

Zulassungsverfahren für Studierende fair gestalten

- ✓ **Fairness – Was heißt das?**
- ✓ **Wie kann Fairness überhaupt erfasst werden?**
- ✓ **Keine relevante Teilgruppe benachteiligen**
- ✓ **Fairere Tests für Nichtmuttersprachler:innen**
- ✓ **Mehr Fairness mit Blick auf Testtrainings**
- ✓ **Gleiche Chancen unabhängig vom sozioökonomischen Status**
- ✓ **Fazit**

Fairness – Was heißt das?

Fairness ist ein Begriff, mit dem jeder von uns etwas anfangen kann, wenngleich unsere Vorstellungen von Fairness sehr unterschiedlich ausfallen können. Auch im Bereich der Testentwicklung wird eine Vielzahl unterschiedlicher Definitionen und Fairnesskonzepte diskutiert. Seit den 70er Jahren besteht die Forderung nach Fairness bei Verfahrens- und Testentwicklung als weiterem Qualitätsmerkmal (sog. Nebengütekriterium). Besonders vor dem Hintergrund der Eignungs- und Intelligenzdiagnostik wurde und wird der Fairnessbegriff viel diskutiert und löste eine Kaskade an Forschungsfragen und empirischen Studien aus.

Worin sich die verschiedenen Fairnessmodelle und -definitionen einig sind, ist die Tatsache, dass allen Teilnehmenden unvoreingenommen und respektvoll begegnet werden muss. Alle Teilnehmenden müssen – unabhängig von ihrer Gruppenzugehörigkeit – die gleiche Chance auf eine gültige Interpretation ihrer Testwerte bekommen.

Laut einer allgemeinen Definition gilt ein Test oder ein Verfahren dann als fair, wenn die daraus gezogenen Schlüsse nicht zu einer systematischen Benachteiligung von Teilgruppen führen. Oder anders ausgedrückt: Die Zugehörigkeit zu einer bestimmten Teilgruppe darf nicht dazu führen, dass die zukünftige Leistung durch das Testergebnis über- oder unterschätzt wird.

Welche Teilgruppen sind bei der Entwicklung fairer Tests relevant?

In der Praxis ist es natürlich unmöglich, alle denkbaren Teilgruppen im Testkonstruktions- und Evaluationsprozess zu berücksichtigen. Daher fokussiert man sich häufig auf Gruppen, die in der Vergangenheit oder Gegenwart das Ziel von Diskriminierung waren bzw. sind. Dazu gehören z. B. Frauen, Menschen mit Behinderung, ethnische oder religiöse Minderheiten, Menschen mit Migrationshintergrund oder Familien mit niedrigem sozio-ökonomischem Status.

Wie kann nun beantwortet werden, ob ein Verfahren diese oder andere Teilgruppen benachteiligt, und wie lässt sich ein Verfahren in solchen Fällen fair(er) gestalten?

Wie kann Fairness überhaupt erfasst werden?

1. Fairness als Einheitlichkeit der Vorhersage

Nach einem gängigen Modell ist ein Verfahren oder Test dann fair, wenn die Kriteriumswerte für keine Subgruppe systematisch über- oder unterschätzt werden. Mit Kriterium wird dabei das bezeichnet, was mit den Testwerten oder dem Verfahren vorhergesagt werden soll. Im Rahmen der Studierendenauswahl ist dies häufig der Studienerfolg, der z. B. über Zwischenprüfungs- und Abschlussnoten gemessen werden kann.

Ein Beispiel: Wenn ein Mitglied der Gruppe A und ein Mitglied der Gruppe B dieselbe Abschlussnote im Studium haben, sollten sie auch dasselbe Ergebnis in einem Studieneignungstest haben. Wenn das Mitglied der Gruppe A ein niedrigeres Testergebnis hat, wurde sein Erfolg durch den Test unterschätzt.

Solange der Zusammenhang zwischen Testergebnis und Studienerfolg für alle Teilgruppen gleich ist und die Regressionsgeraden identisch sind, ist ein Verfahren nach diesem Modell fair (Cleary, 1968).

2. Fairness durch Maximierung der Validität

Andere Ansätze rücken die Bedeutung der Validität für die Fairness ins Zentrum. Die Validität ist ein Hauptgütekriterium psychologischer Tests und gibt an, wie gut ein Test in der Lage ist, das jeweils beabsichtigte Merkmal zu messen. Beispielsweise sollte ein Test zur Studieneignung nur Merkmale erfassen, die auch wirklich für eine erfolgreiche Bewältigung des entsprechenden Studiums wichtig sind.

In der Praxis kann die Validität eines Tests jedoch durch unterschiedliche Fehlerquellen (z. B. finanzielle Investitionen in die Vorbereitung, Zugang zum Test, Lesbarkeit der Unterlagen) beeinträchtigt werden. Wenn solche irrelevanten Merkmale sich auf die Testergebnisse auswirken und sich systematisch zwischen bestimmten Teilgruppen unterscheiden, geht eine Einschränkung der Validität mit einer Einschränkung der Fairness einher. Anders herum nimmt mit zunehmender Validität die Wahrscheinlichkeit ab, dass ein Test oder Verfahren unfair ist. Ein Beispiel: Wenn ein tiefes Verständnis grammatikalischer Regeln in der Testsprache hilft, ein gutes Testergebnis zu erreichen, aber irrelevant ist für den Studienerfolg in einem ingenieurwissenschaftlichen Studienprogramm, dann könnte dieser Aspekt zu einer Benachteiligung von Nichtmuttersprachler:innen oder Menschen mit niedrigem sozio-ökonomischem Status führen.

Der wesentliche Anspruch besteht nach diesem Ansatz daher darin, die (faire) Validität zu maximieren und gleichzeitig (potenziell unfaire) nicht valide Fehlerquellen zu minimieren.

ABER: Punkteunterschiede zwischen Teilgruppen sind kein Indikator für Unfairness.

Eine häufig verbreitete Annahme ist, dass ein Test unfair ist, wenn eine Teilgruppe in dem Test im Mittel weniger Punkte erreicht als eine andere Teilgruppe. Diese Schlussfolgerung ist unzulässig. Ein Test oder Verfahren kann auch unter den geschilderten Umständen fair sein, wenn sich die beiden Teilgruppen tatsächlich in dem zu messenden Merkmal unterscheiden.

Im Umkehrschluss sind nicht bestehende Gruppenunterschiede damit auch kein Fairnessbeleg. So könnte es sein, dass sich die betrachteten Gruppen in der Realität eigentlich in dem zu messenden Merkmal unterscheiden, diese Unterschiede jedoch durch das Verfahren nicht aufgedeckt werden. In diesem Fall würde die Gruppe mit der höheren Merkmalsausprägung benachteiligt werden.

Auswahltests sind fair...

... wenn keine relevante Teilgruppe benachteiligt wird.

Auswahlverfahren sollten regelmäßig evaluiert und überprüft werden. Dazu gehört die Überprüfung von Reliabilität (Messgenauigkeit), Validität, Akzeptanz und eben auch Fairness, z. B. in Hinblick auf Geschlecht, Ethnie, ...

Doch welche Möglichkeiten gibt es, wenn sich herausstellt, dass das Verfahren eine oder mehrere Teilgruppen benachteiligt, wenn wir beispielsweise unterschiedliche Regressionsgeraden für verschiedene Gruppen finden? Sicherlich ist es in dem Fall zunächst sinnvoll, sich das gesamte Verfahren genauer anzusehen und ggf. infrage zu stellen. Hat sich das Verfahren jedoch grundsätzlich bewährt und als ansonsten valide und geeignet herausgestellt, so gibt es auch Möglichkeiten, das Verfahren mit kleineren Anpassungen fair zu gestalten.

3 grundsätzliche Wege zu mehr Fairness:



1. Austausch von „unfairen“ Aufgaben oder Aufgabenformaten

Mit statistischen „DIF-Analysen“ (differential item functioning) lassen sich einzelne Aufgaben identifizieren, die in verschiedenen Gruppen unterschiedlich funktionieren (Zumbo & Thomas, 1997). Ein Beispiel: Eine Aufgabe ist schwieriger als erwartet für Mitglieder der Gruppe A und leichter als erwartet für Mitglieder der Gruppe B (z. B. lösen starke Mitglieder der Gruppe B aber nicht der Gruppe A die Aufgabe mit hoher Wahrscheinlichkeit). Diese Aufgabe kann eliminiert werden.

Es ist auch möglich, ganze Subtests bzw. Module zu eliminieren. Wenn eine Evaluation einen Subtest identifiziert, der unfair für eine Teilgruppe ist, kann dieser Subtest eliminiert oder durch einen anderen ersetzt werden. Ein Beispiel: In einem Test für Rechtswissenschaften unterschätzte der Subtest „Schlussfolgerungen“ die späteren LL. B. Noten von Frauen nach dem Modell von Cleary (1968). Dieser Subtest wurde aus dem Test entfernt (Bergholz & Stegt, 2018).



2. Einführung separater Normen

In bestimmten Kontexten ist es notwendig, dass das Auswahlverfahren unverändert bleibt, z. B. wegen des hohen Anforderungsbezugs. In diesem Fall bietet es sich an, für die relevanten Teilgruppen eigene, separate Normen zu berechnen. In der Praxis wird darauf aber selten zurückgegriffen.



3. Einführung von Quoten

Eine Alternative zur Einführung von Normen sind Quoten. Statt Standardwerte anhand unterschiedlicher Normen zu berechnen, wird eine bestimmte Prozentzahl bzw. Anzahl z. B. von Studienplätzen an Bewerber:innen einer Teilgruppe vergeben. Die Quote sollte dabei in Relation zu einem relevanten Merkmal stehen.

... wenn Nichtmuttersprachler:innen nicht systematisch benachteiligt werden.

Die große Herausforderung bei der Gestaltung des Auswahlverfahrens ist hier, die Balance zu finden zwischen einerseits der Chancengleichheit von Muttersprachler:innen und Nichtmuttersprachler:innen, und andererseits der Berücksichtigung der sprachlichen Anforderungen des Studiums.

Diese Balance kann von Studienfach zu Studienfach anders aussehen. Es ist offensichtlich, dass die Anforderungen z. B. die deutsche Sprache zu beherrschen für ein Germanistikstudium höher sind als für ein Studium der Chemie.

3 Wege zu mehr Fairness mit Blick auf Nichtmuttersprachler:innen:



1. „Sprachlast“ des Auswahltests an die Anforderungen anpassen

Der Einfluss der Sprachbeherrschung auf das Testergebnis kann über die Wahl der Aufgabenformate und die Sprachkomplexität (z. B. Wortwahl, Satzbau) innerhalb der Aufgaben reguliert werden. Beides sollte die Anforderungen des Studiums widerspiegeln und diese keinesfalls überschreiten. Ein Unterschreiten der Anforderungen des Studiums ist möglich, wenn man davon ausgeht, dass fehlende Sprachkenntnisse aufgeholt werden können.



2. Sprachauswahl ermöglichen

Der Auswahltest wird in verschiedenen Sprachen angeboten, z. B. neben der oder den Landessprachen bzw. der Hauptunterrichtssprache zusätzlich in weit verbreiteten Sprachen wie Englisch oder Französisch.



3. Sprachkenntnisse und Sprachlernfähigkeit erfassen

Eine Messung der Sprachkenntnisse ermöglicht eine Analyse, welches Sprachniveau Teilnehmende haben sollten, um ihr wahres Potenzial zu zeigen. Mithilfe spezieller Aufgabenformate, wie wir sie beispielsweise im Sprachlernfähigkeitstest SLFT nutzen, kann zudem die Fähigkeit erfasst werden, Fremdsprachen zu lernen.

... wenn teure Testtrainings keinen Vorteil bringen.

Auswahlverfahren für ein Studium sollten die für das Studium notwendigen Fähigkeiten messen und nicht, wie viel Geld in die Vorbereitung investiert wurde. Gleichzeitig gibt es zahlreiche Unternehmen auf dem Markt, die Vorbereitungskurse und -materialien für Auswahlverfahren anbieten – von zirka 30 Euro für Online-Testsimulationen bis hin zu vierstelligen Beträgen für individuelle Coachings findet man fast alles.

Doch wie trainierbar sind Studieneignungstests wirklich?

- Die Testergebnisse in fähigkeitsbasierten, fachspezifischen Auswahltests (ohne Wissensanteile) können nur in sehr geringem Maße durch Vorbereitung und finanzielle Investments verbessert werden.
- Komplexe, besonders fachspezifische Aufgabengruppen sind robuster gegenüber Vorbereitungsmaßnahmen als weniger fachspezifische.
- Für die Vorbereitung ist es am effektivsten, sich wenige Stunden mit dem Test und den Aufgabenformaten vertraut zu machen (z. B. über Bücher oder Online-Vorbereitungstests mit alten Originalaufgaben).
- Nach den ersten investierten Stunden wird zunehmend mehr Zeit benötigt, um überhaupt noch Vorbereitungseffekte zu erzielen.

Daraus ergeben sich **4 Wege zu mehr Fairness mit Blick auf Testtrainings:**



1. Einsatz von fähigkeitsbasierten Auswahltests

Fähigkeitsbasierte Testverfahren setzen auf langfristig entwickelte kognitive Kompetenzen, die nur sehr eingeschränkt kurzfristig trainiert werden können.



2. Testmodule mit komplexen fachbezogenen Problemlösungsaufgaben

Je komplexer und fachbezogener Aufgaben sind, desto eher entziehen sie sich dem Einfluss von Trainings. Daher sollte diese Form der Aufgaben in Auswahltests integriert werden. Eine Studie zur Trainierbarkeit des Tests für Medizinische Studiengänge (TMS) zeigt, dass zeitliche und finanzielle Investitionen in die Testvorbereitung keine oder nur minimale Auswirkungen bei diesen Arten von Testaufgaben haben (Weppert et al., 2023).



3. Reduktion der Wissensbestandteile

Wissensmodule sollten nur dann zum Einsatz kommen, wenn sich Wissen in bestimmten Bereichen als erfolgskritische Anforderung herausstellt. Das zu erfassende Wissen sollte nicht über das eigentlich voraussetzende Schulwissen hinausgehen und sich auf studienrelevante Gebiete beschränken, und es sollte ausreichend frei zugängliche Lernmaterialien geben.



4. Bereitstellung von kostenlosen Vorbereitungsmaterialien

Um Bewerbenden zu ermöglichen, sich mit dem Test, den Aufgabenformaten und Anforderungen vertraut zu machen, sollten kostenlose (oder mindestens kostengünstige) Vorbereitungsmaterialien bereitgestellt werden. Diese Materialien sollten neben Informationen zu Ablauf und Rahmenbedingungen auch Beispielaufgaben zu jeder Aufgabengruppe enthalten.

... wenn der sozioökonomische Status der Teilnehmenden keine Rolle spielt.

Die Bildungschancen für Schüler:innen aus Familien mit geringem sozioökonomischen Status werden immer wieder thematisiert. Zulassungstests mit hohen Testgebühren und hoher Trainierbarkeit könnten die Benachteiligung dieser Gruppe sogar noch verschlimmern. Auf der anderen Seite könnten sie helfen, die Bildungsgerechtigkeit für diese Personengruppe zu verbessern, wenn die richtigen Tests verwendet und richtig eingesetzt werden.

3 Wege zu mehr Fairness mit Blick auf den sozioökonomischen Status:



1. Möglichst niedrige Gebühren für die Prüfungsteilnehmenden

Die Gebühren für Auswahltests liegen bei deutschen Anbietern meist deutlich unter denjenigen der internationalen Konkurrenz. Ohne die Unterstützung von Hochschulen oder öffentlicher Träger ist dies nur schwer umzusetzen, da die Kosten für die Bereitstellung nicht deutlich niedriger sind.



2. Geringe Trainierbarkeit

Durch die geringe Trainierbarkeit werden keine Vorteile aus kostenintensiven Vorbereitungsmaßnahmen gezogen.



3. Begrenzte Wiederholbarkeit

Eine Begrenzung der Anzahl der Testbearbeitungen durch die Bewerbenden (z. B. einmal im Kalenderjahr) verhindert selbst bei geringer Trainierbarkeit des Tests, dass Bewerbende Vorteile dadurch erlangen, dass sie den Test häufiger durchführen können und die Wahrscheinlichkeit auf ein besseres Ergebnis erhöhen.

Fazit

Die genannten Beispiele bieten selbstverständlich keine ausschöpfende Übersicht aller Fairness-Aspekte. So konnten wir bei weitem nicht alle relevanten Fairness-Modelle thematisieren, und auch die Behandlung von in diesem Zusammenhang zentralen statistischen Methoden (z. B. Differential Item Functioning-Analysen oder Messinvarianzanalysen) hätte den Rahmen dieses Whitepapers gesprengt.

Zusammenfassend sei jedoch auf die folgenden grundlegenden Prinzipien verwiesen, die herangezogen werden können, um die Fairness eines Tests oder eines Verfahrens zu verbessern:

- **Die Fairness eines Auswahlverfahrens und seiner Bestandteile (z. B. Tests) sollte empirisch evaluiert werden.**
- **Alles, was zur validen Messung des beabsichtigten Merkmals erforderlich ist, sollte in einen Test oder ein Verfahren einbezogen werden; und alles, was nicht der Messung des beabsichtigten Merkmals dient, sollte vermieden werden.**
- **Allen teilnehmenden Gruppen sollte mit Respekt begegnet werden und alle Teilnehmenden sollten die gleiche Chance auf eine valide Interpretation ihrer Testergebnisse erhalten.**

Was zunächst einfach klingen mag, ist in der Praxis jedoch alles andere als leicht umzusetzen. Dennoch darf das Thema Fairness unter keinen Umständen ignoriert werden, da entsprechende Maßnahmen nicht nur eine fairere Auswahl ermöglichen, sondern zeitgleich auch die Validität eines Tests bzw. Zulassungsverfahrens sowie die Diversität im Studium erhöhen. Dabei erfordert der Prozess sorgfältige Aufmerksamkeit während der Konzeption, Aufgabenkonstruktion, Durchführung, Auswertung und Evaluation eines Auswahlverfahrens und seiner Bestandteile.

Literatur:

- Bergholz, L. & Stegt, S. J. (2018). Validität und Fairness eines Studierfähigkeitstests für Rechtswissenschaften. *Zeitschrift für Hochschulentwicklung*, 13(4), 57-79.
- Cleary, T. A. (1968). Test bias: prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Weppert, D., Amelung, D., Escher, M., Troll, L., Kadmon, M., Listunova, L., & Montasser, J. (2023). The impact of preparatory activities on the largest clinical aptitude test for prospective medical students in Germany. *Frontiers in Education*, 8, 1104464. <https://doi.org/10.3389/educ.2023.1104464>
- Zumbo, B. & Thomas, D. (1997). *A measure of effect size for a model-based approach for studying DIF*. Prince George, Canada: University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science.

Über das ITB

Wir unterstützen Sie bei der Entwicklung, Implementierung, Durchführung und Auswertung von Auswahlverfahren für Studiengänge, spezifische Studienprogramme und

Stipendien – mit unserer Expertise, unserer Erfahrung, unseren Konzepten und Produkten und all unseren Ideen stehen wir Ihnen gern zur Verfügung.

Kontakt

Denken Sie über die Einführung von Studieneignungstests nach? Dann besuchen Sie uns auf unserer Website <https://itb-academic-tests.org/> oder schreiben Sie uns unter info@itb-consulting.de.