

Im Maschinenraum der Differenzierung

Lehren – Prüfen – Bewerten in der Geschichtswissenschaft

Herbert Kalthoff  · Fabian Kölsch 

Angenommen: 9. September 2025 / Online publiziert: 15. September 2025
© The Author(s) 2025

Zusammenfassung Mit einem anerkennungstheoretischen Begriff der Leistung untersucht der Aufsatz die Phasen des Lehrens, Prüfens und Bewertens als transssituative Formen universitärer Instruktion und Evaluation. Er dokumentiert die Vergegenwärtigung der je anderen Phase durch zeitliche Vor- und Rückgriffe, analysiert deren situativen Vollzug sowie ihre soziale Bedingtheit durch die lokale Konstellation. Im Zentrum steht die Verwandlung von Tätigkeiten und Aktivitäten der Student:innen über Zeiten, Personen und Räume hinweg in individuell zurechenbare Leistungen – durch de/legitimierende Sprechakte in der Lehre, schriftliche Klausuraufgaben, Formen der Klausur und die Be/Wertung mündlicher oder schriftlicher Arbeiten oder Beteiligungen. Die zugeschriebene und fabrizierte Leistung resultiert aus kollaborativen Bewertungsketten, die sozial ratifiziert und abgesichert werden.

Schlüsselwörter Soziologie der Be/Wertung · Bildungssoziologie · Ethnografie · Geschichtswissenschaft · Transssituative Konstellation

✉ Herbert Kalthoff · Fabian Kölsch
Institut für Soziologie, Johannes Gutenberg-Universität, Jakob-Welder-Weg 12, 55128 Mainz,
Deutschland
E-Mail: herbert.kalthoff@uni-mainz.de

Fabian Kölsch
E-Mail: f.koelsch@uni-mainz.de

In the engine room of differentiation

Teaching—Testing—Assessing in history

Abstract Using a recognition-theoretical concept of performance, this article examines the phases of teaching, testing, and assessment as trans-situational forms of university instruction and evaluation. It documents the evocation of each respective phase through temporal anticipations and retrospections, analyzes their situational enactment, and explores their social conditionality within local constellations. At the core is the transformation of students' activities and actions across time, individuals, and spaces into individually attributable performances—through legitimizing or delegitimizing speech acts in teaching, written exam tasks, exam formats, and the evaluation of oral or written work or participation. The attributed and constructed performance emerges from collaborative chains of evaluation that are socially ratified and secured.

Keywords Sociology of e/valuation · Sociology of education · Ethnography · History · Trans-situational constellation

1 Einleitung

Die Unterscheidung, Kategorisierung oder Klassifizierung von Menschen in sozialen Begegnungen oder institutionellen Kontexten ist seit vielen Jahren Gegenstand der sozial- und kulturwissenschaftlichen Differenzierungs- und Bewertungsforschung – sei es das Geschehen im Gerichtssaal (Heffer 2007) oder im Gesundheitswesen (Lambrix 2022), seien es die katholische Beichte (Hahn 1982), alltägliche Geschmackentscheidungen (Bourdieu 1984) oder Praktiken der Menschensortierung (Hirschauer 2021). Ein paradigmatischer Fall einer systematischen und wiederholten Klassifikation von Menschen ist die universitäre Bewertung von Student:innen.¹ Während die Forschungspraxis seit vielen Jahren Gegenstand wissenschaftssoziologischer Untersuchungen ist (klassisch: Latour und Woolgar 1979; Knorr-Cetina 1981), wurde das vom universitären Lehrpersonal praktizierte Lehren, Prüfen und Bewerten bislang wenig erforscht (siehe Kingston 2001; Pabian 2014; Decker und Hoffmann 2023). Diese Asymmetrie des Forschungsinteresses lässt sich u. a. auf den Umstand zurückführen, dass Universitäten in aller Regel als Organisationen innerhalb des Wissenschaftssystems (siehe Frickel und Moore 2006) und nicht als Orte einer spezifischen Bewertungspraxis konzipiert werden. Nun ist in den vergangenen Jahren die universitäre Bewertungspraxis verstärkt in den Fokus der sozialwissenschaftlichen Forschung gerückt, wobei nicht wenige Studien messtheoretisch ausgerichtet sind. Aus ihrer Sicht generieren Hochschulbewertungen (mehr oder weniger)

¹ Dieser Aufsatz verwendet den Begriff der *Prüfung* für formale Situationen (etwa der Leistungsüberprüfung in der Klausur), den Begriff der *Wertung* für die qualitative Einschätzung einer Antwort oder Klausur und den der *Bewertung* für die darauf aufbauende, in eine Entscheidung überführte Benotung (Krüger und Reinhart 2016; Wiggins 1998). Zur begrifflichen Differenz von Unterscheidung, Kategorisierung und Klassifikation siehe Hirschauer (2021, S. 157ff.).

objektive und valide Ergebnisse und ermöglichen damit eine 1:1-Abbildung der Leistungen der Student:innen (Ioannidou 1997; von Stumm et al. 2011; Nieminen et al. 2023). Ein wesentliches Merkmal dieses Ansatzes ist die Annahme, dass die Bewertenden neutrale Instanzen eines Prozesses der institutionalisierten Klassifikation von Menschen sind. Andere Studien stellen diese Perspektive in Frage und betonen den kontingenten Charakter der Bewertung, indem sie das Bewerten als soziale Praxis untersuchen. So wird bspw. gezeigt, wie (Alltags-)Theorien lokale Bewertungsprofile formen (Stratton und Mannix 2005), wie das jeweilige Selbstverständnis des Faches das Bewertungsgeschehen rahmt (Tsarouha 2019), wie das Hintergrundwissen über Student:innen im Benotungsprozess abgerufen wird (Holstein 1983), aber auch wie die universitäre Bewertung das Selbstbild der Student:innen tangiert (Nieminen 2021; Nieminen und Yang 2024). Andere Studien arbeiten heraus, dass Hochschul-lehrer:innen auf den (institutionellen) Ruf nach größerer Transparenz vor allem mit Multiple-Choice-Klausuren reagieren, die sie fach- und situationsspezifisch ausgestalten (Jedemark und Londos 2021), und darüber hinaus ihre Bewertung auch an Diskursen orientieren, in denen gesellschaftliche Akteure ihre Erwartungen an den universitären Abschluss zum Ausdruck bringen (Forde-Leaves et al. 2023). Tyagunova (2019) zeigt, wie das Lernen mit Altklausuren u.a. die Interaktion zwischen Student:innen und Lehrenden verändert. Am Fall der mündlichen Prüfung arbeitet Meer (1998; 2019) heraus, wie das *Enactment* des Wissens durch Student:innen mit ihrer sozialen Positionierung verhandelt wird, während Saner (2019) darlegt, wie die Prüfer:innen die Passung von Student:innen und Hochschule in sachliche Bewertungen übersetzen. Schließlich fragt Rabenstein (2022) nach den sozialen Folgen der universitären „Dauerbewertungen“, die sich in eine Vielzahl unterschiedlicher Prüfungsformen ausdifferenzieren und so das lokale Geschehen verändern.²

Bislang vorliegende soziologische Studien betonen drei Funktionen von Prüfungen. *Erstens* wird die Herrschaftsfunktion von Prüfungen hervorgehoben. So unterscheidet Kvale (1972) zwischen manifesten und latenten Formen von Herrschaft: Manifeste Formen kontrollieren und regulieren die Aneignung orthodoxen Wissens, latente Formen selektieren akademische Eliten, kodifizieren das Prüfungsritual und passen Student:innen geltenden Normen an (siehe auch Pahl 1976). Für Bourdieu (1992, S. 353ff.) wiederum sind Prüfungen der zentrale Mechanismus einer sozialen wie kulturellen Reproduktion, die Klassenverhältnisse stabilisiert. *Zweitens* wird auf die didaktische Funktion von Prüfungen verwiesen. Prüfungen stellen Lernstandserhebungen als eine auf Dauer gestellte Feedback-Schleife dar (Flehsig 1976; Trembl 2000; Prange und Strobel-Eisele 2006). *Drittens* diskutieren Studien verschiedener theoretischer Provenienz die Funktion von Wissensüberprüfungen für die Selbstreproduktion des universitären Systems (Stichweh 2013; Wagner et al. 2023). Bewertungen („Noten“) fungieren in systemtheoretischer Perspektive als Grundlage für studentische Karrieren, da sie eine Arbeit oder Tätigkeit als „Leistung“ sichtbar machen (Luhmann 2002; Albrecht 2024).³ Ferner sind sie ein Medium der „Universi-

² Für Universitäten verweisen Bewertungen nicht nur auf Ergebnisse von Prüfungen, sondern in ihrer aggregierten Form von Kennzahlen auf die Performance ihrer Institute (Stevens 2007).

³ Auf die Frage, wie studentische Karrieren und universitäre Bewertungen auf den Übergang in das professionelle Leben vorbereiten, gehen wir in diesem Aufsatz nicht ein; siehe aber Schröer et al. (2013).

tätserziehung“, da sie eine auf Dauer gestellte Kommunikation zwischen Lehrbetrieb und Prüfungsorganisation der Universitäten ermöglichen (Stichweh 2013). Hieran anschließend betonen Wagner et al. (2023), dass die Bewertungspraxis als Effekt der universitären Organisation zu betrachten ist, während ein situationistischer Ansatz die Selbstständigkeitsfiktion der Bewertenden zu reifizieren droht.

Trotz der skizzierten Überlegungen und Forschungsergebnisse bleibt empirisch wie konzeptionell unklar, wie die Phasen von Lehren, Prüfen und Bewerten einander bedingen und sich in ihrem praktischen Eigensinn dennoch aufeinander beziehen. Der Aufsatz verfolgt daher zwei Ziele: Es soll expliziert werden, wie die Bewertung von Student:innen fabriziert wird. Hierzu schließt er an neuere bewertungssoziologische Forschungen an, die die transsituative Formatierung von Bewertungen sowie die soziale Praxis des Prüfens und Bewertens beschreiben und analysieren (bspw. Meier et al. 2016; Nicolae et al. 2019; Meier und Peetz 2021). Die Herausforderung besteht aus Sicht dieses Aufsatzes darin, die organisatorische Konstellation und den situativen Vollzug des Lehrens, Prüfens und Bewertens dialektisch aufeinander zu beziehen. In diesem Zusammenhang zeigen schon frühe Studien, dass auch das solitäre Bewerten am Schreibtisch transsituativ vermittelt ist, etwa wenn Lehrende diffuse Eindrücke aus der Lehre in die Bewertung von Arbeiten einfließen lassen (Holstein 1983). Wir knüpfen damit an das Konzept der Bewertungskonstellation an (Meier et al. 2016), geben ihm aber eine empirisch-analytische Wendung. Ferner untersucht der Aufsatz am Beispiel der geschichtswissenschaftlichen Lehre an deutschen Universitäten die Verschränkung ihrer Elemente (Lehren, Prüfen, Bewerten) durch die transsituative Mobilisierung und Zirkulation von Dokumenten, Wissensobjekten und institutionellen Rollen. Hierzu substituiert er eine zeitlich-lineare Sicht auf das Lehren, Prüfen und Bewerten als eine simple Chronologie von Ereignissen durch ein rekursiv-reflexives Verhältnis, und zwar als eine beobachtbare Praxis des rückgreifenden Vorgriffs oder des vorgeifenden Rückgriffs.⁴ Hiermit fassen wir ein Handeln, das bspw. Prüfungsfragen zu Beginn der Vorlesung aus dem Vorlesungsskript generiert bzw. dann Themen in der Vorlesung behandelt, wenn sie sich (später) gut prüfen lassen. Hierzu gehört auch die Praxis, aus alten Prüfungsfragen neue zu entwickeln. Zu fragen ist also, wie die Phasen des Lehrens, Prüfens und Bewertens einander informieren und füreinander gleichursprünglich sind. Wie also ist in der Praxis des Lehrens, Prüfens und Bewertens die je andere Phase gegenwärtig, und wie wird sie bearbeitet? Mit diesem Design will der Aufsatz das für Soziolog:innen in Hochschulen so selbstverständliche Tun in einen soziologischen Gegenstand verwandeln (siehe Weber 1988, S. 502).

Die o.g. Phasen implizieren, dass dem Prüfen und Bewerten etwas vorausgeht, zeitlich wie gegenständlich, und zwar das wiederholende Einüben bzw. Einstudieren u. a. von körperlichen Bewegungen, von Fachwissen sowie des schriftlich-mündlichen Gebrauchs des Fachvokabulars. Diese kognitiv wie praktisch ausgerichteten Unterrichtungen und Einübungen finden in je spezifischen Lehrformaten statt und werden von Selbst- und Fremdeinschätzungen begleitet: Es sind situative Ad-hoc-Kommentare (etwa in Seminaren), die den Wert einer Äußerung oder einer Praxis,

⁴ Auf die hier zugrunde liegende Zeittheorie gehen wir im Rahmen dieses Aufsatzes nicht weiter ein; siehe etwa Heidegger (2004), Landwehr (2016).

auf die sie sich richten, einordnen. Ferner gehen der Bewertung auch jene Bereiche voraus, in denen das, was bewertet wird (schriftliche Arbeiten, mündliche Antworten), hergestellt und vollzogen wird. Diese Situationen werden klassischerweise als mündliche oder schriftliche Prüfungen verstanden, die sich von Lehrsituationen durch ihre verstärkte Organisation, Verrechtlichung sowie durch eine rigide Interaktionskontrolle unterscheiden. Sie werden in Zeitintervallen relevant, systematisch organisiert und am Ende der Semester in Form schriftlicher Arbeiten (Klausuren, Hausarbeiten etc.) oder mündlicher Verfahren durchgeführt. Schließlich gehören u. a. die Koordination und Kategorisierung der Beteiligten, administrative Listen und (strikte) Verhaltensregeln, Zeitangaben und Infrastrukturen (Räume, PCs, Personal etc.), rechtliche Kodierungen sowie Klausuraufgaben, die die behandelten Lehrinhalte selektieren und in einer spezifischen Weise darstellen, zur Konstellation des Prüfungsgeschehens.

In Bezug auf die Transsituativität gehen wir von drei Praxiskontexten aus, die für sich existieren, aber temporär und wiederholt aufeinander bezogen sind und Informationen etwa über Prüfungstermine, Zeiträume oder Personen (Lehrpersonal, Student:innen etc.) austauschen (s. o.). Materialisiert wird diese kollaborative Kommunikation durch Dokumente (Formulare, Unterlagen etc.), die in der Prüfungsorganisation entwickelt und entworfen, an die Lehrenden weitergereicht und schließlich von diesen verwendet werden. Als „Format eines (...) realisierbaren Schemas, eines Lückentextes“ (Vismann 2000, S. 160) greifen sie unterschiedlich intensiv in das Prüfungsgeschehen ein, indem sie es formieren und formalisieren. Sie werden relevant, verwendet und ausgefüllt – und gehen an die universitäre Prüfungsorganisation zur weiteren administrativen Verarbeitung der dokumentierten Ergebnisse zurück (etwa in Form von Protokollen). Gegen eine Perspektive, die von einem Durchgriff der Organisations- auf die Interaktionsebene ausgeht (etwa Wagner et al. 2023), vertreten wir eine praxistheoretisch informierte Wissenssoziologie: Wir nehmen an, dass diese Praxiskontexte in ihrem eigenen Recht und in ihrer eigenen Logik existieren, aber kollaborieren, ohne sich gegenseitig in ihrer Praxis zu determinieren oder ihre Eigenständigkeit zu verlieren. Transsituativität besteht somit in diesem reziproken und temporären Verweisungszusammenhang, durch den das Prüfungsgeschehen geschleust wird (siehe etwa Meier et al. 2016).

Grundlage des Aufsatzes ist ein Anerkennungstheoretischer Leistungsbegriff. Dieses Verständnis setzt voraus, dass eine Arbeit, eine Praxis, ein Geschehen, ein Projektantrag oder ein Individuum etc. von beobachtenden Dritten, die mit der symbolischen Macht einer Institution ausgestattet sind, bewertet wird. Im Falle hochschulischer Organisationen sind dies folgenreiche Bewertungen, da sie über biografische Zukünfte der Studierenden mitbestimmen. Wir gehen daher davon aus, dass Leistung keine der Bewertung vorgängige Entität ist, die von den Bewertenden vorgefunden wird, sondern Leistung entsteht erst durch den Akt der Bewertung selbst, d. h. durch die (negative) Anerkennung der Bewertenden. Die Resultate dieser Prozesse (etwa Prädikate, Ziffernnoten) sind in einem hohen Maße von Kontingenz geprägt, die u. a. von Maßstabsverschiebungen und Neigungen der Bewertenden bestimmt wird (ausführlich hierzu Banerjee et al. 2025; Kalthoff 2017; Kalthoff und Kölsch 2025⁵).

⁵ Zu einer Theorie von Anerkennungsverhältnissen siehe Honneth (1992).

Der Aufsatz ist wie folgt strukturiert: Zunächst stellen wir das methodische Design der Studie vor und reflektieren die Fallspezifität des Gegenstandes (2.). Im sich anschließenden Analysekapitel rekonstruieren wir die verteilte Praxis der universitären Bewertung. Der Aufsatz fragt, wie in der universitären Lehre Wissen interaktiv hervorgebracht und in Klausuraufgaben formalisiert wird (3.). Auf der Grundlage von Prüfungsbeobachtungen analysieren wir, wie die Individualisierung der Prüflinge über eine dialektische Verschränkung von Kontrolle und Konzentration durchgesetzt wird (4.). Auf der Grundlage von Beobachtungen mündlicher Prüfungen und Klausurkorrekturen rekonstruieren wir zwei Objektivierungsmodi: die maschinell feststellende und die sozial aushandelnde Form (5.). Zum Abschluss skizziert der Aufsatz drei Anschlussmöglichkeiten für eine transsituative Soziologie des Bewertens (6.).

2 Die empirische Forschung

Der Aufsatz stützt sich auf ethnografische Forschungen in geschichtswissenschaftlichen Instituten deutscher Universitäten, ihrer Lehre und ihren Prüfungen. Das Studium der Geschichtswissenschaft zeichnet sich durch eine Orientierung an der inhaltlichen Aufarbeitung, Kontextualisierung und kritischen Analyse von historischen Quellen aus. Angesichts der großen Diversität des Quellenmaterials deckt die geschichtswissenschaftliche Lehre ein breites Themenspektrum ab: von der Analyse von Schriftzeugnissen (Berichte, Tagebücher etc.), Fotografien oder Videoaufnahmen über das Einüben kultureller Fertigkeiten (etwa das Lesen fremder Schriftarten) bis zum Besuch von Gedenkstätten, Museen und Archiven. In Bezug auf die Wissensüberprüfung, die auf die universitäre Lehre folgt, wird angesichts großer Kohorten von Student:innen eine Spannung innerhalb des Faches deutlich, die sich in einem eher traditionellen Bildungsverständnis auf der einen und einem formalisierten Klausurmanagement auf der anderen Seite artikuliert.

Empirisch haben wir in einem Zeitraum von sechs Wochen zehn Lehrveranstaltungen (Vorlesungen, Seminare, Tutorien, Übungen, Kolloquien usw.), drei schriftliche Prüfungen und sechs mündliche Modulabschlussprüfungen an einem Institut für Geschichtswissenschaft teilnehmend beobachtet. Weiter haben wir neunundzwanzig ethnografische Interviews als „friendly conversation“ (Spradley 1979) mit wissenschaftlichen Mitarbeiter:innen sowie neun leitfadengestützte Interviews mit Historiker:innen (Professor:innen; Postdocs, Doktorand:innen) verschiedener Gebiete (Mittelalterliche, Neuere und Neueste sowie Außereuropäische Geschichte) verschiedener Universitäten geführt. Ethnografische Interviews, mit denen aktuelle Beobachtungen verbalisiert und weitere Informationen und Deutungen generiert werden konnten, boten sich etwa vor allem im Anschluss an Seminare, bei der Prüfungsvorbereitung und beim Bewerten von Klausuren an – sei es in Seminarräumen oder in den Büros der Mitarbeiter:innen. Sie ermöglichten es vor allem, aktuelle Beobachtungen zu verbalisieren und weitere Informationen und Deutungen zu generieren. Die universitäre Lehre und ihre Prüfungen produzieren darüber hinaus eine ganze Reihe (digitaler) Schriftstücke, die das Geschehen u. a. strukturieren und auch als Selbstdarstellung fungieren (etwa von Lehrenden oder Instituten). Entsprechend

haben wir verschiedenste Dokumente des Feldes (Selbstdokumente) gesammelt bzw. fotografiert (Fremddokumente) und dokumentenanalytisch untersucht.

Die Interviews wurden transkribiert und die ethnografischen Protokolle redigiert, und beide Datensorten wurden offen kodiert (Strauss 1998; Emerson et al. 1995). Wir verstehen offenes Kodieren als einen Prozess der Übersetzung von empirischen Beobachtungen in analytische Kategorien. Das heißt, dass die Analyse der empirischen Daten mit theoretischen oder analytischen Überlegungen ins Gespräch gebracht wird, ohne die Daten theoretischen Überlegungen anzupassen. Das offene Kodieren wurde im Forschungsprojekt und in Datasessions des Instituts umgesetzt, wobei wir emische wie analytische Begriffe als Codes verwendet haben. Die von uns als (potenziell) zentral für die Analyse angesehenen Codes haben wir an anderen Protokollen und Interviews geprüft und präzisiert. Auf diese Weise kamen bspw. die Codes „Darstellungsmedien“, „Autorschaft“, „Kontrolle“, „Hermeneutik der Quellen“ in den analytischen Fokus. Der Prozess des Kodierens wurde immer wieder durch das Schreiben von *analytical notes* bzw. Memos unterbrochen: ersten konzeptionellen Ordnungen und Kombinationen der erschlossenen Codes. Sie dienten einer Systematisierung der Einsichten, der Generierung weiterer Fragen und der Formulierung weiterführender Überlegungen. Die Analyse der gesammelten Dokumente konzentrierte sich auf drei Dimensionen: auf den Inhalt des Dokuments (was?), auf seine Verfügbarkeit (wer ist Autor:in, Leser:in; wie zirkuliert das Dokument?) und auf die Vermittlung (wie stellt es etwas dar?) (Atkinson und Coffey 2011; Prior 2004). Die Feldforschung als auch die Analyse der Daten erforderten dabei ein zumindest partielles Verständnis der historischen Inhalte und des fachlichen Vorgehens, was teils durch die o.g. Interviews mit den Teilnehmer:innen, teils durch weitere Lektüre (partiell) erreicht werden konnte.

3 Lehrszenen. Elizitieren und Vorbereiten

Ein Effekt der Bologna-Reformen und der Modularisierung des Studiums ist eine stärkere Verschränkung von Lehre und Prüfung.⁶ Im Folgenden gehen wir dieser Verschränkung am Beispiel von Vorlesungen nach. Trotz zyklisch wiederkehrender Abgesänge hat die Vorlesung im Curriculum der Geschichtswissenschaft einen festen Platz. In der sozialwissenschaftlichen Literatur werden Vorlesungen zumeist als ein Lehrformat betrachtet, das von den Prüfungsprozessen weitgehend entkoppelt ist (Goffman 1981; Bourdieu 1985; Masschelein und Simons 2013). Allerdings dokumentieren unsere Beobachtungen wiederkehrende Momente, in denen Lehre und Prüfung in einem Verweisungszusammenhang stehen. Zur Verbindung von Lehrinhalt und Klausurstoff sagte ein Professor:

⁶ Einige Autor:innen deuten diese Entwicklung als starke Verschulung des Studiums, andere verbinden damit die Vorstellung einer erhöhten Berechen- und Kontrollierbarkeit von Bildungsprozessen (Pfaller 2010; Kühl 2012, S. 342).

„Ich habe das ausformulierte Manuskript immer als Rückfalloption da, dass ich weiß, was ich vermittelt habe, das ist mir ganz wichtig, weil das klausurbehaftet ist. Und ich muss mir am Ende sicher sein, dass das, was ich in der Klausur abfrage, dass ich das auch erzählt habe.“⁷

In aller Regel fungiert ein Vorlesungsskript als standardisierter Vortrag, der Abweichungen nicht unmöglich, aber unwahrscheinlicher werden lässt. Es fungiert, wie das Zitat verdeutlicht, als inhaltliche Rückversicherung *nach* der Vorlesung, damit eine Passung zwischen mündlichem Vortrag und Klausurfragen hergestellt werden kann. Demzufolge ist das Manuskript als Erinnerung an das mündlich Vorgetragene eine absichernde Vorlage für die Klausur. Umgekehrt können Klausuren Vorlesungen präfigurieren, etwa dann, wenn sich das Manuskript an Klausuraufgaben orientiert, die sich gut stellen lassen. Neben Manuskripten wird auch ausgiebig von Power-Point-Folien (i.W.: PP-Folien) Gebrauch gemacht, die den Student:innen digital zur Verfügung gestellt werden. Diese Dokumentensorte steht für eine Kommunikation, die von der physischen Anwesenheit der Student:innen und von der konkreten Performanz der Vorlesung absieht. Darüber hinaus nehmen solche Folien eine Selektion der Inhalte vor, indem sie sie auf zu Lernendes reduzieren und damit zugleich ihre diesbezügliche Relevanz markieren. Dahinter steht die Annahme, dass jeder Sprechakt der Vorlesenden und jede Information auf einer Folie potenziell Gegenstand der Klausur werden können. Dies zeigt sich auch in autoritativen Sprechakten, die öffentlich markieren, was relevantes Prüfungswissen ist und was nicht; entsprechend werden selektierte Teile der Vorlesung zu Prüfungswissen. Ein Beispiel:

Der Dozent steht vorne vor den Bänken (...) und sagt: „An der Stelle würde ich nun gerne in die Bulle *Exsurge siquidem*⁸ reingucken und Sie fragen, die ganz klassischen Fragen, wer, wie, was wann, wo? (...).“ Im Raum herrscht Stille. Einige der Student:innen blicken gemeinsam auf das vorab ausgeteilte Dokument; andere schauen auf ihre Bildschirme; in den letzten Reihen wird getuschelt. Nach kurzer Zeit hakt der Dozent nach: „Wer macht den Anfang?“ Wieder herrscht Stille. Dann sagt er: „Fangen wir mal bei dem ersten an, also bei dem Wer.“ Ein Student aus den vorderen Reihen hebt seine Hand, wird vom Dozenten aufgefordert und sagt kurzatmig: „Urheber ist Bermudo.“ Der Dozent bestätigt die Aussage durch zwei zustimmende „Mhms“, um dann ebenso kurzsilbig weiter zu fragen: „An wen gerichtet?“ Wieder eine Meldung von vorne: „An Marco Pablos gerichtet. Der Herrscher von Castels.“ „Sehr gut!“, meint der Dozent und schließt eine Frage an: „Können Sie uns noch die Lebensdaten berichten?“. Der Student antwortet: „Ehm (2) Nein.“ Der Dozent: „Okay ((lacht und wendet sich zur Menge)). Wenn Sie die Quelle in der Klausur beschreiben, dann sollten Sie das tun.“

Der Ausschnitt zeigt eine für die Geschichtswissenschaft charakteristische Vorgehensweise. Im Zentrum steht ein historisches Dokument, welches als Quelle auf

⁷ Der Aufsatz nutzt folgende Transkriptionszeichen: (...): Auslassungen; (2): Sprechpause (in Sek.); [lachen]: Kommentar des Protokollanten.

⁸ Alle orts- und inhaltsbezogenen Angaben im Text wurden pseudonymisiert bzw. verfremdet.

seine Entstehungsbedingungen hin befragt wird. Vollzogen wird hier *erstens* keine hermeneutische Analyse des Dokuments selbst, vielmehr wird am Dokument die Arbeitsweise historischen Forschens aufgezeigt. Deren (wiederholtes) Einüben – und hierzu gehören auch die „klassischen Fragen“ – ist ein relevanter Bestandteil des geschichtswissenschaftlichen Curriculums. *Zweitens* wird die Vorlesung in ein Unterrichtsgespräch mit eindeutiger Rollenverteilung von Fragenden (der Dozent) und Antwortenden (Student:innen) verwandelt – eine Modulation, die später im Interview als Ergebnis didaktischer Erwägungen ausgewiesen wird. Deutlich wird an diesen (schulisch induzierten) Frage-Antwort-Kommentar-Sequenzen (hierzu McHoul 1978) auch, dass eine Bewertung der studentischen Antworten impliziert ist, die zunächst folgenlos bleibt (auch Luhmann 1996: 288). *Drittens* ist der Hinweis auf die historischen Lebensdaten insofern relevant, als er die Anforderung formuliert, diese Daten zu kennen und im Rahmen der Klausur wiedergeben zu können. Hierin zeigt sich also einerseits die Erwartung, das Sprachspiel des Faches zu kennen, andererseits wird deutlich, wie die Inhalte der Vorlesung in den Sprechakten der Vortragenden auf die Klausur bezogen sind.

Wendet man sich von dieser Überlegung ausgehend nun der Klausur zu, dann zeigt sich, wie die Prüfung an den Lehrstoff anschließt und diesen formalisiert. In der Klausur wird der oben erwähnte historische Fall als Klausurfrage wie folgt formuliert:

„Die Bulle Exsurge siquidiem von 1181 spielt für die kastilianische Vorherrschaft eine zentrale Rolle. Bringen Sie die Bedeutung der Bulle in Stichworten auf den Punkt (8P).“

In ihrer ganzen Knappheit abstrahiert die Formulierung der Klausurfrage von der konkreten Vorlesungssituation und ruft sie doch mit der Bezeichnung der Bulle wieder auf. Auch unterscheidet sich die Klausuraufgabe, die auf die Produktion adäquater Stichworte abzielt, von der tastenden Annäherung in der Vorlesung. Während im Vorlesungsgespräch durch Fragen, Antworten und Rückfragen das Soziale als Interaktion mobilisiert wird, reduziert die Klausur im Modus von Frage-Antwort alles auf das individuelle Wissen der Student:innen. Auffällig ist in diesem Fall, dass die geschlossene Frageformulierung die Komplexität des Falls (erneut) reduziert. So wird die Frage nach der Bedeutung der Bulle, die hier durch einige Informationen ergänzt wird (Jahreszahl, Region), nicht offen, sondern bezogen auf einen spezifischen historischen Kontext gestellt und richtet so schon die Antwort der Student:innen aus. Ferner bestärkt die Klausuraufgabe die Wissensautorität des Lehrpersonals: Relevant ist nicht das, was die Student:innen gehört und gegebenenfalls mitgeschrieben haben, sondern das, was die Lehrkraft gesagt und welche Themen sie für die Prüfung ausgewählt hat (Elwick 2023). Schließlich schreibt die Nennung der maximalen Punktzahl der Klausur ein kalkulatives Prinzip ein und macht den je spezifischen Wert der Fragen transparent (Vatin 2013). Mit Blick auf die Verschränkung von Lehren, Prüfen und Bewerten rahmt diese Valorisierung der Klausuraufgaben die später erfolgende Be/Wertung und organisatorische Objektivierung durch das universitäre Personal (siehe Abschnitt 5). Lehre und Prüfung stehen somit in einem organisationalen und fachlichen Verweisungszusammenhang, diffe-

renzieren sich jedoch hinsichtlich ihrer Modi und Adressat:innen der Zurechnung von Wissen und Nicht-Wissen.

4 Prüfungsszenen. Kontrolle und Autorschaft

Bisher hat der Aufsatz Lehre und Prüfung hinsichtlich ihrer Verschränkung resp. Verknüpfung in der universitären Lehrpraxis untersucht; er wechselt jetzt zur Analyse des Prüfungsgeschehens. Das damit verbundene Handlungsproblem des universitären Lehrpersonals besteht darin, sicherzustellen, dass das Wissen der Student:innen, das diese in der Klausur dokumentieren, als deren originäre Leistung gelten werden kann (oder eben nicht). Ein Auszug aus einem Beobachtungsprotokoll:

„Es ist still im Hörsaal. Zu vernehmen ist nur das knarrende Geräusch von Stiften auf Holztischen. Während die studentischen Hilfskräfte Sarah und Jens durch die Reihen gehen und nervöse Blicke auf die Student:innen werfen, lehnt Ulf, der Professor, lässig an der Wand und tippt auf seinem Handy. Die Student:innen blicken konzentriert auf das Papier. Zwischen ihnen ist jeweils ein freier Platz. Wenn sie mal ihre Köpfe heben, dann schauen sie expressiv in die Luft oder an die Decke und machen dabei nachdenkliche, konzentrierte Posen. Ich erwische mich dabei, wie ich die kontrollierenden Blicke der Hilfskräfte imitiere und ebenfalls schaue, ob ich einzelne Student:innen bei auffälligen Blicken oder Bewegungen erwische. Ab und zu trifft sich mein Blick mit dem der Student:innen, woraufhin diese dann rasch wieder zurück auf ihr Papier schauen. Nach einer Weile tritt Sarah vor die Menge ans Pult und unterbricht die konzentrierte Stille des Raumes mit einer Ansage: ‚Bitte lassen Sie Ihre Augen auf Ihrem Papier. Das muss sonst als Täuschungsversuch gewertet werden.‘“

Der Ausschnitt zeigt eine typische Prüfungsszene, die verdeutlicht, wie die Interaktion zwischen Student:innen untereinander und zwischen Student:innen und Lehrpersonal auf ein Minimum beschränkt ist, sodass jede abweichende Bewegung und jeder verstohlene Blick auf das Papier eines anderen Prüflings auffällig wird, weshalb Student:innen darum bemüht sind, Interaktionslosigkeit darzustellen. Öffentliche Sprechakte des Lehrpersonals beschränken sich auf autoritative Hinweise und Sanktionsandrohungen. Gestützt wird diese Form der Aufsicht bzw. Kontrolle durch die Platzierung der Student:innen und die normierte Blickordnung (Foucault 1977). Dieses institutionelle Arrangement von räumlicher Ordnung, regulierten Blickkonventionen, kontrollierendem Lehrpersonal und mündlichen Hinweisen zielt darauf ab, die Zurechenbarkeit der Leistung auf individuelle Student:innen sicherzustellen (Meadmore 1993). Diese Zurechenbarkeit ist die Bedingung für den weiteren Prozess der Bewertung und eine ihrer zentralen Annahmen.

Darüber hinaus geht es auch darum, Bedingungen für ein konzentriert-aufmerksames Denken und Schreiben zu schaffen. In Anlehnung an die leibphänomenologische Diskussion zum Topos der Aufmerksamkeit lässt sich diese Dialektik von Kontrolle und Konzentration, von Achtung und Aufmerksamkeit spezifizieren: „Aufzumerken bedeutet, zugleich auf ein Objekt und auf sich selbst ausgerichtet zu sein“ (Rost 2017, S. 67). Eine Klausur aufmerksam zu bearbeiten, lebt in diesem Sinne von

einem Changieren zwischen Text und Selbst, das seinen leiblichen Ausdruck im Wechsel von konzentriertem Lesen und flüchtigem Flottieren der Augen und Gedanken findet. Gleichzeitig sind studentische Blicke in der Prüfung, wie erörtert, prekär und unterliegen einer Verdachtshermeneutik des Prüfungspersonals. Aus diesem Spannungsfeld erwächst die besondere Expressivität der Zurschaustellung der Blicke der Prüflinge, die nicht zu den Nachbarn, sondern an Decken und Wände gerichtet sind.

Schließlich gibt es neben diesen situativen *Performances* auch eine Kontrolle durch die Ordnung der Klausurfragen. Ein Beobachtungsprotokoll:

(...) Nach der Klausur kommen die Aufsichtspersonen zusammen und tauschen sich aus. Ulf äußert schelmisch-grinsend, dass die Student:innen aus der letzten Reihe ihre Augen „nicht bei sich halten“ konnten. Sarah bestätigt dies und reklamiert, dass diese „ganz sicher gespickt“ hätten. Daraufhin wieder Ulf: „(...) bei den Multiple-Choice-Aufgaben hätte man schon spicken können.“ Aber die wenigen MC-Punkte, so der Professor, können die Prüflinge auch nicht retten.

Die Unterstellung, dass einige Student:innen abgeschrieben haben könnten und damit die Annahme der Zurechenbarkeit möglicherweise unterliefen, wird durch die Gewissheit, dass diese Subversion sie „nicht retten“ wird, neutralisiert und als Gelassenheit aufgeführt. Es ist die geschlossene Liste der Klausurfragen (u. a. ihre Anzahl, ihre Form, ihr Gewicht), mit denen das Lehrpersonal einem potenziellen Abschreiben von Student:innen vorgreift und ihm damit zeitlich voraus ist. Deutlich wird hier, wie (Alltags-)Theorien über das Verhalten und Ansinnen von Student:innen in ein schriftlich gefasstes Dokument des Feldes eingebaut sind: Die Kontrolle des Aufsichtspersonals wird durch eine Delegation von Kontrolle an die Klausur und ihren Aufbau sowie die Gewichtung der Fragen ergänzt, um die Individualitätsannahme der Prüfung zusätzlich zu sichern.

Neben dieser sozialen Kontrolle gibt es aber auch formal-organisatorische Kontrollen in Form von Unterschriftenlisten. Am Umgang mit solchen Listen lässt sich beobachten, wie verschiedene Instanzen und Akteure an der Organisation der universitären Prüfung und Bewertung beteiligt sind. Universitäre Prüfungssituationen zeichnen sich dadurch aus, dass sie – in einem wesentlich stärkeren Maße als die universitäre Lehre – durch organisatorische Artefakte gerahmt werden (u. a. durch Klausurdokumente, Protokollvorlagen, Anwesenheitslisten), die ihrerseits anzeigen, dass Prüfungen ein Verwaltungsakt sind, dessen Durchführung zu dokumentieren ist. Verschiedene Studien (Goody 1977; Brückner und Wolff 2015; Fludernik 2016) haben die ordnende Funktion von Listen hervorgehoben. Sie kombinieren und ordnen Dinge (etwa Matrikelnummern, Unterschriften), stellen Übersichtlichkeit und Vergleichbarkeit her und machen hierdurch Differenzen leicht erkennbar. In Bezug auf universitäre Prüfungen übernehmen Listen weitere Funktionen. Sie dokumentieren etwa die physische Anwesenheit von Student:innen und das temporäre Verlassen des Klausurraumes, sind zugleich aber auch ein Mittel, um die geordnete Durchführung der Prüfung zu belegen (etwa im Falle studentischer Klagen), und stehen so für die rechtlich-administrative Seite des universitären Prüfungsmanagements.

Somit stehen Lehrende nicht allein für die juristisch-administrative Seite des universitären Prüfungsmanagements, sondern gehen bei den obligatorischen Kontrollen

durchaus eigensinnig mit diesen Listen um. Ein Ausschnitt aus einem Beobachtungsprotokoll:

„Ein Student kommt nach vorne, legt seinen Klausurbogen auf den Abgabestapel, nimmt die daneben liegende Anwesenheitsliste in seine linke Hand und sucht mit ausgestrecktem Zeigefinger seinen Namen. (...) Gegenüber dem Dozenten, der herbeieilt, sagt er: ‚Ich stehe nicht auf der Liste.‘ Der Dozent nimmt die Liste und stellt erstaunt fest: ‚Stimmt, Sie stehen wirklich nicht drauf.‘ Dann überlegt er kurz und fragt, ob er sich denn angemeldet habe. Im ersten Moment bejaht dieser, schiebt dann aber hinterher, dass er sich jetzt nun auch nicht mehr sicher sei. Er zückt sein Smartphone. Der Dozent entgegnet entspannt, dass man sicher eine Lösung finden werde und dass er zur Sicherheit seinen Namen und seine Unterschrift auf die Rückseite der Liste schreiben solle. Der Student tippt aber weiter auf seinem Handy herum, um dann zu äußern, dass er jetzt gerade nicht reinkomme. Der Dozent wiederholt mit beruhigender Stimme, dass das schon werde und er sich via Mail melden werde. In einer krakeligen Schrift und mit zitternden Fingern notiert der sichtlich nervöse Student seinen Namen auf der Rückseite.“

In dieser Szene geht es um die berechtigte Teilnahme an der Klausur. Nicht auf der Liste zu stehen, bedeutet für Student:innen tendenziell, nicht über einen legitimen Status als Klausurteilnehmer:in zu verfügen. Die studentische Sorge wird jedoch konterkariert durch die Souveränität des Dozenten im Umgang mit der Liste. Sie ist für ihn nicht sakrosankt, sondern ein pragmatisches Mittel zur Dokumentation, in dem digital erzeugte und handschriftliche Einträge einander ergänzen. Hieran lässt sich sehen, wie formale Dokumente bisweilen unterlaufen werden, um Interaktionsmanagement zu betreiben und die geregelte Ordnung der Klausur am Laufen zu halten. Dabei ist das nachträgliche Hinzufügen der Unterschrift ein *workaround*: Es stellt keine Reparatur dar, sondern temporalisiert das Problem insofern, als alle hinzugefügten Listeneinträge später noch einmal kontrolliert werden (Schabacher 2017). Während die Liste für das Klausurpersonal ein gestaltbares Dokument ist, ist sie für die Prüflinge Zeichen eines machtvollen bürokratischen Ordnungsprinzips, dem sie sich unterwerfen müssen. Als ein Schreibobjekt ist die Liste also zunächst unabschlossen; das gilt auch im Hinblick darauf, dass sie erst durch den Akt der Signatur der Prüfenden und des Studenten vervollständigt wird. Die handschriftliche Signatur sichert die Zeugenschaft im Verfahren und so die Identifikation von Klausur und Student:in formal ab (Widmer 2010).

Das Zusammenspiel der heterogenen Elemente (Blicke, Aufgaben, Architektur, Unterschriften etc.) in der Prüfungssituation zeigt, wie Lehrpersonal und universitäre Prüfungsorganisation das praktische Problem der Herstellung unabhängiger Autorschaft systematisch bearbeiten (Stevens 2007). Die lokale Ordnung der Prüfung mit ihrem Fokus auf der eigenständigen Darstellung und Reproduktion von Wissen steht so im Gegensatz zur verteilten Elitzitation von richtigen Antworten in der Lehre. Klausuren lassen sich somit als Orte verstehen, die über heterogene Kontrollpraktiken institutionalisiert sind, die temporär aktualisiert werden. Das Gelingen der Kontrolle stellt schließlich die Grundlage für jenes Vertrauen dar, das die Bewertenden ihren Bewertungsobjekten im Anschluss entgegenbringen (Williams 2007):

Das Misstrauen in der Prüfung wechselt über in das Vertrauen auf die eigenständige Leistung bei der Bewertung.

5 Bewertungsszenen. Korrektur und Kritik

An die Praxis des Lehrens und Prüfens schließt sich die Korrektur und Bewertung schriftlicher Arbeiten oder mündlicher Äußerungen an, in einem Fall als solitäres Lesen, Markieren und Benoten von Texten am Schreibtisch, im anderen Fall als dyadisch strukturierte Diskussion von Eindrücken im Anschluss an mündliche Prüfungen. Der subjektivierenden Zuschreibung von Antworten auf Student:innen in der Lehre steht mit dem Bewerten schriftlicher und mündlicher Leistungen eine Objektivierung gegenüber, die wir anhand von zwei exemplarischen Formen analysieren: Bei der ersten Form geht es um die Korrektur von Klausuren, die vom Lehrpersonal oder von technischen Artefakten mehrfach vorgenommen wird. Die Objektivierung der Bewertung vollzieht sich hier als eine Korrektur von Korrekturen (5.1). Die zweite Form besteht in der Verhandlung von Noten in mündlichen Prüfungen, die auf das Bezugsproblem der Flüchtigkeit von Antworten reagieren (5.2). Beide Formen dokumentieren, wie in diesen Fällen das Bewerten als ein raum-zeitlich, materiell und interpersonell verteilter Prozess vollzogen wird, womit sie der Vorstellung solitärer Bewertungssubjekte widersprechen.

5.1 Bewertungsketten als Modus sozialer Objektivierung

Entgegen der Annahme, dass die Korrektur von Klausuren zumeist von einzelnen Lehrpersonen vorgenommen wird, finden sich in unseren Beobachtungen vielfältige Formen des kollaborativen Korrigierens. Dabei handelt es sich zumeist um arbeitsteilig organisierte und hierarchisch gegliederte Formen der Vor- und Nachkorrektur. Diese „Bewertungsketten“ (Saner 2019) sind einerseits das Ergebnis pragmatischer Erwägungen im Angesicht großer Student:innenzahlen und andererseits ein Mittel zur Einhegung subjektiver Einschätzungen. Ein Beispiel aus einer Korrekturbeobachtung im Büro einer Professorin:

Auf einigen Klausurseiten haften bunte Post-its. Frau Marx erklärt mir, dass ihre Hilfskräfte die Klausuren bereits in einem ersten Schritt vorkorrigiert haben, insbesondere bei Multiple-Choice-Fragen und bei solchen, die sich lediglich auf die Reproduktion von Daten oder Orten beziehen. Die Post-its markieren jene Stellen, bei denen sich die Hilfskräfte unsicher waren und daher eine endgültige Bewertung nicht vornehmen konnten oder wollten. An diesen markierten Stellen muss sie selbst nochmals eine abschließende Korrektur vornehmen. Frau Marx konzentriert sich jedoch nicht ausschließlich auf die markierten Stellen, sondern überprüft stichprobenartig auch die unmarkierten Abschnitte, die von den Hilfskräften bereits vollständig bearbeitet wurden. In den meisten Fällen bestätigt sie die vorgenommenen Vorkorrekturen stillschweigend. Einmal stoppt sie abrupt und flüstert die Antwort eines Studenten leise vor sich hin. Mit einem leicht verwunderten Ausdruck auf ihrem Gesicht äußert sie plötz-

lich Unverständnis über die Entscheidung der Hilfskräfte: „Was haben die denn da gemacht?“ Mit schrillum Tonfall („1817 ist doch völlig richtig!“), der ihre Verwunderung erkennen lässt, schreibt sie die korrigierte Punktzahl an den Rand.

Der Ausschnitt gibt Einblick in die kollaborative Arbeit an Klausuren. Erstens zeigt sich eine unterschiedliche Gewichtung von Fragetypen und eine damit variierende Autorisierung ihrer Bewertung. Standardisierte Fragetypen können von studentischen Hilfskräften gewertet werden, während die Freitextaufgaben durch die Professorin selbst gewertet werden. Die Markierungen der Hilfskräfte stehen wiederum für ein Hochzonen der Verantwortung an die Professorin, d.h. an die professionelle Bewerterin. Der Umstand, dass nicht nur die Antworten der Kandidat:innen, sondern auch die Bewertungsfähigkeit ihrer Hilfskräfte geprüft wird, verweist darauf, dass die Authentifizierung der Bewertung durch die Professorin entscheidend ist. Die zur Disposition stehende Wertung der Hilfskräfte wird umgewandelt in ein von der Professorin ratifiziertes und zu verantwortendes Ergebnis. Neben der Klassifikation von Student:innen erfüllt das Bewertungsgeschehen eine weitere Funktion. Es führt den akademischen Nachwuchs in einen Rollenwechsel ein und macht die Prüfung zu einem Bestandteil seiner Instruktion. Ein ähnlich gelagerter Fall einer Objektivierung durch Bewertungsketten zeigt sich auch im Umgang mit der technisierten Bewertung von MC-Klausuren und der Arbeit am Schnitt. Ein Historiker:

„Die Klausur wird am PC durchgeführt und das System korrigiert das automatisch. Also Multiple-Choice, das ist ja klar. Und die Lückentexte sind so eingerichtet, dass auch Antwortmöglichkeiten richtig sind, wo mal ein Buchstabendreher oder Ähnliches drin ist. Und dann wird eben händisch nochmals nachkorrigiert, dass man guckt, okay, da ist jetzt jemand, der hat nur, sagen wir, 32 % der Punkte, kann man da jetzt bei dem Lückentext noch irgendwo einen Begriff gelten lassen und dann sagen, okay, Schwamm drüber. Da kann man händisch nochmals nachhaken. Und da gucke ich dann natürlich nochmals drauf, insbesondere über die Zweifelsfälle.“

Die maschinelle Bewertung setzt eindeutige Frage-Antwort-Schemata und damit eine Formalisierung des Fachwissens voraus. Es gilt, dass alles, was formalisierbar ist, auch mechanisiert und damit durch ein Rechenprogramm bearbeitet werden kann (Heintz 2001). Gegen diese Mechanisierung – die numerische Form der Objektivierung – bringt der Historiker eine soziale Form der Objektivierung ins Spiel, mit der nachträglich sichergestellt wird, dass alle Bewertungsverbesserungen ausgeschöpft worden sind. Er sieht ein zweistufiges Verfahren vor: Während wissenschaftliche Mitarbeiter:innen die Lückentexte daraufhin durchsehen, ob alles gewertet und berücksichtigt wurde, entscheidet der Professor die kritischen Einzelfälle. Im Kern bedeutet dies, dass auch fachlich unklare oder fast falsche Antworten als punktefähige Antworten anerkannt werden können („Schwamm drüber“). In diesem Akt einer substanziellen Verwandlung der Antworten wird der Maßstab geändert. Nicht allein die fachliche Richtigkeit zählt, sondern die anerkennende Markierung der Antwort als eine richtige Antwort, die ihrerseits das Klausurergebnis verbessert.

Der Rückgriff auf maschinelle Tools findet sich auch in jenen Momenten, in denen Student:innen Antworten produzieren, deren Richtigkeit vom Prüfungspersonal nicht ad hoc geprüft werden kann. Ein Beispiel:

Einzelne Fragen kommentiert Frau Marx mit dem Eingeständnis, dass sie selbst unsicher ist, ob die Antwort des Studenten richtig sei. Zwar entspricht die Antwort nicht der erwarteten Lösung, doch ob sie dennoch richtig sei, kann sie nicht sofort beurteilen. In solchen Momenten greift sie mehrfach zu ihrem Smartphone, um entsprechende Informationen zu suchen. Einmal stellt sich heraus, dass die gefundene Information nicht mit der Antwort des Studenten übereinstimmt. Zufrieden sagt Frau Marx: „Hab ich’s doch gewusst!“

In diesem Fall geht es nicht um die nochmalige Durchsicht der maschinellen Vorkorrektur, sondern um Momente der Unsicherheit, in denen die Prüfenden, um sich abzusichern, eine weitere Instanz (Bücher, das Internet etc.) zurate ziehen. Dies geschieht hauptsächlich bei Antworten, die jenseits der Erwartungen des Prüfungspersonals liegen. Es handelt sich dabei um Details wie Geburtsorte oder Verwandtschaftsverhältnisse von Herrschern, die von den Student:innen erwähnt werden, obwohl sie nicht explizit abgefragt wurden. Mit ihrer inhaltlichen Vergewisserung gelingt der Historikerin eine Aktualisierung ihrer fachlichen Souveränität („Hab ich’s doch gewusst“). In anderen Worten: Der maschinellen und der nicht-authentifizierten Korrektur durch Mitarbeiter:innen pflöpft sich die verantwortende, professorale Bewertung auf – und sei es durch digitale Faktenchecks. Das Vorgehen dokumentiert eine verteilte Absicherung durch das Bewertungsteam sowie eine verantwortende Logik der Hierarchisierung.

5.2 Deliberative Be/Wertung und die mündliche Prüfung als Instruktion

Ein zentrales Bezugsproblem der Bewertungspraxis in mündlichen Prüfungen bildet der Umstand, dass sich die Bewertung auf eine situative *Performance* bezieht. Im Gegensatz zu schriftlichen Hausarbeiten oder Klausuren, die als Texte vorliegen und als solche begutachtet werden können, müssen die Prüfer:innen in mündlichen Prüfungen Wege finden, um mit der Flüchtigkeit von Sprechakten umzugehen. Die Beobachtungen der Bewertungspraxis zeigen, wie die flüchtige Mündlichkeit der Prüfung durch das Protokoll und die Erzählungen des Prüfungspersonals, mit dem es auf das Gehörte und Protokollierte referiert, substituiert wird. Das Beratungsgespräch, in dem leistungsorientierte Wertungen expliziert und kommuniziert werden, folgt dabei seiner eigenen Verfahrensrationalität und reduziert sukzessive die Komplexität des mündlichen Prüfungsgeschehens, das schließlich in objektivierte Notenform übersetzt wird. Im Folgenden dokumentieren wir drei Beratungsszenen, die wir auf der Grundlage unserer Empirie als charakteristisch für mündliche Benotungsprozesse in der Geschichtswissenschaft ansehen:

Beratungsszene 1

Der Prüfling verlässt den Raum und schließt die Tür hinter sich.

Protokollant: Das war gut.

Prüferin: Nee, das war sehr gut.

Protokollant: Das war sehr gut, ja.

Prüferin: Also ich sehe keinen Grund, irgendeinen Abstrich zu machen. Was denkst du?

Protokollant: Nee. Also das Einzige, was jetzt hier war, war die Nachfrage mit den Quäkern. Da ist sie gestolpert–

Prüferin: Nein. Das ist eine 1,0.

Kennzeichnend für diese Beratungsszene ist erstens die Durchsetzung einer sich selbst gewissen Bewertung und zweitens ihre hierarchische Form, die sich in diesem Fall am Status der Akteure festmacht (Professorin vs. Wissenschaftlicher Mitarbeiter). Aus einer sich herantastenden, eher alltagsweltlichen Einschätzung („war gut“) wird eine an der hochschulischen Klassifikation orientierte Bewertung („nee, sehr gut“), die keine Einwände oder Widersprüche zulässt („keine Abstriche“, „nein, das ist“). Der genauen Beobachtung des Protokollanten weiß die Prüferin nur ihre Devaluierung entgegenzusetzen, die das Verfahren ohne weitere Begründung abkürzt: eine bewertende, auf hierarchischer Differenz fundierende Entscheidung, die Gegenteiliges ausschließt und sich auf keine Diskussion einlässt.

Ist die Vergabe von Bestnoten oft in wenigen Sekunden erledigt, erfordern andere Prüfungsverläufe mitunter einen etwas höheren zeitlichen und diskursiven Aufwand; ihr Merkmal sind eine stärker egalitäre Kommunikation und eine anfängliche Unsicherheit darüber, wie die Prüfung einzuordnen und zu bewerten ist. Ein weiteres Beispiel:

Beratungsszene 2

Protokollant: Ja, das war schon komisch.

Prüferin: Und Quellen konnte sie auch nicht richtig benennen. Vielleicht 2,7?

Protokollant: Das ist aber Friedenau-Spezial.

Prüferin: Ich kenne das Problem auch noch von woandersher.

Protokollant: Ja, da kann sie auch nix für. Forschungsliteratur konnte sie schon gut benennen.

Prüferin: Ja. Ich denke wir können auch 2,3 machen. Ich denke, dass sie das insgesamt gut findet. Weil ich denke, dass sie das, was das Modul leisten soll, dass man sich damit auseinandersetzt und einen kritischen Blick gewinnt, dass sie das schon geleistet hat. Mit allen Abstrichen, die es sonst so gibt, ist da eine 2,3 gut.

Um die Unklarheit („komisch“), wie die mündliche Prüfung einzuschätzen ist, bearbeiten zu können, folgt diese Beratung der Logik einer Liste, mit der im Hinblick auf bestimmte Kriterien geklärt wird, was die Studentin konnte und was nicht („Quellen“, „Forschungsliteratur“). Mit ihrer Einordnung in eine Stadtgesellschaft („Friedenau-Spezial“) dämpfen die Prüfenden ihre Erwartungen, relativieren aber zugleich ihre stille Übereinkunft, das stadträumliche Herkunftseinschließen auf die potenzielle Qualität einer Studentin zulassen („auch nix für“). Aus einer ersten Einordnung („2,7“) wird ein Schätzwert („auch 2,3“) generiert und mit der Modulbeschreibung legitimiert. Markant ist schließlich die wechselnde Verwendung von Personalpronomina, mit der die Prüferin zwar eine soziale Distanz zum Protokollanten markiert („ich denke“) und einen Perspektivenwechsel inszeniert („findet sie

gut“), dabei aber doch auffällig abwägend bleibt („schon“) und abschließend die Bewertung noch einmal begründet. Es ist eine unsichere, sich herantastende Kommunikation, in der Sicherheit über die Bewertung – im Gegensatz zum ersten Fall – erst hergestellt werden muss, indem alternative Einschätzungen, die hier zunächst stärker präsent sind als oben, ausgeschlossen werden.

Ein dritter Fall, bei dem es um die Vergabe einer schlechten Note geht, ist hingegen wieder, äquivalent zur Vergabe von Bestnoten, von Gewissheit geprägt; der zentrale Unterschied liegt hier jedoch in einer Entkopplung von Beratung und Bewertung.

Beratungsszene 3

Prüferin: Das war gar nix. Naja, ist zweifellos bestanden. Aber ist jetzt im Vergleich mit den anderen schon sehr schwach.

Protokollant: Also ich mein, sie hat ja schon einiges gesagt. Also zur Architektur der Stadt.

Prüferin: Aber es war eigentlich fast alles ziemlich falsch, was sie gesagt hat.

Protokollant: Ich habe die Bücher nicht gelesen.

Prüferin: [Beide lachen.] Okay. Ja, also bei Benevolo geht es nicht um den Aufbau der mittelalterlichen Stadt, also das ist das, was er äußerlich untersucht, aber er interessiert sich dafür, wie die Gesellschaft funktioniert. Also die ganze Theorie dahinter hat sie gar nicht gesehen. Also auch das mit den Quellen, dass sie da mathematische Modelle nennt...

Protokollant: Ja, da kam auch nix weiter außer archäologische Befunde.

Prüferin: Auch überhaupt Fragen zu verstehen, hat auch nicht funktioniert. Eine 3,3?

Trotz dargestellter Gewissheit durch eine Wertung der mündlichen Antworten („alles ziemlich falsch“, „kam auch nix“), ist diese schnelle Entscheidungsfindung problematisch, denn der Protokollant gibt sich als unwissend zu erkennen („nicht gelesen“). Diese Unkenntnis wird nicht weiter problematisiert, sondern als erwartbar hingenommen. Zugleich wird der Protokollant – um die Legitimität des Verfahrens zu sichern und die Begründung zu plausibilisieren – historisch unterwiesen („Aufbau der mittelalterlichen Stadt“) und in die Bewertung eingebunden („eine 3,3?“). Ferner gibt der Hinweis auf Student:innen, die zuvor bewertet worden sind und vergleichend als „exemplary goods“ (Dekker 2016) gelten, die Möglichkeit an die Hand, die Prüfungsleistung abzuschätzen, was aber im Gegenzug die Bewertungsleistung selbst unmarkiert lässt. Zwei weitere Merkmale sind kennzeichnend für die Beratungsszene: Erstens ist es ein ambivalentes Geschehen, das sowohl Prüfung einer Studentin als auch Einübungszeit für Wissenschaftliche Mitarbeiter:innen ist; zweitens verbindet sich die Wertung der *Performance* („war nix“) kaum mit ihrer Bewertung („zweifellos bestanden“), sichert aber einen formalen Abschluss des Verfahrens.

6 Schluss

Dieser Aufsatz behandelte die Praxis einer universitären Unterrichtung und Differenzierung von Menschen, d. h. Student:innen. Am Beispiel der Geschichtswissenschaft wurde die zeitliche Verschränkung des Lehrens, Prüfens und Bewertens sowie ihre organisationale Verknüpfung über flüchtige Sprechakte, Notizen und eine Vielzahl von Dokumenten (Listen, (digitale) Klausurbögen, Protokolle etc.) analysiert. Das heißt, dass Transsituativität in zwei Formen praktiziert wird: zum einen als gegenwärtige bzw. ineinander gefaltete Präsenz der Phasen des Lehrens, Prüfens und Bewertens, die daher nicht strikt voneinander zu trennen sind, zum anderen als Zirkulation organisatorischer Formate und Personen. Somit stehen diese Phasen für je eigene Bereiche und Dynamiken und bilden doch einen Zusammenhang, der über sie hinausweist: eine Praxis des Diskurses, des Einübens und des Zeigens, und zwar mit dem Ziel, eine individuelle Zurechnungsadresse für das Prüfen und Bewerten zu etablieren und zu garantieren.

Die universitären Kontexte sind u. a. von Tätigkeiten der Lehrenden bestimmt, die – organisatorisch und rechtlich gesetzt – zu Entscheidungen über schriftliche Arbeiten oder mündliche Äußerungen führen müssen, die von den Student:innen getan wurden. Auf schriftliche und mündliche Prüfungen, d. h. auf Ankreuzungen, (kurze) Texte und Äußerungen, folgen also Aktivitäten, mit denen die Bewertenden die schriftlichen bzw. mündlichen Antworten zunächst qualifizieren und somit werten. Dies geschieht als eine verteilte und interpretative soziale Praxis: etwa als Nachkorrektur einer softwaregestützten Bewertung, die Fachwissen mobilisiert; oder als situative Mobilisierung von Einschätzungen, die ein gemeinsam geteiltes Bild des mündlichen Geschehens erzeugen. Hierauf folgt die Bewertung als Bestimmung des Prädikats (der Zensur oder der Note) – sei es als berechnetes Ergebnis einer MC-Klausur oder als eine eingeschätzte Platzierung der Student:innen auf der Notenskala. Es handelt sich um eine hierarchische Absicherung der (kontingenten) Bewertung sowie die Verwandlung studentischer Tätigkeiten in eine zertifizierte Leistung, die in der Organisation zirkuliert, berechnet und in Zeugnissen etc. dokumentiert wird.

Die Schriftdokumente, die in diesem Prozess zwischen der universitären Prüfungsorganisation und den Prüfenden zirkulieren, authentifizieren die Bewertenden ebenso wie die Zu-Bewertenden und lösen damit (partiell) das Problem der Zurechenbarkeit von Leistung, denn die universitäre Differenzierung setzt voraus, dass Student:innen als menschliche Individuen und als Zurechnungsadresse isoliert werden können. Prüfen und Bewerten werden auf diese Weise nicht allein als ein situatives Tun einzelner Lehrender, sondern als kollaborative Herstellung von Leistung verstehbar. Aus der hier eingenommenen Perspektive der Transsituativität, die den Dualismus von theoretischem Funktionalismus einerseits und mikrosoziologischer Fokussierung auf singuläre Situationen andererseits aufhebt, stellt die eindeutige Zurechenbarkeit von „Leistungen“ auf individuelle Student:innen jedoch in gewisser Weise eine Fiktion dar, denn sie setzt voraus, dass menschliche Individuen als Zurechnungsadresse isoliert werden können und Prüfende mit ihrer Tätigkeit des Lehrens, Prüfens und Bewertens unmarkiert bleiben. In anderen Worten: Es gehört zur Funktionsbedingung der Universität und seiner meritokratischen Leitidee, dass

die kollaborative Dimension der Herstellung studentischer Leistung in den Hintergrund rückt.

Schließlich weist der Aufsatz auf drei konzeptionelle Überlegungen hin, an die die weitere Forschung anschließen kann: *Erstens* führt er mit Blick auf die Fiktion der eindeutigen Zurechnung und der Bewertungskontingenz Perspektiven ein, mit der die involvierten Anerkennungstheorien komparativ erforscht werden können. Wie bspw. werden die Annahmen der Bildungsorganisationen in anderen Disziplinen erzeugt, aufgeführt und stabilisiert? *Zweitens* hat der Aufsatz auf die Beteiligung nicht-menschlicher Akteure an der Bewertung hingewiesen (etwa bei E-Klausuren). Wer oder was bewertet, wenn elektronische Systeme oder zukünftig KI-Anwendungen die Bewertung durchführen? Wie also wird die kollaborative Hervorbringung der Bewertung durch Technologien transformiert, die in diese Prozesse eingebunden werden? *Drittens* hat der Aufsatz die Bedeutung der Prüfung als Bestandteil einer Professionalisierung universitärer Noviz:innen angedeutet. Die Annahme, dass das Prüfungsgeschehen an Universitäten auch Bestandteil einer professionellen Sozialisation ist, lenkt die theoretische wie empirische Aufmerksamkeit auf die Situationen und Verfahren der Einübung eines Fachhabitus.

Förderung Dem Aufsatz liegt ein Forschungsprojekt zugrunde, das durch die DFG gefördert wurde (KA 1235/10-1).

Funding Open Access funding enabled and organized by Projekt DEAL.

Einhaltung ethischer Richtlinien

Interessenkonflikt H. Kalthoff und F. Kölsch geben an, dass kein Interessenkonflikt besteht.

Ethische Standards Das universitäre Lehrpersonal, das wir in ihren Lehrveranstaltungen und Prüfungen beobachtet haben, stimmten der Beobachtung zu. Zu keinem Zeitpunkt wurden im Rahmen des Forschungsprojekts individuelle soziodemografische Daten über diese Teilnehmer erhoben.

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

- Albrecht, Kathrin. 2024. *Universität im Bologna-Prozess. Form und Re-Form der universitären Erziehung*. Weilerswist: Velbrück.
- Atkinson, Paul, und Amanda Coffey. 2011. Analysing documentary realities. In *Qualitative research. Theory, method and practice*, Hrsg. David Silverman, 56–75. London: SAGE.
- Banerjee, Mita, Gregor Feindt, Herbert Kalthoff, und Benjamin Wihstutz. 2025. Leistung im Vergleich. Objektivierungen trotz Differenzierungskontingenz. In *Humandifferenzierung im Vergleich*, Hrsg.

- Anne Friedrichs, Stefan Hirschauer, Peter Hofmann, und Gabriele Schabacher, 236–266. Weilerswist: Velbrück.
- Bourdieu, Pierre. 1984. *Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft*. Frankfurt a.M.: Suhrkamp.
- Bourdieu, Pierre. 1985. *Sozialer Raum und Klassen. Leçon sur la leçon*. Frankfurt a.M.: Suhrkamp.
- Bourdieu, Pierre. 1992. *Homo academicus*. Frankfurt a.M.: Suhrkamp.
- Brosziewski, Achim. 2023. *Lebenslauf, Medien, Lernen. Skizzen einer systemtheoretischen Bildungssoziologie*. Weinheim: Beltz.
- Brückner, Fabian, und Stephan Wolff. 2015. Die Listen der Organisation. Und der Blick zwischen die Zeilen. In *Formalität und Informalität in Organisationen*, Hrsg. Victoria von Groddeck, Sylvia Marlene Wilz, 341–370. Wiesbaden: Springer VS.
- Decker, Kris, und Christoph Hoffmann. 2023. Training scenes. Taking science studies to the classroom. *Social Studies of Science* 53(3):402–426.
- Dekker, Erwin. 2016. Exemplary goods. Exemplars as judgment devices. *Valuation Studies* 4(2):103–124.
- Elwick, James. 2023. Knowing the same things. Mass examinations, credentials, and infrastructures of shared knowledge. *Interdisciplinary Science Reviews* 48(2):202–215.
- Emerson, Robert M., Rachel I. Fretz, und Linda L. Shaw. 1995. *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.
- Flechsig, Karl-Heinz. 1976. Prüfungen und Evaluation. *Blickpunkt Hochschuldidaktik* 40:303–336.
- Fludernik, Monika. 2016. Descriptive lists and list descriptions. *Style* 50(3):309–326.
- Forde-Leaves, Natalie, Jack Walton, und Ken Tann. 2023. A framework for understanding assessment practice in higher education. *Assessment & Evaluation in Higher Education* 48(8):1076–1091.
- Foucault, Michel. 1977. *Überwachen und Strafen*. Frankfurt a.M.: Suhrkamp.
- Frickel, Scott, und Kelly Moore (Hrsg.). 2006. *The new political sociology of science: Institutions, networks, and power*. Madison: University of Wisconsin Press.
- Goffman, Erving. 1981. *Forms of talk*. Oxford: Basil Blackwell.
- Goody, Jack. 1977. *The domestication of the savage mind*. Cambridge: Cambridge University Press.
- Hahn, Alois. 1982. Zur Soziologie der Beichte und anderer Formen institutionalisierter Bekenntnisse: Selbstthematization und Zivilisationsprozess. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 34:407–434.
- Heffer, Chris. 2007. Judgement in court: Evaluating participants in courtroom discourse. *Language and the law: International outlooks* 16:145–179.
- Heidegger, Martin. 2004. *Der Begriff der Zeit*. Frankfurt a.M.: Klostermann.
- Heintz, Bettina. 2001. *Die Innenwelt der Mathematik. Zur Kultur und Praxis einer beweisenden Disziplin*. Wien: Springer.
- Hirschauer, Stefan. 2021. Menschen unterscheiden. Grundlinien einer Theorie der Humandifferenzierung. *Zeitschrift für Soziologie* 50(3–4):155–174.
- Holstein, James A. 1983. Grading practices: the construction and use of background knowledge in evaluative decision-making. *Human Studies* 6:377–392.
- Honneth, Axel. 1992. *Kampf um Anerkennung. Zur moralischen Grammatik sozialer Konflikte*. Frankfurt/Main: Suhrkamp.
- Ioannidou, Mary A. 1997. Testing and life-long learning: Open-book and closed-book examination in a university course. *Studies in Educational Evaluation* 23(2):131–139.
- Jedemark, Marie, und Mikael Londos. 2021. Four different assessment practices: how university teachers handle the field of tension between professional responsibility and professional accountability. *Higher Education* 81:1293–1309.
- Kalthoff, Herbert. 2017. Kontingenz und Unterwerfung. Die organisierte Humanevaluation der Schule. In *Un/doing Differences. Praktiken der Humandifferenzierung*, Hrsg. Stefan Hirschauer, 259–284. Weilerswist: Velbrück.
- Kalthoff, Herbert, und Fabian Kölsch. 2025. Before e/valuating. Student testing in history and engineering. *British Journal Sociology of Education* 46(3):323–340.
- Kingston, Paul W. 2001. The unfulfilled promise of cultural capital theory. *Sociology of Education* 74:88–99.
- Knorr-Cetina, Karin. 1981. *The manufacture of knowledge*. Oxford: Pergamon Press.
- Krüger, Anne K., und Martin Reinhart. 2016. Wert, Werte und (Be)Wertungen. Eine erste begriffs- und prozesstheoretische Sondierung der aktuellen Soziologie der Bewertung. *Berlin Journal für Soziologie* 26(3–4):485–500.

- Kühl, Stefan. 2012. Zur Einführung einer neuen ‚Kunstwährung‘ an den Hochschulen – Die Vergleichs- und Tauschfunktion von ECTS-Punkten. In *Hochschule als Organisation*, Hrsg. Uwe Wilkesmann, Christian Schmid, 325–344. Wiesbaden: Springer VS.
- Kvale, Steinar. 1972. *Prüfung und Herrschaft*. Weinheim: Beltz.
- Lambrix, Philip. 2022. Erstrebenswerte Unselbständigkeit? Die Ambivalenz der Pflegebegutachtung. *Zeitschrift für Soziologie* 51(3):227–243.
- Landwehr, Achim. 2016. *Die anwesende Abwesenheit der Vergangenheit*. Frankfurt a.M.: Fischer.
- Latour, Bruno, und Steve Woolgar. 1979. *Laboratory life: The construction of scientific facts*. London: SAGE.
- Luhmann, Niklas. 1996. Takt und Zensur im Erziehungssystem. In *Zwischen System und Umwelt. Fragen an die Pädagogik*, Hrsg. Niklas Luhmann, Karl E. Schorr, 279–294. Frankfurt a.M.: Suhrkamp.
- Luhmann, Niklas. 2002. *Das Erziehungssystem der Gesellschaft*. Frankfurt a.M.: Suhrkamp.
- Masschelein, Jan, und Maarten Simons. 2013. The university in the ears of its students. On the power, architecture and technology of university lectures. In *Die Idee der Universität – revisited*, Hrsg. Norbert Ricken, Hans-Christoph Koller, und Edwin Keiner, 173–192. Wiesbaden: Springer VS.
- McHoul, Alexander. 1978. The organization of turns at formal talk in the classroom. *Language in Society* 7:183–213.
- Meadmore, Daphne. 1993. The production of individuality through examination. *British Journal of Sociology of Education* 14(1):59–73.
- Meer, Dorothee. 1998. *Der Prüfer ist nicht der König. Mündliche Abschlussprüfungen in der Hochschule*. Tübingen: Niemeyer.
- Meer, Dorothee. 2019. Der Prüfer ist immer noch nicht der König. Nachdenken über mündliche Hochschulprüfungen der Gegenwart. In *Studentische Praxis und universitäre Interaktionskultur: Perspektiven einer praxeologischen Bildungsforschung*, Hrsg. Tanya Tyagunova, 173–200. Wiesbaden: Springer VS.
- Meier, Frank, und Thorsten Peetz (Hrsg.). 2021. *Organisation und Bewertung*. Wiesbaden: Springer VS.
- Meier, Frank, Thorsten Peetz, und Désirée Waibel. 2016. Bewertungskonstellationen. Theoretische Überlegungen zur Soziologie der Bewertung. *Berliner Journal für Soziologie* 26(3):307–328.
- Nicolae, Stefan, Martin Endreß, Oliver Berli, und Daniel Bischof (Hrsg.). 2019. *(Be)Werten. Beiträge zur sozialen Konstruktion von Wertigkeit*. Wiesbaden: Springer VS.
- Nieminen, Juuso H. 2021. Beyond empowerment: student self-assessment as a form of resistance. *British Journal of sociology of Education* 42(8):1246–1264.
- Nieminen, Juuso H., und Lili Yang. 2024. Assessment as a matter of being and becoming: theorising student formation in assessment. *Studies in Higher Education* 49(6):1028–1041.
- Nieminen, Juuso H., Margaret Bearman, und Joanna Tai. 2023. How is theory used in assessment and feedback research? A critical review. *Assessment & Evaluation in Higher Education* 48(1):77–94.
- Pabian, Petr. 2014. Ethnographies of higher education: introduction to the special issue. *European Journal of Higher Education* 4(1):6–17.
- Pfalter, Robert. 2010. Der Kampf gegen die Fortentwicklung der Universität zur repressiven Attrappe. In *Was passiert? Stellungnahmen zur Lage der Universität*, Hrsg. Unbedingte Universitäten, 51–54. Zürich: Diaphanes.
- Prahl, Hans-Werner. 1976. *Hochschulprüfungen. Sinn oder Unsinn?* München: Kösel.
- Prange, Klaus, und Gabriele Strobel-Eisele. 2006. *Die Formen des pädagogischen Handelns*. Stuttgart: Kohlhammer.
- Prior, Lindsay. 2004. Doing things with documents. In *Qualitative research: theory, method and practice*, Hrsg. David Silverman, 76–94. London: SAGE.
- Rabenstein, Kerstin. 2022. Einblicke in die soziale Praxis des Prüfens an Hochschulen. Innen(ein)sichten. *Journal für LehrerInnenbildung* 22(1):46–55.
- Rost, Katharina. 2017. *Sounds that matter. Dynamiken des Hörens in Theater und Performance*. Bielefeld: transcript.
- Saner, Philippe. 2019. Das Unbestimmbare bewerten. Aufnahmeverfahren an Kunsthochschulen aus rechtfertigungstheoretischer Perspektive. In *Bildung und Konventionen*, Hrsg. Christian Imdorf, Regula J. Leemann, und Philipp Gonon, 179–203. Wiesbaden: Springer VS.
- Schabacher, Gabriele. 2017. Im Zwischenraum der Lösungen. Reparaturarbeit und Workarounds. *ilinx – Berliner Beiträge zur Kulturwissenschaft* 4:XIII–XXVIII.
- Schröder, Wolfgang, Barbara Stauber, Andreas Walther, Lothar Böhnisch, und Karl Lenz (Hrsg.). 2013. *Handbuch Übergänge*. Weinheim: Beltz Juventa.
- Spradley, James. 1979. *The Ethnographic Interview*. New York: Holt, Rinehart and Winston.

- Stevens, Mitchell L. 2007. *Creating a class. College admissions and the education of elites*. Cambridge: Harvard University Press.
- Stichweh, Rudolf. 2013. *Wissenschaft, Universität, Professionen. Soziologische Analysen*. Frankfurt a.M.: Suhrkamp.
- Stratton, Julius A., und Loretta H. Mannix. 2005. *Mind and hand: the birth of MIT*. Cambridge: MIT Press.
- Strauss, Anselm. 1998. *Grundlagen qualitativer Sozialforschung. Datenanalyse und Theoriebildung in der empirischen soziologischen Forschung*. München: Fink.
- von Stumm, Sophie, Benedikt Hell, und Tomas Chamorro-Premuzic. 2011. The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science* 6(6):574–588.
- Trembl, Alfred K. 2000. *Allgemeine Pädagogik. Grundlagen, Handlungsfelder und Perspektiven der Erziehung*. Stuttgart: Kohlhammer.
- Tsarouha, Elena. 2019. *Prüfungspraktiken an deutschen Hochschulen*. Wiesbaden: Springer VS.
- Tyagunova, Tanya. 2019. Lernen mit Altklausuren. Informelle Praktiken der Studierenden bei der Prüfungsvorbereitung. In *Studentische Praxis und universitäre Interaktionskultur: Perspektiven einer praxeologischen Bildungsforschung*, Hrsg. Tanya Tyagunova, 201–239. Wiesbaden: Springer VS.
- Vatin, François. 2013. Valuation as evaluating and valorizing. *Valuation Studies* 1(1):31–50.
- Vissmann, Cornelia. 2000. *Akten. Medientechnik und Recht*. Frankfurt a.M.: Fischer.
- Wagner, Gabriele, Juan S. Guse, und Monika Hasenbruch. 2023. „Eigentlich war es immer sonnenklar.“ Zur Invisibilisierung von formaler Organisation in Bewertungspraktiken. *Berliner Journal für Soziologie* 33:69–97.
- Weber, Max. 1988. Der Sinn der Wertfreiheit der soziologischen und ökonomischen Wissenschaften. In *Gesammelte Aufsätze zur Wissenschaftslehre*, 489–540. Tübingen: Mohr.
- Widmer, Peter. 2010. *Der Eigenname und seine Buchstaben: psychoanalytische und andere Untersuchungen*. Bielefeld: transcript.
- Wiggins, Grant P. 1998. *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.
- Williams, Bronwyn T. 2007. Trust, betrayal, and authorship: Plagiarism and how we perceive students. *Journal of Adolescent & Adult Literacy* 51(4):350–354.

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

Herbert Kalthoff Professur für Bildungs- und Wissenssoziologie, Qualitative Methoden; Johannes Gutenberg-Universität Mainz; Forschungsschwerpunkte: Soziologie der Bildung und Bewertung, der Robotik und der Mensch-Maschine-Interaktion. Ausgewählte Veröffentlichung: Hybrid imbalance. Collaborative fabrication of digital teaching and learning material. In: *Qualitative Sociology* 46(3) (2023): 403–428 (mit Torsten Cress); Von der Handskizze zum Algorithmus. Zur intermediären Praxis der symbolischen Robotik. In: *Zeitschrift für Soziologie* 54(4) (2025) (mit Hannah Link; erscheint).

Fabian Kölsch wissenschaftlicher Mitarbeiter am Institut für Soziologie der Johannes Gutenberg-Universität Mainz. Forschungsschwerpunkte: Bewertungs- und Bildungssoziologie, Humandifferenzierungsforschung und Methoden der qualitativen Sozialforschung. Ausgewählte Veröffentlichung: Before (e)valuating: Student Testing in History and Engineering, in: *British Journal of Sociology of Education*, 2025 (mit Herbert Kalthoff).