

Schulnoten - wenig aussagekräftig, aber (noch) notwendig?

Das System der Ziffernnoten ist seit vielen Jahren umstritten. Jun.-Prof. Nicolas Hübner erläutert im Interview, warum er es trotz begründeter Kritik derzeit nicht für ratsam hält, Schulnoten flächendeckend abzuschaffen.

Michael Klitzsch

24.06.2024 Lesezeit 9 min

Trotz nachweislich mangelnder Aussagekraft hält das deutsche Bildungssystem an Ziffernnoten fest. Doch was bedeutet eine Zwei in Deutsch oder eine Fünf in Mathe eigentlich? Worin liegen die Probleme mangelnder Vergleichbarkeit von Schulnoten begründet? Und gibt es überhaupt sinnvolle Alternativen? Über das und mehr spricht Jun.-Prof. Nicolas Hübner von der Universität Tübingen im Interview.

Redaktion: Herr Professor Hübner, in Ihrer aktuellen Arbeit „Alles eine Frage des Bundeslandes? Eine mehrebenenanalytische Betrachtung der eingeschränkten Vergleichbarkeit von Schulnoten“ haben Sie die mangelhafte Vergleichbarkeit von Schulnoten als „ein sehr vielschichtiges Problem“ beschrieben. Können Sie kurz erläutern, was Sie damit meinen?

Jun.-Prof. Nicolas Hübner: In dem Beitrag haben meine Kolleginnen und Kollegen Malte Jansen, Petra Stanat, Thorsten Bohl, Wolfgang Wagner und ich uns mit der häufig in bildungspolitischen Debatten vorgetragenen These auseinandergesetzt, dass sich mit einer Abschaffung föderaler und der Einführung zentraler Strukturen, also einer Abschaffung der Kulturhoheit der Länder im Bildungswesen, auch viele Probleme in Bezug auf die mangelnde Vergleichbarkeit von Schulnoten auflösen würden. Hierzu haben wir auf Basis der Daten von rund 55.000 Schülerinnen und Schülern aus den IQB-Bildungstrends 2015 und 2018 untersucht, auf welchen Ebenen im Bildungswesen – innerhalb von Schulen, zwischen Schulen innerhalb von Bundesländern oder zwischen Bundesländern – vermutlich besonders bedeutsame Faktoren zu verorten sind, die die mangelnde Vergleichbarkeit von Schulnoten erklären können. Es zeigte sich, dass 21 Prozent der Unterschiede in den Schulnoten durch Unterschiede in einem standardisierten Kompetenztest erklärt werden konnten. Die übrigen Unterschiede verteilten sich vor allem auf Unterschiede innerhalb von Schulen (85 Prozent), gefolgt von Schulunterschieden innerhalb von Bundesländern (12 Prozent) und schließlich Bundeslandunterschieden (3 Prozent). Diese Ergebnisse sprechen gegen die Annahme, dass die mangelnde Vergleichbarkeit von Schulnoten ganz wesentlich auf unterschiedliche Praktiken in den Bundesländern zurückzuführen ist.

Redaktion: Es gibt also keine einfache Antwort auf die Frage nach der Ursache der mangelnden Vergleichbarkeit von Noten.

Hübner: Richtig, die Vielschichtigkeit ergibt sich aus der Summe der möglichen Faktoren auf den einzelnen Ebenen, die zur mangelnden Vergleichbarkeit beitragen können. Diese reichen von Unterschieden zwischen Schülerinnen und Schülern über Beobachtungs- und Beurteilungsfehler und Unterschiede in Bezugsnormorientierungen von Lehrpersonen bis hin zu Unterschieden in den intendierten und implementierten Curricula und gesetzlichen Vorgaben. Die Ergebnisse legen nahe, dass eine Abschaffung föderaler Strukturen und eine Vereinheitlichung der gesetzlichen Vorgaben in allen Bundesländern allein vermutlich nur einen sehr kleinen Beitrag zur Lösung des Problems einer

mangelnden Vergleichbarkeit von Schulnoten leisten würden.

Redaktion: Woran liegt es, dass viele der Unterschiede in der Notengebung innerhalb oder zwischen Schulen zu finden sind?

Hübner: Betrachtet man mögliche Faktoren, die zur mangelnden Vergleichbarkeit von Schulnoten beitragen, so fällt schnell auf, dass die meisten dieser Faktoren auf unteren Ebenen im Bildungswesen zu verorten sind. Bereits innerhalb von Schulen, aber ebenso zwischen Einzelschulen der gleichen Schulform im selben Bundesland, ist die Heterogenität in der Schülerschaft und der Lehrerschaft oft beträchtlich. Hinzu kommen Unterschiede zwischen Schulen, die beispielsweise durch unterschiedliche Schulcurricula erklärt werden können oder durch Unterschiede in der wahrgenommenen Bedeutsamkeit dieser Thematik durch die erweiterte Schulleitung und diesbezüglich initiierten Schulentwicklungsprozessen. Darüber hinaus werden an vielen Schulen Orientierungsrahmen zur Regelung von Leistungsbeurteilungs- und Benotungsprozessen – zum Beispiel auf Ebene der Fächer – festgelegt. Diese können sich zwischen Schulen substantiell unterscheiden und somit Unterschiede in der Benotung von Schülerinnen und Schülern zwischen Einzelschulen verstärken. In der Summe scheinen die Faktoren auf unteren Ebenen im Bildungswesen, also innerhalb und zwischen Schulen innerhalb von Bundesländern, so bedeutsam zu sein, dass Bundeslandunterschiede nur noch einen vergleichsweise geringfügig zur Erklärung der mangelnden Vergleichbarkeit von Noten beitragen.

Redaktion: Sie schreiben auch mit Verweis auf die Forschung von Prof. Eckhard Klieme, es bestehe möglicherweise gar keine Notwendigkeit, Vergleichbarkeit für pädagogische Prozesse einzufordern, da es „praktisch kaum möglich“ sei, Noten vollends „vergleichbar zu machen“. Inwiefern können Noten dennoch nützlich sein, auch wenn sie nicht oder nur bedingt vergleichbar sind?

Hübner: Noten vollständig „vergleichbar zu machen“ ist praktisch utopisch. Gleichzeitig könnte man, wie Eckhard Klieme in einem Beitrag zur Vergleichbarkeit von Abiturnoten kürzlich andeutete, hinterfragen, ob eine universelle Vergleichbarkeit überhaupt immer notwendig ist. Lassen Sie mich diesen Punkt an einem Beispiel verdeutlichen:

Angenommen, an einer Schule gibt es zwei fünfte Klassen: eine leistungsstärkere und eine leistungsschwächere. Wenn die Lehrkräfte, die beide Klassen unterrichten, nach der sozialen Bezugsnorm Noten vergeben, das heißt Schülerinnen und Schüler innerhalb einer Klasse in eine Leistungsreihe bringen und dann benoten, werden die Noten der beiden Klassen wenig vergleichbar sein. Bei gleicher Leistung erhalten Schülerinnen und Schüler in der leistungsstärkeren Klasse eine schlechtere Note und in der leistungsschwächeren Klasse eine bessere Note. Gleichzeitig spiegelt diese Note, unabhängig von ihrer mangelnden Vergleichbarkeit, trotzdem die relative Leistung der Schülerinnen und Schüler in der jeweiligen Lerngruppe wider und kann daher zur Diagnose und Quantifizierung der Leistung in dieser Lerngruppe und zur Kommunikation mit den Eltern verwendet werden und möglicherweise auch als motivationaler Anreiz dienen. Wichtig ist, dass Noten allein sehr informationsarm sind und kaum Hinweise für konkrete Verbesserungen liefern. Deswegen verwenden einige Lehrpersonen, mit denen ich mich zu diesem Thema austausche, hybride Formen, die neben den Noten auch formative Elemente beinhalten, die stärker den Lernprozess fokussieren, beispielsweise durch regelmäßige Lernentwicklungsgespräche oder elaboriertes individuelles Feedback.

Redaktion: Sie schreiben auch von der Überlegung, ob man Noten an wichtigen

Gelenkstellen nicht einfach durch standardisierte Tests ersetzen könnte. Können Sie diese Idee genauer erläutern?

Hübner: Bei den ganzen Problemen von Schulnoten, könnte man leicht auf die Idee kommen, als ersten Lösungsansatz standardisierte Tests einzufordern, und zwar an den Stellen im Bildungswesen, an denen vergleichbare Leistungsindikatoren benötigt werden. Standardisierte Testergebnisse sind idealerweise hoch objektiv, valide und reliabel und daher zwangsläufig über Klassenkontexte hinaus sinnvoll miteinander vergleichbar. Die Leitidee einer stärkeren Standardisierung ist schon alt und findet sich beispielsweise auch im Zentralabitur wieder oder im gemeinsamen Pool der Abituraufgaben der Länder.

Redaktion: Welche Veränderungen, Vorteile und Risiken bringt ein vermehrter Einsatz von standardisierten Tests mit sich?

Hübner: Probleme einer einseitigen oder zu starken Fokussierung auf standardisierte Tests konnte ich kürzlich mit meinem Kollegen Daniel Koretz von der Harvard Graduate School of Education besprechen. Er forscht bereits lange zum Thema „Score Inflation“, also der künstlichen Verbesserung der Ergebnisse von Schülerinnen und Schülern in standardisierten Tests bei fehlender tatsächlicher Leistungsverbesserung. Dieses Phänomen lässt sich unter anderem durch eine verstärkte oder überwiegende Konzentration auf Testinhalte im Unterricht erklären. Am Ende kann eine solche Fokussierung zu einer starken Einengung der Inhalte des Unterrichts auf Testinhalte führen: Anstatt sich an der Breite des Curriculums zu orientieren, konzentrieren sich Lehrpersonen dann möglicherweise primär auf Testinhalte. Darüber hinaus besteht natürlich die Gefahr, dass Testinhalte, die oft am mittleren Leistungsniveau der Schülerinnen und Schüler ausgerichtet sind, für leistungsstarke Klassen viel zu einfach und für leistungsschwache Klassen viel zu schwierig sind. Die Flexibilität der Lehrperson, die Inhalte und das Niveau des Unterrichts und der Prüfungen an den Lernstand der Klasse anzupassen, würde also möglicherweise entfallen, wenn es nur noch einen Test für alle gäbe, und dies könnte insbesondere für benachteiligte Kinder mit Lern- und Sprachschwierigkeiten mit großen Hürden einhergehen. Außerdem ist es naheliegend, dass Schülerinnen und Schüler, deren Unterricht und vorherige Prüfungen stärker mit den Testinhalten übereinstimmen, besser in diesen Tests abschneiden, was Fragen in Bezug auf die Instruktionssensitivität der Tests aufwerfen würde. Auf diese und weitere „Nebenwirkungen“ weisen auch aktuelle Debatten in den USA hin. An vielen Standorten lässt sich dort eine Rückkehr zur beziehungsweise eine stärkere Berücksichtigung von Noten erkennen.

Redaktion: Was lehren uns moderne, erfolgreiche Bildungssysteme im Ausland über den Einsatz von Schulnoten? Haben diese Länder bessere Alternativen zu Schulnoten gefunden?

Hübner: Wichtig ist zunächst: Was anderswo funktioniert, muss nicht automatisch bei uns funktionieren. Die Identifikation und Verwendung bestimmter Elemente aus leistungsstarken Bildungssystemen als eine Art „Blueprint“ für unser System ist verlockend, wird der Komplexität der Unterschiede aber nur selten ohne Weiteres gerecht. Darüber hinaus ist wichtig zu berücksichtigen, dass Schulsysteme im Ausland häufig eingliedrig strukturiert sind. Daraus ergeben sich auch andere Diskussionen sowie andere Herausforderungen und Chancen in Bezug auf Noten und alternative Formen der Leistungsbeurteilung. Mein genereller Eindruck ist, dass erfolgreiche Bildungssysteme häufig einen stärkeren Fokus auf individuelle Förderung und den Lernprozess selbst legen und Noten eine eher untergeordnete Rolle spielen. Darüber hinaus finden sich an Gelenkstellen im

Bildungswesen im Ausland oft standardisierte Testoptionen. Besonders bekannt sind beispielsweise der American College Test und der Scholastic Assessment Test in den USA. Man versucht hier also mögliche Probleme in Bezug auf die Vergleichbarkeit von Noten zu überwinden, indem Zulassungsentscheidungen stärker an die Ergebnisse aus standardisierten Testverfahren gekoppelt werden. Auch in Deutschland findet sich inzwischen eine zunehmende Anzahl von Studiengängen, die individualisierte Auswahlgespräche oder sogar die Teilnahme an speziellen Studieneignungstests empfehlen oder einfordern. Ein besonders prominentes Beispiel ist hier der Mediziner-test. In anderen Ländern, beispielsweise den Niederlanden, spielt die Note wiederum kaum eine Rolle bei der Studienaufnahme. Schülerinnen und Schüler können dort – mit nur sehr wenigen Ausnahmen – Studiengänge also überwiegend frei aussuchen. Eine perfekte Lösung für die Probleme von Noten in Deutschland, die einfach zu implementieren wäre, habe ich auch im Ausland noch nicht gesehen.

Redaktion: Sie beschreiben einen Mix aus Lernstandserhebungen, standardisierten Aufgabenpools oder zentralen Abiturprüfungen als „einen vielversprechenden Mittelweg, um Unterricht an der kriterialen Bezugsnorm auszurichten“. Können Sie genauer erläutern, was Sie damit meinen? Was würde ein solcher Wandel im System konkret für die Arbeit der Schulleitungen und Lehrkräfte vor Ort bedeuten?

Hübner: Damit ist ein Mittelweg aus standardisierten und nicht standardisierten Elementen gemeint. Ich halte wenig davon, die Schule zu einer Testmaschinerie umzuwandeln, und auch die Belastung vieler Lehrkräfte, mit denen ich spreche, liegt derzeit am absoluten Maximum. Für erfolgreiche Reformen und Schulentwicklungsprozesse benötigen Lehrpersonen, neben gut erprobten, belastbaren Konzepten und guten Fortbildungs- und Unterstützungsstrukturen, vor allem viel zusätzliche Zeit. Ein vielversprechenderer Weg bestünde aus meiner Sicht derzeit darin, eine Stärkung formativer Elemente im Unterricht und eine stärkere Fokussierung auf den Lernprozess „trotz Noten“ anzugehen. Lassen Sie mich dies kurz erläutern: Verschiedene Studien legen nahe, dass Feedback von Lehrpersonen an Schülerinnen und Schüler insbesondere dann seine Wirksamkeit entfalten kann, wenn es nicht lediglich eine Rückmeldung zur Korrektheit der Antwort, sondern, in Abhängigkeit vom Vorwissen der Lernenden, weitere Informationen, beinhaltet wie Lösungsstrategien, Hinweise zu Fehlvorstellungen oder zum Arbeitsprozess. Gleichzeitig zeigen Studien, dass Lehrpersonen, beispielsweise in Arbeitsphasen, häufig primär Feedback zur Korrektheit der Antwort geben. Sicherzustellen, dass Lehrkräfte solche Lerngelegenheiten erkennen, und Strukturen zu schaffen, in denen sie wirklich genutzt werden können, halte ich für einen wichtigen Schritt.

Redaktion: Welche Schulnote würden Sie dem deutschen Schulnotensystem heute geben? Und warum?

Hübner: Noten sind ja, wie ich oben ausgeführt habe, für sich genommen wenig informativ. Das deutsche Notensystem ist in erster Linie effizient. Mit ihm wird es möglich, Millionen von Schülerinnen und Schülern durch das Schulsystem zu lenken und den Aufwand für Lehrkräfte im Bereich des Machbaren zu halten. Insbesondere in Zeiten von Lehrkräftemangel und einer zunehmenden Heterogenität der Schülerschaft sehe ich zumindest aktuell kaum praktikable Alternativen, die Noten flächendeckend ersetzen könnten. Umso wichtiger erscheint es mir aber daher auch, dass wir systematisch praxistaugliche Konzepte entwickeln und wissenschaftlich begleitet erproben. Aktuell erforsche ich in einem Projekt mit Kolleginnen und Kollegen aus der Computerlinguistik und der Fachdidaktik, ob Lehrpersonen mithilfe von KI-Tools bei Korrekturaufgaben entlastet werden können und Korrektur- und Benotungsprozesse standardisierter und damit vergleichbarer ablaufen. Solche Tools besitzen großes Potenzial, aber bis sie ausgereift sind und wirklich in der breiten Praxis eingesetzt werden können, wird es noch eine ganze Weile dauern.

Redaktion: Eine Abschaffung der Noten würden Sie also nicht empfehlen?

Hübner: Ein zentrales Argument, das Befürworterinnen und Befürworter einer Abschaffung von Noten häufig ins Feld führen, ist, dass die Prüfungskultur einen ganz wesentlichen Einfluss auf die Lernkultur hat. Eine gute Lernkultur könne nur dann entstehen, wenn die „schädliche“ Prüfungskultur abgeschafft würde. Auch wenn dieser Punkt umstritten ist – hier fehlen meines Erachtens wirklich belastbare Studien – bin ich davon überzeugt, dass eine zumindest bessere Lernkultur, um in dieser Denklogik zu bleiben, in vielen Klassenzimmern auch „trotz Noten“ entstehen kann, wenn das Wissensspektrum zu lernförderlichen Leistungsrückmeldungen systematischer genutzt wird. Nötig sind hierfür praktikable Konzepte, wie formative Elemente stärker im Unterrichtsalltag integriert werden können, eine wirklich belastbare Erprobung und eine empirische Absicherung dieser Konzepte. Darüber hinaus braucht es auch erheblich mehr Zeit für Lehrpersonen, um sich intensiver mit diesen Konzepten zu beschäftigen und diese auch praktisch einzuüben und umzusetzen. Damit praxistaugliche Konzepte flächendeckend realisiert werden können, sind also auch die Bildungspolitik in den Ländern und die Schuladministration gefordert, diesen für die Qualität unseres Bildungswesens sehr bedeutsamen Innovationsprozess engagiert und nachhaltig zu unterstützen. Hierzu laufen bereits einige vielversprechende Forschungsprojekte. In Baden-Württemberg findet beispielsweise derzeit ein einzigartiger Schulversuch statt, dessen Ergebnisse für die Beantwortung dieser Frage wegweisend sein werden. Die Befunde dieser sehr wertvollen Initiativen sollte man abwarten, bevor die Vor- und Nachteile einer flächendeckenden Abschaffung mit der notwendigen Sorgfalt, Sachlichkeit und auf Basis der aktuellen Erkenntnisse diskutiert werden können.

Redaktion: Herr Juniorprofessor Hübner, wir danken Ihnen für dieses Gespräch.



Zur Person

Nicolas Hübner ist Juniorprofessor für Schulpädagogik mit einem besonderen Schwerpunkt im Bereich Quantitative Forschungsmethoden an der Universität Tübingen. In seiner Forschung beschäftigt er sich u. a. mit Eigenschaften von lernförderlichen Formen der Leistungsbeurteilung sowie mit der Vergleichbarkeit und Bedeutung von schriftlichen und mündlichen Schulnoten.