

Teacher Bias: Wie voreingenommen sind deutsche Lehrkräfte?

Bekommen die schlankeren Mädchen aus den wohlhabenden Familien die besseren Noten? Richard Nennstiel und Sandra Gilgen erklären im Interview die Ergebnisse ihrer Studie zur Voreingenommenheit von Lehrkräften.

[Daniel Lauris](#)

22.10.2024 Lesezeit 5 min

Stimmt es wirklich, dass skinny Sophie bessere Noten bekommt als chubby Can? Diese Frage haben sich Dr. Richard Nennstiel (Universität Bern) und Dr. Sandra Gilgen (Universität Zürich) gestellt - und sie sind durch die Untersuchung von Datenmaterial aus Deutschland (9. Jahrgangsstufe) zu dem Ergebnis gekommen, dass die Voreingenommenheit der Lehrkräfte statistisch nachweisbar ist.

Redaktion: Könnten Sie für unsere Leserinnen und Leser bitte einmal zusammenfassen, worum es in Ihrer Studie geht?

Dr. Richard Nennstiel: Es hat uns interessiert, zu untersuchen, inwieweit die Notengebung von askriptiven Merkmalen der Schülerinnen und Schülern beeinflusst wird, also von Merkmalen, die nicht mit der eigentlichen Leistung zusammenhängen, wie etwa das Geschlecht, die Religionszugehörigkeit oder die Hautfarbe. Wir haben also, einmal vereinfacht ausgedrückt, durch Lehrkräfte individuell gegebene Noten mit standardisierten Leistungstests verglichen und dann geschaut, ob die Schülerinnen und Schüler - beispielsweise in einem Fach wie Mathematik - von den Lehrkräften unterschiedliche Noten erhalten, obschon sie bei einer standardisierten Bewertung die gleichen Ergebnisse erzielten.

Für unsere Untersuchung haben wir auf die Daten der 4. Startkohorte des Nationalen Bildungspanels (NEPS) zurückgegriffen. Die dadurch relevanten circa 14.000 Schülerinnen und Schüler waren zum Messzeitpunkt in der 9. Klasse im Schuljahr 2010/2011.

Redaktion: Standardisierte Leistungstests sind in diesem Fall also Prüfungen, die nicht von einer Lehrkraft der Schülerinnen und Schüler korrigiert werden, sondern die entweder von einem Computer oder anonymisiert, also ohne sichtbarem Namen, auf dem Prüfungsbogen von nicht mit den Schülerinnen und Schülern in Kontakt stehenden Personen korrigiert werden: Ist das richtig?

Nennstiel: Ja, das ist richtig. Es werden da also entweder technische Geräte mitgebracht, sodass die Kinder die Tests auf diesen Geräten ausführen können, die dann auch von den Geräten automatisch korrigiert werden. Oder aber die Tests werden von Forscherinnen und Forschern, die die Kinder nicht kennen, ausgewertet.

Redaktion: Und im anderen Fall haben wir also die die Schülerinnen und Schüler unterrichtenden Lehrkräfte, die die Schülerinnen und Schüler dann beim Korrigieren vor dem geistigen Auge haben - und dadurch fließt, so besagt Ihre Studie, eben auch ein gewisser Anteil an Subjektivität in die Bewertungen ein. In welcher Hinsicht treten Unterschiede auf?

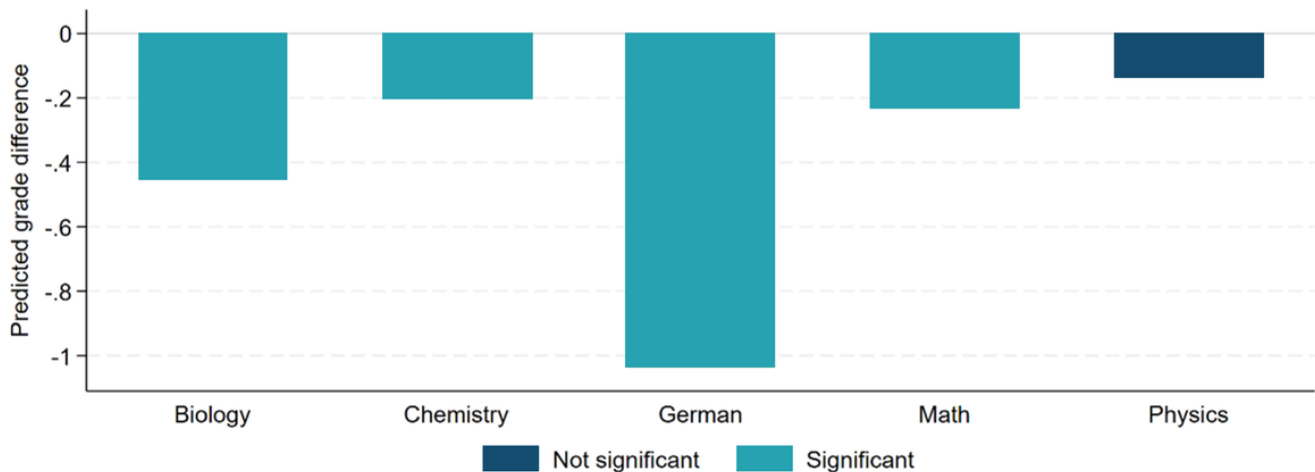
Dr. Sandra Gilgen: Wir sehen in allen von uns getesteten Faktoren Unterschiede: Wir haben auf das Geschlecht geachtet, auf die sozioökonomische Herkunft, auf den ethnischen Hintergrund und auch – was nicht so häufig gemacht wird – auf das Körpergewicht. Und wir haben bei all diesen Faktoren Effekte in der Benotung festgestellt, die eigentlich nichts in den Noten zu suchen haben.

Und in diesem Zusammenhang auch sehr interessant ist die Tatsache, dass sich diese Effekte akkumulieren. In der Realität kann es ja zum Beispiel vorkommen, dass Schülerinnen und Schüler einen Migrationshintergrund haben, aus einem sozial schwächeren Familienhintergrund kommen und auch noch übergewichtig sind: Die Notenunterschiede werden durch diese kumulierten Effekte dann tendenziell noch größer.

Hier reichten die Abweichungen bis hin zu einer ganzen Schulnote. Wenn man die beiden Extreme skinny Sophie und chubby Can miteinander vergleicht, beträgt der Unterschied circa eine Note.

Comparison of school grades

Turkish low SES overweight/obese boy ('Can') vs majority high SES non-overweight/obese girl ('Sophie')



Note: Estimates based on models controlling for domain-specific test scores, general cognitive test scores, and attended school track.
Source: NEPS Starting cohort 4, weighted data, our own calculations.

Abbildung 1: Ein Vergleich der Schulnoten zwischen einem türkischen Jungen mit niedrigem sozioökonomischen Status (SES), der übergewichtig/adipös ist, und einem Mädchen mit hohem SES aus der Mehrheitsbevölkerung, das nicht übergewichtig/adipös ist.

Redaktion: Sind die Verzerrungs-Effekte, von denen Sie sprechen, bei allen für die Untersuchung zur Verfügung stehenden Lehrkräften und bei allen Kriterien, die hier eine Rolle spielen, gleich stark? Also beim Migrationshintergrund, bei der sozialen Situation der Familie, beim Körpergewicht und beim Geschlecht?

Gilgen: Während die Verzerrungen in Bezug auf die sozioökonomische Herkunft in vielen Schulklassen recht ähnlich waren, gab es bezüglich des Körpergewichts – und insbesondere auch des Geschlechts der Schülerinnen und Schüler – deutlichere Unterschiede zwischen den einzelnen Schulklassen.

Redaktion: *Das sind ja durchaus dramatische Ergebnisse. Was kann man denn tun, damit sich dieser Bias, diese Voreingenommenheit, in den Köpfen bzw. im Unterbewusstsein der Lehrkräfte nicht so gravierend auswirkt?*

Nennstiel: Diese Frage war jetzt nicht unbedingt unsere Aufgabe. Es ging uns primär darum, aufzuzeigen, wie groß der Bias bzw. die Verzerrungen in der Notengebung sein können. Folglich haben wir uns jetzt weniger mit den Gründen für diesen Bias beschäftigt. Das können wir auf wissenschaftlicher Basis aber auch nicht tun, denn wir haben von den Lehrkräften bzw. auch über die Lehrkräfte keine Informationen zu dieser Frage.

Interessant ist in diesem Zusammenhang aber beispielsweise der Umstand, dass sich im deutschen Kontext Benachteiligungen für Kinder mit einem türkischen Namen zeigen. In dieser Forschungsrichtung wird als Ursache dafür häufig diskutiert, dass hier implizite Stereotype wirken. Die Lehrpersonen aus dem Datenmaterial, das wir verwendet haben, kennen die Schülerinnen und Schüler jedoch bereits mindestens ein Schuljahr lang. Da würde man dann eigentlich schon erwarten, dass diese Lehrkräfte irgendwann nicht mehr auf diese impliziten Stereotype zurückgreifen. Unsere Ergebnisse könnten so interpretiert werden, dass dies trotzdem der Fall ist – und das hat uns dann schon erstaunt.

Um Ihnen doch noch eine Idee in Bezug auf die Frage nach Mitteln und Methoden gegen die Verzerrungen bei der Benotung zu geben: Ein Ansatz wäre hier beispielsweise, die Bewertungskriterien bei der Bewertung von Prüfungen und Tests noch enger und konkreter zu fassen – insbesondere in den sprachlichen Fächern, in denen sich bei unserer Untersuchung ja auch größere Verzerrungen gezeigt haben.

Redaktion: *Sehen Sie in der Künstlichen Intelligenz (KI) eine Chance bzw. Möglichkeit, diesen Bias zu beseitigen? Also stellen wir uns doch einfach einmal vor, die KI würde in Zukunft die Prüfungen der Schülerinnen und Schüler korrigieren – und nicht mehr die Lehrkräfte.*

Gilgen: Also in Bezug auf einen möglichen Einsatz der Künstlichen Intelligenz für die Bewertung von Prüfungen und Tests wäre ich jetzt sehr kritisch, denn damit wurden ja auch bereits negative Erfahrungen gemacht – insbesondere auch, weil sich die KI ja auch wiederum auf bestehende Daten stützt – und diese Daten häufig bereits voller Verzerrungen sind. Das kann dann dazu führen, dass auch Beurteilungen durch die KI wieder von diesem Bias geprägt sind.

Redaktion: *Und wenn wir der KI die Namen der Schülerinnen und Schüler und weitere Merkmale und Daten über sie gezielt vorenthalten, die Prüfungen und Tests also anonymisieren? Könnte die Gefahr einer verzerrten Bewertung dadurch nicht verhindert werden?*

Nennstiel: Ja – wenn es uns gelingt, die KI so für den Schulkontext zu trainieren, dass sie bias-frei bleibt und automatisiert Rückmeldungen gibt, dann könnte das funktionieren. Dazu müsste die KI über die entsprechenden Informationen verfügen, wie beispielsweise ein guter Aufsatz aussehen oder wie ein Lösungsweg in Mathematik beschrieben sein soll. Das könnte die Lehrkräfte durchaus ein Stück weit entlasten, die ja auch unter Leistungs- und Zeitdruck stehen. Die KI könnte hier zumindest unterstützend wirken.

Redaktion: Herr Doktor Nennstiel, Frau Doktorin Gilgen, wir danken Ihnen für dieses Gespräch.



Zur Person

Richard Nennstiel ist Soziologe und arbeitet an der Abteilung für Bildungssoziologie an der Universität Bern. Seine Forschungsschwerpunkte sind: Ungleichheiten im Bildungssystem, Übergänge von der Schule in den Beruf, soziale Mobilität und Kompetenzentwicklung.



Zur Person

Sandra Gilgen ist Soziologin und arbeitet aktuell sowohl an der Universität Zürich im Data Centre des Universitären Forschungsschwerpunktprogramms Human Reproduction Reloaded als auch an der Universität Bern als wissenschaftliche Mitarbeiterin an der Abteilung Bildungssoziologie. Ihre Forschungsschwerpunkte sind: Reproduktive Gerechtigkeit, Verteilungsgerechtigkeit, Diskriminierung und Bildungssoziologie