



Universität Potsdam

Diether Hopf

## Zensuren und Tests zur Hochschulauswahl

first published in:

Flitner, Andreas (Hrsg.): Der Numerus Clausus und seine Folgen. Klett, Stuttgart 1976, S. 100-107

Postprint published at the Institutional Repository of Potsdam University:

In: Postprints der Universität Potsdam

Humanwissenschaftliche Reihe ; 87

<http://opus.kobv.de/ubp/volltexte/2009/3590/>

<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-35901>

Postprints der Universität Potsdam

Humanwissenschaftliche Reihe ; 87

**Diether Hopf**

## **Zensuren und Tests zur Hochschulauswahl**

Seit dem Staatsvertrag über die Vergabe von Studienplätzen von 1973 unterliegen die Abiturienten, die studieren wollen (Ausländer, Härtefälle etc. ausgenommen), einem einheitlichen Auswahlverfahren, wenn die Nachfrage das Studienplatzangebot übersteigt: Aus den Noten des Reifezeugnisses und den nicht im Reifezeugnis enthaltenen Noten der am Ende des

11. und 12. Schuljahres abgeschlossenen Fächer wird eine Durchschnittsnote gebildet; die Bewerber mit dem niedrigsten Durchschnittswert werden zugelassen. Welcher Mindestwert zur Aufnahme genügt, richtet sich, von Semester zu Semester wechselnd, vor allem nach der Anzahl freier Plätze und nach der Bewerberzahl.

Das Hochschulrahmengesetz vom Januar 1976 enthält flexiblere und teilweise neuartige Regelungen, von denen ich hier lediglich diejenigen diskutieren möchte, in denen Schul- oder Testleistungen im Vordergrund stehen. § 32 beschreibt das „allgemeine Auswahlverfahren“, dessen Kern die Abiturnoten bleiben, das jedoch auch solche Leistungen zu gewichten ermöglicht, „die über die Eignung für den jeweiligen Studiengang besonderen Aufschluß geben können“. Die Länder sollen im übrigen für die Vergleichbarkeit der Anforderungen und Bewertungen innerhalb eines Landes und zwischen den Ländern sorgen. § 33 regelt das „besondere Auswahlverfahren“, welches immer dann an die Stelle des allgemeinen Auswahlverfahrens treten soll, wenn es dort zu „unvertretbar hohen Anforderungen an den Grad der Qualifikation“ kommen würde. Bei dem sogenannten Feststellungsverfahren, welches zum Abitur hinzukommen oder, teilweise, an seine Stelle treten kann, denkt man insbesondere an Tests und an bewertete praktische Tätigkeiten.

Ich will im folgenden Grundlagen und Konsequenzen des mit dem Numerus clausus verbundenen Selektionsvorganges betrachten, mich aber dabei auf die Diskussion einiger Probleme beschränken, die bei den beiden wichtigsten Ausleseverfahren, Abiturnoten und Tests, auftreten.

Jede Selektion von Bewerbern um einen Ausbildungsplatz, um eine Stellung o. ä. soll dazu führen, daß sich unter den Angenommenen eine möglichst hohe Proportion an „Erfolgreichen“ befindet. Ausleseentscheidungen beinhalten demnach eine Prognose künftigen Verhaltens, die auf der Diagnose gegenwärtigen oder vergangenen Verhaltens beruht.

Die Sicherheit von Prognosen hängt *erstens* von der Qualität der Auswahlverfahren ab – Verfahren, die unzuverlässig sind, von verschiedenen Personen unterschiedlich ausgewertet werden und deren Ergebnisse bei verschiedenen Bewerbern nicht miteinander vergleichbar sind, dürften für eine Auslese untauglich sein. *Zweitens* muß das vorhergesagte Verhalten, beispielsweise die Bewährung in Studium oder Beruf, verläßlich gemessen werden können. *Drittens* muß nachweislich eine enge Beziehung zwischen dem Abscheiden im Auswahlverfahren und der späteren Bewährung bestehen, damit nicht unvertretbar viele Fehler bei der Auslese auftreten, viele Bewerber also zu Unrecht abgewiesen werden, obwohl sie hätten Erfolg haben können, und zugleich viele zu Unrecht aufgenommen werden, die sich dann nicht bewähren. Neue Auswahlverfahren können vorhandene

nur dann ergänzen, wenn dadurch die Anzahl der Auslesefehler merkbar zurückgeht.

Als Auswahlverfahren kommen vor allem Abiturnoten und in Zukunft möglicherweise Tests in Frage. Was die Qualität der Abiturnoten als Auswahlverfahren betrifft, bedarf es keiner ausführlichen Erörterung, da die oft bestätigten Forschungsbefunde, wonach Schulzensuren weder objektive noch verlässliche Maße darstellen, in letzter Zeit mehrfach dargestellt worden sind. Allenfalls bleibt hinzuzufügen, daß der Versuch, die Leistungen durch ein zentrales Abitur vergleichbar zu machen, zum Scheitern verurteilt ist, solange nicht für gleiche Lernbedingungen in den vorhergehenden Schuljahren gesorgt ist (vgl. hierzu auch die Ausführungen von D. Lenzen).

Tests lassen sich so konstruieren, daß sie objektivere und zuverlässigere Ergebnisse bringen als Zensuren. Ob sie das auch messen, was sie messen sollen, und ob sie eine bessere Prognose auf künftige Leistungen als Zensuren erlauben, wird später zu fragen sein.

Die zweite Voraussetzung, die für eine sichere Prognose gegeben sein muß, nämlich die verlässliche Messung des Verhaltens, welches man vorhersagen will, wirft die Vorfrage auf, was man eigentlich prognostizieren sollte: Sollen diejenigen Bewerber ausgewählt werden, welche gute Zensuren im Vor- oder Abschlußexamen ihres Studiums erwarten lassen, oder diejenigen, die sich im Beruf bewähren werden, oder sollen nur die möglichen Studienabbrecher ausgeschlossen werden? Berufserfolg wäre gewiß das befriedigendste Kriterium, will man doch vor allem diejenigen zum Studium zulassen, die später einmal gute Ärzte, Lehrer, Juristen sein werden. Aber alle Versuche, Berufserfolg zu definieren oder gar zu messen, sind bislang fehlgeschlagen – was nicht weiter verwunderlich ist angesichts der Schwierigkeit, Übereinstimmung darüber zu erzielen, woran er sich bemißt: Läßt er sich am Einkommen ablesen? Oder an der Berufszufriedenheit? Oder am Kollegenurteil? Soll man Berufsbewährung bald nach Studienabschluß bestimmen oder nach 10 Jahren oder am Ende eines Berufslebens? Trotz vielfältiger Forschungsbemühungen in diesem Problemfeld ist man bislang noch weit davon entfernt, überzeugende Indikatoren für dieses Kriterium zu finden. So kann es auch nicht weiter verwundern, daß es bislang nicht gelungen ist, auch nur partiell den Berufserfolg aus den Examensnoten im Studium vorherzusagen, wie beispielsweise die Arbeit von Hoyt (1965) belegt. So wünschenswert es wäre, über Auswahlverfahren zu verfügen, mit deren Hilfe man diejenigen Bewerber herausfinden und zum Studium zulassen könnte, die sich später im Beruf bewähren, die Voraussetzungen zu ihrer Konstruktion fehlen gegenwärtig und können erst in langfristigen Forschungsarbeiten, welche auch die wechselnden Anforderungen sich wandelnder Berufe berücksichtigen müssen, geschaffen werden. Man muß daher

jetzt und in den kommenden Jahren auf das optimale Kriterium, die Berufsbewährung, verzichten und sich vorläufig mit anderen Kriterien begnügen.

Hier zeigt sich nun, daß in der bisher vorliegenden Forschungsliteratur nahezu ausschließlich der Studienerfolg, und das heißt vor allem die Zensuren nach dem ersten Studienjahr oder im Vorexamen, in Einzelfällen auch im Abschlußexamen, als Kriterien gewählt worden sind, Leistungsurteile also, denen eine ähnliche Unsicherheit anhaftet wie den Schulzensuren. Dieser Befund ist um so erstaunlicher, als es gute Gründe gibt, die dafür sprechen, ein anderes Kriterium zu wählen, nämlich Studienabbruch (gegenüber Studienabschluß), wird doch von den Hochschullehrern auch einem Studenten mit der Abschlußnote 4 bescheinigt, daß er die Voraussetzungen für seinen Beruf besitzt; und die Examenszensuren lassen, wie erwähnt, keinen Schluß auf die Bewährung im Berufsleben zu. Könnte man mit einer gewissen Sicherheit herausfinden, welcher Bewerber vermutlich sein Studium nicht zum Abschluß bringen wird, so hätte man damit eine bessere Basis für die individuelle Beratung oder Entscheidung (oder könnte über Merkmale der Institution nachdenken, die Studienabbruch fördern).

Geht man nun der für jede Auslese zentralen Frage nach der Beziehung zwischen dem Abschneiden der Bewerber im Auswahlverfahren und ihrer späteren Bewährung nach, so folgt schon aus der Diskussion der ersten beiden Punkte, daß es erstaunlich wäre, wenn Zensuren im Abitur Zensuren im Universitätsexamen sicher voraussagten. Und in der Tat sind die bisher gefundenen Zusammenhänge sehr schwach ausgeprägt, so daß die Auswahl von Studienbewerbern mittels Durchschnittszensur (oder auch mittels gewichteter Einzelzensuren) mit einer hohen Fehlerquote verbunden ist.

Das Abitur stellt demnach nicht einmal dann ein brauchbares Auswahlverfahren dar, wenn man sich mit dem Studienerfolg als Kriterium begnügt. Darüber hinaus bleibt es eine offene Frage, ob man nicht hinsichtlich der Berufsrekrutierung gerade das Falsche getan hat: Werden die besonders guten Latinisten oder Mathematiker auch gute Lehrer sein? Welche Folgen wird es haben, wenn eine halbe Generation lang diejenigen Schüler Ärzte oder Psychologen werden, die der schulischen Spitzengruppe angehört und sich in einer Situation bewährt haben, welche zumindest auf den ersten Blick andere Qualifikationen erfordert als ihr Beruf? Gerade angesichts des mit dem Numerus clausus gegebenen, zerstörerischen Leistungsdrucks auf die Schulen (vgl. u. a. Blankertz und Lempp) wird man annehmen müssen, daß Qualitäten wie Kooperationsbereitschaft, Erfindungsreichtum im Gestalten interpersoneller Beziehungen, Initiative, Empathie,

kreative Ausgestaltung einer Rahmenvorgabe usw., Qualitäten also, die in manchen Akademikerberufen eine zentrale Rolle spielen, weniger noch als in den Zeiten des freien Hochschulzugangs in den Abiturnoten ihren Niederschlag finden.

Nach den Vorstellungen des Hochschulrahmengesetzes sollen nun in denjenigen Studienfächern, in welchen das Mißverhältnis von Angebot und Nachfrage sehr kraß ist, insbesondere Testverfahren zu einer Verbesserung der Auslese führen. Entwicklungsaufträge zur Testentwicklung für Medizin und Zahnmedizin, darüber hinaus aber auch für neuere Philologien, Psychologie, Naturwissenschaften und weitere Studienfelder wurden folgerichtig bereits erteilt; die Arbeiten sind, teilweise schon seit geraumer Zeit, voll im Gange. Mit der Möglichkeit zum erstmaligen Einsatz von Tests bei der Auslese wird zum Wintersemester 1978/79 gerechnet.

Die mit der Testkonstruktion betrauten Wissenschaftler stehen hier vor einer außerordentlich schwierigen Aufgabe, zumal das Hochschulrahmengesetz bestimmt, daß „grundsätzlich nicht die Kenntnisse festgestellt werden (sollen), die bereits Gegenstand der Bewertung in der Hochschulzugangsberechtigung sind“ (§ 33, 2). In den Projekten wird grundsätzlich die Berechtigung der Forderung anerkannt, die prognostische Gültigkeit der zu konstruierenden Verfahren nachzuweisen; dementsprechend sind Längsschnittuntersuchungen geplant oder bereits im Gang. Gleichwohl besteht keine Möglichkeit, Daten über den Prognosewert der Tests zu dem Zeitpunkt zu erhalten, in dem diese bereits die Auslese mitbestimmen sollen; auch wenn man sich auf den Studienerfolg als Kriterium beschränkt, müssen ja zunächst die Testverfahren entwickelt sein, bevor sie, einzeln und zusammen mit den Abiturnoten, auf ihre prognostische Gültigkeit geprüft und im Anschluß daran revidiert werden können. Streng genommen muß man während der Längsschnittuntersuchungen sogar denjenigen Bewerbern eine Bewährungschance geben, die nach ihren Testwerten nicht aufgenommen werden dürften, um überprüfen zu können, ob der Test nicht Geignete ausgeschlossen hätte.

In einer solchen Situation bietet sich als Ausweg der Versuch an, die Tests auf der Grundlage sorgfältiger Vorüberlegungen über das Kriteriumverhalten, beispielsweise also über die „Studierfähigkeit“ im Medizinstudium, unter Berücksichtigung einschlägiger wissenschaftlicher Befunde, nach Auskünften von Experten und Praktikern sowie aufgrund von Arbeitsplatzanalysen zu konstruieren – ein höchst diffiziler Forschungsprozeß, in den ständig korrigierende Befunde aus der Erprobung von Testaufgaben und aus der theoretischen Weiterarbeit am Konzept der „Studierfähigkeit“ eingehen, und der prinzipiell unabschließbar ist; für die praktische Anwendung solcher Tests muß ein Zeitpunkt gesucht werden, zu dem sie für den

jeweiligen Verwendungszweck als hinreichend gesichert gelten. Für ein einzelnes Fach entsteht auf diese Weise eine ganze Testbatterie. Beispielsweise kommen für das Feststellungsverfahren in Medizin Tests zu folgenden Merkmalen in Frage: manuelle Geschicklichkeit, optisches und akustisches Wahrnehmungsvermögen, räumliches Vorstellungsvermögen, Farbdifferenzierung, Fähigkeit zu genauer Beobachtung, naturwissenschaftliches Verständnis, Aufnahmefähigkeit für Information, Gedächtnis für Namen und Formen, Konzentration und Arbeitsorgfalt, Einfühlungsvermögen, Interesse an medizinischen Problemen, logisches Denken, Arbeitsplanung<sup>1</sup> – Tests also, die zumindest teilweise Verhalten messen, welches in die Abiturnoten kaum eingegangen sein dürfte.

Das aufgeführte Beispiel hilft, die bei der Verwendung solcher Tests zur Selektion auftretenden Probleme deutlicher zu sehen.

Zunächst müßte, nachdem die Testkonstruktion abgeschlossen ist, überprüft werden, ob es berechtigt ist, etwa 20jährige Bewerber nach den aufgeführten Dimensionen – z. B. ihrer manuellen Geschicklichkeit – auszulesen, oder ob es sich nicht teilweise um *üb bare* Verhaltensweisen handelt, die sinnvollerweise erst während des Medizinstudiums aufgebaut oder verbessert werden. Wenn dies der Fall ist – Forschungen würden darüber Aufschluß geben –, wäre es problematisch, auf einer solchen Grundlage zu selektieren.

Ein zweites Problem läßt sich ebenfalls erst in Angriff nehmen, wenn die Einzeltests vorliegen: Nach welchem Modell soll die Entscheidung über Aufnahme oder Zurückweisung erfolgen? Können sehr niedrige Testergebnisse in bestimmten Tests ausgeglichen werden durch besonders hohe Werte in anderen (kompensatorisches Modell)? Oder müssen in sämtlichen Tests bestimmte Minimalwerte erreicht werden (additives Modell)? Wie lassen diese Minimalwerte sich festsetzen und begründen? Welches Gewicht sollen einzelne Abiturnoten oder die Durchschnittszensur in dem Gesamtverfahren erhalten? Um derartige Fragen zu beantworten, gibt es bewährte – allerdings zeitaufwendige – Verfahren der empirischen Überprüfung, mit deren Hilfe sich angeben läßt, mit welchen Kombinationen und Gewichtungen von Testergebnissen sich der Studienerfolg am ehesten vorher-sagen läßt. Darüber hinaus bedarf es auch der inhaltlichen Entscheidung darüber, an welcher Stelle bei bestimmten Tests bestimmte Leistungen nicht unterschritten werden sollten; hierzu sind Arbeitsplatz- und Berufsanalysen die geeignete Voraussetzung.

Drittens besteht die Schwierigkeit, daß auch das Kriterium Studienerfolg von Universität zu Universität etwas Verschiedenes beinhalten kann. Vielleicht ist zur Exemplifizierung dieses Problems die Medizin weniger geeignet als andere Studienfächer, in denen von Ort zu Ort recht unterschiedliche Leistungen honoriert werden. (In den USA ist es ein alltäglicher Be-

fund, daß dieselben Testergebnisse und Schulnoten in der einen Universität mit einer bestimmten Wahrscheinlichkeit einen erfolgreichen Abschluß erwarten lassen, in einer anderen dagegen nicht.) Auch hier läßt sich nur durch weitere Forschung Klärung herbeiführen.

Schließlich sei auf den bei der Testkonstruktion üblichen Befund hingewiesen, daß bestimmte Untergruppen der Bewerber sich systematisch in bestimmten Tests voneinander unterscheiden. So hat auch Trost (1975, S. 174) bei seinem „Test der akademischen Befähigung“ herausgefunden, daß Männer und Frauen signifikant verschiedene Testpunktwerte erreichen. Derartige Ergebnisse erfordern einerseits die weitere Klärung der Ursachen und der Folgen solcher Unterschiede, andererseits aber in jedem Falle eine Entscheidung darüber, wie man bei der Auslese verfahren will: Legt man die absoluten Testergebnisse ohne Rücksicht auf die Zugehörigkeit zu den Untergruppen zugrunde, so wird die eine Gruppe unter den erfolgreichen Bewerbern regelmäßig unterrepräsentiert sein; verwendet man unterschiedliche Normen für jede Gruppe, kann man zwar die gleichen Proportionen von Bewerbern rekrutieren, muß aber die unterschiedlichen absoluten Standards begründen.

Zusammenfassend läßt sich feststellen, daß innerhalb des begrenzten Zeitraums, in welchem eine spürbare Diskrepanz zwischen Studienplatzangebot und -nachfrage bestehen bleibt (sofern die Universitäten sich nicht öffnen), zwar Tests entwickelt werden können, die auf einem hohen theoretischen und konstruktionstechnischen Niveau stehen, und die auch für die Studienberatung außerordentlich hilfreich sein könnten, daß es aber trotz aller derzeit in den laufenden Projekten versammelten wissenschaftlichen Kompetenz kaum gelingen wird, eine tragfähige Grundlage für Selektionsentscheidungen zu gewinnen. Selbst wenn sich nach Ablauf von Längsschnittuntersuchungen in einigen Jahren herausstellen sollte, daß die verwendeten Tests – allein oder auch zusammen mit den Abiturnoten – den Studien Erfolg mit einer gewissen Wahrscheinlichkeit vorherzusagen erlauben, werden die bei der Selektion notwendig auftretenden Fehler (zu Unrecht abgewiesene und zu Unrecht aufgenommene Bewerber) so zahlreich bleiben, daß keiner der Beteiligten befriedigt sein wird.

Der faszinierte Verweis auf die breite und anscheinend erfolgreiche Testverwendung in den USA übersieht, daß es sich dort um eine unvergleichbare Situation handelt: Tests dienen dort der Klassifikation, nicht aber der Selektion, da, abgesehen von Medizin und Jura, jeder Studierwillige auch einen Studienplatz findet und nur zur Debatte steht, an welcher Universität er am besten aufgehoben ist. Außerdem beziehen sich die Tests dort auf ca. 75 % der Altersgruppe, welche die Sekundarschule abschließen, nicht aber auf eine hochselektierte Gruppe wie unsere Abiturienten, denen



die allgemeine Hochschulreife ja nicht ohne Begründung zugesprochen wurde. Abgesehen davon aber gibt es auch in den USA eine zunehmende Kritik am Testwesen, in der die Ungleichbehandlung der Bewerber (schwarz/weiß, männlich/weiblich, arm/reich etc.), vor allem aber die negativen Rückwirkungen der Tests auf die vorausgehenden Schuljahre wiederkehrende Themen sind. Darüber hinaus hat insbesondere McClelland (1973) in konstruktiver Weise die Diskussion um die Partialität derjenigen Merkmale, welche mit den gebräuchlichen Tests gemessen zu werden pflegen, wiederbelebt. Seine Argumente führen auf die oben aufgeworfene Frage zurück, ob die Zensuren im Hochschulexamen ein geeignetes Kriterium sind oder ob nicht vielmehr der Versuch unternommen werden sollte, potentielle Studienabbrecher frühzeitig zu identifizieren (wobei es, wie gesagt, offen ist, welche Maßnahmen dann zu treffen wären). Dazu dürften aber weder Abiturnoten noch die üblichen Testergebnisse brauchbare Anhaltspunkte liefern, da Studienabbruch sich nur zum geringen Teil aus Leistungsdefiziten erklärt und ganz andere Persönlichkeitsmerkmale die entscheidende Rolle zu spielen scheinen (Dörner, 1967). Man muß also nicht nur Zweifel daran haben, daß sich die Hoffnungen erfüllen, mit Hilfe der Tests eine bessere Auslese durchzuführen, sondern es stellt sich darüber hinaus die im Zusammenhang der Testkonstruktion zentrale Frage, ob die notwendigen Längsschnittuntersuchungen nicht auf ein fragwürdiges Kriterium gerichtet wären, wenn sie die Qualität der Auslese an den Studienzensuren überprüfen, nicht aber den Gründen für Studienabbruch (und schon gar nicht der Berufsbewährung) nachgehen.

In einer Situation, wo von einem so kompetenten Gremium wie dem Wissenschaftsrat die Öffnung der Hochschule (außer für Medizin) auch in den kommenden, schwierigen Jahren für möglich erklärt wird, ist der Versuch, das Heil in der Perfektionierung der Selektion zu suchen, besonders fehl am Platz; vielmehr sollten die Anstrengungen darauf gerichtet werden, diejenigen Bedingungen zu schaffen, die bei der Nachfrageorientierung der Universität gegeben sein müssen.

Dies bedeutet nicht, daß die Entwicklung von Tests wieder eingestellt werden sollte; im Gegenteil können Tests wertvolle Zusatzinformationen zur Beratung von Studienanfängern bereitstellen, auf deren Grundlage der einzelne besser über seinen Berufsweg entscheiden kann.

#### *Anmerkung*

<sup>1</sup> Aus: information bildung wissenschaft (hrsg. v. Pressereferat des BMBW), Nr. 1/76 vom 22. 1. 76, S. 8.