

## Studentische Lehrveranstaltungsevaluation - und wodurch sie verzerrt wird

(Dr. Birgit Hawelka im Gespräch mit Prof. Dr. Bob Uttli)

Birgit Hawelka:

Hallo und herzlich willkommen zur heutigen Folge unseres Podcasts. Ich bin Birgit Hawelka und bei mir ist heute Bob Uttli. Bob ist Professor für Psychologie an der Mount Royal University in Calgary, Kanada. Sein Fachgebiet ist kognitive Psychologie, wobei er sich auf Bereiche wie Gedächtnis, Intelligenz, Altern, Bewertung und Psychometrie konzentriert. Bob hat an mehreren bedeutenden Forschungsarbeiten mitgewirkt, insbesondere an der Untersuchung der Gültigkeit und der Verzerrungen von studentischen Bewertungen der Lehrqualität. Seine Arbeit in diesem Bereich hat die Aufmerksamkeit verschiedener Medien auf sich gezogen, darunter Inside Higher Ed, University Herald, The Guardian und andere. Bob, herzlich willkommen, schön, dass du hier bist.

Bob Uttli:

Vielen Dank für die Einladung.

Birgit Hawelka:

Lass uns direkt beginnen. Bob, in deiner Arbeit Lehrveranstaltungsevaluation durch Studierende hast du die Verwendung standardisierter Instrumente zur Messung der Effektivität der Lehre kritisiert. Kannst du erläutern, welche Argumente und Erkenntnisse dazu geführt haben?

Bob Uttli:

Ja. Das erste Problem, das wir haben, ist, dass es keine anerkannte Definition von effektiver Lehre gibt. Wir wissen also nicht, was wir eigentlich messen wollen. Und viele Leute haben darüber nachgedacht, es gibt viele Experten, die dieses Thema diskutieren. Und im Grunde genommen sind sie sich alle einig, dass wir nur wissen, dass effektive Lehre zu einem höheren Lernerfolg der Studierenden führen sollte. So kam man auf die Idee, diese mehrteiligen Studien durchzuführen, um die Gültigkeit der SETs, d. h. der Bewertungen der Lehre durch die Studierenden, im Verhältnis zum Beitrag der Professorinnen und Professoren zum Lernen der Studierenden zu untersuchen. Im Grunde genommen hat man also eine große Lehrveranstaltung, bei der man die Studierenden in verschiedene Gruppen einteilt. Die

Lehrpersonen werden nach dem Zufallsprinzip zu den Gruppen eingeteilt und unterrichten mit demselben Lehrplan, demselben Buch, denselben Prüfungen usw. Gegen Ende des Kurses werden dann die SETs ermittelt. Und wenn die Professoren den Lernerfolg der Studenten effektiv steigern können, sollte die Gruppe dieser Lehrpersonen besser abschneiden als die anderer Lehrpersonen, so dass es eine Korrelation zwischen den SETs, den durchschnittlichen SETs der Gruppe und der durchschnittlichen Note der Abschlussprüfung geben sollte.

Birgit Hawelka:

Okay. Und ist das der Fall?

Bob Uttl:

Nun, Cohen sagte 1981 in der berühmten Meta-Analyse, dass dies der Fall ist. Er sagte, dass die Korrelation von etwa 0.41 oder 0.43 eine ausreichende Korrelation ist, um Lehrpersonen danach zu beurteilen. Dies erklärt aber natürlich nur etwa 16 % der Varianz. Aber es gibt ein grundlegendes Problem mit dieser Meta-Analyse: Die Verzerrung durch kleine Studien wurde nicht berücksichtigt. Wenn man also die Meta-Analyse erneut durchführt, was wir getan haben, findet man heraus, dass die Korrelation zwischen SET und dem Lernerfolg der Studenten, die dem Professor zuzuschreiben ist, in mehrteiligen Studien gleich Null ist, nachdem man das Vorwissen und die Intelligenz der Studenten kontrolliert hat.

Birgit Hawelka:

Ok, würdest du also sagen, Intelligenz ist der wichtigste Faktor, der die Lernergebnisse beeinflusst?

Bob Uttl:

Ich bin mir nicht sicher, ob es der wichtigste Faktor ist. Sie ist sicherlich ein sehr wichtiger Faktor. Aber es gibt noch einen anderen wichtigen Faktor, und das ist die Motivation der Studierenden. Wie sehr ist ein Studierender daran interessiert, ein bestimmtes Thema in einer bestimmten Lehrveranstaltung zu studieren? In der kognitiven Psychologie hat man festgestellt, dass wir etwas viel schneller lernen, wenn wir daran interessiert sind, als wenn wir es nicht lernen wollen.

Birgit Hawelka:

Wäre es aber möglich, dass Lehrpersonen das Interesse der Studierenden an dem Teilfach, das sie unterrichten, beeinflussen und somit indirekt für die Lernergebnisse verantwortlich sind oder sie beeinflussen können?

Bob Uttl:

Das könnte sein, aber wir wissen aus der Forschung, dass das Interesse der Studierenden an einem Kurs, bevor sie ihn belegen, eine der besten Determinanten dafür ist, wie die Studierenden die Lehre bewerten.

Birgit Hawelka:

Wie wir schon sagten, liegen die Verzerrungen in den Fähigkeiten oder der Motivation der Studierenden. Gibt es noch andere Verzerrungen, die die Ergebnisse der studentischen Bewertung beeinflussen oder verzerren könnten?

Bob Uttl:

Ja. Im Grunde wissen wir, dass die Validität gleich Null ist, oder? Die zweite Möglichkeit, die Validität zu untersuchen, besteht darin, zu prüfen, ob diese SETs von verschiedenen Faktoren beeinflusst werden, die nichts mit der Lehrperson zu tun haben. Man kann zum Beispiel eine Seminargröße haben, man bekommt kleine Klassen zugewiesen, sagen wir bis zu 20 Personen. Du wirst mehr Zeit haben, mit jedem einzelnen Studierenden zu interagieren. Du wirst in der Lage sein, die Namen aller Studierenden zu lernen. Wenn du dagegen eine Veranstaltung mit 300 Studierenden zugeteilt bekommen, hast du Pech gehabt. Wenn du nicht eine besondere Person bist, die sich 300 Namen merken kann, hast du keine Chance, oder? Wenn du einen Kurs mit zehn Teilnehmenden hast, dann kennst du natürlich alle und hast mit allen zu tun, richtig? Die Kursgröße macht also den Unterschied.

Eine weitere große Auswirkung hat - wir sprachen auch über die Fähigkeiten der Schüler und die Motivation, aber auch über das Themengebiet. Wenn du also heute einen quantitativen Kurs unterrichten sollst, dann wünsche ich dir viel Glück. Denn fast niemand in deiner Lehrveranstaltung ist wirklich daran interessiert, diesen Kurs zu belegen. In der Psychologie sind Statistik und Forschungsmethoden Pflichtfächer, weil wir natürlich darauf angewiesen sind. Und wir haben eine Umfrage unter den Studierenden durchgeführt, in der wir sie gefragt haben, wie interessiert sie an den verschiedenen Lehrveranstaltungen sind, die unser Fachbereich anbietet. Und Lehrveranstaltungen, die Statistik oder Forschungsmethoden beinhalten, lagen sechs Standardabweichungen unter dem Mittelwert unserer normalen inhaltlichen Kurse. Und natürlich sind die Leute am meisten daran interessiert, etwas wie abnorme Psychologie, Sozialpsychologie usw. zu belegen. Eine Lehrperson, die abnorme Psychologie unterrichtet, hat es also mit Studierenden zu tun, die im Großen und Ganzen sehr an diesem Thema interessiert sind. Eine Lehrperson, die einführende Statistik unterrichtet, hat es mit Studierenden zu tun, die nicht dabei sein wollen.

Birgit Hawelka:

Das würde auf Pflichtkurse zutreffen, denke ich. Sie sollten schlechter bewertet werden als Kurse, die die Studierenden selbst und nach ihren Interessen wählen können, oder?

Bob Uttl:

Das ist richtig, ja. Es gibt einige Untersuchungen, die sich damit befassen. Natürlich werden Wahlfächer, die man belegen möchte, in der Regel besser bewertet als Pflichtfächer.

Birgit Hawelka:

Okay. Wir haben also einige Studierendenmerkmale, wir haben einige Kursmerkmale, die die Bewertungen der Studierenden beeinflussen könnten. In einem deiner Texte erwähnst du eine Studie, die darauf hindeutet, dass Anreize wie das Verteilen von Schokolade die

Bewertung der Studierenden beeinflussen können. Wie würdest du diese seltsamen Ergebnisse erklären? Und welche psychologischen Faktoren könnten in diesem speziellen Fall die Meinung der Studierenden beeinflussen.

Bob Uttl:

Ich glaube nicht, dass das ein sehr merkwürdiges Ergebnis ist. Wir haben zwei randomisierte Studien, die zeigen, dass Schokolade und Schokoladenkekse die Bewertungen der Lehrpersonen durch die Studierenden um etwa 0.3 bis 0.5 Standardabweichungen verbessern.

Birgit Hawelka:

Das ist eine ganze Menge.

Bob Uttl:

Das ist eine Menge, ja. Das macht den Unterschied, oder? Und es gibt eine einfache Erklärung. Wenn du etwas mitbringst, was die Studierenden mögen, werden sie dir höhere Bewertungen geben. Wenn man sagt, dass SETs im Grunde genommen ein Maß für die Zufriedenheit der Studierenden sind, und die Definition im Wörterbuch lautet: glückliche, zufriedene Gefühle aufgrund von etwas, das einem widerfahren ist, richtig? Alles, was die Zufriedenheit der Studierenden steigert, führt also wahrscheinlich zu einer besseren Bewertung der Lehre durch die Studierenden. Geringe Arbeitsbelastung, gute Noten, Schokolade, selbstgebackene Kekse, alles, was man tun kann, um die Zufriedenheit der Studierenden zu erhöhen.

Wir haben nicht darüber gesprochen - aber eine Frage ist natürlich: Wie werden Evaluationen eingesetzt? In einigen oder vielen Fällen entscheidet ein Komitee oder vielleicht eine Institutsleitung, dass alle, die über dem Durchschnitt liegen, zufriedenstellend sind und alle, die unter dem Durchschnitt liegen, nicht zufriedenstellend sind. So entsteht ein Wettlauf, bei dem man in der oberen Hälfte der Professorenschaft sein muss, um zufriedenstellend zu sein, ohne dass man mit seinem Lehrstuhl oder Dekan über die schlechte Bewertung der Lehre durch die Studierenden sprechen muss.

In einem kommerziellen System wird sogar empfohlen, dass nur die besten 30 % der Lehrpersonen zufriedenstellend sind. 70 % sind also per Definition nicht zufriedenstellend, egal was sie tun. Du magst ein Olympionike sein, aber du bist trotzdem nicht zufriedenstellend, weil du zu den unteren 70 % gehörst. Das führt zu einem Wettlauf mit sehr, sehr hohem Einsatz, bei dem alles möglich ist: Also Kekse, die Standards senken, Noten umsonst geben. Wir haben einen massiven Anstieg der Noteninflation. Heute ist die Durchschnittsnote im Grunde eine Eins, was vor 40 Jahren noch undenkbar gewesen wäre, oder? All diese Faktoren beeinflussen also die SETs. Die Professoren lernen, was die SETs beeinflusst. Bevor wir also die Lehrveranstaltungsevaluation online hatten, konnte man Professorinnen und Professoren sehen, die mit Tüten voller Schokolade oder Schokoriegeln herumliefen, Evaluationsbögen in Papierform verteilten und Schokoriegel dazu gaben, um die Studierenden glücklich zu machen.

Birgit Hawelka:

Wenn ich für den Moment zusammenfasse, kann man sagen, dass es keine objektive Methode gibt, um die Lehrqualität zu messen. Aber die Lehrveranstaltungsevaluationen messen dann vielleicht, wie gut es den Lehrenden gelingt, die Erwartungen der Studierenden an eine gute Lehrqualität zu erfüllen?

Bob Uttl:

Ich würde nicht sagen, dass sie die Erwartungen der Studierenden an die Qualität der Lehre messen. Ich würde sagen, dass sie die Zufriedenheit der Studierenden messen. Und die Zufriedenheit der Studierenden ist sehr umfassend, nicht wahr? Wenn ich also eine gute Note haben möchte und du mich zufrieden stellst, dann bin ich glücklicher, als wenn du mir eine Vier gibst.

Birgit Hawelka:

Kennen die Studierenden ihre Noten in Kanada oder an deiner Universität schon vor der Lehrveranstaltungsevaluation?

Bob Uttl:

Nein, das tun sie nicht. Aber im Grunde genommen doch, denn man hat ja normalerweise mindestens eine Zwischenprüfung. Und die Lehrveranstaltungsevaluation finden natürlich nach der Zwischenprüfung statt, damit man eine Vorstellung davon bekommt. Und wenn du zum Beispiel eine Lehrperson bist, die die Studierenden durch mehrere Aufgaben oder vielleicht mehrere Tests bewertet, dann wissen die Studierenden im Grunde genommen schon, wie ihre Note lautet, wenn sie zur Evaluation kommen.

Birgit Hawelka:

Sowohl die Amerikanische als auch die Deutsche Gesellschaft für Soziologie schlagen vor, Lehrveranstaltungsevaluation nicht zur summativen Bewertung, sondern zur formativen Bewertung zu verwenden. Sie empfehlen sie als eine Möglichkeit, das Feedback der Studierenden zu sammeln, ihre Sichtweise zu verschiedenen Aspekten eines Kurses. Würdest du einem solchen Szenario zustimmen?

Bob Uttl:

Ja, ich denke, es ist nichts Falsches daran, wenn ein Professor seine eigenen Studierenden befragt, um herauszufinden, wie sie verschiedene Aspekte einer Lehrveranstaltung wahrnehmen. Das scheint nicht nur vernünftig, sondern auch wünschenswert zu sein. Aber es muss eine Umfrage sein, die der Professor unter den Studierenden seines Kurses durchführt und die niemand sonst sieht. Denn dann weiß der Professor, der den Kurs tatsächlich unterrichtet, was los war, welche Fragen er oder sie tatsächlich stellen will, die irgendwie wichtig sind. Vielleicht unterrichten sie den Kurs zum ersten Mal. Sie werden nicht wissen, womit die Studierenden die meisten Schwierigkeiten hatten, also können sie sie tatsächlich fragen. Vielleicht verwenden sie eine andere Software und erhalten ein Feedback zu dieser Software. Das

sind also die wichtigen Verwendungszwecke von Studierendenfeedback, aber nicht für die Bewertung der Effektivität der Lehre.

Birgit Hawelka:

Wie wir erfahren haben, und um das Wesentliche deiner Ergebnisse zusammenzufassen, scheinst du zu dem Schluss gekommen zu sein, dass es aus verschiedenen Gründen kaum möglich ist, die Effektivität einer Lehrperson zu messen. Aber drehen wir das Ganze um: Lassen sich Verhaltensweisen von Lehrenden erkennen und messen, die eindeutig auf Ineffizienz hindeuten? Gibt es also Anzeichen, die man verwenden könnte, um zu sagen: Oh, das ist ein kritischer Punkt, das könnte ziemlich ineffektiv in der Lehre sein?

Bob Uttl:

Ja. Eine Reihe von Leuten weist darauf hin, dass, wenn wir versuchen, darüber nachzudenken, was effektive Lehre ist, die Leute sich nicht einig sind. Das heißt, sie können sich einfach nicht einigen. Weißt du, um dir ein Beispiel zu geben: Jemand könnte sagen, nun, es ist ein effektiver Unterricht, wenn man die Studierenden während der Vorlesung auffordert, eine Frage zu beantworten. Andere Professoren werden sagen: Nein, das erzeugt Ängste, das setzt die Studierenden in Zugzwang. Es ist erniedrigend. Es ist negativ. Stimmt's? Aber fast jeder würde zustimmen, dass es keine effektive Lehre ist, wenn man nicht in der Veranstaltung erscheint. In der Veranstaltung aufzutauchen und ein Kapitel aus dem Lehrbuch wortwörtlich vorzulesen, ist ebenfalls kein effektiver Unterricht. Es gibt also eine Reihe von Dingen, denen viele Leute zustimmen würden, oder? Selbst wenn du also sagst, du gibst den Studierenden Feedback. Studierenden zwei Monate später ein Feedback zu geben, ist nicht sehr effektiv. Ich habe in meinem letzten Beitrag eine solche Umfrage vorgeschlagen, in der man die Studierenden fragt, ob bestimmte Dinge, objektiv beobachtbare Dinge, tatsächlich in der Lehrveranstaltung passiert sind.

Sagen wir, ich habe innerhalb von drei Werktagen eine Rückmeldung oder Antwort auf meine E-Mail erhalten. Wenn nun 50 % der Studierenden sagen, der Professor sei nicht zum Unterricht erschienen, und die anderen 50 % sagen, er sei erschienen, dann sagt jemand sozusagen nicht die Wahrheit. Und jetzt kannst du nachprüfen, was tatsächlich passiert ist. Nutze dies als Alarmzeichen für ineffektive Lehre. Und in der Tat, wenn du ein oder zwei Jahrzehnte an der Universität bist, wirst du feststellen, dass diese ineffektiven Lehrveranstaltungen wirklich in diese Kategorien fallen. Du weißt schon, Leute, die nicht zur Lehrveranstaltung erscheinen, die eine Veranstaltung auf halbem Weg abbrechen oder so etwas in der Art, richtig? Sie geben kein Feedback, die Studierenden bekommen nie eine Note, vielleicht die letzte, aber das war's. Das sind also ineffektive Verhaltensweisen im Unterricht.

Birgit Hawelka:

Okay. Und nach deinen Erkenntnissen könnte es hilfreich sein, solche Punkte oder Signale zu identifizieren. In deinem letzten Text fasst du einige von ihnen als Indikator für ineffektive Lehre zusammen. Das könnte die Grundlage für die Dekane sein, um darüber zu sprechen, was in dieser Lehrveranstaltung passiert, oder um ineffektive Dozierende zu identifizieren, oder?

Bob Uttl:

Das ist richtig. Es wäre einfach nur ein Alarmzeichen, richtig? Es sollte nicht als Bewertung verwendet werden. Sondern einfach: Okay, die Studierenden sind in der Lehrveranstaltung. Wir müssen auch die Tatsache in Betracht ziehen, dass, wenn wir sagen, dass Studierende in einer Lehrveranstaltung sind, einige es sind und einige nicht. Es ist also wichtig, sich das vor Augen zu halten, oder? Aber wenn man dann einige dieser objektiv beobachtbaren Verhaltensweisen hat und 90 % der Studierenden sagen, ja, er ist die meiste Zeit nicht aufgetaucht. Dann sollte man sich ansehen, was passiert ist, und dann feststellen, was tatsächlich passiert ist. Und dann hast du eine Grundlage für eine Bewertung oder eine Korrekturmaßnahme oder etwas Ähnliches.

Birgit Hawelka:

Okay, also anstatt zu versuchen, die Effektivität des Unterrichts zu messen, würdest du vorschlagen, ineffektive Lehre zu identifizieren. Um es also umzudrehen.

Bob Uttl:

Richtig, ja. Man sucht also nach ineffektiver Lehre und danach, was man dagegen tun kann. Es gibt noch eine andere Methode, die ich in dieser Veröffentlichung besprochen habe und die an einer Universität in Alberta angewandt wurde. Bei der sie im Grunde eine statistische Definition eines Ausreißers verwenden. Also solange man sich innerhalb des großen Felds der Professorenschaft befindet und kein Ausreißer ist, man nicht der Nachzügler am Ende ist, ist alles in Ordnung. Aber wenn man ein Ausreißer ist, dann löst das ein Alarmzeichen. Es bedeutet nicht, dass man unfähig ist, aber es ist ein Alarmzeichen. Jemand sollte sich ansehen, was mit dieser Lehrveranstaltung los ist. Und vielleicht wurde sie zum ersten Mal unterrichtet. Vielleicht hattest du einen wirklich schlechten Kurs. Vielleicht hat man dich in einen Raum mit 20 Stühlen gesetzt und du hattest 30 Studierende, also mussten zehn von ihnen stehen. Auch so etwas kommt vor, oder?

Birgit Hawelka:

Okay, Bob, vielen Dank, dass du dir die Zeit genommen hast, hier bei uns zu sein, und dass du uns deine wertvollen Einblicke in die Bewertung von Studierenden gegeben hast.

Bob Uttl:

Vielen Dank für die Einladung.