

Robin GLEESON¹ (Wien)

Warum die Fokussierung auf den „gender gap“ in Aufnahmetests mehr schadet, als sie nützt

Zusammenfassung

Im öffentlich-medialen Diskurs über Geschlechtsunterschiede in Aufnahmeverfahren rückt häufig der sogenannte *gender gap* in den Fokus. Die psychometrische Fachwelt definiert Fairness hingegen mit dem Modell gleicher Erfolgswahrscheinlichkeiten, das den Studienerfolg der Subgruppen mitberücksichtigt. Der Artikel zeigt anschaulich anhand von Echtdateien, dass die „öffentlich-mediale Fairnessdefinition“ bestehende Diskriminierungen zuungunsten von Frauen (oder Männern) nicht nur unzureichend erfasst, sondern mitunter sogar zu falschen Schlüssen führen kann. Anschließend wird diskutiert, warum das Modell nur so valide sein kann wie das Erfolgskriterium, auf dem es beruht. Gerade beim Erfolgskriterium Studienerfolg steht eine umfassende, allgemein akzeptierte Operationalisierung noch aus.

Schlüsselwörter

Fairness, Aufnahmeverfahren, Studienerfolg, Lehramt

¹ E-Mail: robin.gleeson@univie.ac.at



Why closing the gender gap in college admission tests is not a proper safeguard against discrimination

Abstract

In the media and in public discourse, fairness in college admissions tests is often understood as the absence of mean differences between subgroups, especially when talking about the gender gap. Psychometricians, however, use the model of equal success probabilities to define fairness by also taking academic access into account. Using real-world data, this paper shows how focusing on the absence of a gender gap in test results can hide actual gender discrimination and even lead to incorrect conclusions. The paper then discusses why the model of equal success probabilities can only be as good as the operationalization of academic success, which remains an ongoing challenge.

Keywords

fairness, college admission test, academic success, teaching

1 Einleitung

1.1 Fairness als Abwesenheit von Mittelwertsunterschieden

Die Fairness von Aufnahmeverfahren wird im öffentlich-medialen Diskurs häufig als bloße Abwesenheit von Mittelwertsunterschieden zwischen unterschiedlichen Personengruppen verstanden. Besonders markant ist dieses intuitive Verständnis von Testfairness bei der Diskussion von Geschlechtsunterschieden: Wenn im Test beide² Geschlechter gleich gut abschneiden, also kein *gender gap* vorliegt, so die Annahme, habe der Test wohl nicht unzulässig diskriminiert. Deutlich wird diese

² Das dritte Geschlecht für „nicht eindeutig männliche oder weibliche“ Personen wurde in dieser Publikation aufgrund der weitaus kleineren Gruppengröße nicht berücksichtigt.

Ansicht etwa im folgenden Zitat einer österreichischen Ministerin aus dem Jahr 2016:

Es freut mich, dass zehn Jahre nach der Einführung der Medizin-Aufnahmetests endlich faire Rahmenbedingungen für Frauen und Männer geschaffen wurden. Der kleinere Gender Gap zeigt, dass die systematische Benachteiligung von Frauen bei den Aufnahmetests mit den genderneutralen Testverfahren weitgehend beseitigt werden konnte. (APA, 2016)

Daraus folgt dann auch die Forderung, dass der Anteil der zum Studium zugelassenen Studienbewerber/innen dem Anteil der zum Test angetretenen Studienbewerber/innen entsprechen müsse: Die „Erfolgsquoten“ der Geschlechter werden zum Maßstab für die Fairness eines Aufnahmeverfahrens (derStandard/APA 2017; kurier.at, 2015).

Bei diesem Verständnis von Fairness wird jedoch implizit angenommen, dass in Bezug auf das maßgebliche Erfolgskriterium (Studien- oder Berufserfolg) keine Fähigkeitsunterschiede zwischen den Personengruppen bestehen – denn wenn Frauen oder Männer für ein bestimmtes Studium besser geeignet wären, müssten diese natürlich auch im Aufnahmetest für dieses Studium besser abschneiden. Bei der Betrachtung von Geschlechtsunterschieden ist die Annahme gleicher Eignung aber ohne vorherige Prüfung kaum haltbar: Selbst wenn wir annehmen, dass Männer und Frauen prinzipiell für jedes erdenkliche Hochschulstudium gleich gut geeignet sind, müssten wir zudem davon ausgehen, dass sich sämtliche Männer und Frauen eines Jahrgangs völlig zufällig für ein bestimmtes Studium entscheiden (können und wollen), um Selektionseffekte ausschließen zu können. Einige Indizien sprechen jedoch gegen diese Annahme: So ist etwa bereits die Population der Studienberechtigten unterschiedlich stark vorselektiert. In Deutschland lag die Studienberechtigtenquote³ der Frauen im Jahr 2016 bei 58 %, jene der Männer aber nur bei 47 % (STATISTISCHES BUNDESAMT, 2017, S. 111). In Österreich ist

³ Anteil der studienberechtigten Schulabgänger/innen zwischen 18 und 20 (in Österreich 19) Jahren an der der gleichaltrigen Wohnbevölkerung

der Unterschied sogar noch ausgeprägter, 50 % der Frauen stehen nur 36 % der Männer gegenüber (STATISTIK AUSTRIA, 2017, S. 41). Während in Österreich 2016 also jede zweite Frau der Kohorte ein Studium beginnen konnte, traf dies nur auf jeden dritten Mann zu. Auch die enormen Unterschiede im Geschlechterverhältnis verschiedener Studienrichtungen⁴ machen deutlich, dass sich die Population nicht rein zufällig auf die einzelnen Studienrichtungen verteilt, sondern auch andere Mechanismen eine Rolle spielen müssen (ZAUSSINGER et al., 2015, S. 27).

Es wäre daher durchaus plausibel anzunehmen, dass sich etwa besonders fähige Frauen oder besonders fähige Männer überproportional häufig für ein bestimmtes Studium entscheiden, und dann müsste ein faires Aufnahmeverfahren natürlich auch einen *gender gap* aufweisen, um diesen Fähigkeitsunterschied adäquat abzubilden.

1.2 Prognostische Fairness: das Modell gleicher Erfolgswahrscheinlichkeiten

Die Ausführungen machen deutlich, warum Fairness auch in Bezug auf das Geschlecht nicht als Abwesenheit von Mittelwertsunterschieden aufgefasst werden sollte, ohne diese Annahme anhand eines Erfolgskriteriums zu überprüfen.

Testfairness im Sinne von prognostischer Fairness⁵ wird deshalb seit Jahrzehnten über das Modell gleicher Erfolgswahrscheinlichkeiten definiert: Demzufolge ist ein Aufnahmeverfahren dann fair, wenn der Studienerfolg aufgrund der Ergebnisse im Aufnahmeverfahren für keine Subgruppe (kein Geschlecht) systematisch über- bzw. unterschätzt wird (KUBINGER & PROYER, 2010; SPIEL, LITZENBER-

⁴ In manchen Studien sind mehr als 90 % der Studierenden entweder Frauen oder Männer.

⁵ Eine umfassende Definition von Testfairness findet sich in den „Standards for Educational and Psychological Testing“ (American Educational Research Association, American Psychological Association & National Council on Measurement in Education (AERA, APA & NCME, 2014).

GER & HAIDEN, 2007, S. 48). Diese Definition lässt also Mittelwertsunterschiede zwischen den Geschlechtern zu, sofern sich diese auch im Erfolgskriterium Studienerfolg wiederfinden, oder anders gesagt: Falls sich zeigen sollte, dass Frauen (oder Männer) im betrachteten Studium wesentlich erfolgreicher studieren, dann sollten diese auch im Aufnahmetest wesentlich erfolgreicher abschneiden. Der Ansatz geht auf die Psychologin T. ANNE CLEARLY zurück, die ihr Modell 1968 publizierte, und ist in modifizierter Form in der Fachwelt heute weit verbreitet.⁶

Anhand aktueller Echt Daten von Aufnahmeverfahren der Universität Wien wird im Folgenden anschaulich gezeigt, dass nur das Modell gleicher Erfolgswahrscheinlichkeiten eine bestehende Benachteiligung eines Geschlechts (*test bias*) aufzeigen kann, die bei der bloßen Betrachtung von Mittelwertsunterschieden unsichtbar geblieben wäre, bzw. dass die Fokussierung auf Mittelwertsunterschiede sogar zu völlig falschen (weil verkehrten) Schlussfolgerungen führen kann.

2 Methode

Daten aus dem Eignungstest für das Lehramtsstudium der Universität Wien aus dem Jahr 2015 werden analysiert. 2618 Personen (1654 Frauen und 964 Männer) nahmen am Eignungstest teil.⁷

Getrennt je Geschlecht wird der Mittelwert der Testergebnisse berechnet, grafisch visualisiert und statistisch überprüft, ob ein signifikanter Mittelwertsunterschied vorliegt.

Anschließend werden die Testergebnisse je Person mit den innerhalb der ersten drei Studienjahre absolvierten Credit Points (ECTS)⁸ und der Modulnote der Stu-

⁶ Eine aktuelle Darstellung der eingesetzten Regressionsmodelle findet sich etwa bei WAHL & WALENTA, 2017.

⁷ Durch Pseudonymisierung der Datensätze wurde der Datenschutz gewahrt.

⁸ Soweit vorhanden. Fehlende Werte wurden aus der Analyse ausgeschlossen.

dieneingangs- und Orientierungsphase (STEOP)⁸ als Studienerfolgskriterium verknüpft und in einem Scatterplot visualisiert. Dazu wird eine hierarchische multiple lineare Regressionsanalyse der Erfolgskriterien auf das Ergebnis im Eignungstest sowie das Geschlecht als Moderatorvariable berechnet. Zur Visualisierung der Ergebnisse werden Regressionsgeraden getrennt nach Geschlecht über den Scatterplot gelegt.

3 Ergebnisse

Ein Mittelwertsvergleich der Ergebnisse im Eignungstest zeigt, dass Frauen im Schnitt ein um knapp 1 % besseres Ergebnis erzielen, der Unterschied ist aber statistisch nicht signifikant⁹. Auch die übereinandergelegten Dichteverteilungen zeigen, dass sich die Testergebnisse beider Gruppen kaum unterscheiden, es liegt also kein *gender gap* vor:

⁹ T-Test für unabhängige Stichproben, $\alpha = 0,05$.

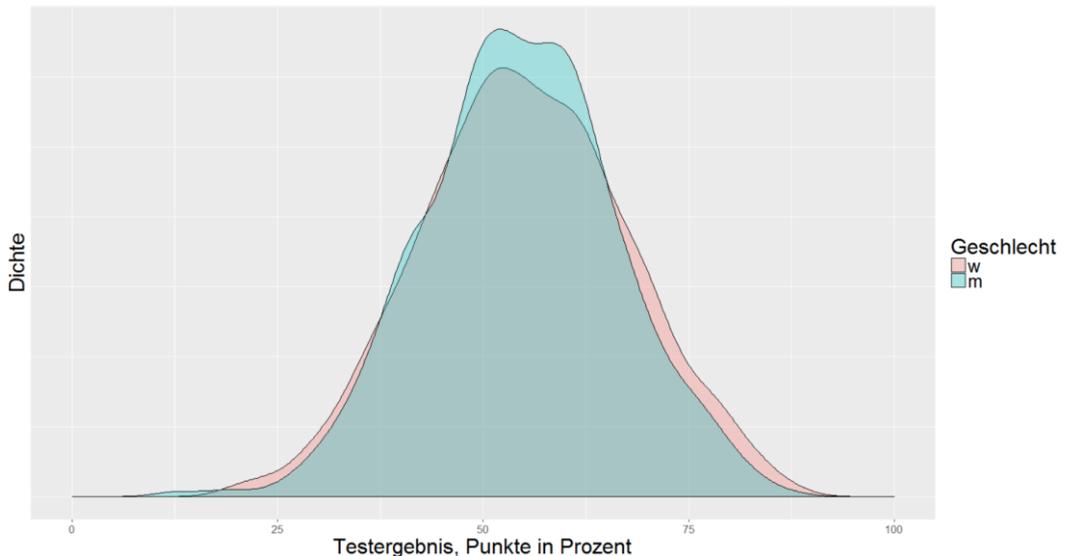


Abb. 1: Übereinandergelegte Dichteverteilungen der Testergebnisse, getrennt nach Geschlecht.

Eine schrittweise durchgeführte multiple lineare Regression auf das Studienerfolgskriterium „ECTS nach drei Studienjahren“ ergibt zwei statistisch signifikante Haupteffekte ($p < .001$) auf die Prädiktoren Ergebnis im Eignungstest und Geschlecht; die Regressionsgleichung ist signifikant mit $F(2, 1410) = 66.09$, $p < .001$, mit einem R^2 von 0.08571.

Die prognostizierten ECTS entsprechen $27.9636 + 1.1720 \cdot \text{Testergebnis} + 12.6344 \cdot \text{Geschlecht}_{[\text{weiblich} = 1, \text{männlich} = 0]}$

Eine schrittweise durchgeführte multiple lineare Regression auf das Studienerfolgskriterium „STEOP-Note“ ergibt einen statistisch signifikanten Haupteffekt ($p < .001$) für den Prädiktor „Ergebnis im Eignungstest“ sowie einen statistisch signifikanten Wechselwirkungseffekt ($p < .001$) zwischen dem Ergebnis im Eignungs-

test und dem Geschlecht; die Regressionsgleichung ist signifikant mit $F(3, 1877) = 143.7$, $p < .001$, mit einem R^2 von 0.1868

Die prognostizierte STEOP-Note entspricht $4.62924 - 0.02506 * \text{Testergebnis} - 0.011771 * \text{Testergebnis} * \text{Geschlecht}_{[\text{weiblich} = 1, \text{männlich} = 0]}$

In den folgenden zwei Abbildungen wird der Zusammenhang zwischen dem Ergebnis im Eignungstest und ECTS nach drei Studienjahren bzw. STEOP-Note¹⁰ in zwei Scatterplots grafisch dargestellt. Jeder rote Punkt entspricht dabei einer Studentin, jeder blaue Punkt einem Studenten. Die Ergebnisse im Eignungstest je Person finden sich auf der horizontalen Achse, die nach drei Jahren erbrachten ECTS bzw. die STEOP-Note auf der vertikalen Achse.

In die Scatterplots wurden jeweils zwei Regressionsgeraden eingezeichnet, die anhand der Daten vorhersagen, welcher Wert im Studienerfolgskriterium (ECTS oder STEOP-Note) erwartet wird, wenn eine Person einen bestimmten Wert im Eignungstest erreicht. Der 95%-Konfidenzbereich der Regressionsgeraden ist jeweils grau unterlegt.

Da die Regressionsgeraden für Männer (blau) und Frauen (rot) getrennt dargestellt werden, kann gezeigt werden, ob das Modell für beide Geschlechter vergleichbare Werte vorhersagt. Bei idealer Testfairness sollten beide Geraden übereinanderliegen. Weichen die Kurven signifikant voneinander ab, bedeutet dies, dass das Regressionsmodell für Männer und Frauen bei gleichem Ergebnis im Eignungstest einen unterschiedlich hohen Studienerfolg voraussagt.¹¹

¹⁰ An österreichischen Universitäten reicht die Notenskala von 1 (Sehr gut) bis 5 (Nicht genügend).

¹¹ Das R-Package ggplot2 bietet mit der Funktion `geom_smooth()` die Möglichkeit, bereits vor einer eingehenden regressionsanalytischen Prüfung eine erste Sichtung statistischer Zusammenhänge vorzunehmen, und unterstützt dabei auch Schätzmethoden zur Visualisierung nichtlinearer Zusammenhänge.

Die schwarzen (horizontalen und vertikalen) Linien dienen dazu, den Zusammenhang anhand eines Beispiels zu veranschaulichen.

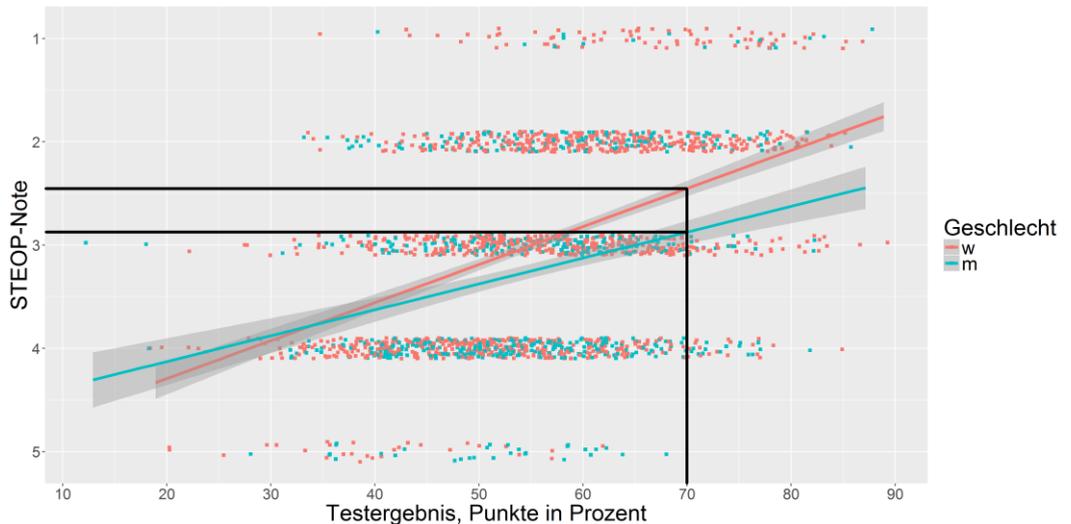


Abb. 2: Zusammenhang von Testergebnis und STEOP-Note, getrennt nach Geschlecht.¹²

Abbildung 2 zeigt, dass das Regressionsmodell ab einem Testergebnis von knapp 50 % der Punkte unterschiedliche STEOP-Noten für Männer und Frauen vorhersagt, weil sich die Geraden nicht mehr überlappen. Vergleichen wir beispielsweise einen Mann und eine Frau, die beide im Eignungstest 70 % der Punkte erreicht haben, so sagt unser Regressionsmodell voraus, dass die Frau im Schnitt eine Note von 2,5 erreichen wird, während der Mann nur auf eine Note von 2,9 kommt.

¹² Zur anschaulicheren Visualisierung wurden die Punkte auf der vertikalen Achse mit der ggplot2-Funktion `position_jitter()` um die wahren Werte gestreut. Tatsächlich liegen die Punkte natürlich genau auf den diskreten Notenausprägungen 1 bis 5.

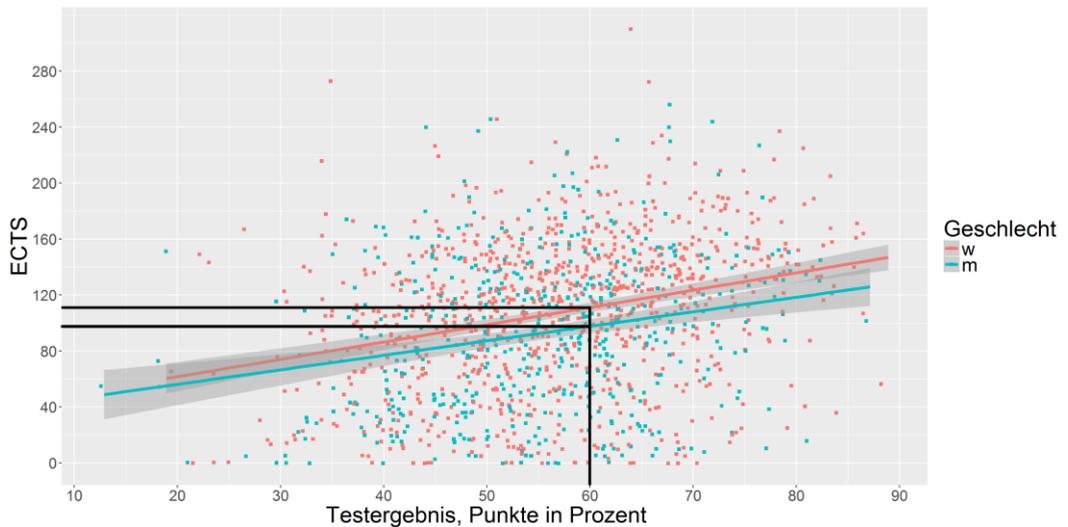


Abb. 3: Zusammenhang von Testergebnis und ECTS, getrennt nach Geschlecht.

Ein ähnliches Bild ergibt sich bei den ECTS (Abbildung 3). Vergleichen wir etwa einen fiktiven Mann und eine fiktive Frau, die im Eignungstest 60 % der Punkte erzielt haben, so sagt das Regressionsmodell für den Mann nach drei Jahren 98 ECTS voraus, für die Frau hingegen 111 ECTS.

4 Diskussion

Aufnahmeverfahren sind ein für die Hochschulen sensibles Thema, das oft im Fokus der Öffentlichkeit steht. Auch medial werden insbesondere Fragen der Fairness diskutiert – nicht nur, aber auch in Bezug auf das Geschlecht. Üblicherweise steht dabei die Frage nach einem etwaigen *gender gap* im Vordergrund. Wenn Frauen im Aufnahmetest schlechter abschneiden, so die gängige Annahme, stelle dies eine unzulässige Benachteiligung von Frauen dar. Umgekehrt wird kein Handlungsbedarf gesehen, wenn die Ergebnisse keinen *gender gap* aufweisen.

Diese Sichtweise setzt aber voraus, dass sich beide Geschlechter hinsichtlich ihrer durchschnittlichen Studieneignung nicht unterscheiden. Unterschiedlich hohe Studienberechtigtenquoten und nach wie vor unterschiedliche studienspezifische Interessenlagen legen aber nahe, dass dies nicht immer der Fall sein muss. Es ist daher wichtig, auch den späteren Studienerfolg differenziert zu erfassen, um eine Aussage über die Fairness eines Verfahrens treffen zu können. Das auf Regressionsanalysen basierende „Modell gleicher Erfolgswahrscheinlichkeiten“ erfüllt diesen Anspruch.

Im vorliegenden Artikel wurde anhand von Echtdateen aus dem Eignungstest für das Lehramtsstudium an der Universität Wien (Kohorte 2015) demonstriert, dass zwar kein *gender gap* bei den Testergebnissen vorliegt, der Test aber dennoch den „Studienerfolg“ der Frauen (operationalisiert durch die STEOP-Note und Studiergeschwindigkeit) im oberen Leistungsbereich systematisch unterschätzt hat, weil sie bei gleichen Testergebnissen etwas bessere Noten erzielen und im Studium etwas schneller vorankommen. Der Eignungstest hat also zuungunsten der Frauen diskriminiert, deren höhere Studierfähigkeit er nicht in ausreichendem Maße vorhersagen konnte, wobei die Unterschiede im vorliegenden Fall zwar statistisch signifikant, praktisch aber wenig bedeutsam sind.

Die Ergebnisse dieser Regressionsmodelle stellen eine wichtige Informationsquelle für die Testentwicklung und -überarbeitung dar. Durch eine differenzierte Betrachtung einzelner Aufgaben können etwa in weiterer Folge besonders problematische Aufgaben identifiziert und überarbeitet werden.¹³ Auch die Gewichtung einzelner Testteile oder das Testkonstrukt selbst kann angepasst werden, um die Testfairness zu erhöhen.

Der in diesem Artikel diskutierte Eignungstest konnte seit 2015 bereits mehrfach überarbeitet werden, wobei auch die beschriebenen Regressionsanalysen berücksichtigt wurden. Da der Eignungstest bisher als positiv absolviert galt, wenn die

¹³ Für weitere Informationen zum sogenannten *Differential item functioning* (DIF) siehe etwa HOLLAND & WAINER, 2012.

Kandidatin/der Kandidat mindestens 30 % der Punkte erreicht hatte¹⁴, die beobachteten Abweichungen im Regressionsmodell aber erst ab knapp 50 % der Punkte auftraten, wirkte sich der *test bias* auch nicht negativ auf die Kohorte 2015 aus.

Die diskutierten Ergebnisse zeigen aber eindrucksvoll, wie wichtig das Modell gleicher Erfolgswahrscheinlichkeiten ist, um die Fairness eines Verfahrens adäquat beurteilen zu können. Wie der (statistisch zwar nicht signifikante) höhere Mittelwert der Frauen beim Eignungstest zeigt, wäre sogar ein Szenario denkbar, bei dem man nur aufgrund eines Mittelwertsvergleichs zu verkehrten Schlussfolgerungen gelangt und fälschlich annimmt, dass der Test Frauen (oder Männer) benachteiligt, obwohl es sich eigentlich genau umgekehrt verhält. Würde man daraufhin den Test umgestalten, um jeden gender gap zu eliminieren, hätte man bestehende Diskriminierungen damit nur noch vergrößert.

Die diskutierte Methode ist nicht auf das Geschlecht beschränkt, sondern lässt sich auf beliebige Teilungskriterien anwenden. Gerade so wichtige Merkmale wie Bildungs- oder Migrationshintergrund lassen sich kaum ohne Studienerfolgsdaten analysieren. Dass etwa Personen mit schlechten Deutschkenntnissen in einem deutschsprachigen Studium im Schnitt weniger erfolgreich studieren werden als Personen mit deutscher Erstsprache, ist naheliegend, und daher wäre wohl auch niemand überrascht, wenn diese Personengruppe auch im Aufnahmetest schlechtere Ergebnisse erzielt. Wie möchte man aber beurteilen, ob der Aufnahmetest Bewerber/innen mit schlechten Deutschkenntnissen zu sehr, zu wenig oder in angemessener Weise schlechter einstuft als Personen mit deutscher Erstsprache? Das Modell gleicher Erfolgswahrscheinlichkeiten kann helfen, diese Fragen beantworten, stößt dabei aber auch an seine Grenzen (MEADE & TONIDANDEL, 2010).

Einschränkend muss zudem gesagt werden, dass auch diese Methode nur so valide sein kann wie das Erfolgskriterium, auf dem es beruht. Studien- oder Berufserfolg

¹⁴ Personen, die weniger als 30 % der Punkte erreichten, mussten an einem verpflichtenden Eignungs- und Beratungsgespräch teilnehmen und durften sich anschließend ebenfalls zum Studium zulassen.

sind tatsächlich kaum zu operationalisieren. Eine Operationalisierung mittels Prüfungsnoten oder ECTS bildet diese Konstrukte nur mangelhaft ab und muss sich grundsätzlich den Vorwurf gefallen lassen, dass hier nicht die besten, sondern nur bequem quantifizierbare Indikatoren herangezogen wurden. (Und kaum jemand würde wohl behaupten, dass die mittels ECTS gemessene Studiargeschwindigkeit die beste Methode ist, um im Sinne der Berufseignung gute Lehrer und Lehrerinnen zu identifizieren.) Im Sinne einer umfassenden psychometrischen Definition von Fairness muss diese daher stufenweise überprüft werden, wobei die prognostische Fairness, wie oben skizziert, erst dann empirisch überprüft werden kann, wenn u. a. die gültige Skalierung des Aufnahmeverfahrens und die Validität des Erfolgskriteriums belegt werden konnte (vgl. ARENDASY, SOMMER, GUTIÉRREZ-LOBOS, PUNTER, 2016; BORSBOOM, ROMEIJN & WICHERTS, 2008; MEADE & TONIDANDEL, 2010;). Eine allgemein akzeptierte Definition des Studien Erfolgs, die sich auch operationalisieren lässt, liegt allerdings noch in weitere Ferne (SCHMIDT, 2017), weshalb die psychometrische Praxis hier vor nach wie vor großen Herausforderungen steht.

Schließlich kann auch hinterfragt werden, ob ein Aufnahmeverfahren ausschließlich den späteren Studienerfolg, wie in diesem Artikel dargestellt, zum Ziel haben sollte, oder ob als Diversity-Maßnahme auch eine positive Diskriminierung benachteiligter Bevölkerungsgruppen erfolgen darf, die im Widerspruch zum Modell gleicher Erfolgswahrscheinlichkeiten steht. Dies wäre allerdings eine politische Entscheidung, die nicht von Psychometrikerinnen/Psychometrikern getroffen werden muss. Deren Aufgabe ist es vielmehr, den Entscheidungsträgerinnen und Entscheidungsträgern solide und anschaulich visualisierte Daten als Entscheidungsgrundlage zur Verfügung zu stellen.

5 Literaturverzeichnis

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (AERA, APA, & NCME) (2014). *Standards for educational & psychological testing*. Washington, DC: American Educational Research Association.

APA (2016). *Oberhauser zu Medizin-Aufnahmetest: Kleinerer Gender Gap bedeutet mehr Fairness*.

https://www.ots.at/presseaussendung/OTS_20160809_OTS0067/oberhauser-zu-medizin-aufnahmetest-kleinerer-gender-gap-bedeutet-mehr-fairness, Stand vom 27. Juli 2018.

Arendasy, M. E., Sommer, M., Gutiérrez-Lobos, K. & Punter, J. F. (2016). Do individual differences in test preparation compromise the measurement fairness of admission tests? *Intelligence*, 55, 44-56.

Borsboom, D., Romeijn, J.-W. & Wicherts, J. M. (2008). Measurement invariance versus selection invariance: Is fair selection possible? *Psychological Methods*, 13(2), 75-98.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.

derStandard/APA (2017). *Medizin-Aufnahmetest: Gender-Gap heuer wieder etwas größer*. <https://derstandard.at/2000062402011/Medizin-Aufnahmetest-Gender-Gap-heuer-wieder-etwas-groesser>, Stand vom 27. Juli 2018.

Kubinger, K. D. & Proyer, R. (2010). Gütekriterien. In K. Westhoff, C. Hagemeyer, M. Kersting, F. Lang, H. Moosbrugger, G. Reimann & G. Stemmler (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (3. überarbeitete Aufl., S. 173-180). Lengerich: Pabst.

kurier.at (2015). *Medizin-Aufnahmetest: Gender Gap bei Ergebnissen*. <https://kurier.at/politik/inland/medizin-aufnahmetest-geschlechtsunterschiede-bei-ergebnissen/145.522.410>, Stand vom 27. Juli 2018.

Meade, A. W. & Tonidandel, S. (2010). Not seeing clearly with Cleary: What test bias analyses do and do not tell us. *Industrial and Organizational Psychology*, 3(2), 192-205.

Schmidt, U. (2017). Monitoring an der Johannes Gutenberg-Universität Mainz. *Proceedings des HRK (Hochschulrektorenkonferenz) Projekt nexus*. <https://www.hrk-nexus.de/aktuelles/tagungsdokumentation/monitoring-iii/>, Stand vom 27. Juli 2018.

Spiel, C., Litzenberger M. & Haiden, D. (2007). Bildungswissenschaftliche und psychologische Aspekte von Auswahlverfahren. In C. Badelt, W. Wegscheider & H. Wulz (Hrsg.), *Hochschulzugang in Österreich* (S. 253-327). Graz: Grazer Universitätsverlag.

Statistik Austria (2017). *Bildung in Zahlen 2015/16. Schlüsselindikatoren und Analysen*. https://uniko.ac.at/modules/download.php?key=13584_DE_O&cs=4BAE, Stand vom 27. Juli 2018.

Statistisches Bundesamt (2017). *Bildung und Kultur. Nichtmonetäre hochschulstatistische Kennzahlen. Fachserie 11 Reihe 4.3.1*. <https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/KennzahlenNichtmonetaer2110431167004.pdf>, Stand vom 27. Juli 2018.

Wahl, I. & Walenta, C. (2017). Evaluierung des Aufnahmeverfahrens eines Fernstudiengangs anhand des Studienerfolges. Theorie, Methoden und Praxis der Evaluation. *Zeitschrift für Evaluation*, 2017(01), 73-96.

Zaussinger, S., Unger, M., Thaler, B., Dibiasi, A., Grabher, A., Terzieva, B., Litofcenko, J., Binder, D., Brenner, J., Stjepanovic, S., Mathä, P. & Kulhanek, A. (2015). *Studierenden-Sozialerhebung 2015. Bericht zur sozialen Lage der Studierenden. Band 2: Studierende*. Wien: Institut für Höhere Studien.

Autor



Mag. Robin GLEESON || Universität Wien, Center for Teaching and Learning || Universitätsstraße 5, A-1010 Wien

<https://ctl.univie.ac.at>

robin.gleeson@univie.ac.at