum: Eine ethnographische Studie zur schulischen Bewertungspraxis. *Zeitschrift für Soziologie*, 25(2), 106–124.
- Kalthoff, H., & Dittrich, T. (2016). Unterscheidung und Härtung. Bewertungs- und Notenkommunikation in Lehrerzimmer und Zeugniskonferenz. *Berliner Journal für Soziologie*, 26(3), 459–483.
- Kalthoff, H., & Kölsch, F. (2025a). Before (e)valuating: student testing in History and Engineering. *British Journal of Sociology of Education*, 46(3), 323–340.
- Kalthoff, H., & Kölsch, F. (2025b). Im Maschinenraum der Differenzierung. *Österreichische Zeitschrift für Soziologie*, 50(1), 1–44.
- Kehm, B. M. (2001). Oral examinations at German universities. *Assessment in Education: Principles, Policy and Practice*, 8(1), 25–31.
- Kvale, S. (1972). *Prüfung und Herrschaft. Hochschulprüfungen zwischen Ritual und Rationalisierung*. Deutscher Studienverlag.
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38(1), 201–221.
- Latour, B. (2001). Eine Soziologie ohne Objekt? Anmerkungen zur Interobjektivität. *Berliner Journal für Soziologie*, 11, 237–252.
- Leonhard, T., Müller, A., & Herzmann, P. (2025). Worauf es wirklich ankommt. Mündliche Prüfungen zwischen akademischem Anspruch und beruflicher Eignung. In T. Leonhard (Hrsg.), *Lehrer: in werden. Trajektorien in den Lehrberuf*. Klinkhardt.
- Luhmann, N. (1998). *Die Gesellschaft der Gesellschaft*. Suhrkamp.
- Luhmann, N. (2002). *Das Erziehungssystem der Gesellschaft*. Suhrkamp.
- Luhmann, N., & Schorr, K. E. (1988). *Reflexionsprobleme im Erziehungssystem*. Suhrkamp.
- Mead, G. H. (1973). *Geist, Identität und Gesellschaft*. Suhrkamp.
- Meer, D. (1998). *Der Prüfer ist nicht der König. Mündliche Abschlussprüfungen in der Hochschule*. Max Niemeyer Verlag.
- Meer, D. (2019). Der Prüfer ist immer noch nicht der König. Nachdenken über mündliche Hochschulprüfungen der Gegenwart. In T. Tyagunova (Hrsg.), *Studentische Praxis und universitäre Interaktionskultur. Perspektiven einer praxeologischen Bildungsforschung* (S. 173–199). VS.
- Meier, F., Peetz, T., & Waibel, D. (2016). Bewertungskonstellationen. Theoretische Überlegungen zur Soziologie der Bewertung. *Berliner Journal für Soziologie*, 26(3–4), 307–328.
- Mummelthey, S., & Rabenstein, K. (2024). Vergleichen, Differenzieren, Kategorisieren und das Problem der Fairness. (Sport-)Unterricht als instruktiver Fall einer Soziologie der Bewertung. In R. Schmidt, D. Kempf & M. Weigelin (Hrsg.), *Leistungsvergleiche und evaluative Praktiken. Sport als instruktiver Fall für eine Soziologie der Bewertung* (S. 145–172). transcript.
- Muniesa, F., & Helgesson, C. F. (2013). Valuation studies and the spectacle of valuation. *Valuation Studies*, 1(2), 119–123.
- Nicolae, S., Endreß, M., Berli, O., & Bischur, D. (Hrsg.). (2019). *(Be)Werten. Beiträge zur sozialen Konstruktion von Wertigkeit*. VS.
- Prahl, H.-W. (1976). *Hochschulprüfungen, Sinn oder Unsinn? Sozialgeschichte und Ideologiekritik der akademischen Initiationskultur*. Kösel.
- Prange, K., & Strobel-Eisele, G. (2006). *Die Formen des pädagogischen Handelns*. Kohlhammer.

- Prior, L. (2003). *Using documents in social research*. SAGE.
- Ricken, N., & Reh, S. (2017). Prüfungen – Systematische Perspektiven der Geschichte einer pädagogischen Praxis. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 63(3), 247–258.
- Schmidt, R. (2022). Toward a culture-analytical and praxeological perspective on decision-making. *Human Studies*, 45, 653–671.
- Solem, M. S., Landmark, A. M. D., Stokoe, E., & Skovholt, K. (2024). Assessment in practice: achieving joint decisions in oral examination grading conversations. *Scandinavian Journal of Educational Research*, 68(7), 1522–1539.
- Sommer, A. (2018). Mündliche Einzelprüfung. In A. Gerick, A. Sommer & G. Zimmermann (Hrsg.), *Kompetent Prüfungen gestalten* (S. 140–143). Waxmann.
- Sowada, M., & Peetz, T. (2019). Schulen beurteilen. Grammatik und Pragmatik der Bewertung in Schulsinspektionsteams. In C. Imdorf, R. J. Leemann & P. Gonon (Hrsg.), *Bildung und Konventionen* (S. 205–227). VS.
- Strauss, A. L. (2004). Analysis through microscopic examination. *Sozialer Sinn*, 5(2), 169–176.
- Syverud, M. S., & Prøitz, T. S. (2025). Co-assessment at the boundary: judges and advocates grading students' performances in oral exams. *Education Inquiry*, , 1–19.
- Tyagunova, T. (2021). Prüfungskompetenz: Interaktive Steuerung von Wissensdarstellungen in mündlichen Universitätsprüfungen. *Sozialer Sinn*, 22(1), 185–221.
- Tyagunova, T. (2023). Heterogene Wissensordnungen: Modi des Umgangs mit Wissen in mündlichen Lehramtsprüfungen. *Zeitschrift für Pädagogik*, 5, 599–617.
- Verheyen, N. (2018). *Die Erfindung der Leistung*. Hanser.
- Verkuyten, M. (2000). School marks and teachers' accountability to colleagues. *Discourse Studies*, 2(4), 452–472.
- Vismann, C. (2002). Action Writing: Zur Mündlichkeit im Recht. In F. Kittler, T. Macho & S. Weigel (Hrsg.), *Zwischen Rauschen und Offenbarung. Zur Kultur- und Mediengeschichte der Stimme* (S. 133–151). Akademie Verlag.
- Wildt, B., & Redecker, D. (2008). Prüfungscoaching' – zur Vorbereitung von Studierenden auf die mündliche Prüfung: Ein Werkstatt-Konzept. In S. Dany, B. Szczyrba & J. Wildt (Hrsg.), *Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen* (S. 200–217). Bertelsmann.
- Wittgenstein, L. (1984). *Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchungen*. Suhrkamp.

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.