

# „Die juristische Prüfung ist eine Blackbox“ – Interview mit Jun.-Prof. Dr. Judith Brockmann

Das Interview führte **Selma Gather**, Wissenschaftliche Mitarbeiterin an der Freien Universität Berlin und Vorsitzende des Arbeitsstabs Ausbildung und Beruf im djb.

**Liebe Judith, Du beschäftigst dich schon lange mit rechtswissenschaftlicher Prüfungsorschung. Du hast eine Juniorprofessur ausdrücklich auch für rechtswissenschaftliche Fachdidaktik und bist Mitherausgeberin der Zeitschrift für Didaktik der Rechtswissenschaft (ZDRW). 2015 hast du den Band „Prüfen in der Rechtswissenschaft: Stand und Perspektiven der disziplinären Forschung“ mitherausgegeben. Wie bist Du zur Fachdidaktik gekommen?**

Mein Ansatz war sehr stark ausgehend von den eigenen Herausforderungen in der Lehre und dem ernüchternden Befund, dass ich in der Rechtswissenschaft selbst wenig Antworten gefunden habe – zumindest im ersten Zugriff. Dann habe ich mich erst einmal der allgemeinen Hochschuldidaktik zugewendet, als einer Disziplin, die diese Antworten bereithielt. Darüber kam im Wesentlichen auch mein Zugang zur juristischen Prüfung.

**Wie groß ist die Resonanz für diese Forschung? Interessiert sich die Rechtswissenschaft genügend für ihr Prüfungswesen?**

Wir sind zwar alle sehr viel mit Prüfen befasst, trotzdem kümmern wir uns verhältnismäßig wenig um die Qualität dieses Systems.

**Was macht denn nach allgemeinen Qualitätsmaßstäben eine gute Prüfung aus?**

Allgemeine Qualitätskriterien sind beispielsweise die Objektivität der Prüfung, die Reliabilität – also Verlässlichkeit – und die Validität. Ganz wesentlich ist es, den Prüfungsmaßstab festzulegen. Welches sind die Beobachtungskriterien? Essentiell ist auch ein Niveaustufenmodell, also abstrakt definierte Leistungsstufen. Das heißt: Was muss ich konkret leisten, um eine bestimmte Notenstufe zu erreichen? Das ist eine hochkomplexe Aufgabe. In Jura haben wir zum Beispiel eine sehr ausgeprägte Notenskala, das gibt uns ein sehr großes Differenzierungspotential in der Bewertung. Wir wissen aber, dass dieses typischerweise nicht ausgeschöpft wird.

**Du hast „Objektivität“ als ein Kriterium genannt. In der Tat gelten die juristischen Staatsexamina gemeinhin als objektiv und fair. Wird die Prüfung diesem Ruf gerecht?**

Was man konstatieren kann: Die juristische Prüfung ist eine Blackbox. Gemeinhin wird gesagt, dass die Prüfungen im Ergebnis ein geeignetes Mittel zur Leistungsmessung seien. Und tatsächlich gibt es Selbstversuche von Prüfer\*innen, die zeigen,



▲ (Foto: Thomas Morgner)

1996 – 2001	Studium der Rechtswissenschaft in Hamburg und 1. Staatsprüfung
2001 – 2002	Studium der Rechtswissenschaft an der Université Paris XI mit Abschluss Maîtrise en droit, mention droit social; Eintritt in den djb
2002 – 2005	Referendariat in Hamburg, Paris und Lomé (Togo) und 2. Staatsprüfung
2005 – 2011	Selbständige Rechtsanwältin mit Tätigkeitsschwerpunkt im Arbeits- und Sozialrecht, zeitweilig hauptberuflich und in Sozietät
2005 – 2008	Wissenschaftliche Mitarbeiterin am Lehrstuhl für Öffentliches Recht und Sozialrecht an der Fakultät für Rechtswissenschaft der Universität Hamburg von Prof. Dr. Dagmar Felix
2010	Promotion zur Doktorin der Rechtswissenschaft, Fakultät für Rechtswissenschaft der Universität Hamburg mit der Arbeit „Ambulante Krankenversorgung ohne Kassenärztliche Vereinigungen? – Ein deutsch-französischer Rechtsvergleich zur Rechtsstellung niedergelassener Ärzte im System der Versorgung gesetzlich krankenversicherter Patienten“, ausgezeichnet mit dem Promotionspreis des Vereins zur Förderung sozialrechtlicher und sozialpolitischer Forschung e.V.
seit 2011	Juniorprofessorin für Arbeitsrecht mit sozialrechtlichen Bezügen und rechtswissenschaftliche Fachdidaktik an der Fakultät für Rechtswissenschaft der Universität Hamburg; Mitglied des dortigen Zentrums für rechtswissenschaftliche Fachdidaktik; Mitherausgeberin der Zeitschrift für rechtswissenschaftliche Fachdidaktik (ZDRW)
seit 2015	Stellvertretendes Mitglied im Ausschuss für Gleichstellung des Akademischen Senats der Universität Hamburg
seit 2017	Gastprofessuren an den Universitäten Lumière Lyon 2, Bordeaux, Paris Nanterre

dass das Bewertungsverhalten der Prüfer\*innen insgesamt kohärent ist. Also da ist Stabilität im System, es gibt implizite geteilte Kriterien. Aber: Sie sind eine Blackbox, wir kennen sie nicht.

Das bedeutet, dass die Prüferinnen und Prüfer zwar tendenziell zu einer ähnlichen Bewertung kommen mögen – z.B. das war jetzt VB oder das war nur ausreichend –, aber die Begründung, die sie liefern, wird sehr unterschiedlich sein. Es kann aber für eine Staatsprüfung nicht angehen, dass die Bewertungskriterien

letztlich subjektiv sind. Um die Prüfung objektiver zu machen, müsste ausbuchstabiert werden, was bislang nur als stabiles implizites Wissen vorhanden ist.

**2016 zeigte die breit angelegte Studie „Geschlechts- und Herkunftseffekte bei der Benotung juristischer Staatsexamen“ für NRW, dass es in der Benotung in den Staatsexamina zu Geschlechts- und Herkunftseffekten kommt, d.h. dass Frauen und Menschen mit (vermutetem) Migrationshintergrund statistisch schlechter abschneiden als „deutsche“ Männer. Dies gilt ganz besonders in der mündlichen Prüfung. Worauf könnten, Deiner Meinung nach, diese Effekte zurückzuführen sein?**

Darüber habe ich natürlich nur Vermutungen. In aller Regel werden das keine bewussten Diskriminierungen sein. Ich gehe davon aus, dass hier gewisse Verzerrungseffekte am Werk sind, die die Bewertung der Prüfung beeinflussen, aber auch schon die Prüfungsgestaltung, also die Frage, wie ich ein Prüfungsgespräch führe. Die mündliche Prüfung ist eine interaktive Kommunikationssituation zwischen Menschen. Darin liegen viele Chancen, aber es spielen eben Verzerrungseffekte eine Rolle – also Sympathie, individuelle Leistungserwartungen der Prüferinnen und Prüfer, Pfadabhängigkeiten und Ankereffekte, beispielsweise weil es keine doppelt blinden Bewertungen gibt. Einflussreich ist auch das „Drumherum“, ob z.B. ein Vorgespräch stattfindet oder nicht.

Schließlich vermute ich, dass eine Form der Ähnlichkeitselektion eine Rolle spielt – wenn der\* die Prüfer\* in den\* die Kandidat\* in als ähnlich wahrnimmt in irgendeiner Hinsicht, hat das eine stärkere Identifikation zur Folge. Das wird das Prüfungs- und Bewertungsverhalten beeinflussen können. Übrigens gilt das auch umgekehrt für die Identifikation der Kandidat\*innen mit dem\* der Prüfer\* in. Verzerrungseffekte zulasten von Menschen mit Migrationshintergrund oder Frauen lassen sich so überzeugend erklären. Und das Gleiche gilt für den Befund, dass die Effekte verschwinden, sobald nur eine Frau in der Prüfungskommission sitzt.

**Das klingt erst einmal, als wären das Probleme, die jede mündliche Prüfung betreffen. Warum treten die Effekte bei der mündlichen Prüfung aber in der universitären Schwerpunktprüfung nicht auf?**

Spiegelbildlich zu den intransparenten Bewertungskriterien besteht in den Staatsprüfungen ein hohes Maß an Erwartungsunsicherheit bei den Kandidatinnen und Kandidaten im Hinblick auf die konkreten Prüfungsanforderungen. Angst und Verunsicherung führen in der Regel dazu, dass das Leistungsvermögen nicht ausgeschöpft werden kann. Typischerweise kennen die Kandidat\*innen ihre Prüfer\* in vor der universitären mündlichen Prüfung schon, das nimmt ganz viel Verunsicherung und Angst.

**Das heißt, die Anonymität kann tatsächlich ein Problem sein?**

Das Entscheidende ist die Verunsicherung. Und die ist eben kleiner, wenn man das Gegenüber schon kennt. Das spricht erst einmal für die sogenannten Vorgespräche, aber in der Abwägung birgt das Vorgespräch meines Erachtens doch mehr Gefahren, als es Vorteile bietet. Dort wird in einer kurzen Zeit ein persönlicher Eindruck gewonnen. Meine Erfahrung ist: Wo

bestimmte stereotype Vorstellungen bestehen, wurden diese in erster Linie im Vorgespräch gewonnen.

Auch die Kenntnis der Vornoten und der Studienleistung ist ein Problem, weil dann auf eine bestimmte Note geprüft wird. Die Kandidat\*innen äußern ja im Vorgespräch schon häufig, was sie anstreben. Die werden dann „auf ein VB“ geprüft. Und die Person, die im Vorgespräch den Mund vielleicht nicht so voll nimmt, hat diese Chance nicht unbedingt und wird vielleicht nur auf ihr Bestehen hin geprüft. Wenn es um stereotype Verhaltensmuster geht, ist es eine naheliegende Vermutung, dass Frauen sich im Vorgespräch tendenziell vorsichtiger und bescheidener äußern.

**Das ist deswegen eine interessante These, weil die NRW-Studie gezeigt hat, dass Männer eher über die relevanten Notenschwellen geprüft werden als Frauen. Wenn ich Dir so zuhöre, scheint mir auch eine weitere These plausibel: Sind diese impliziten Kriterien, von denen du gesprochen hast, gerade weil sie nicht expliziert oder standardisiert sind, besonders anfällig für den Einfall stereotyper Zuschreibungen und (unbewusste) stereotypenbasierte Bewertungen?**

Ich halte das zumindest für sehr plausibel. Man müsste das empirisch untersuchen. Darüber hinaus gibt es natürlich organisatorische Maßnahmen, die ergriffen werden könnten. Zum Beispiel, dass in jeder Prüfungskommission eine Frau vertreten ist – das würde den Geschlechtereffekt ja nivellieren. Gleichzeitig kann man Verfahrensvorschläge machen, um die Objektivität zu steigern. Es wird z.B. in der mündlichen Prüfung schon sehr unterschiedlich gehandhabt, wann die Einzelnotenvorschläge gemacht werden und eine Bewertung stattfindet, ob direkt im Anschluss an die einzelnen Prüfungen oder erst am Ende. Ebenso uneinheitlich wird gehandhabt, wer eigentlich in welcher Reihenfolge den Vorschlag macht. Einheitliche Vorgaben würden für mehr Objektivität sorgen.

**Welche institutionellen Player siehst Du besonders in der Pflicht?**

Das sind natürlich in erster Linie die Justizprüfungsämter (JPAs). Die unterstehen den Justizminister\*innen. Wenn es um kleinere Maßnahmen geht, wie die Gestaltung von Musterleitfäden für mündliche Prüfungen, wären die JPAs in der Pflicht. Das wäre auch – trotz knapper Ressourcen – eigentlich ziemlich leicht hinzubekommen und würde schon viel bringen. Man müsste daneben aber breit Fortbildungen für Prüfer\*innen anbieten. Auch die Hochschulen, die Professor\*innen könnten da vielleicht Druck aufbauen, weil sie eine Öffentlichkeit generieren können.

Wir als Prüfer\*innen könnten aber auch anfangen, darüber zu reden, wie wir prüfen: defizit- oder kompetenzorientiert? Was sind meine Kriterien? Wo bin ich voreingenommen? Es erhöht natürlich die Objektivität, wenn sich Prüfer\*innen darüber im

1 Towfigh, Emanuel/Traxler, Christian/Glöckner, Andreas (2014): Zur Benotung in der Examensvorbereitung und im ersten Examen. Eine empirische Analyse. In: Zeitschrift für Didaktik der Rechtswissenschaft (ZDRW) 2014, Heft 1, S. 8–27. Im Internet abrufbar unter: [http://www.zdrw.nomos.de/fileadmin/zdrw/doc/2014/Aufsat\\_ZDRW\\_14\\_01\\_Towfigh\\_u.a.pdf](http://www.zdrw.nomos.de/fileadmin/zdrw/doc/2014/Aufsat_ZDRW_14_01_Towfigh_u.a.pdf), vgl. Interview mit Prof. Towfigh auf S. XX dieser Ausgabe.

Klaren sind, was die Faktoren sind, die sie beeinflussen. Ich wäre nicht aufrichtig, wenn ich nicht sagen würde, dass das Geschlecht kein Faktor ist, der eine Rolle spielt. Sich diesen Prozessen zu stellen und diese Mechanismen zu reflektieren, sollte im Übrigen auch in der Ausbildung zum\* zur Richter\* in einer Rolle spielen.

Wir müssen uns im Klaren sein: Der Druck, der derzeit noch auf dem Staatsexamen liegt, ist riesig – zumindest solange es nicht flächendeckend einen Bachelor gibt. Wir sollten daher ein sorgfältiges Augenmerk auf die Qualität der Prüfung legen und ihre Bedeutung mitdenken, wenn es darum geht, sie zu verbessern.

#### Was kann der djb dabei tun?

Wir können uns einerseits im Hands-on-Bereich einbringen, also in die Prüfungsvorbereitung, indem wir z.B. Coachings anbieten für Kandidatinnen. Wir sollten auch in den Erfahrungsaustausch gehen: Wie bin ich eigentlich zu meinem Beruf gekommen, wie

bin ich beispielsweise als Juristin „ohne Prädikatsexamen“ erfolgreich geworden? Empowerment spielt eine große Rolle für die Frage nach dem Umgang mit verunsichernden Situationen.

Dann kann eine Aufgabe des Verbandes sein, die Qualität und Diskriminierungsfreiheit der juristischen Prüfung immer wieder auf der politischen Agenda zu halten. Die große Stärke des djb ist immer gewesen, dass er Juristinnen in der Breite der Tätigkeitsfelder und auch in der politischen Ausrichtung zusammenbringt – das ist genau das, was wir in der Frage der Prüfungen brauchen. Und natürlich sind alle Mitglieder aufgerufen, als Prüferinnen tätig zu werden und dafür zu werben. Man könnte auch Fortbildungen innerhalb des djb anbieten. Aber auch jede Einzelne kann die Frage nach den möglichen Verzerrungseffekten, der eigenen Voreingenommenheit bei einem Tee oder einem Glas Rotwein diskutieren. Schon dann würde etwas vorangehen.

#### Herzlichen Dank für das Gespräch!

DOI: 10.5771/1866-377X-2020-1-4

## „Ergebnisunterschiede sind Ausdruck eines fundamentalen gesellschaftlichen Problems“ – Interview mit Prof. Dr. Emanuel Towfigh

Das Interview führte **Nora Wienfort**, Wissenschaftliche Mitarbeiterin an der HU Berlin, Mitglied des Arbeitsstabs Ausbildung und Beruf und der Kommission Europa- und Völkerrecht.

**Lieber Herr Towfigh, Sie haben zusammen mit Christian Traxler und Andreas Glöckner 2018 eine Studie über Geschlechts- und Herkunftseffekte bei der Benotung juristischer Staatsprüfungen<sup>1</sup> veröffentlicht. Dieser Studie liegen die Examensnoten von fast 20.000 Personen aus NRW zugrunde. Sie konnten zeigen, dass Frauen und Menschen mit Migrationshintergrund in der mündlichen Prüfung im zweiten Examen im Vergleich schlechter abschneiden als männliche Kandidaten ohne Migrationshintergrund, die dieselben Abitur- und schriftlichen Vornoten haben. Wie erklären Sie sich das?**

Wichtig ist zunächst: Es handelt sich um Korrelationen; Kausalitäten haben wir nicht nachgewiesen. Dafür muss man genauer hinsehen und die hinter den Ergebnissen liegenden Mechanismen identifizieren. Es gibt verschiedene Erklärungsansätze: Es kann sich um (wahrscheinlich unbewusste) Diskriminierung in der Prüfung handeln. Es kann auch sein, dass Frauen und Kandidat\*innen mit Migrationshintergrund Opfer des sogenannten *stereotype threat* werden: Sie versuchen in



▲ (Foto: privat)

der Prüfung unbewusst, sich vermeintlich männlichen bzw. „deutschen“ Verhaltensmustern anzupassen. Das bindet kognitive Ressourcen, die für die Beantwortung der Prüfungsaufgaben fehlen. Das wäre dann also keine Diskriminierung durch die Prüfenden.

In einer früheren Studie haben wir gesehen, dass der Lernfortschritt von Frauen in der Examensvorbereitung im Durchschnitt langsamer ist als der von Männern.<sup>2</sup> Das spricht gegen eine Erklärung, die allein auf eine Diskriminierung durch die mündliche Prüfungskommission abstellt. Es gibt Erklärungsansätze, denen zufolge Frauen schlechter im Examen abschneiden, weil sie weniger ehrgeizig oder wettbewerblich seien als Männer. Manche meinen auch, dass Frauen im Abitur bevorzugt sind, also das Staatsexamen die „neutralere“ Prüfung sei. Meiner Meinung nach deuten die vielfältigen Befunde ungleicher Behandlung und Bewertung auf ein fundamentales gesellschaftliches Problem hin, dessen Wurzeln wahrscheinlich schon in der Sozialisation im Kindergarten, in der Schule und in der universitären Ausbildung liegen und das sich im Prüfungswesen besonders markant zeigt.

1 Towfigh/Traxler/Glöckner ZDRW 2018, 115.

2 Glöckner/Towfigh/Traxler InstrSci 41 (2013), 989.