

**Stefan ZIMMERMANN<sup>1</sup>, Dietrich KLUSMANN & Wolfgang HAMPE  
(Hamburg)**

## **Angleichung von Schulnoten für die Studierendenauswahl**

### **Zusammenfassung**

Schulabschlussnoten sind prädiktiv für die Leistung in vielen Studienfächern, unterscheiden sich jedoch oftmals z. B. zwischen Schulformen oder Regionen. Das deutsche Bundesverfassungsgericht fordert vom Gesetzgeber aktuell eine Angleichung der bundeslandspezifischen Noten. Eine Umrechnung von Abiturnoten in bundeslandspezifische Prozentränge führt zur Angleichung der Zulassungschancen. Zusätzlich wird dadurch die prädiktive Validität des Schulerfolgs in Bezug auf den Erfolg im Hamburger Medizinstudium angeglichen. Prozentrangtransformationen können auch genutzt werden, um Abschlussnoten bei anderen Gruppen wie z. B. den Absolventinnen/Absolventen unterschiedlicher Studiengänge zu vergleichen.

### **Schlüsselwörter**

Schulabschlussnoten, Fairness, Studierendenauswahl, Studienerfolg, Medizinstudium

---

<sup>1</sup> E-Mail: [st.zimmermann@uke.de](mailto:st.zimmermann@uke.de)



## Adjusting school-leaving grades for student selection

### Abstract

School-leaving grades are predictive for study success in many subjects, although grading standards often differ between regions or school types. The German Constitutional Court compelled the legislature to implement a procedure for adjusting the grades between federal states in Germany. We propose using state-specific percentile ranks, as we can show that this homogenises the chance of admission between federal states. In addition, the percentile ranks also lead to a more similar predictive validity in terms of study success in undergraduate medicine in Hamburg. Furthermore, the percentile ranks would make it possible to adjust the grades of other groups whose grades are difficult to compare.

### Keywords

GPA, fairness, student selection, study success, medical studies

## 1 Ausgangssituation

Weltweit übersteigt in vielen Studienfächern die Zahl der Bewerber/innen die Studienplatzzahl. Wesentliches Kriterium bei der deshalb erforderlichen Bewerberauswahl ist meist die Studieneignung, wobei jedoch auch andere Kriterien wie z. B. eine angemessene Berücksichtigung unterschiedlicher Bewerbergruppen beachtet werden (PATTERSON et al., 2018). Für die Ermittlung der Studieneignung werden oft Vorleistungen z. B. aus der Schule, allgemeine oder studienfachspezifische Testverfahren oder Interviews herangezogen (HELL, TRAPMANN, & SCHULER, 2007; PATTERSON et al., 2016).

Schulabschlussnoten können Beurteilungen durch viele Lehrer/innen über einen Zeitraum von mehreren Jahren widerspiegeln. In vielen Staaten und über viele Studienfächer hinweg sind sie der beste Prädiktor für den Studienerfolg (TRAPMANN, HELL, WEIGAND & SCHULER, 2007), obwohl die Differenzierung der

Absolventinnen und Absolventen im oberen Leistungsbereich in den letzten Jahren durch häufigere Vergabe von Bestnoten eingeschränkt wird. Zusätzliche Probleme können sich durch eine Differenzierung des Schulsystems ergeben. So können die Leistungsanforderungen zwischen unterschiedlichen Schularten oder Regionen differieren. Ein Beispiel hierfür ist das Schulsystem in Großbritannien mit der Unterscheidung in staatliche und (teure) private Schulen. Absolventinnen und Absolventen staatlicher Schulen mit denselben, in nationalen A-Level-Examen gemessenen Schulnoten haben einen höheren Erfolg im darauffolgenden Medizinstudium (MCMANUS, DEWBERRY, NICHOLSON & DOWELL, 2013). Solche Probleme können dazu führen, dass, wie für das Medizinstudium in Österreich, Schulabschlussnoten trotz ihres Potenzials nicht für die Studienbewerberauswahl genutzt werden.

## **1.1 Unterschiede zwischen Abiturnoten deutscher Bundesländer**

In Deutschland sind die Abiturnoten dagegen das wichtigste Kriterium. Schulnoten erfassen – wie auch Kenntnistests – das in formalen Bildungsprozessen akkumulierte Wissen und die darauf basierenden Fertigkeiten. Sie sind in etwa gleich gut geeignet, den kognitiven Studienerfolg und die Studienadhärenz in den ersten Semestern der aktuellen Curricula vorherzusagen (PATTERSON et al., 2016; RINDERMAN & OUBAID, 1999; TRAPMANN et al., 2007). Beim Zugang zum besonders begehrten Medizinstudium werden die Studienplätze auf unterschiedliche Quoten aufgeteilt: 20 % werden ausschließlich aufgrund der Schulnoten, 60 % der Plätze aufgrund anderer Kriterien wie Tests und Interviews in Kombination mit der Abiturnote, die dabei jedoch das höchste Gewicht haben muss, vergeben (HAMPE et al., 2009; SCHWIBBE et al., 2018). In Deutschland hat das Bundesverfassungsgericht Ende 2017 wesentliche Teile der Studierendenauswahlgesetze für nicht verfassungskonform erklärt. Unter anderem wird ein Ausgleichsmechanismus für Abiturnoten aus den Bundesländern gefordert, deren Schulsysteme und -gesetze sich unterscheiden (BUNDESVERFASSUNGSGERICHT, 2017).

## 1.2 Ungleiche Notenverteilungen

Problematisch sind die großen Unterschiede zwischen den Bundesländern in der Vergabe der Noten und in den Abiturientenquoten (Abb. 1): 2015 lag der Anteil der Schulabgänger/innen mit Hochschulzugangsberechtigung bezogen auf die gleichaltrige Bevölkerung zwischen 32 % und 58 % (STATISTISCHES BUNDESAMT, 2017).

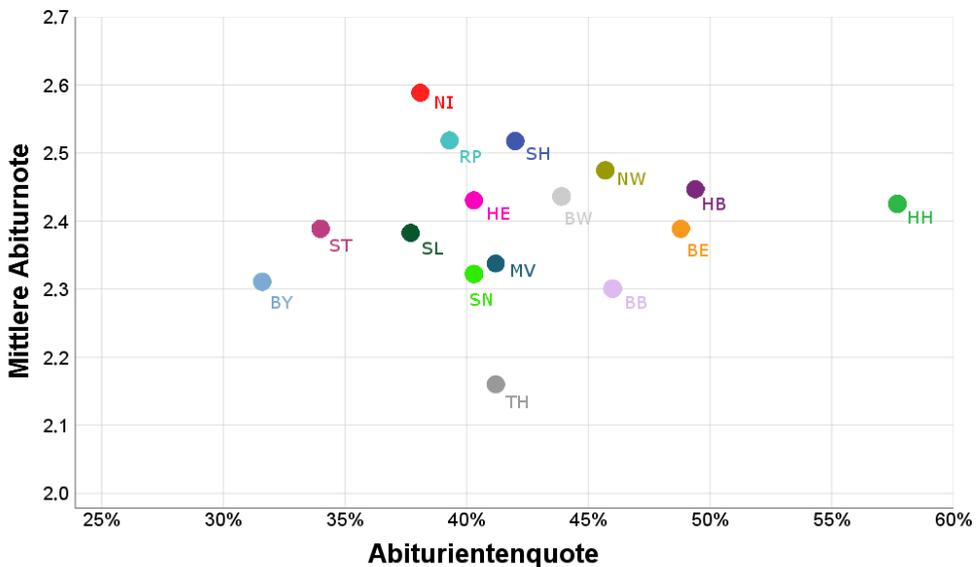


Abb. 1: Mittlere Abiturdurchschnittsnoten und Abiturientenquoten der 16 Bundesländer für das Schuljahr 2015/2016 in Deutschland.

BB: Brandenburg, BE: Berlin, BW: Baden Württemberg, BY: Bayern, HB: Bremen, HE: Hessen, HH: Hamburg, MV: Mecklenburg-Vorpommern, NI: Niedersachsen, NW: Nordrhein-Westfalen, RP: Rheinland-Pfalz, SH: Schleswig-Holstein, SL: Saarland, SN: Sachsen, ST: Sachsen-Anhalt, TH: Thüringen.

Die Bestnote 1.0 erreichten 2015 länderspezifisch zwischen 0.8 % und 3.1 % der Abiturientinnen und Abiturienten (KULTUSMINISTERKONFERENZ, 2016). Es ist schwer einzuschätzen, in welchem Ausmaß diese Unterschiede die Qualität der schulischen Bildung oder unterschiedliche gesetzliche Regelungen des Zugangs und der Leistungsbewertung widerspiegeln.

### **1.3 Einflussfaktoren auf die Notengebung**

Die 16 Bundesländer unterscheiden sich in ihren Bildungssystemen zum Teil sehr deutlich: Einerseits unterscheidet sich die Schulstruktur, die mehr oder weniger Wege zur Hochschulzulassung erlaubt und den Übergang auf eine entsprechende Sekundarschule liberaler oder strenger reguliert (TILLMANN, 2013). Andererseits ist die gymnasiale Oberstufe teilweise sehr unterschiedlich ausgestaltet, da den Schülerinnen und Schülern Kurswahlfreiheiten in unterschiedlichem Ausmaß gewährt werden (HELBIG & NIKOLAI, 2015).

Welche Faktoren führen zu einer ungleichen Notenvergabe zwischen den Bundesländern? Schulnoten stellen immer auch eine relative Bewertung dar, da die Lehrerinnen und Lehrer insbesondere auch immer einen pädagogischen Ermessensspielraum in der Bewertung haben, wobei dieser in den föderalen Schulgesetzen geregelt wird. Neben einer kriterienorientierten Bewertung, die sich z. B. an den Bildungsstandards orientiert, können auch intraindividuelle oder interindividuelle Unterschiede Eingang in die Bewertung finden. Das führt dann möglicherweise dazu, dass bestehende Leistungsunterschiede zwischen den Bundesländern, wie sie bspw. in den PISA-Studien im Bereich der naturwissenschaftlichen (RÖNNEBECK, SCHÖPS, PRENZEL & HAMMANN, 2008) bzw. mathematischen (FREY, ASSEBURG, EHMKE & BLUM, 2008) Kompetenz festgestellt werden, sich nicht unmittelbar in Schulnotenunterschieden niederschlagen. Das zeigt auch ein Vergleich mittels standardisierter Schulleistungstests: Die gleiche Schulleistung in den Fächern Mathematik und Englisch führt in Abhängigkeit des Bundeslandes zu einer unterschiedlichen Note, auch wenn dieses Phänomen weniger stark beim „Zentralabitur“ zum Tragen kommt (NEUMANN, NAGY, TRAUTWEIN & LÜDTKE, 2009).

Aktuell versuchen die Bundesländer die Anforderungen an das Abitur anzugleichen (INSTITUT ZUR QUALITÄTSENTWICKLUNG IM BILDUNGSWESEN, 2018; KULTUSMINISTERKONFERENZ, 2017). Durch gemeinsame Bildungsstandards und einen Pool von gemeinsam entwickelten Abituraufgaben soll sich die Notenvergabe nicht nur im Abitur, sondern auch in den letzten beiden Schuljahren angleichen, deren Ergebnisse großen Anteil an der Abiturdurchschnittsnote haben. Die formative Evaluation dieses Ansatzes durch das Institut zur Qualitätsentwicklung im Bildungswesen steht noch aus.

Um aktuell die Abiturnoten aus unterschiedlichen Bundesländern zu vergleichen, muss daher ein Ausgleichsmechanismus eingesetzt werden, da andernfalls davon auszugehen ist, dass die Prognosefairness der Abiturnote verletzt wird. Dies wird z. B. daran deutlich, dass in den USA die vorhergesagte Studienleistung von Schülerinnen und Schülern an schlechter ausgestatteten Schulen überschätzt wird, während die vorhergesagte Studienleistung von Schülerinnen und Schülern an besser ausgestatteten Schulen unterschätzt wird (ZWICK & HIMELFARB, 2011).

## 2 Fairness von Abiturnotenausgleichsmechanismen

Im Hinblick auf die Fairness von Auswahltests unterscheiden bereits WOTTAWA & AMELANG (1980) drei verschiedene „Fairnessdefinitionen“:

- 1) Die mittleren Ergebnisse relevanter Gruppen sollen gleich sein.
- 2) Unter Kontrolle des Fähigkeitsniveaus soll sich kein Unterschied zwischen relevanten Gruppen zeigen (kein *differential test functioning* soll vorliegen).
- 3) Die Vorhersage des Studienerfolgs soll innerhalb relevanter Gruppen gleich sein (Prognosefairness).

In der wissenschaftlichen Literatur wird der erste Definitionsansatz als naiv bewertet (AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN

PSYCHOLOGICAL ASSOCIATION & NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION, 2014; WOTTAWA & AMELANG, 1980). Unterschiede zwischen Minoritäts- und Majoritätsgruppen können Ausgangspunkt für eine gründliche Untersuchung der Testfairness sein, man kann jedoch nicht automatisch schlussfolgern, dass ein Test unfair ist. Unter Testfairness wird zudem verstanden, dass das intendierte Konstrukt für alle Personen gleichermaßen gemessen wird und gleichzeitig ein Test derart gestaltet wird, so dass konstruktirrelevante Aufgabenmerkmale, die bestimmte Gruppen bevorzugen bzw. benachteiligen, möglichst vermieden werden (AMERICAN EDUCATIONAL RESEARCH ASSOCIATION et al., 2014; MISLEVY et al., 2013). Soll ein Test bspw. Intelligenz erfassen, sollten die Testteilnehmer/innen den Test jeweils in ihrer Muttersprache absolvieren. Andernfalls wird die Validität der Interpretation der Testwerte korrumpiert, da dann für eine Minoritätsgruppe von niedrigen Ergebnissen fälschlicherweise auf niedrige Intelligenz geschlossen wird. Ansätze wie *universal design* und *evidence centered design* können Hilfestellung geben, so dass die Aufgabenbearbeitung und die Verständlichkeit für einen breiten Kreis an Testteilnehmer zugänglich werden, ohne dass sich das zu messende Konstrukt ändert. Möglicherweise sind für eine faire Messung von Menschen mit Lernbehinderung aber auch darüberhinausgehende Testanpassungen, wie bspw. die Verlängerung der Testzeit, notwendig, die dann möglicherweise auch das Konstrukt leicht verändern. In diesen Fällen muss die Testanpassung und deren Fairness empirisch untersucht werden.

Die Abiturnoten selbst erfüllen zumindest offensichtlich die naive Fairnessdefinition nicht, so dass eine vertiefte Analyse unternommen werden sollte: Frauen, Personen ohne Migrationshintergrund, Kinder von Akademikerinnen/Akademikern und, wie oben gezeigt, Menschen aus einigen Bundesländern haben im Mittel bessere Abschlüsse und erreichen häufiger das Abitur (BLOSSFELD et al., 2009). Ausgleichsmechanismen können auf unterschiedliche Ausgangsgruppen angewendet werden. Hier beschreiben wir eine Angleichung der Abiturnoten aus unterschiedlichen deutschen Bundesländern.

## 2.1 Bundeslandspezifische Ranglisten

Um einen Vergleich der Abiturnoten aus unterschiedlichen Bundesländern zu vermeiden, können bundeslandspezifische Ranglisten gebildet werden. Dieses Verfahren wird aktuell für die Studienplatzvergabe in der Abiturbestenquote in Deutschland genutzt. Hierbei wird zunächst die Gesamtzahl der verfügbaren Studienplätze auf die Bundesländer aufgeteilt. Dabei wird mit einem Gewicht von einem Drittel die Bewerberzahl und mit einem Gewicht von zwei Dritteln die Einwohnerzahl des jeweiligen Bundeslandes und somit indirekt die Abiturientenquote berücksichtigt. Die Bewerber/innen werden anschließend aufgrund ihrer Abiturnoten in bundeslandspezifische Listen eingereiht. Entsprechend der dem Bundesland zustehenden Studienplatzzahl erhalten die ranghöchsten Bewerber/innen auf jeder Bundeslandliste eine Zulassung. So konkurrieren jeweils nur Bewerber/innen aus einem Bundesland miteinander um die ihrem Bundesland zugeteilten Plätze. Die Chance auf einen Studienplatz hängt nicht von der absoluten Abiturnote ab, sondern vom Rangplatz dieser Abiturnote innerhalb eines Bundeslandes.

Dieses Verfahren hat sich für die deutschlandweite Vergabe von Plätzen in der Abiturbestenquote bewährt und ist verfassungskonform. Bei einer Bewerbung an einzelnen Hochschulstandorten, wie sie für 60 % der Medizinstudienplätze im „Auswahlverfahren der Hochschulen“ erfolgt, ist sie jedoch nicht übertragbar, da sich hier die Anteile von Bewerberinnen/Bewerbern aus einzelnen Bundesländern z. B. aufgrund der Entfernung zu der einzelnen Hochschule sehr stark unterscheiden. Eine Aufteilung der Studienplätze anhand der Einwohnerzahl eines Bundeslandes würde daher zu starken Verzerrungen führen. Eine Aufteilung nur nach hochschulspezifischen Bewerberzahlen würde dagegen die unterschiedlichen Abiturientenquoten der Bundesländer nicht berücksichtigen.

## 2.2 Prozentrangtransformation

Wir schlagen eine andere Regelung vor, die in der psychometrischen Nomenklatur als Equiperzentil Linking bezeichnet wird (KOLEN, 1984; KOLEN & BRENNAN, 2014; LIVINGSTON, 2004; RYAN & BROCKMANN, 2009): Eine

voraussetzungsvolle Annahme ist, dass sich die objektive Leistungsverteilung der Schülerinnen und Schüler zwischen den Bundesländern nicht unterscheidet. Um die Leistungen der Abiturientinnen und Abiturienten zu vergleichen, werden in der Folge bundeslandspezifische Prozentränge gebildet. Dabei wird für jeden Schulabgänger ein Prozentrang zwischen 0 und 100 ermittelt, der angibt, wie viele andere Schüler/innen der gesamten Jahrgangskohorte<sup>2</sup> des Bundeslandes eine bessere Abiturnote aufweisen. In jedem Bundesland wird den 1.0er- Abiturientinnen/Abiturienten der Prozentrang 0 zugeordnet. Abhängig von der Anzahl der 1.0er- Abiturientinnen und Abiturienten unterscheidet sich der Prozentrang der 1.1er- Abiturientinnen/Abiturienten: 2015 lag er zwischen 0.3 % im „strengen“ Niedersachsen und 1.3 % im „großzügigen“ Hamburg. Die 4.0er- Abiturientinnen/Abiturienten erhalten den Prozentrang, der der Abiturientenquote ihres Bundeslandes entspricht (Abb. 2).

Bei dieser Transformation werden zunächst wie bei den bundeslandspezifischen Ranglisten die Abiturientinnen/Abiturienten innerhalb eines Bundeslandes in eine Rangreihenfolge gebracht. Anschließend wird ihnen dann jedoch ein Prozentrang zugewiesen, der anstelle der Abiturnote für die Auswahlverfahren an den einzelnen Hochschulen verwendet werden kann. Weil in die Berechnung des Prozentranges die Gesamtzahl der Einwohner/innen einer Jahrgangskohorte der Bundesländer eingeht, bildet er indirekt auch die Abiturientenquote ab. Bei der Verwendung der so gebildeten Prozentränge erhält somit derselbe Anteil der Einwohner eines Bundeslandes und nicht derselbe Anteil der Abiturienten die Chance auf einen Studienplatz.

Die Zuordnung von Prozenträngen zu Abiturnoten unterscheidet sich zwischen den Bundesländern: in „großzügigen“ Ländern, die häufig sehr gute Noten vergeben

---

<sup>2</sup> Würde man sich hingegen auf den Anteil der Abiturientinnen und Abiturienten beziehen, müsste man postulieren, dass in Hamburg im Vergleich zu Bayern relativ fast doppelt so viele geeignete Schülerinnen und Schüler vorhanden sind (siehe Abbildung 1). Ergebnisse aus Bildungsstudien lassen diesen Schluss nicht zu.

oder hohe Abiturientenquoten aufweisen, führt dieselbe Abiturnote zu schlechteren Prozenträngen als in „strengen“ Bundesländern mit einer restriktiveren Noten- und Abiturvergabe. So liegt der Prozentrang für die Abiturnote 1.1 zwischen 0.3 % im „strengen“ Niedersachsen und 1.3 % im „großzügigen“ Hamburg, für die Note 1.8 lag die Spanne zwischen 4.2 % und 9.7 %. Der Anteil der Abiturientinnen/Abiturienten mit diesen Noten an der Gesamtbevölkerung unterschied sich also jeweils etwa um den Faktor 3.

### **2.3 Chancengleichheit in der Vorauswahl**

Um die Auswirkungen dieser Transformation auf die Studierendenauswahl abzuschätzen, haben wir sie auf die Bewerber/innen für das Medizinstudium in Hamburg im Jahr 2015 angewendet. In diesem Jahr bewarben sich 2420 Bewerber/innen mit 1. Ortspräferenz auf die 212 Studienplätze, die im Auswahlverfahren der Hochschule vergeben wurden. Hiervon wurden 1304 Bewerber/innen mit einer Abiturnote von 1.8 oder besser zum Vor-Ort-Auswahlverfahren vorausgewählt, das den Naturwissenschaftstest HAM-Nat (HISSBACH, KLUSMANN & HAMPE, 2011; ZIMMERMANN, KLUSMANN & HAMPE, 2017) und, für einige Bewerber/innen in einer 2. Stufe, das Multiple Mini-Interview HAM-Int umfasst (HISSBACH, SEHNER, HARENDZA & HAMPE, 2014; KNORR et al., 2018).

Von den Bewerberinnen/Bewerbern stammten 206 aus Niedersachsen und 585 aus Hamburg. Bei der Vorauswahl wurden 2015 alle Bewerber/innen bis zu einer Note von 1.8 berücksichtigt. In Niedersachsen entsprach die Abiturnote 1.8 dem Prozentrang 4.2 %, in Hamburg dem Prozentrang 9.7 %.

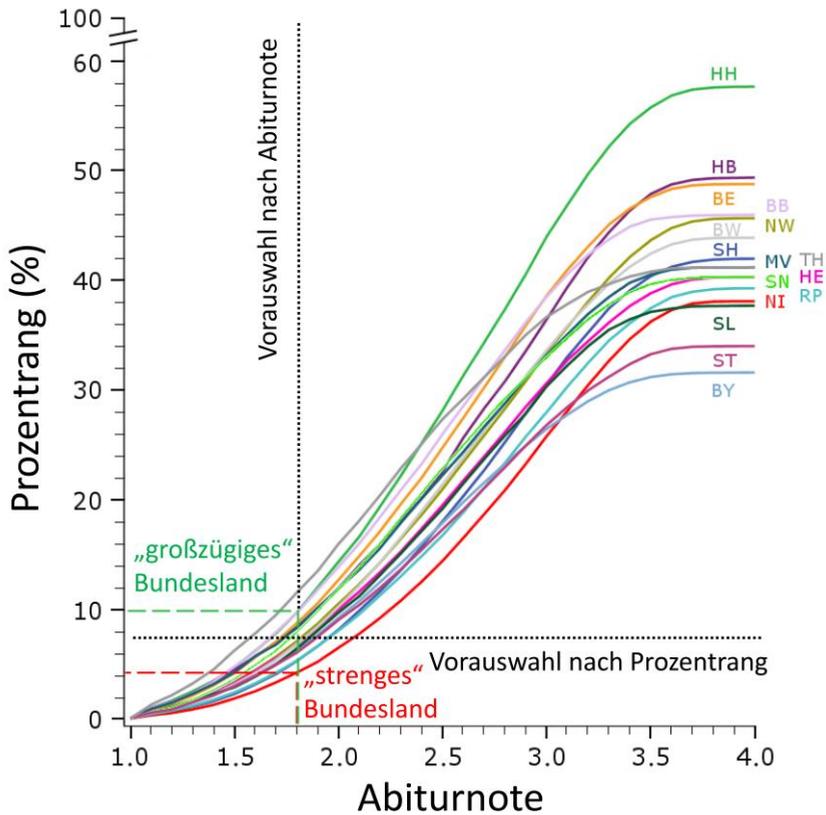


Abb. 2: Zuordnung von Prozenträngen zu Abiturnoten für die 16 Bundesländer.  
 BB: Brandenburg, BE: Berlin, BW: Baden Württemberg, BY: Bayern, HB: Bremen, HE: Hessen, HH: Hamburg, MV: Mecklenburg-Vorpommern, NI: Niedersachsen, NW: Nordrhein-Westfalen, RP: Rheinland-Pfalz, SH: Schleswig-Holstein, SL: Saarland, SN: Sachsen, ST: Sachsen-Anhalt, TH: Thüringen.

Mit dem vorgeschlagenen Verfahren der Vorauswahl nach bundeslandspezifischen Prozenträngen wären dagegen alle Bewerber/innen bis zu einem bestimmten Prozentrang, in diesem Falle 7.8 %, berücksichtigt worden. Aus Niedersachsen wären dann 197 statt 129 Bewerber/innen eingeladen worden (+53 %), aus Hamburg nur 225 statt 334 (-33 %). Die Grenznote hätte dann für das Niedersachsen bei 2.1 gelegen und für Hamburg bei 1.7. Wenn man über die Prozenträge und nicht über die Noten die Chancen auf einen Studienplatz verteilt, kann dies für einen Studienort mit vielen Bewerberinnen/Bewerbern aus sehr strengen und sehr großzügigen Bundesländern also zu erheblichen Verschiebungen führen. Für die meisten Bundesländer, die eher durchschnittliche Abiturientenquoten und Abiturnoten aufweisen, und die meisten auswählenden Universitäten mit Bewerberinnen/Bewerbern aus überwiegend durchschnittlichen Bundesländern werden die Auswirkungen jedoch mäßig sein.

## **2.4 Vorhersage des Studienerfolgs**

Am Beispiel von 772 Studierenden des Hamburger Medizinstudiums, die in den Jahren 2012-2015 zugelassen wurden, haben wir untersucht, wie die Korrelation der Abiturnote mit dem Studienerfolg in den ersten Semestern vom Bundesland abhängt, in dem das Abitur erworben wurde. Sowohl für Abiturientinnen und Abiturienten aus dem „strengen“ Niedersachsen (N=116) als auch aus dem „großzügigen“ Hamburg (N=219) korrelierte die Abiturnote mit dem Studienerfolg (Abb. 3a). Allerdings hatten Bewerber/innen aus Niedersachsen bei identischer Abiturnote von z. B. 1.8 einen höheren Studienerfolg. Für denselben Studienerfolg müssten Schüler/innen aus Hamburg eine Abiturnote mitbringen, die um 0.4 Notenschritte besser ist: 1.4. Wenn wir in der Analyse die Abiturnote durch den bundeslandspezifischen Prozentrang ersetzen, dann verringert sich dieser Unterschied und ist nicht mehr signifikant: Gleiche Prozenträge korrespondieren mit gleichen Studienleistungen (Abb. 3b). Ein bestimmter Prozentrang repräsentiert auch über die Landesgrenzen hinweg Schulleistung, die zum gleichen Studienerfolg führt.

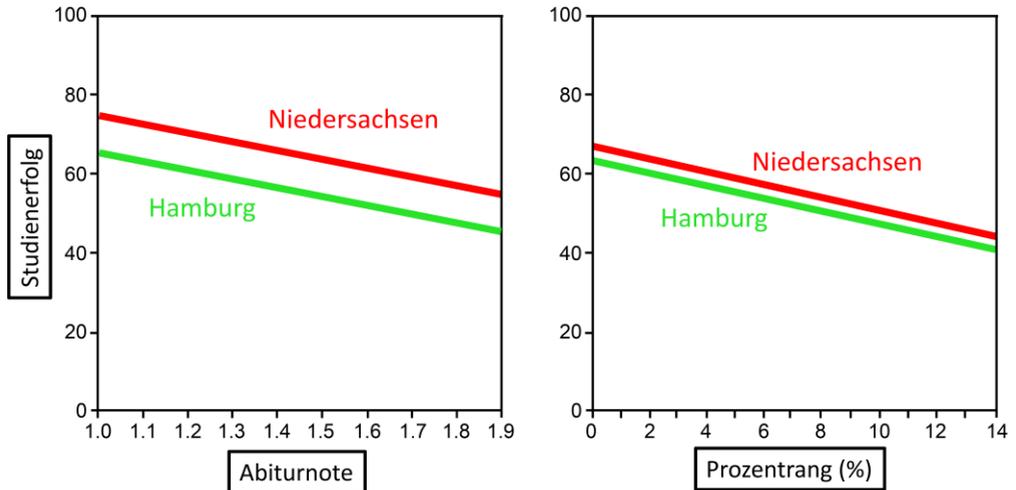


Abb. 3: Regressionsgeraden von a) bundeslandspezifischer Abiturnote und b) Prozentrang mit dem Erfolg im Medizinstudium (durchschnittliche Punktzahl der Module in den ersten drei Semestern) am Universitätsklinikum Hamburg-Eppendorf.

### 3 Diskussion

Unsere Ergebnisse sprechen dafür, dass das hier vorgestellte Prozentrangverfahren der Prognosefairness (AMERICAN EDUCATIONAL RESEARCH ASSOCIATION et al., 2014; WOTTAWA & AMELANG, 1980) besser entspricht als die Abiturnote: Sowohl die Chancen als auch die Vorhersage des Studiererfolgs werden für Bewerber/innen aus unterschiedlichen Bundesländern angeglichen. Andernfalls unterschätzt man systematisch die vorhergesagte Studienleistung von Bewerberinnen/Bewerbern aus Bundesländern mit strengen Bewertungsmaßstäben, während

man die Studienleistung von Bewerberinnen/Bewerbern aus Bundesländern mit großzügigeren Bewertungsmaßstäben systematisch überschätzt.

Wesentlich für den Erfolg des Verfahrens ist der Einbezug der Abiturientenquoten, der durch den Bezug des Prozentranges auf die Einwohnerzahl eines Bundeslandes erreicht wird. Es ist möglich, ähnlich wie bei der vorgestellten Ranglistenbildung eine Bezugsgröße zu wählen, die zu einem Drittel auf dem Bewerberanteil und zu zwei Dritteln auf der Einwohnerzahl eines Bundeslandes beruht. Solch ein Verfahren würde dem in Deutschland bereits akzeptierten Angleichungsverfahren für die Abiturnoten in der Abiturbestenquote entsprechen und ebenfalls die Abiturientenquoten berücksichtigen.

Das Prozenrangverfahren kann auch eingesetzt werden, um Unterschiede der Leistungsbewertung zwischen anderen Gruppen auszugleichen. Denkbar sind z. B. die separate Bildung von Ranglisten und damit Prozenträgen für unterschiedliche Schulformen oder die beiden Geschlechter. Nicht sinnvoll ist es dagegen, dieses Verfahren auf kleine Bewerbergruppen anzuwenden, um z. B. Unterschiede der Notengebung zwischen einzelnen Schulen, gewählten Kursen oder Klassenlehrern auszugleichen, da hier die Gruppengrößen für eine zuverlässige Statistik nicht ausreichen. Das Problem der Vergleichbarkeit von Abiturnoten betrifft nicht nur das Medizinstudium, sondern auch andere Studienfächer mit einem großen Bewerberüberhang wie z. B. Zahnmedizin oder Psychologie. Auch dort könnten die Prozenträge anstelle der Abiturnoten für die Studierendenauswahl herangezogen werden.

Auch im Bereich der Zulassung zum Masterstudium spielt die Vergleichbarkeit von Noten eine wichtige Rolle. Hier kann sich die Notengebung in Bachelorstudiengängen an unterschiedlichen Standorten unterscheiden. Insbesondere weisen verschiedene Fachdisziplinen unterschiedliche Bewertungstraditionen auf (GAENS, 2015; MÜLLER-BENEDICT & GAENS, 2015; MÜLLER-BENEDICT & TSAROUHA, 2011), so dass bei Bewerbungen auf interdisziplinäre Masterstudiengänge Absolventinnen und Absolventen bestimmter Bachelorstudiengänge möglicherweise bevorzugt werden. Auf der anderen Seite wirkt die Angleichung der Noten immer nivellierend, auch falls Institutionen verschiedener Qualität mit-

einander verglichen werden. Das Projekt Egracons (European Grade Conversion System, <http://egracons.eu>) benutzt einen ähnlichen Ansatz, um die Wertigkeit von Noten zu kommunizieren: Zusätzlich zur vergebenen Note soll auch immer die Verteilung der Noten auf dem Zeugnis ausgewiesen werden, indem der jeweilige prozentuale Anteil mitgeteilt wird.

### 3.1 Fazit

Gerechtigkeit bei der Studierendenauswahl ist nicht einfach zu definieren, da unterschiedliche Aspekte wie Chancengleichheit, die Anerkennung bisheriger Leistung und das Potenzial zu zukünftiger Leistung abgewogen werden müssen. Wir halten die Verwendung von gruppenspezifischen Prozenträngen für gerechter als die der Abiturnoten selbst, da hierdurch Unterschiede der Bewertung bisheriger Leistungen z. B. von aus unterschiedlichen Bundesländern stammenden Bewerberinnen/Bewerbern angeglichen werden.

## 4 Literaturverzeichnis

**American Educational Research Association, American Psychological Association, & National Council on Measurement in Education** (2014).

*Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

**Blossfeld, H.-P., Bos, W., Lenzen, D., Hannover, B., Müller-Böling, D., Prenzel, M. & Wößmann, L.** (2009). *Geschlechterdifferenzen im Bildungssystem – die Bundesländer im Vergleich: Fakten und Daten zum Jahresgutachten 2009*.

München: Vereinigung der Bayerischen Wirtschaft.

**Bundesverfassungsgericht** (2017). Urteil des Ersten Senats vom 19. Dezember. [https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2017/12/ls20171219\\_1bvl000314.html](https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2017/12/ls20171219_1bvl000314.html), Stand vom 12. Juli 2018.

**Frey, A., Asseburg, R., Ehmke, T. & Blum, W.** (2008). Mathematische Kompetenz im Ländervergleich. In PISA-Konsortium Deutschland (Hrsg.), *PISA*

2006 in Deutschland: Die Kompetenzen der Jugendlichen im dritten Ländervergleich (S. 127-148). Münster: Waxmann.

**Gaens, T.** (2015). Noteninflation an deutschen Hochschulen – Werden die Examennoten überall immer besser? *Beiträge zur Hochschulforschung*, 37(4), 8-35.

**Hampe, W., Hissbach, J., Kadmon, M., Kadmon, G., Klusmann, D. & Scheutzel, P.** (2009). Wer wird ein guter Arzt? Verfahren zur Auswahl von Studierenden der Human- und Zahnmedizin. [Who will be a good physician?]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 52(8), 821-830. <https://doi.org/10.1007/s00103-009-0905-6>

**Helbig, M., & Nikolai, R.** (2015). *Die Unvergleichbaren. Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Bad Heilbrunn: Klinkhardt.

**Hell, B., Trapmann, S. & Schuler, H.** (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. [A metaanalytic investigation of discipline-specific admission tests in German-speaking countries]. *Empirische Pädagogik*, 21(3), 251-270.

**Hissbach, J., Klusmann, D., & Hampe, W.** (2011). Dimensionality and predictive validity of the HAM-Nat, a test of natural sciences for medical school admission. *BMC Medical Education*, 11(83).

**Hissbach, J., Sehner, S., Harendza, S. & Hampe, W.** (2014). Cutting costs of multiple mini-interviews – changes in reliability and efficiency of the Hamburg medical school admission test between two applications. *BMC Medical Education*, 14.

**Institut zur Qualitätsentwicklung im Bildungswesen** (2018). Gemeinsame Abituraufgabenpools der Länder. <https://www.iqb.hu-berlin.de/abitur>, Stand vom 16. April 2018.

**Knorr, M., Schwibbe, A., Ehrhardt, M., Lackamp, J., Zimmermann, S. & Hampe, W.** (2018). Validity evidence for the Hamburg multiple mini-interview. *BMC Medical Education*, 18(1).

- Kolen, M. J.** (1984). Effectiveness of Analytic Smoothing in Equipercetile Equating. *Journal of Educational Statistics*, 9(1), 25-44.
- Kolen, M. J. & Brennan, R. L.** (2014). *Test Equating, Scaling, and Linking: Methods and Practices*. New York: Springer.
- Kultusministerkonferenz** (2016). Abiturnoten 2015 an Gymnasien, Integrierten Gesamtschulen, Fachgymnasien, Fachoberschulen und Berufsschulen. [https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Aus\\_Abiturnoten\\_2015.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Aus_Abiturnoten_2015.pdf), Stand vom 15. April 2018.
- Kultusministerkonferenz** (2017). FAQs: Gemeinsamer Abituraufgabenpool der Länder. <https://www.kmk.org/fileadmin/Dateien/pdf/Bildung/AllgBildung/FAQs-Abiturpool.pdf>, Stand vom 16. April 2018.
- Livingston, S. A.** (2004). *Equating Test Scores (Without IRT)*. <https://www.ets.org/Media/Research/pdf/LIVINGSTON.pdf>
- McManus, I. C., Dewberry, C., Nicholson, S. & Dowell, J. S.** (2013). The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools. *BMC Medicine*, 11.
- Mislevy, R. J., Haertel, G., Cheng, B. H., Ructtinger, L., DeBarger, A., Murray, E., ... Vendlinski, T.** (2013). A "conditional" sense of fairness in assessment. *Educational Research and Evaluation*, 19(2-3), 121-140.
- Müller-Benedict, V. & Gaens, T.** (2015). Sind Examensnoten vergleichbar? Und was, wenn Noten immer besser werden? *Die Hochschule*, 2015(2), 79-93.
- Müller-Benedict, V. & Tsarouha, E.** (2011). Können Examensnoten verglichen werden? Eine Analyse von Einflüssen des sozialen Kontextes auf Hochschulprüfungen. *Zeitschrift für Soziologie*, 40(5), 388-409.
- Neumann, M., Nagy, G., Trautwein, U. & Lüdtke, O.** (2009). Vergleichbarkeit von Abiturleistungen: Leistungs- und Bewertungsunterschiede zwischen Hamburger und Baden-Württemberger Abiturienten und die Rolle zentraler Abiturprüfungen. *Zeitschrift für Erziehungswissenschaft*, 12, 691-714.

**Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F. & Cleland, J.** (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, 50, 36-60.

**Patterson, F., Roberts, C., Hanson, M. D., Hampe, W., Ponnampereuma, G., Eva, K., ... Cleland, J.** (2018). 2018 Ottawa Consensus Statement: Selection and Recruitment in the Healthcare Professions. *Medical Teacher*.  
<https://doi.org/10.1080/0142159X.2018.1498589>

**Rindermann, H. & Oubaid, V.** (1999). Auswahl von Studienanfängern durch Universitäten – Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20, 172-191.

**Rönnebeck, S., Schöps, K., Prenzel, M. & Hammann, M.** (2008). Naturwissenschaftliche Kompetenz im Ländervergleich. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2006 in Deutschland: Die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 67-94). Münster: Waxmann.

**Ryan, J. & Brockmann, F.** (2009). *A Practitioner's Introduction to Equating with Primers on Classical Test Theory and Item Response Theory*.  
<https://files.eric.ed.gov/fulltext/ED544690.pdf>

**Schwibbe, A., Lackamp, J., Knorr, M., Hissbach, J., Kadmon, M. & Hampe, W.** (2018). Medizinstudierendenauswahl in Deutschland – Messung kognitiver Fähigkeiten und psychosozialer Kompetenzen. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 61(2), 178-186.

**Statistisches Bundesamt** (2017). Bildung und Kultur – Allgemeinbildende Schulen, Schuljahr 2015/2016; Tabelle 6.7  
<https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/AllgemeinbildendeSchulen2110100167004.pdf>, Stand vom 15. April 2018.

**Tillmann, K.-J.** (2013). Schulstrukturen in 16 deutschen Bundesländern: Zur institutionellen Rahmung des Lebenslaufs. [https://www.neps-data.de/Portals/0/Working%20Papers/WP\\_XXVIII.pdf](https://www.neps-data.de/Portals/0/Working%20Papers/WP_XXVIII.pdf), Stand vom 20. Juli 2018.

**Trapmann, S., Hell, B., Weigand, S. & Schuler, H.** (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21, 11-27.

**Wottawa, H. & Amelang, M.** (1980). Einige Probleme der „Testfairness“ und ihre Implikationen für Hochschulzulassungsverfahren. *Diagnostica*, 26, 199-221.

**Zimmermann, S., Klusmann, D. & Hampe, W.** (2017). Correcting the predictive validity of a selection test for the effect of indirect range restriction. *BMC Medical Education*, 17.

**Zwick, R. & Himelfarb, I.** (2011). The Effect of High School Socioeconomic Status on the Predictive Validity of SAT Scores and High School Grade-Point Average. *Journal of Educational Measurement*, 48(2), 101-121.

## Autoren



Dr. Stefan ZIMMERMANN || Universitätsklinikum Hamburg-Eppendorf, Institut für Biochemie und Molekulare Zellbiologie || Martinistraße 52, D-20246 Hamburg

[st.zimmermann@uke.de](mailto:st.zimmermann@uke.de)



Dr. Dietrich Klusmann || Universitätsklinikum Hamburg-Eppendorf, Institut für Biochemie und Molekulare Zellbiologie || Martinistraße 52, D-20246 Hamburg

[klusmann@uke.de](mailto:klusmann@uke.de)



Prof. Dr. Wolfgang HAMPE || Universitätsklinikum Hamburg-Eppendorf, Institut für Biochemie und Molekulare Zellbiologie || Martinistraße 52, D-20246 Hamburg

[hampe@uke.de](mailto:hampe@uke.de)