

Leserbrief zum Artikel „Vom Sinn und Unsinn der Lehrevaluationen an deutschen Hochschulen – Über den Missbrauch von Statistik“ von Frank Brand in DNH 1/2015

Problemlösung statt Methodenfundamentalismus

In Heft 1/2015 der Zeitschrift Die Neue Hochschule reflektiert Frank Brand über den „Missbrauch von Statistik“. Das Hauptargument besteht darin, dass Schulnoten ordinalskaliert sind, weshalb das arithmetische Mittel und die Standardabweichung keine zulässigen Berechnungsmethoden seien. Der Autor versteht die Noten als Transkription der Verbalisierung, z. B. in den „Noten eins und zwei die Abkürzungen für sehr gut und gut ...“, sodass „die verwendeten Zahlen ... nichts weiter als Labels ...“ seien (S. 37). Realiter sollte aber umgekehrt die Verbalisierung ein Label für eine Leistungszahl sein, eine Verständnishilfe. Manche Menschen können sich unter einer Bezeichnung wie „gut“ eben mehr vorstellen als unter einem Zahlenwert, zumal dieser auch anders belegt werden kann, vgl. die umgekehrte Notenskala in der Schweiz.

Nimmt man z. B. eine 60-Punkte-Klausur, verbunden mit einer gängigen Konvention wie etwa der, dass davon 50 Prozent für das Bestehen erforderlich sind, und teilt die Punkte in äquidistante Abschnitte ein (ab 39 Punkte = 3; 48 Punkte = 2; 57 Punkte = 1), dann sind Noten lediglich mathematisch transformierte Leistungspunkte, die Verbalisierung ist irrelevant. Dass diese Gruppierung am oberen und unteren Rand ungleiche Klassenbreiten aufweist und „eigentlich“ nur eine Ordinalskala liefert, mag den Mathematiker stören. In der Lebensrealität entspricht dies der Einsicht, dass es obere und untere Leistungsgrenzen gibt, die keiner weiteren Differenzierung bedürfen.

Man mag problematisieren, ob es Prüfern möglich ist, Aufgaben aus tatsächlich gleichgewichtigen Leistungspunkten zusammenzustellen, ob unterschiedliche Prüfer gleich vorgehen, dass es Unterschiede zwischen Hochschulen und Ländern gibt etc. Statt übliche Berechnungsmethoden mit Begriffen

wie „peinlich“ und „sinnlos“ zu belegen, bestünde Lösungsorientierung darin, an der Vereinheitlichung von Punkteskalen und Leistungsstandards zu arbeiten. Wo dies nicht fruchtet, muss man Noten mit der gebotenen kritischen Distanz betrachten, wie dies in der Personalselektion längst üblich ist – entbehrlich werden sie dadurch nicht. Sicher hat eine zweite Nachkommastelle wenig Bedeutung, aber entgegen der Meinung des Autors ist ein Notenunterschied von 0,45 keineswegs uninterpretierbar. Verbessert sich z. B. der „Notendurchschnitt“ zwischen Semestern um solch einen Wert, dann ist dies ein Anlass für Ursachenforschung und damit sinnvoller Bestandteil des QM. Haben die Studierenden in anderen Fächern ebenfalls besser abgeschnitten, könnte es sich um eine leistungsfähigere Semestergruppe handeln. Sind die anderen Noten gleich geblieben, wird sich der an seinen Studierenden interessierte Dozent selbst fragen, ob es an der Lehre gelegen hat, ob die Klausurfragen diesmal leichter waren u. a. m.

Der Vorschlag, die Analyse auf Mediane zu stützen, löst nicht die Aufgabenstellung. Der Median hat seinen Nutzen besonders bei breiten Skalen, bei denen Extremwerte das arithmetische Mittel stark verzerren können. Innerhalb der nur fünfstufigen Notenskala tritt dieses Problem gar nicht auf. Bei den zunehmend ins Positive tendierenden Hochschulnoten wird der Median meist bei „2“ liegen – so gewinnt man zu wenig verwertbare Information (noch ärmer ist der Informationsgehalt der „Spannweite“ als Verteilungsmaß innerhalb der Fünferskala). Der Autor zitiert selbst ein Ranking mit dem Fazit: „Die Plätze 3 bis 14 müssten bei korrekter statistischer Vorgehensweise exakt einem Platz zugeordnet werden.“ Wenn aber der präferierte Indikator dazu führt, dass man kaum noch Unterschiede feststellen kann, dann ist er unbrauchbar.

Brand bezeichnet das Marketing als eines der „Gebiete mit falscher statistischer Vorgehensweise“ und konstatiert dort „ein geschlossenes System von nicht zutreffenden Annahmen“, wenn Noten wie intervallskalierte Daten verarbeitet werden und mehrere Quellen es rechtfertigen, Faktorenanalysen durchzuführen. Man wird jedoch in der empirischen Praxis feststellen, dass solche Verfahren gegenüber dem, was der Mathematiker als „Verstoß“ identifiziert, robust sind. Versuche mit verschiedenen Skalierungen (6-, 10-, 100-stufig mit und ohne Verbalisierung) zeigen, dass die Urteile der Konsumenten sich nicht signifikant unterscheiden, wobei die mit multivariater Statistik ermittelten Positionierungsmodelle skalierungsunabhängig ein identisches Bild der Marktsituation liefern. Das ist es, was die Praxis von einer anwendungsorientierten Wissenschaft fordert: reliable, d. h. methodenunabhängig stabile Ergebnisse, die als Basis für konkretes Managementhandeln dienen können. Mit dem Hinweis, dass man dies alles hätte gar nicht berechnen dürfen, ist niemandem geholfen.

*Prof. Dr. Joachim Riedl
Leiter des Masterstudiengangs
Marketing Management
Hochschule Hof
joachim.riedl@hof-university.de*

*Dr. Sebastian Zips
Leiter Research
AccessMM e.V.
dr.zips@accessmm.de*

Leserbriefe geben nicht notwendigerweise die Meinung der Redaktion wieder. Eventuelle Kürzungen behält sich die Redaktion vor.