

Geschlechts- und Herkunftseffekte bei der Benotung juristischer Staatsexamen*

Ergebnisse einer Studie für das Ministerium der Justiz des Landes Nordrhein-Westfalen

Andreas Glöckner, Professor für Psychologie, FernUniversität Hagen

Emanuel V. Towfigh, Professor für Öffentliches Recht, Empirische Rechtsforschung und Rechtsökonomik, EBS Universität Wiesbaden

Christian Traxler, Professor für Ökonomie, Hertie School of Governance

I. Hintergrund

Im Nachgang zu einer Studie aus dem Jahr 2014, die auf systematische Notenunterschiede im ersten juristischen Staatsexamen aufmerksam machte (Towfigh, Traxler, Glöckner 2014), hat das Ministerium der Justiz des Landes Nordrhein-Westfalen eine umfangreiche Untersuchung initiiert. Dazu wurde den Autoren ein umfangreicher Datensatz mit bis zu 20.000 Ergebnissen der ersten und zweiten juristischen Staatsprüfungen in NRW aus den Jahren 2006 bis 2016 zur Verfügung gestellt.

Untersucht wurde u.a., ob die ursprünglichen Befunde robust sind, ob sich in den Prüfungen (systematische) Geschlechts- und Herkunftseffekte zeigen, und ob sich dafür Ursachen benennen lassen, insbesondere ob sich Diskriminierung als Ursache belegen oder ausschließen lässt.

II. Datensatz

Vom Landesjustizprüfungsamt NRW (LJPA) wurde ein anonymisierter Datensatz bereitgestellt, der alle elektronisch erfassten Noten der zweiten juristischen Staatsprüfung der Abschlussjahrgänge 2006 bis 2016 in NRW abbildet.

Vor der Anonymisierung wurde der Datensatz im LJPA einer onomastischen Analyse unterzogen, bei der aus der Kombination aus Vor- und Nachnamen die Herkunftsregionen der Personen abgeleitet wurden. Gemeinsam mit Informationen zum Geburtsland bzw. zur Staatsangehörigkeit der Personen lassen sich damit Indikatoren für einen Migrationshintergrund bilden. Daneben werden auch noch das Geschlecht, das Alter sowie Informationen über die Zusammensetzung der Prüfungskommission ausgewertet.

Der Datensatz umfasst Noten von 19.883 Personen, die über 26.000 Prüfungsversuche im zweiten Examen durchliefen. 52,9% der Personen sind Frauen. 9,1% wurden im Ausland geboren, 2,8% haben keine deutsche Staatsbürgerschaft und für etwa 14% der Personen weist das Onomastik-Verfahren auf eine nicht-deutsche Herkunftsregion hin.

Die unten besprochenen Resultate für das zweite Examen basieren auf der Analysen der „relevanten“ Prüfungsversuche, d.h. entweder dem direkt bestandenen Versuch oder — falls dadurch eine Notenverbesserung erzielt wurde — dem relevanten Wiederholungs- bzw. Verbesserungsversuch.

Für etwa 10.000 Personen werden Noten aus dem ersten Examen, für etwa 4.600 Beobachtungen auch die Abiturnote berücksichtigt.

* Zusammenfassung der Studie „Empirische Untersuchung zur Benotung in der staatlichen Pflichtfachprüfung und in der zweiten juristischen Staatsprüfung in Nordrhein-Westfalen von 2006 bis 2016“, von Andreas Glöckner, Emanuel Towfigh und Christian Traxler. Die vollständige Studie (62 Seiten) kann hier heruntergeladen werden: https://www.justiz.nrw/JM/schwerpunkte/juristenausbildung/180331-v_fin-Abschlussbericht-korr1.pdf

III. Ergebnisse

1. Geschlechtsunterschiede

1.1 Noten im ersten Examen

Die zentralen Befunde von Towfigh et al. (2014) konnten – nun mit einer umfangreicheren Datenbasis – bestätigt werden:

Gesamtnoten. Die Abschlussnoten von Frauen sind durchschnittlich etwa 0,3 Punkte (3,6%) schlechter als jene von Männern. Zusätzliche Analysen zeigen auf, dass dieser Notenunterschied vor allem im staatlichen, nicht aber im universitären Teil der Prüfung entsteht.

Prädikatsnoten. Die Unterschiede in den durchschnittlichen Noten verbergen deutlich ausgeprägte Effekte am oberen Ende der Notenverteilung. So zeigt sich etwa, dass Frauen eine um 5,3 Prozentpunkte niedrigere Rate an „Prädikatsnoten“ (9,0 Punkte oder besser) haben (Männer: 31,5% Frauen: 26,2%). Mit anderen Worten haben Frauen relativ zu Männern eine etwa 17% geringere Rate an Prädikatsnoten.

Zur Relevanz der Notenschwellen: Einer empirischen Studie des Deutschen Instituts für Wirtschaftsforschung (DIW) hat aufgezeigt, dass bereits in den ersten Jahren nach dem Studium ein Gehaltsunterschied von nahezu 15% zwischen Personen mit einem Prädikatsexamen und solchen, die kein Prädikat erreichen konnten, auftritt.

Wenn in der Analyse Abiturnoten, Altersunterschiede sowie der Prüfungszeitpunkt berücksichtigt werden, d.h., wenn zwei Prüflinge mit gleichem Alter, gleicher Abiturnote, und gleichem Prüfungszeitpunkt, die sich jedoch im Geschlecht unterscheiden, verglichen werden, so sind die gemessenen Geschlechtsunterschiede stärker.

1.2 Noten im zweiten Examen

Gesamtnoten. Auch im zweiten Examen zeigen sich signifikante Geschlechtsunterschiede in den Abschlussnoten. Frauen schneiden durchschnittlich um 0,14 Notenpunkte (1,9%) schlechter ab als Männer.

Prädikatsnoten. Diese Zahl verbirgt jedoch deutlich stärkere Unterschiede hinsichtlich dem Erreichen von wichtigen Notenschwellen. So zeigt sich etwa, dass Frauen eine um 3,6 Prozentpunkte geringere Rate an Prädikatsexamen (Männer: 29,3%; Frauen: 25,7%). Das entspricht einem relativen Geschlechtsunterschied von über 12%. Den Schwellenwert für eine „gute“ Note (11,5 Notenpunkte) erreichen 4,7% der Männer, aber nur 3,2% der Frauen.

Wenn Unterschiede in Abiturnoten, im Alter oder im Prüfungszeitpunkt berücksichtigt werden, steigen auch hier die gemessenen Geschlechtsunterschiede weiter an. Es gibt jedoch keine Evidenz dafür, dass sich die Geschlechtsunterschiede zwischen dem ersten und dem zweiten Examen verstärken. D.h., die Notenunterschiede im zweiten Examen können im Wesentlichen über die im ersten Examen beobachteten Geschlechtsunterschiede erklärt werden.

In den mündlichen Prüfungsteilen wird ein etwas stärkerer Geschlechtsunterschied beobachtet (die durchschnittlichen Benotungen von Frauen fällt hier etwa 2,4% schlechter aus als für Männer) als in den anonymen, schriftlichen Prüfungen (1,9%).

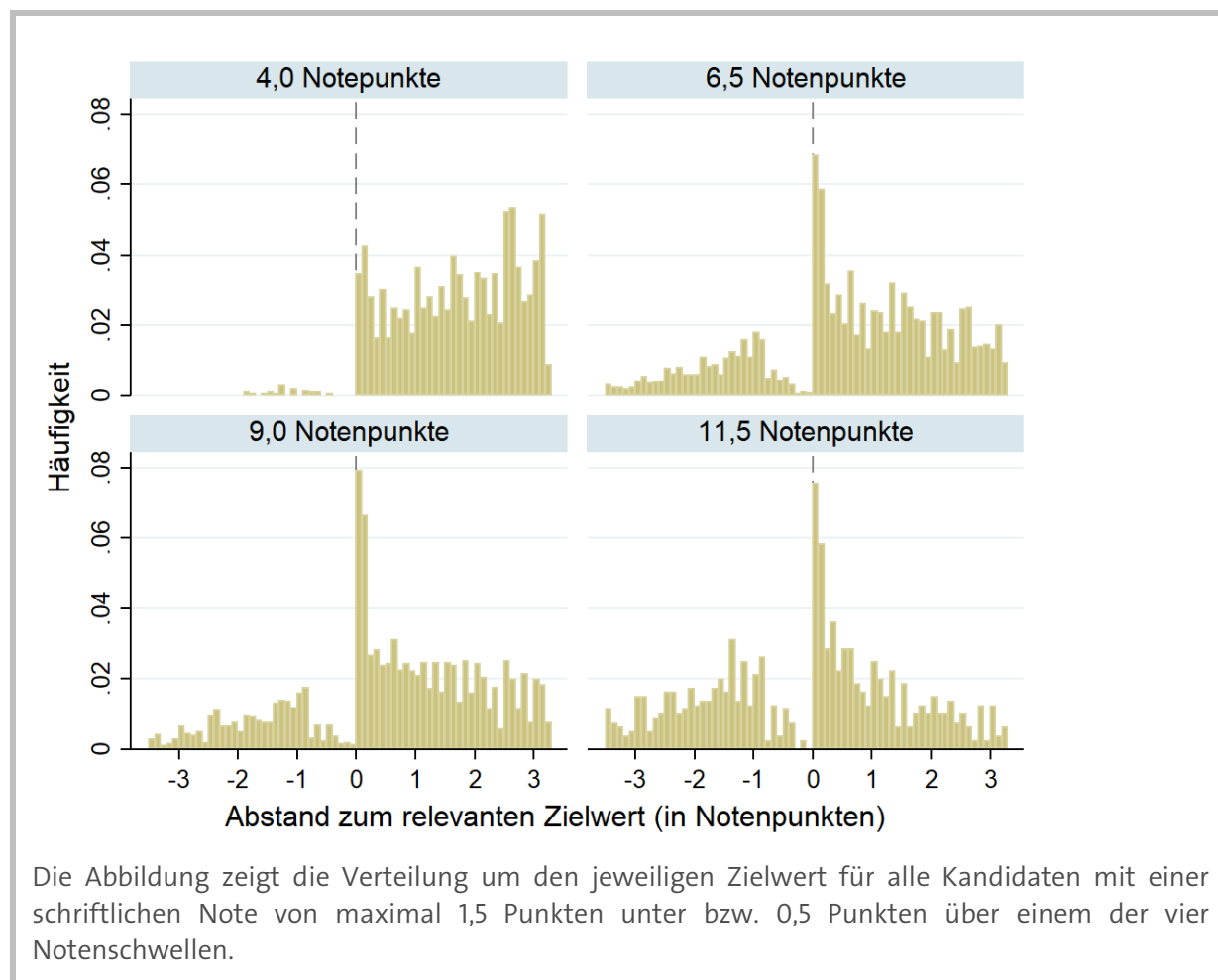
Bei den mündlichen Prüfungsteilen besteht – selbst nach Kontrolle für die Vornoten aus den schriftlichen Teilen (und/oder Kontrolle für die Noten aus dem ersten Examen) – ein statistisch signifikanter Geschlechtsunterschied: Vergleicht man zwei „gleich gute“ Rechtsreferendarinnen und Rechtsreferendare – gemessen anhand einer identischen Note im schriftlichen Prüfungsteil des zweiten Examins – so erzielt die Frau durchschnittlich eine um etwa 0,2 Notenpunkte schlechtere Note in den mündlichen Prüfungsteilen.

Eine genauere Analyse der schriftlichen Prüfungen in den unterschiedlichen Rechtsgebieten zeigt, dass die Geschlechtsunterschiede vor allem im Bereich Zivilrecht konzentriert sind.

1.3 Vornotenorientierte Notenvergabe und Zusammensetzung der Prüfungskommission

Ob zumindest eine weibliche Prüferin Teil der Kommission ist, hat *keinen* Einfluss auf die *durchschnittliche* Note in den mündlichen Prüfungsteilen – weder auf die Durchschnittsnote von weiblichen noch auf jene von männlichen Studierenden. Der Fokus auf die *durchschnittlichen* Noten verbirgt jedoch einen genauso systematischen wie substantiellen Effekt, der erst bei einer differenzierteren Betrachtung sichtbar wird. Dabei zeigt sich, dass die Notenvergabe in den mündlichen Prüfungen systematisch mit den Vornoten aus den schriftlichen Prüfungsteilen variiert:

Die Mitglieder der Prüfungskommissionen verwenden ihr Wissen über die schriftlichen Vornoten, um über Anpassungen der mündlichen Noten „eindeutige“ Gesamtnoten herzustellen – d.h. solche, die nicht knapp unter relevanten Notenschwellen (4,0 / 6,5 / 9,0 bzw. 11,5 Notenpunkte) liegen. Dieses Verhalten lässt sich am besten grafisch dokumentieren. Dazu wurden für alle Rechtsreferendarinnen und Rechtsreferendare die fiktiven „Zielwerte“ berechnet, die sie – gegeben ihre schriftlichen Noten – in der mündlichen Prüfung erzielen müssten, um die nächste relevante Notenstufe zu erreichen. In einem zweiten Schritt wird dann die Notenverteilung aus der mündlichen Prüfung dargestellt, normiert um eben jene spezifische Zielwerte. Der Nullpunkt stellt dabei exakt die mündliche Note dar, die für das Erreichen der jeweiligen Notenstufe notwendig wäre.



In diesem Schritt der vornotenorientierten Notenvergabe scheint die Dynamik innerhalb der Prüfungskommissionen hinsichtlich der Behandlung von Frauen und Männern von der Zusammensetzung der Kommissionen abzuhängen.

Die Studie dokumentiert deutliche Geschlechtsunterschiede in der Wahrscheinlichkeit, eine der vier Notenschwellen zu erreichen. Vergleichen man eine Rechtsreferendarin und einen Rechtsreferendar mit den *gleichen* schriftlichen Vornoten, so hat erstere eine um 2,3 Prozentpunkten *geringere* Wahrscheinlichkeit als ihr männlicher Kollege, die nächsten Notenschwelle zu erreichen, wenn die Prüfungskommission aus drei Männern gebildet wird. Am oberen Ende der Notenverteilung (d.h., bei den Schwellwerten von 9,0 und 11,5 Notenpunkten) ist dieser Unterschied noch deutlich stärker ausgeprägt: hier liegt - in ausschließlich mit Männern besetzten Prüfungskommissionen - der Geschlechtsunterschied bei etwa 6 Prozentpunkten.

Ist jedoch zumindest eine Frau Teil der Prüfungskommission, so verschwindet dieser Geschlechtsunterschiede. Relativ zu einer rein männlich besetzten Kommission erhöht sich die Wahrscheinlichkeit für Frauen, die nächste Notenschwelle zu erreichen, während die Wahrscheinlichkeit für Männern fällt. In (zumindest) teilweise „weiblichen“ Prüfungskommissionen verbleiben damit *keine* messbaren Geschlechtsunterschiede.

Bei den Prädikatsnoten steigt die Wahrscheinlichkeit für Frauen, die Notenschwellen 9,0 bzw. 11,5 Punkte zu erreichen, um etwas mehr als 3 Prozentpunkte, während die Wahrscheinlichkeit für Männer um 3 Prozentpunkte zurückgeht. Es kommt somit zu einer vollständigen Nivellierung.

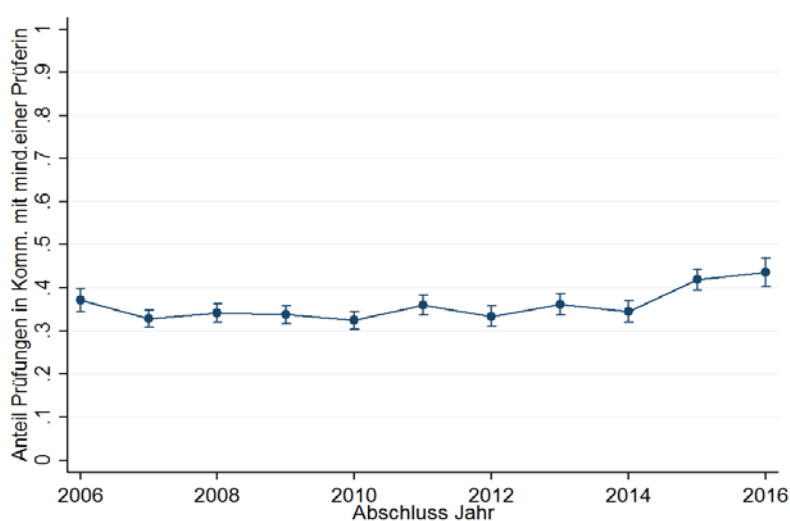
Diese Befunde bleiben weitgehend unverändert, wenn für zusätzliche Faktoren (etwa die Note aus dem ersten Examen) kontrolliert wird.

Bei diesen Ergebnissen gilt zu berücksichtigen, dass hier Frauen und Männer mit *gleichen* schriftlichen Vornoten verglichen werden. D.h., ein Teil der beobachteten Geschlechterunterschiede – nämlich die schlechteren schriftlichen Vornoten der Frauen – wird hier absorbiert („herausgerechnet“). Die schlechteren schriftlichen Noten sind jedoch sehr wohl eine zusätzliche Quelle für die beobachteten Geschlechtsunterschiede in der Wahrscheinlichkeit, eine bestimmte Notenschwelle zu erreichen. Der Effekt aus „männlichen“ Kommissionen ist eine *zusätzliche* Quelle für die Geschlechtsunterschiede.

Hängt der „Nivellierungs-Effekt“ von der Zahl der Frauen in der Kommission ab? Die vorliegenden Daten lassen es nicht zu, eine aussagekräftige Antwort zu geben, da kaum Beobachtungen mit mehr als einer Frau pro Kommission vorliegen: 65% der Prüfungen werden vor rein männlich besetzten, 30% vor Kommissionen mit genau einer Frau, 5% vor Kommissionen mit zwei Prüferinnen und nur 0,3% vor Kommissionen mit drei Prüferinnen abgelegt.

Der Anteil der mündlichen Prüfungen mit mindestens einer Frau in der Kommission ist in den Jahren 2015 und 2016 deutlich angestiegen. Vor dem Hintergrund der oben berichteten Befunde erscheint dieser Trend höchst begrüßenswert.

Berücksichtigt man jedoch, dass 52% der in der Studie analysierten Prüfungen von Rechtsreferendarinnen stammen, so erscheint es unverhältnismäßig, dass noch weit mehr als jede zweite Prüfungskommission mit drei männlichen Prüfern besetzt ist.



2. Herkunftseffekte

Methodisch muss festgehalten werden, dass die vorliegende Studie – im Unterschied zu Towfigh et al. (2014) – nicht nur onomastische Indikatoren zur Namensherkunft, sondern auch „harte“ Indikatoren für einen Migrationshintergrund (Geburtsland, Staatsbürgerschaft) auswerten konnte. Es wurde sowohl der jeweilige Einfluss dieser primären Indikatoren wie auch der Einfluss von abgeleiteten Indikatoren untersucht, die auf einen Migrationshintergrund der ersten bzw. einer späteren Migrations-Generation schließen lassen.

2.1 Noten im ersten Examen

Gesamtnoten. Die Analyse zeigt substantielle Notenunterschiede auf. Studierende mit einer nicht-deutschen Namensherkunft, einem nicht-deutschen Geburtsort und ohne deutsche Staatsangehörigkeit erreichen durchschnittlich eine Gesamtnote von lediglich 6,51 Punkten. Ein „deutscher“ Prüfling (d.h. mit Null-Werten in allen drei Indikatoren) erzielt durchschnittlich 7,93 Notenpunkte. Personen mit Migrationshintergrund erreichen also im ersten Examen durchschnittlich bis zu 18% (1,42 Punkte) schlechtere Noten.

Prädikatsnoten. Im Hinblick auf das Erreichen einer Prädikatsnote zeigt sich ein noch deutlicherer Unterschied. Personen, bei denen alle drei Indikatoren (Geburtsland, Staatsangehörigkeit, Namenherkunft) auf einen Migrationshintergrund schließen lassen, haben eine 25 Prozentpunkte niedrigere Wahrscheinlichkeit als „deutsche“ Referendare, eine Prädikatsnote zu erreichen. Die Wahrscheinlichkeit eine Prädikatsnote bei Personen mit Migrationshintergrund zu beobachten liegt damit um bis zu 80% niedriger als die „deutsche“ Vergleichsgruppe.

Weiterführende Analysen zeigen auf, dass die Notenunterschiede für Personen, die im Ausland geboren wurden und über keine deutsche Staatsbürgerschaft verfügen („1.Generation“), deutlich stärker ausgeprägt sind als für Personen, die zwar in Deutschland geboren wurden, aber entweder einen Namen ausländischen Ursprungs oder eine nicht-deutsche Staatsbürgerschaft haben („höhere Generation“). Die Resultate deuten damit auf einen positiven Integrationseffekt hin, der zu einer Verminderung der Notenunterschiede zu führen scheint.

2.2 Noten im zweiten Examen

Gesamtnoten. Im zweiten Examen zeigen sich ähnliche Notenunterschiede. Personen mit einem Migrationshintergrund haben um bis zu 17% (1,31 Punkte) schlechtere Note als „deutsche“ Rechtsreferendare. Selbst wenn man Rechtsreferendare mit identischen Noten im ersten Examen vergleicht, so erzielen jene mit einem nicht-deutschen Namen bzw. einem nicht-deutschen Geburtsort im zweiten Examen systematisch schlechtere Noten.

Prädikatsnoten. Konsistent mit den vorigen Ergebnissen zeigt sich auch hinsichtlich der Wahrscheinlichkeit, eine Note von 9,0 oder höher zu erreichen, ein starken Einfluss der unterschiedlichen Migrationsindikatoren. Die Wahrscheinlichkeit, eine Prädikatsnote zu erreichen, liegt für Personen mit Migrationshintergrund bei 6,7%. In der „deutschen“ Vergleichsgruppe ist dieser Wert um 24 Prozentpunkte höher (30,9%).

Bei der Differenzierung zwischen den schriftlichen und mündlichen Prüfungsteilen des zweiten Examens ergeben sich quantitativ unterschiedliche Befunde. In den schriftlichen Prüfungen erzielen Personen mit Migrationshintergrund um bis zu 26% schlechtere Noten, während der Unterschied in den mündlichen Prüfungen „nur“ 15% ausmacht.

Im Gegensatz zu den Resultaten für das erste Examen lässt sich kein systematischer „Vorteil“ der höheren gegenüber der ersten Migrantengeneration feststellen. Zwar deuten einige Befunde für die schriftlichen Noten in diese Richtung, jedoch findet sich hinsichtlich der schriftlichen Noten teilweise widersprüchliche Evidenz. Sollten sich positive Integrationseffekte in einem sichereren mündlichen Sprachgebrauch zeigen, so schlagen sich diese zumindest nicht in besseren mündlichen Noten im zweiten Examen nieder.

Um die Migrationseffekte zwischen den einzelnen Rechtsgebieten zu differenzieren, wurden die schriftlichen Teilnoten des zweiten Examens getrennt für diese untersucht. Die unterschiedlichen Analyseansätze liefern hier keine Evidenz für differenzierte Herkunftseffekte, die nach Rechtsgebieten unterschiedlich ausgeprägt sind.

IV. Interpretationen und Implikationen

1. Ursachen: Diskriminierung?

Die Befunde aus der Vorgängerstudie haben sich auch angesichts der erweiterten Datenbasis als robust erwiesen und bestätigt. Es werden quantitativ starke Notenunterschiede zwischen gesellschaftlich wichtigen Gruppen aufgezeigt. Anhand der empirischen Befunde lassen sich jedoch nur wenige Schlussfolgerungen zu den Ursachen der beobachteten Notenunterschiede ziehen. Zumindest kann Diskriminierung als Faktor für die beobachteten Bewertungsunterschiede aufgrund Geschlecht oder Herkunft nach wie vor nicht ausgeschlossen werden.

Insbesondere die Befunde zur Zusammensetzung der Kommission sprechen jedoch klar für eine (ggf. unbewusste bzw. „indirekte“) Diskriminierung von Frauen.

2. Zusammensetzung der Prüfungskommissionen

Die Befunde zum Einfluss der Zusammensetzung der Prüfungskommissionen lassen eine klare, kausale Schlussfolgerung zu. Die Partizipation von Prüferinnen ist hier wichtig für eine geschlechterneutrale Beurteilung. Damit gibt es eine klare Forderung: In jeder Prüfungskommission sollte zumindest eine Prüferin teilnehmen.

3. Einfluss der Vornoten

Die Diskontinuität in der Notenverteilung rund um die relevanten Notenschwellen, die über die vornotenorientierte Punktvergabe in der mündlichen Prüfung entsteht, bekommt im Kontext der ungleichen Wahrscheinlichkeiten für das Erreichen der Notenschwellen einen schalen Beigeschmack.

Eine weitere Anpassung im Prüfungsverfahren, die mit nur geringem Aufwand umsetzbare wäre, bestünde darin, die Noten aus den schriftlichen Prüfungen den Kommissionen der mündlichen Prüfungen nicht vorab mitzuteilen.

Eine Benotung der mündlichen Prüfung ohne vorherige Kenntnis der Ergebnisse aus der schriftlichen Prüfung könnte frei von strategischen Überlegungen der Prüfungskommission erfolgen. Die Aggregation von unabhängigen schriftlichen und mündlichen Bewertungen würde zu einem Ausgleich von Fehlern in der Beurteilung und damit auch zu valideren Bewertungen führen, die weniger Raum für Willkür lassen („Hat diese Kandidatin bzw. dieser Kandidat die nächste Notenstufe ‚verdient?‘“).

Weiterführende Literatur:

Emanuel V. Towfigh, Christian Traxler, und Andreas Glöckner, [Geschlechts- und Herkunftseffekte bei der Benotung juristischer Staatsprüfungen](#), Zeitschrift für Didaktik der Rechtswissenschaft 2018, S. 115-142.

Emanuel V. Towfigh, Christian Traxler, und Andreas Glöckner, [Zur Benotung in der Examensvorbereitung und im ersten Examen: Eine empirische Analyse](#), Zeitschrift für Didaktik der Rechtswissenschaft 2014, S. 8-27.