

Über einige Irrtümer bei der Einschätzung der Evaluation an Hochschulen

Gralki, Heinz O.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Gralki, H. O. (2001). Über einige Irrtümer bei der Einschätzung der Evaluation an Hochschulen. *Sozialwissenschaften und Berufspraxis*, 24(4), 397-403. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-41005>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.



Über einige Irrtümer bei der Einschätzung der Evaluation an Hochschulen

Heinz Gralki

Schüler:

Meister, kann man mit einem Lineal die Entfernung zwischen zwei Punkten messen?

Meister I. Grades:

Nein, mein Sohn, das geht nicht, denn die Erde ist eine Kugel!

Schüler:

Aber Meister, warum benutzen denn so viele Menschen ein Lineal?

Meister I. Grades:

Weil sie nur eine methodische Halbbildung haben. Sie wissen nicht, dass man zur korrekten Bestimmung einer Länge auf einer Kugel den Erdradius kennen und sich der Erkenntnisse der Trigonometrie bedienen muss!

Das Dilemma

Kromrey formuliert in seinem Beitrag in „Sozialwissenschaften und Berufspraxis“ 1/99 einige Bedenken zur Praxis der Studentenforschungen an unseren Hochschulen.¹ Um seine Argumentation zu belegen, bezieht er sich in weiten Bereichen auf unsere Arbeit im „Projekt pro Lehre“, dem Evaluationsprojekt der Freien Universität Berlin. Er kennt die

¹ Helmut Kromrey, Von den Problemen anwendungsorientierter Sozialforschung und den Gefahren methodischer Halbbildung, in: Heft 1/99 S. 58ff

Arbeit gut, weil er lange Zeit Mitglied des wissenschaftlichen Beirates in diesem Projekt war und das Projekt in vielfältiger Weise unterstützt hat.²

Sein Beitrag in diesem Heft hinterlässt allerdings eine gewisse Ratlosigkeit. Ganz sicher hat er mit seiner Argumentation in weiten Teilen recht. Genauer: die meisten seiner Sätze stimmen. Und dennoch: seine Argumentation geht dicht am Thema vorbei. Da unsere Arbeit in seinem Beitrag einen prägnanten Stellenwert hat, möchte ich hier – wenn auch spät, denn das Projekt ist an der FU mangels Interesse der Hochschulleitung längst eingestellt worden – auf einige seiner Stellungnahmen eingehen.

Das Problem methodische Halbbildung in der Evaluation

Kromrey beklagt, dass die Evaluation an unseren Hochschulen z.T. von Personen betrieben wird, die häufig nicht über eine genügende Qualifikation verfügen.

Das Problem ist sicher von ihm hübsch beschrieben.

Statistik und empirische Sozialforschung ist an unseren Universitäten nicht besonders beliebt. Kenntnisse in beiden Bereichen können kaum in Grundzügen von den meisten Studierenden – aber auch von Hochschullehrern – erwartet werden. Der sachgerechte Umgang mit Problemen der Evaluation, der Fragebogenentwicklung, statistischen Maßzahlen und Zufallsmodellen gehört nicht zur Allgemeinbildung.

Man kann dies beklagen, wie Kromrey es tut, man kann es aber auch anders ausdrücken:

Statistik und empirische Sozialforschung ist in besonderem Maße eine didaktische Herausforderung an Dozenten. Und dies in einigen Fächern mehr als in anderen. In Pädagogik (wo ich selbst das Glück hatte, fünf Jahre lang Methoden zu lehren) mehr als in BWL, in der Medizin weniger als in der Soziologie. Das hängt natürlich mit den Erwartungshaltungen der Studierenden zusammen – und auch mit ihrer Vorbildung. Wer in der Schule einen Leistungskurs Mathematik besucht hat, kann einen Korrelationskoeffizienten besser verstehen als jemand, der sich in der Schule eher mit philologischen Inhalten beschäftigt hat.

Oder um es anders auszudrücken: In der Soziologie und der Pädagogik muss der Statistikdozent didaktisch viel gewandter sein als in Medizin und BWL. Und das ist nicht immer der Fall.

² PROJEKT PRO LEHRE (PPL): Das Evaluationsprojekt zur Evaluation der Lehre an der Freien Universität Berlin. Verantwortlich: Dr. Heinz Gralki und Heidemarie Hecht. <http://userpage.fu-berlin.de/~pplfu/Welcome.htm> Das Projekt wurde 2000 weitgehend eingestellt, weil die Leitung der FU – im Zusammenhang mit personellen Wechsels im Präsidialamt – das Interesse an einer systematischen Lehrbefragung verloren hatte.

Einer der Gründe für die richtig beobachtete methodische Halbbildung ist sicher auch die didaktische Unbeholfenheit, mit der der Stoff häufig vermittelt wird, ein anderer aber auch die nicht immer vorfindbare Qualität der entsprechenden Lehrbücher.

Und so mag es auch sein, dass das eine oder andere Evaluationsprojekt nicht mit der notwendigen Sachkompetenz bestückt ist – die Regel ist dies nach meinem Eindruck allerdings nicht.

Die drei Fehler, die wir nach Kromreys Auffassung in unserer Arbeit gemacht haben

1. Fehler bei der Informationssammlung

Als Beispiel für unprofessionelles Arbeiten führt Kromrey das von mir mit verantwortete „Studienbarometer“ der FU Berlin an.

Die angesprochenen Erhebungen wurden über einige Jahre von dem „Projekt pro Lehre“ (PPL) durchgeführt und bezogen sich auf 35 Fächer der FU. In jedem Fachbereich wurden jährlich einmal 100 Studierende gebeten, bestimmte Elemente ihres Fachbereichs (z.B. die Bibliothek, die Rechnerausstattung, die Beratung durch die Lehrenden, die didaktische Qualität der Lehre, usw.) auf einer Fünferskala zu beurteilen.

Kromrey hält dieses Instrument für unzulänglich und oberflächlich. Nun gut, darüber mag man streiten – und Alternativen entwickeln.

Unverständlich bleibt aber, warum das Instrument so verkürzt wiedergegeben wird. Es wird einfach unterschlagen, dass das Instrument zwar aus den 17 im Aufsatz wiedergegebenen Vorgaben besteht, dass aber alle Fächer die Möglichkeit hatten, eigene Fragen hinzuzufügen – und davon wurde auch Gebrauch gemacht. Durch diese Option sollte gerade der fachspezifische Bezug hergestellt werden, den Kromrey so schmerzlich vermisst.

Pikanterweise hatten die Soziologen der FU, also die Disziplin, der auch Kromrey angehört, von dieser Möglichkeit keinen Gebrauch gemacht.

Weiter unterschlägt Kromrey, dass das Instrument auch eine Rückseite besitzt, auf denen die Studierenden Lob, Kritik und Verbesserungsvorschläge zum Fach äußern konnten – und dies in einem erstaunlichen Umfang auch getan haben. In der Regel lieferten diese Anmerkungen eine Fülle von wertvollen Hinweisen für die einzelnen Fächer.

2. Fehler bei der Validität unserer Erhebungsinstrumente

Kromrey merkt an, dass wir mit dem genannten Instrument und dem Instrument ELVA (Empirische Lehrveranstaltungsanalyse) die Qualität der Lehre verfehlen und meint, wir würden nur eine subjektive und verallgemeinernde Einschätzung, z.B. der didaktischen

Kompetenz der Lehrenden im jeweiligen Fach, erfragt haben. Diese Anmerkung ist völlig richtig. Um genau diese subjektiven Einschätzungen ging es uns.

Ja, wir haben die Zufriedenheit der Studierenden mit ihrem Fach, mit ihrer Universität und mit ihren Lehrveranstaltungen erfragt. Und wir hielten dies auch für wichtig.

Wir hatten – jedenfalls im Zusammenhang mit PPL – nie einen höheren Anspruch. Man wird vergeblich in allen unseren Dokumenten zur Evaluation der Lehre den Anspruch formuliert finden, wir hätten die Qualität der Lehre messen wollen. Mir erscheint es etwas unfair, unsere Arbeit an einem Anspruch zu messen, den wir nie erhoben haben – den wir im Gegenteil immer zurückgewiesen haben, auch gegenüber höheren Instanzen unserer Universität.

Allerdings bestehen wir darauf, dass die Zufriedenheit der Studierenden mit ihrer Lehre ein wichtiger Bestandteil einer qualitativ hochstehenden Lehre ist. Wir können uns schlichtweg keine gute Lehre vorstellen, mit der die Studierenden unzufrieden sind – und auch keine schlechte, mit der die Studierenden zufrieden sind. Und wir bestehen auch darauf, dass die Antworten der Studierenden legitime Bewertungen der durch unsere Vorgaben angesprochenen Sachverhalte sind – allerdings Bewertungen aus einer studentischen Perspektive.

Da Kromrey uns also explizit bescheinigt, dass unser Instrument für die Erfassung von Zufriedenheitsurteilen sinnvoll ist, können wir unsere Argumentation hier also abbrechen und erhobenen Hauptes die Arena verlassen. Da aber im Aufsatz noch eine Reihe anderer Ungereimtheiten sind, lohnt es sich, hierauf doch noch etwas genauer einzugehen.

3. Fehler bei der Informationsaufarbeitung

Natürlich wird auch in diesem Beitrag wieder auf das Problem der Verwendung von Mittelwerten zurückgegriffen - ein Thema, das in einer Vorlesung zur Statistik ungefähr in der dritten Stunde angesprochen wird und mittlerweile zum Standardrepertoire aller Evaluationskritiker gehört.

Es amüsiert, dass Kritiker der Evaluation immer wieder annehmen, gerade in der dritten Stunde hätten die Evaluatoren in ihrem Studium gefehlt.

Jedem auch nur halbgebildeten Evaluator ist das Argument bekannt, und es ist allen so vertraut, dass man darüber nicht immer wieder Abhandlungen schreiben muss. Allerdings hätte man in einigen Dokumenten von PPL fündig werden können, wo wir dieses Problem auch dem Laien erklären.

Aber Kromrey rüttelt auch an den Grundfesten der quantitativen Sozialforschung und übersieht, dass auch in qualitativ hochstehenden Arbeiten Urteile von Befragten gemittelt werden.

Und dies aus gutem Grund.

Der Mittelwert gibt einen HINWEIS - mehr nicht - auf die zentrale Tendenz einer Häufigkeitsverteilung. Sehen wir uns dann noch die Form unserer Häufigkeitsverteilungen

(z.B. Schiefe und Steilheit) an - und das haben wir selbstverständlich getan -, dann erweisen sich Mittelwerte immer wieder als ein einfaches und robustes Maß zur schnellen ersten Beschreibung einer Datenmenge. Und außerdem als effektiver Anreiz zur Diskussion.

Von „Artefakten“ - diesem immer mit etwas Häme vorgetragenen Vorwurf - kann bei unseren Ergebnissen gar keine Rede sein.

Wir haben nicht über den Zusammenhang von Sonnenflecken und Aktienkursen spekuliert, und auch nicht über den Zusammenhang von Storchenzahl und Geburtenhäufigkeit, diesen beiden, nun schon sehr abgedroschenen, Beispielen für Scheinzusammenhänge und Artefakte.

Gegen den Vorwurf, Artefakte produziert zu haben, spricht auch, dass die von PPL erhobenen Mittelwerte in hohem Maße über die Zeit stabil waren - was sie nicht hätten sein dürfen, wenn Kromrey recht hätte!

Zeigten sich dennoch einmal zwischen zwei Zeitpunkten deutliche Unterschiede, dann konnten wir mit einer hohen Wahrscheinlichkeit davon ausgehen, dass sich in der Wirklichkeit etwas geändert hatte.

Zeigte sich z.B. in unseren Daten, dass die didaktische Kompetenz des Lehrkörpers plötzlich besser bewertet wurde, dann fand sich z.B. die Erklärung, dass einige Lehrende, die nicht zu den didaktischen Koryphäen zählten, aus Altersgründen ausgeschieden waren. Wurde plötzlich die Rechnerausstattung im Fach besser beurteilt, stellte sich z.B. heraus, dass fünf neue Rechnerarbeitsplätze eingerichtet worden waren.

Die Beispiele könnten beliebig erweitert werden.

Wenn Kromrey nun fragt, was es bedeutet, dass das Institut für Soziologie bei der Vorgabe „Hilfsbereitschaft der Verwaltung“ nur einen Wert von 3,68 hat und damit auf Platz 29 aller Fächer lag, so ist die Antwort ganz einfach. Es handelt sich um kein „Artefakt“, sondern die Studierenden bewerteten die Hilfsbereitschaft am Institut für Soziologie schlechter als in anderen Fächern.

Eine sorgfältige Inspektion der offenen Antworten hätte wahrscheinlich schon Klarheit geschaffen und wenn nicht, so hätte man dieser Frage mit einer weiteren Untersuchung nachgehen können, wie wir es in anderen Fächern getan haben - und so z.B. herausgefunden haben, dass das Problem besonders im Prüfungsamt (natürlich ist das nur ein fiktives Beispiel) lag.

Mittelwerte sind eben ein sehr sensibler Anzeiger für Beurteilung von Sachverhalten. Aber selbst wenn man dies bestreiten würde, hätte man uns keine methodischen Vorwürfe machen können: die genannte Kritik am Mittelwert antizipierend, haben wir unseren Ansprechpartnern immer auch die entsprechenden Häufigkeitsverteilungen mitgeteilt. Niemand musste sich mit der Angabe des Mittelwertes begnügen.

Amüsiert haben wir dann aber immer feststellen müssen, dass sich die Diskussion in den Fächern und Fachbereichen doch fast immer nur auf die Mittelwerte bezog. Wer unseren Datensatz darüber hinaus noch weiter erkunden wollte, dem haben wir auch immer noch das entsprechende SPSS-Datenfile zur Verfügung gestellt.

Die Vorwürfe gegen die Verwendung des Mittelwertes lösen sich also bei sorgfältiger Analyse unserer Arbeit in Nichts auf.

Dass schließlich mit der ungenügenden Aussagekraft von Ranking-Positionen argumentiert wird, ist ein wenig boshaft, da wir ausdrücklich davor warnten und in unserer Darstellungsform eine grobe Dreiteilung der Spannweite in positiv / mittelmäßig / negativ gewählt hatten.

Mein Problem mit Kromreys Argumentation

Was schließlich die Kromrey-Grafik betrifft, mit der der Vorwurf der Artefaktproduktion untermauert werden soll, so befinde ich mich in einem Dilemma.

Ich kenne die Grafik schon seit vielen Jahren und habe sie bis heute nicht verstanden. Offenbar sind 10369 Studenten befragt worden. Weder aus der Legende noch aus dem Text wird allerdings klar, worauf sich die Befragung bezieht. Ist es eine Veranstaltung, sind es mehrere Veranstaltungen? Ist es ein Befragungszeitpunkt oder sind es mehrere? Sind Studenten nur einmal oder mehrfach befragt worden?

Damit kann man aber noch leben, völlig unverständlich sind jedoch die verschiedenen Gruppengrößen. Was bedeuten sie? Warum haben verschiedene Gruppen gleiche Linienart (z.B. die 2220er Gruppe und die 477er Gruppe)?

Mein Fragenkatalog ließe sich noch erweitern.

Es wird sicher seine Richtigkeit haben mit der Grafik. Schade nur, dass man das als Leser letztlich nicht beurteilen kann, weil entsprechende Informationen nicht vorhanden sind.

Aus der Grafik wird mir nichts klar, und sie erscheint mir denkbar ungeeignet für die weitere Argumentation.

Wenn diese Grafik allerdings als Beleg für Bankbildungen in den Urteilsstrukturen dienen soll - was sich mit etwas Mühe aus dem Text herauslesen lässt - dann wird damit nur eine Selbstverständlichkeit beschrieben.

Natürlich gibt es Gruppen, die alles positiv und andere, die alles negativ beurteilen. Aber was ist damit gesagt?

Was schließlich die triumphierende Feststellung betrifft, man habe keine Gruppe gefunden, die so urteilt, wie es die Mittelwerte anzeigen, so ergibt sich mir ein zweites Dilemma.

Welcher Methodiker wird denn ernsthaft versuchen, den Mittelwert in der Wirklichkeit wiederzufinden?

Wenn deutsche Studierende im Mittel ein mittleres monatliches Budget von 1132,14 DM haben, wo ist denn der konkrete Student, der genau diese Summe im Geldbeutel hat. Ist die Suche nach dieser Person nicht etwas naiv und verkennt die Bedeutung des arithmetischen Mittels?

Fazit

Wir erkennen die Probleme unserer Arbeit und sind dankbar für jede Kritik.

Mit den Ergebnissen unserer Arbeit, die von Kromrey als „Artefakte“ bezeichnet wurden, sind wir allerdings zufrieden, zeigten sie doch immerhin schnell und präzise Probleme in den Fächern und Fachbereichen auf. Unsere Ergebnisse waren als Diskussionsanreize in den Fachbereichen gut geeignet und haben so manche Veränderung zur Verbesserung der Lehre angeregt.

Nicht zuletzt waren wir auch immer ein wenig stolz darauf, dass andere Evaluationseinrichtungen an deutschen Universitäten sich an unserer Arbeit orientiert haben oder sogar das von uns entwickelte Modell übernommen haben.

Schüler:

Meister, ich habe Hunger, aber mir fehlt ein Messer, um mir vom Brotlaib eine Scheibe abzuschneiden.

Meister III. Grades:

Nimm die Axt!

Schüler:

Aber Meister, das ist doch nicht das richtige Instrument!

Meister III. Grades:

Aber Du wirst satt, mein Sohn!

Heinz Gralki
Institut für Soziologie an der FU Berlin
Garystraße 55
14195 Berlin
Tel.: ++49.30.838.57629
Fax: ++49.30.838.57601
eMail: gralki@web.de

Dr. Heinz Gralki, geb. 1943, Steuerinspektor, Diplomsoziologe, promovierte in Erziehungswissenschaft an der Freien Universität Berlin, war von 1975 bis 1991 in der didaktischen Fortbildung für Dozenten und Lehrende an der Hochschule beschäftigt. Von 1991 bis 2000 führte er im Projekt pro Lehre der FU (PPL) Evaluationen durch. Seit 2000 ist er wissenschaftlicher Mitarbeiter am Institut für Soziologie der FU Berlin. Seine Schwerpunkte in der Lehre sind: Empirische Methoden der Sozialforschung, Entwicklungspolitik, Soziologie des Internets und eLearning