



Wie gut ist meine Lehre (wirklich)? Studentisches Feedback konstruktiv nutzen

B. Schlegel^{1*}, B. Watzka²

¹ Professur für Magnetofluidynamik, Mess- und Automatisierungstechnik, Institut für Mechatronischen Maschinenbau, Fakultät Maschinenwesen, TU Dresden

² Jun. Prof.‘in für Didaktik der Physik, Institut für Physik, Fakultät für Naturwissenschaften, Otto-von-Guericke-Universität Magdeburg

Abstract

Studentisches Feedback ermöglicht einen differenzierten Blick auf die (wahrgenommene) Nützlichkeit von Lehr-Lern-Angeboten, Methoden und Tools bezüglich des Lernfortschrittes bei den Studierenden. Neben fachdidaktischen Prinzipien und Lehr-Methodik lässt sich auf Basis studentischen Feedbacks Lehre bedarfsgerecht gestalten. Evaluation zur Weiterentwicklung der Lehre und Evaluationsforschung stehen neben anderen Arten studentischer Rückmeldung als Werkzeuge zur Verfügung. Beide Varianten von Evaluation dienen nicht der Analyse und Kontrolle der Qualität von Lehre, sondern können als Instrumente von Lehrenden eingesetzt werden, die ihre Lehre weiterentwickeln wollen. Bevor eine entsprechende Erhebung geplant und durchgeführt werden kann, müssen folgende Fragen ausführlich diskutiert werden: Welche konkrete Fragestellung soll mit Hilfe der zu erhebenden Daten beantwortet werden? Welche Einflussfaktoren spielen im Kontext der Fragestellung eine Rolle? Welche Ergebnisse sind denkbar und welche Konsequenzen hätten diese Ergebnisse für die Gestaltung von Lehre? Sind diese Konsequenzen relevant? Aus den Antworten auf diese Fragen lässt sich ableiten, welche Form der Evaluation gewählt wird und wie diese konkret gestaltet werden kann.

Student feedback enables a differentiated view of the (perceived) usefulness of teaching and learning offers, methods and tools with regard to students' learning progress. In addition to didactic principles and teaching methodology, student feedback can be used to tailor teaching to students' needs. Evaluation for the further development of teaching and evaluation research are available as tools alongside other types of student feedback. Both types of evaluation are not used to analyze and control the quality of teaching, but can be used as instruments by teachers who want to further develop their teaching. Before a corresponding survey can be planned and carried out, the following questions must be discussed in detail: What specific question is to be answered with the help of the data to be collected? Which influencing factors play a role in the context of the question? What results are conceivable and what consequences would these results have for the design of teaching? Are these consequences relevant? The answers to these questions can be used to determine which form of evaluation is chosen and how it can be specifically designed.

*Corresponding author: beatrice.schlegel@tu-dresden.de

1. Einleitung

Die Neuerungen in der Lehre – getrieben bzw. potenziert durch die veränderten Bedingungen der Pandemiezeit – sind in den letzten Semestern vielfältig reflektiert, diskutiert, angepasst oder auch verworfen worden. Dass digitale Elemente unsere Lehre bereichern, ist mittlerweile Konsens, auch, dass Lehrveranstaltungen (im Folgenden mit LV abgekürzt) in Präsenz zu guten Studienbedingungen gehören. Welche Elemente in welcher Form zum Einsatz kommen und wirklich tauglich sind, um den Lernprozess der Studierenden voranzubringen, kann man nicht pauschal beantworten und ist in jedem Falle fach- oder sogar modulspezifisch. Um dazu einen differenzierten Einblick zu bekommen, können verschiedene Arten studentischer Rückmeldung eingeholt werden (klassisches Feedback, Test, Evaluation usw.) In Kombination mit der Expertise der Lehrenden können sie entscheidend zu einer Verbesserung der Lehre führen.

In diesem Beitrag geht es um Evaluation und Evaluationsforschung als Mittel, um Lehre besser zu gestalten. Das Forschungsfeld wird aufgezeigt, beginnend mit der reinen Evaluation bis zur empirischen Forschung, die im Vergleich mehr Expertise voraussetzt. Im Abschnitt zwei werden die Funktionen unterschiedlicher Evaluationsansätze beispielhaft erläutert, um dann in den beiden darauffolgenden Abschnitten zum einen auf die Evaluation zur Weiterentwicklung von Lehre und zum anderen auf die Evaluationsforschung genauer einzugehen. Hier wird konkret Bezug genommen auf den gleichnamigen Workshop im Rahmen der Lessons Learned Konferenz 2023. Unsere Vision ist es, dass Lehrende aus unterschiedlichen Fachbereichen gemeinsam Evaluationsbausteine entwickeln und in der eigenen Lehre einsetzen, um Lehre nachhaltig zu verbessern und gegebenenfalls Evaluationsprozesse weiter zu entwickeln.

2. Funktionen von Evaluation

Lehrveranstaltungsevaluation, wie sie beispielsweise an der TU Dresden vom Zentrum für Qualitätsanalyse (ZQA) durchgeführt wird, dient in erster Linie der Qualitätskontrolle von Lehre an einer Universität oder Hochschule,

und zwar für alle fachlichen Bereiche. Die Evaluation, die vom ZQA genutzt wird, basiert auf dem Heidelberger Inventar zur Lehrveranstaltungsqualität [1]. Diese Art der Evaluation ermöglicht eine Erfassung des Ist-Standes und einen Vergleich mit einem Qualitätsstandard, der die Erwartungen an Lehre widerspiegelt. Der Evaluationsbogen [2] beinhaltet Aussagen zur Lehrperson, die auf einer fünf-stufigen Skala von „trifft völlig zu“ bis „trifft überhaupt nicht zu“ von den Studierenden bewertet werden. Beispiele für Aussagen sind, die Lehrperson „...vermittelt die Lehrinhalte verständlich“ oder „...steht bei Bedarf für Rücksprachen zur Verfügung.“ Außerdem werden Einschätzungen zur Lehrveranstaltung wie „Die Vorlesung hat mein Wissen erweitert“ abgefragt sowie Einschätzungen zu den Anforderungen in der LV, Arbeitsaufwand, studentisches Engagement und weitere. Die Ergebnisse dieser Evaluation geben der Lehrperson einen Eindruck der studentischen Sichtweise und erste Anhaltspunkte, in welchem Bereich Verbesserungen möglich sind. Sie lassen sich allerdings nicht dahingehend interpretieren, welche Verbesserung in welcher Art und Weise vorgenommen werden sollten. Dies ist nicht die Zielsetzung dieser Art von Evaluation.

Die in diesem Beitrag vorgestellte Art der Evaluation hat zum Ziel, Anhaltspunkte zur konkreten Verbesserung von Lehre zu finden. Sie ermöglichen im Umkehrschluss wenige Rückschlüsse auf die Qualität der Lehre.

Ein erstes vergleichendes Beispiel soll diesen Unterschied illustrieren. In der Lehrveranstaltungsevaluation des ZQA werden die Studierenden unter der Überschrift „Einsatz digitaler Lehrformate“ gefragt, wie sie die in der LV eingesetzten digitalen Formate bewerten (Seite 3 [2]). Die zu bewertenden Formate sind Livestream, Videoaufnahmen, Powerpoint-Präsentationen, Diskussionsforen und andere Formate. In der Evaluation zur Weiterentwicklung der Lehre wird nach der bevorzugten Variante für die LV gefragt mit der Option, eine der folgenden Antworten auszuwählen:

- ausschließlich Vorlesung in Präsenz,
- ausschließlich Videos,
- ausschließlich Live-Übertragung über Youtube,

- Hybrid-Variante I: Vorlesung in Präsenz, Online Teilnahme,
- Hybrid-Variante II: Vorlesung in Präsenz, Videos,
- Hybrid-Variante III: Vorlesung in Präsenz, Online-Teilnahme, Videos,
- Sonstiges (mit der Möglichkeit etwas einzugeben).

Aus den Ergebnissen der Befragung des ZQA lassen sich Aussagen zur Qualität der digitalen Formate ableiten. Die Ergebnisse der Evaluation zur Weiterentwicklung der Lehre liefern einen Hinweis zur Entscheidung, welches Format für die Studierenden passender ist.

Eine wesentliche Limitation von Evaluationen besteht darin, kausale Zusammenhänge herzustellen. Obwohl Evaluationen korrelative Beziehungen identifizieren können, erlauben sie keine verlässlichen Aussagen über Ursache und Wirkung. Als Beispiel: Eine Evaluation könnte zeigen, dass Studierende, die Vorlesungen bevorzugt im Videoformat verfolgen, ihren Wissenserwerb als höher einschätzen. Dies legt einen Zusammenhang zwischen dem Vorlesungsformat „Video“ und der subjektiven Wahrnehmung des Wissenserwerbs nahe, doch die eigentliche Ursache dieses Phänomens bleibt unklar. Um verlässliche Ursache-Wirkungs-Zusammenhänge zu ermitteln, sind spezifische Forschungsdesigns mit Kontrollgruppen erforderlich.

Es sei an dieser Stelle darauf hingewiesen, dass die Entwicklung von Lehre niemals rein auf Basis von Evaluation stattfinden kann. Die Expertise der Lehrperson ist dafür von zentraler Bedeutung. Sie speist sich zum einen aus dem fachlichen Tiefenverständnis und ihren Forschungserfahrungen, zum anderen aus ihrem didaktischen und pädagogischen Knowhow. Erkenntnisse aus (Fach)Didaktik und Pädagogik geben wichtige Anhaltspunkte zur Neuausrichtung bzw. Entwicklung von Lehre. Beispielsweise gibt die Erkenntnis, dass – nach der Selbstbestimmungstheorie nach Deci und Ryan – die Motivation davon abhängig ist, inwiefern die drei psychologischen Grundbedürfnisse nach Kompetenzerleben,

Autonomie und sozialer Eingebundenheit gedeckt sind [3]. Um folglich Studierende zu motivieren, müssen Lehr-Lern-Situationen geschaffen werden, die diese Bedürfnisse decken. Studierende haben allerdings Bedürfnisse in unterschiedlichen Ausprägungen. Dies kann fachabhängig sein und auch nach den zu vermittelnden Inhalten variieren. Um herauszufinden, inwiefern die Gestaltung der Lehr-Lern-Situationen tatsächlich den Studierenden entgegenkommt, sollte evaluiert werden.

3. Lehrveranstaltungsevaluation zur Weiterentwicklung von Lehre

Anhand von zwei Beispielen werden im Folgenden die Möglichkeiten und Grenzen der Evaluation zur Weiterentwicklung von Lehre aufgezeigt.¹

Möglichkeiten und Grenzen

Beispiel 1 – die Zwischenabfrage

Eine Zwischenabfrage in der LV kann helfen, Wissen zu wiederholen und zu festigen. Denkbare Zeiträume für Zwischenabfragen sind der Anfang, die Mitte oder das Ende der LV. Welche Variante von den Studierenden bevorzugt wird, kann durch eine einfache Abfrage im Rahmen einer Evaluation ermittelt werden, wie in der Evaluation des Moduls Mess- und Automatisierungstechnik (im Folgenden abgekürzt mit MAT). Das Ergebnis für das Wintersemester 2022 ist in der Abbildung 1 zu sehen.



Abbildung 1: Evaluationsbeispiel Zwischenabfrage – Frage: Welche Variante bevorzugen Sie? n=122

¹ Die Befragung der Studierenden findet online innerhalb der LV statt.

93 % der befragten Studierenden gaben an, dass sie die Zwischenabfrage in der Mitte der Vorlesung bevorzugen.

Was lässt sich aus dem gewonnenen Ergebnis ableiten?

Daraus lässt sich ein Gestaltungshinweis ableiten, nämlich die Zwischenabfragen in den nächsten Lehrveranstaltungen des Moduls in der Mitte einzubauen. Nicht ableiten lässt sich, warum die Studierenden diese Variante bevorzugen. Möglicherweise deshalb, weil der Prozess des Zuhörens und Mitschreibens unterbrochen wird und sich diese „Pause“ positiv auf die Konzentration auswirkt. Dies ist nur eine Vermutung und wird durch die Evaluation nicht bestätigt oder widerlegt. Ebenfalls lässt sich an den Ergebnissen nicht ablesen, an welcher Stelle generell in LV Zwischenabfragen eingebaut werden sollten oder welche Funktion sie im Lernprozess einnehmen.

Beispiel 2 – Verstehen der Studierenden

Ziel von Lehre ist (unter anderem), dass Studierende fachliche Zusammenhänge verstehen.

Den Prozess des Verstehens zu unterstützen und Lehr-Lern-Angebote zu schaffen, in denen Erkenntnisse möglich sind, ist Aufgabe der Lehrperson. Um das Lehrangebot zu verbessern, kann es nützlich sein herauszufinden, in welcher Lehr-Lern-Situation am meisten verstanden wird. Die Studierenden wurden in der Evaluation des Moduls MAT gefragt, inwiefern sie den folgenden fünf Aussagen auf einer vierstufigen Skala von eins „stimme nicht zu“ bis vier „stimme voll zu“ zustimmen:

- (1) Den Erklärungen der Lehrperson konnte ich ohne Probleme folgen.
- (2) Durch das, was aufgeschrieben wurde, konnte ich alles verstehen.
- (3) Erst beim Durcharbeiten meiner Mitschrift konnte ich Zusammenhänge verstehen.
- (4) Erst durch die Übungsaufgaben habe ich die Zusammenhänge wirklich verstanden.
- (5) Erst durch den Besuch der Übungen habe ich die Zusammenhänge wirklich verstanden.

Die Ergebnisse der Befragung sind in der Abbildung 2 dargestellt.

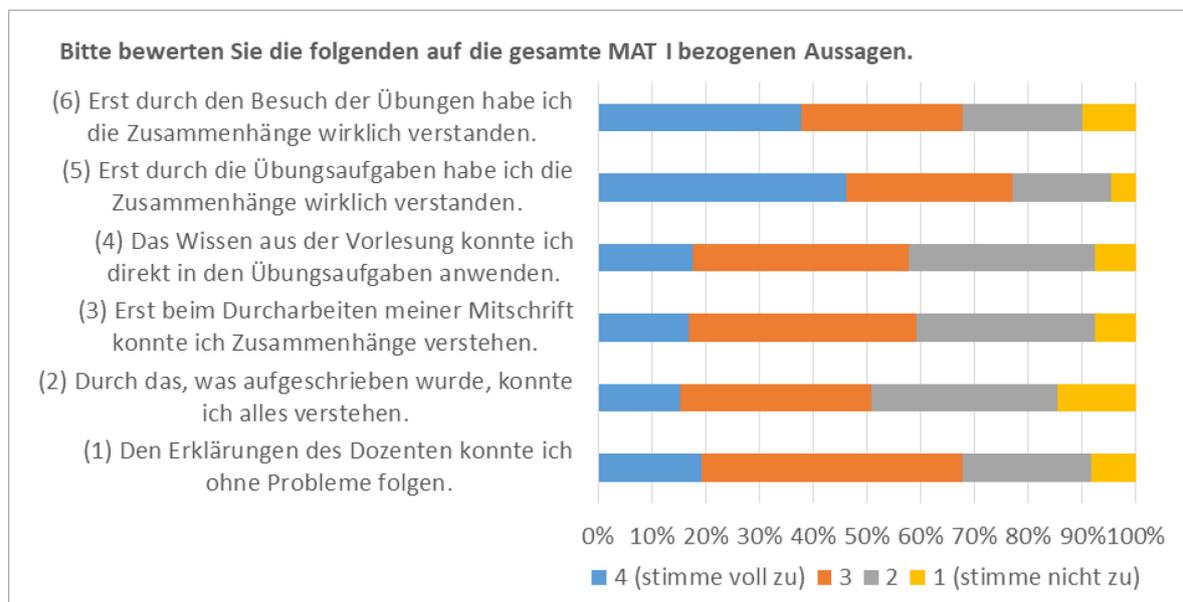


Abbildung 2: Ausschnitt aus den Ergebnissen der Evaluation des Moduls MAT im Wintersemester 2023

Die ersten drei Aussagen zu den Erklärungen der Lehrperson und zu den Mitschriften wurden von weniger als 20 % mit „stimme voll zu“ bewertet. Im Vergleich dazu stimmten doppelt so viele den Aussagen zu den Übungen und Übungsaufgaben voll zu (Aussagen (4) und (5)).

Fasst man die Ergebnisse von den Bewertungen vier (stimme voll zu) und drei zusammen, gleichen sich die Unterschiede teilweise aus. Beispielsweise erreichen die beiden Aussagen zu den Erklärungen (1) und zum Besuch der Übung (5) beide knapp 70 % Zustimmung. Die

geringste Zustimmung (16 % stimme voll zu, vier und drei in Summe: 51 %) erhielt die Aussage „Durch das, was aufgeschrieben wurde, konnte ich alles verstehen“. Die höchste Zustimmung erhielt die Aussage „Erst durch die Übungsaufgaben habe ich die Zusammenhänge wirklich verstanden“ (46 % stimme voll zu, vier und drei in Summe: 77 %).

Die Studierenden, die der Aussage „Den Erklärungen der Lehrperson konnte ich ohne Probleme folgen“ nicht oder wenig zustimmten (eins oder zwei), wurden zusätzlich explizit gefragt, warum sie den Erklärungen wenig oder nicht folgen konnten. Unter anderem wurden folgende Antworten gegeben:

- Komm nicht mit dem Schreiben hinterher,
- Weil ich nur schreiben oder der Vorlesung folgen kann,
- zu schnell,
- Für die Menge des Stoffes war zu wenig Zeit in der Vorlesung eingeplant.

Was lässt sich aus den gewonnenen Ergebnissen ableiten?

Die befragten Studierenden schätzen für das Verstehen von Zusammenhängen die Übungsaufgaben als wirkungsvoller im Vergleich zur Vorlesung ein. Das, was aufgeschrieben wurde, unterstützt den Verstehensprozess deutlich weniger als die anderen Lehr-Lern-Angebote. Es könnte folglich hilfreich sein, mehr Übungsaufgaben einzubinden und Übungsbeispiele in die LV zu integrieren, um das Aufgeschriebene zu illustrieren bzw. anzuwenden. Offenbar ist die Stoffmenge für einige Studierende zu groß, um in der LV parallel mitschreiben und mitdenken zu können. Es könnte hilfreich sein, ein Teil der zu vermittelnden Inhalte in einem Skript anzubieten, so dass nicht alle Inhalte abgeschrieben werden müssen. Es ist außerdem zu überlegen, ob die Stofffülle reduziert werden kann.

Grenzen: Kausale Zusammenhänge zwischen der Gestaltung von Lehr-Lern-Situationen und der Unterstützung studentischer Lernprozesse werden mit dieser Evaluation nicht untersucht. Erkenntnisse, die aufgrund der vertieften Nachfrage von Teil-Gruppen gewonnen wurden, sind nur für die Studierenden-Gruppe gül-

tig. Sie lassen sich nicht verallgemeinern. Kausale Ableitungen und die Verallgemeinerung von Erkenntnissen erlaubt nur die Evaluationsforschung, um die es im Kapitel 4 gehen wird.

Entwicklung eines Evaluationsbausteins

Im letzten Abschnitt wurde exemplarisch gezeigt, welche Potentiale Evaluation für die Entwicklung von Lehre bietet, und welche Grenzen bestehen. Ergebnisse von Evaluation bieten Anhaltspunkte für Entscheidungen und zeigen Stellschrauben auf, an denen Änderungen vorgenommen werden können. Dabei führen globalere Fragen zu Aussagen, die sich auf mehr als eine Lehr-Lern-Situationen beziehen. Konkretere Fragen liefern Aussagen zu einzelnen Aspekten. Um passgenaue Fragen zu stellen, die zu nutzbaren Ergebnissen führen, braucht es einen entsprechenden Entwicklungsprozess, der durch Diskussion folgender Fragen gekennzeichnet ist:

- Um welche Lehr-Lern-Situation, Methode oder um welchen Inhalt geht es? Global oder konkret?
- Was möchte die Lehrperson von den Studierenden wissen?
- Welche Evaluationsergebnisse sind denkbar?
- Was sind die Konsequenzen jedes einzelnen denkbaren Ergebnisses?
- Sind diese Konsequenzen relevant / interessant und lassen sie sich umsetzen?

Sind diese Fragen geklärt, führen folgende Schritte zu einem Evaluationsbaustein:

Schritt 1: Recherche nach entsprechenden Bausteinen in bestehenden Inventaren,

Schritt 2: Anpassung von Bausteinen bzw. Neu-Konstruktion,

Schritt 3: Prüfen, ob Fragestellungen mit dem Baustein tatsächlich beantwortet werden können,

Schritt 4: Test und ggf. Anpassung.

Im Rahmen eines Workshops der Lessons Learned Konferenz 2023 wurde mit einer Gruppe von vier Lehrenden ein Evaluationsbaustein entwickelt, der in Zukunft in ihrer Lehre direkt eingesetzt werden soll. Ziel ist es,

die Ergebnisse der Evaluation gemeinsam auszuwerten, zu interpretieren und zu diskutieren. Durch diese Art der Zusammenarbeit kann Evaluation und Lehrentwicklung konstruktiver erfolgen, da verschiedene Studierenden-Gruppen und unterschiedliche Perspektiven der Lehrenden verglichen werden können. Im Folgenden werden die im Team diskutierten Fragen und der entwickelte Evaluationsbaustein vorgestellt. **Lehrende sind eingeladen, den Baustein in die eigene Evaluation einzubauen.**

Das Team hat entschieden, dass eine Methode zur Festigung und mit hoher Aktivität der Studierenden in den Fokus genommen wird, die in einer LV bis zu zehn Minuten in Anspruch nehmen kann, wie beispielsweise eine Zwischenabfrage. Die Frage, die mit Hilfe der Evaluation beantwortet werden soll, ist: Wie wird

die Methode von den Studierenden angenommen? Unterstützt sie den Verstehensprozess?

Es handelt sich folglich um eine konkrete Fragestellung. Denkbare Ergebnisse sind: (a) die Methode wird gut angenommen, oder (b) die Methode wird nicht gut angenommen, oder (c) die Methode wird teilweise gut angenommen. Aus den einzelnen Ergebnissen wurden folgende Konsequenzen abgeleitet: für das Ergebnis (a) die Methode wird verstärkt eingesetzt, für (b) es wird nachgefragt, woran es liegt und ggf. angepasst, für (c) die Methode wird in passenden LV eingesetzt, aber nicht verstärkt. Diese Konsequenzen wurden von den Lehrenden als relevant und umsetzbar bewertet.

Die Abbildung 3 zeigt den Evaluationsbaustein im Umfragetool Limesurvey, das vom Bildungsportal Sachsen für die TU Dresden bereitgestellt wird [4].

Bewerten Sie den Einsatz der Zwischenabfragen auf einer Skala von 1 bis 4, wobei 1 bedeutet "ich stimme nicht zu" und 4 bedeutet "ich stimme voll und ganz zu".

	1 (stimme nicht zu)	2	3	4 (stimme voll und ganz zu)	Keine Antwort
Die Zwischenabfragen haben mir gefallen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Sie bringen mir nichts.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Sie waren für meinen Verstehensprozess hilfreich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Wie müssten die Zwischenabfragen aus Deiner Sicht gestaltet sein? Wähle eine oder mehrere Antworten aus oder formuliere eine eigene unter Sonstiges. Du kannst die Freitextfelder nutzen, um eine Antwort zu konkretisieren.

! Kommentieren wenn eine Antwort gewählt wird

So, wie sie gestaltet wurden.

Sie müssten länger bzw. kürzer sein.

Es müssten andere Fragen sein, nämlich...

Ich finde andere Methoden sinnvoller, nämlich...

Sonstiges:

Abbildung 3: Evaluationsbaustein Zwischenabfrage

Der Baustein kann Teil einer Lehrveranstaltungsevaluation im Rahmen der Qualitätskontrolle sein oder auch Teil einer umfassenderen Evaluation zur Weiterentwicklung von Lehre. Es ist denkbar, den Baustein als Kurzevaluation in Kombination mit allgemeinen Informationen wie Vertiefungsfach, Semester, ggf. auch Geschlecht und Herkunft zu nutzen, falls entsprechende Vergleiche zwischen Studierenden-Gruppen interessant und hilfreich sind.

4. Evaluationsforschung

Evaluationsforschung bedient sich sowohl quantitativen als auch qualitativen Forschungsmethoden. Dabei kann sie formative Ziele verfolgen, um laufende Prozesse zu verbessern, oder summative Ziele, um abschließende Bewertungen vorzunehmen. Doch Evaluationsforschung geht über einfaches Feedback hinaus. Sie sucht in kontrollierten Umgebungen nach den Ursachen und Mechanismen, die hinter den beobachteten Ergebnissen stehen.

Insbesondere bei der summativen Evaluation stehen zwei Hauptziele im Fokus. Einerseits kann untersucht werden, wie eine Treatmentbedingung (unabhängige Variable: z.B. das Lehrformat) eine Beobachtung / Messung (abhängige Variable: z.B. die Lernleistung) beeinflusst. Andererseits geht es bei summativen Evaluationen darum, Wirkungsanalysen durchzuführen oder Prognosen zu erstellen. Hier könnten beispielsweise Fragen aufkommen wie: "Wie stark beeinflusst eine Erhöhung der wöchentlichen Quizaufgaben um eine Aufgabe die Lernleistung?" oder "Was lässt sich über die Lernleistung vorhersagen, basierend auf der Anzahl der Quizaufgaben?"

Königsweg der Forschung

In Experimenten werden spezifische Bedingungen kontrolliert und Variablen gezielt manipuliert, um kausale Beziehungen zu erforschen. Um sicherzustellen, dass andere unerwartete Variablen nicht für beobachtete Effekte verantwortlich sind, wird häufig ein Prä-Post-Kontrollgruppendesign mit randomisierter Zuordnung der Teilnehmenden verwendet. Bei einem Prä-Post-Kontrollgruppendesign wird die abhängige Variable vor der Intervention

(Prä-Test) und nach der Intervention (Post-test) gemessen, wobei zusätzlich noch ein Abgleich mit einer Kontrollgruppe stattfindet, die keine Intervention erhält. Die Teilnehmenden werden hierbei rein zufällig der Gruppe mit bzw. ohne Intervention zugewiesen. Durch dieses Design lässt sich verhindern, dass nicht nur die unabhängige Variable 'Lehrformat', sondern auch andere Faktoren wie Vorwissen, die abhängige Variable 'Lernleistung' beeinflussen.

Der ideale Forschungsprozess beginnt mit aus der Theorie abgeleiteten Fragen und Hypothesen. Diese bestimmen dann das Forschungsdesign. Nehmen wir an, man möchte den Einfluss von zwei unabhängigen Variablen – 'Lehrformat' (mit den Stufen 'Flipped Classroom' und 'klassisch') und 'Vorwissen' (mit den Stufen 'hoch' und 'niedrig') – auf eine abhängige Variable wie 'Lernleistung' untersuchen. In diesem Fall wäre ein 2x2-Design geeignet, welches insbesondere potenzielle Wechselwirkungen zwischen den Variablen bewerten kann.

		Lehrformat	
		klassisch	Flipped Classroom
Vorwissen	niedrig	ab Var. Lernleistung*	ab Var. Lernleistung*
	hoch	ab Var. Lernleistung*	ab Var. Lernleistung*

* standardisierter Fragebogen

Abbildung 5: Klassisches 2x2-faktorielles Design

Nachdem das Forschungsdesign festgelegt wurde, geht es an die Operationalisierung und die Auswahl geeigneter, idealerweise standardisierter, Messinstrumente, etwa Fragebögen. Empfehlenswerte Quellen für standardisierte Messinstrumente sind Plattformen wie www.testarchiv.eu generell und www.physport.org speziell für physikalische Inhalte.

Anschließend wird der detaillierte Versuchsplan entworfen. Dieser legt fest, wann genau welche Schritte (z. B. Vortest) mit welchen Me-

thoden (z. B. Fragebogen) für welche Probandengruppen (etwa eine Gruppe, die das klassische Lehrformat mit niedrigem Vorwissen erfährt) und über welchen Zeitraum (z. B. 45 Minuten für einen Vortest) durchgeführt werden. Einfachere Versuchspläne, wie sie häufig bei Evaluationen eingesetzt werden, beschränken sich nur auf das Lehrmittel (Treatment) und einen anschließenden Test (Nachtest). Ein Anwendungsbeispiel wäre die Untersuchung von Leistungsunterschieden nach der Anwendung eines bestimmten Lehrformats, wie in Abbildung 6 dargestellt.



Abbildung 6: Einfacher Versuchsplan mit Nachtest

Ein Hauptproblem bei der Bewertung der Lernleistung mittels so einfacher Versuchspläne ist, dass die Lernleistung von weiteren Faktoren, wie beispielsweise dem Vorwissen, beeinflusst werden könnte. Deshalb wird man nicht aus den Ergebnissen schließen können, dass ein bestimmtes Lehrformat direkt zu einer spezifischen Lernleistung führt.

Um dieses Problem anzugehen, setzen viele Forschungsprojekte auf Vor-Nachtest-Designs, wie in Abbildung 7 dargestellt. Allerdings ermöglichen auch diese Designs nicht zwangsläufig einen klaren kausalen Zusammenhang zwischen der unabhängigen Variable oder dem Treatment (hier das Lehrformat) und den festgestellten Veränderungen zwischen den Messungen (z.B. Lernzuwachs). Es könnte immer noch eine nicht kontrollierte dritte Variable, wie kognitive Fähigkeiten, vorliegen, die den Unterschied zwischen den beiden Messzeitpunkten erklärt.



Abbildung 7: Einfacher Versuchsplan mit Vor- und Nachtest.

Um unkontrollierte Drittvariablen auszuschließen, sind Prä-Post-Designs mit Kontrollgruppen in Verbindung mit einer Randomisierung

hilfreich, wie in Abbildung 8 dargestellt. Bei diesem Design erhält die Kontrollgruppe kein Treatment, wird jedoch zu beiden Messzeitpunkten hinsichtlich der abhängigen Variable überprüft. Aber selbst dieses Design ist nicht ohne Einschränkungen, besonders wenn es darum geht, kausale Schlussfolgerungen zu ziehen. Ein möglicher Stolperstein könnte sein, dass die Vorabmessung das anschließende Treatment beeinflusst. Dies könnte zum Beispiel geschehen, wenn den Teilnehmenden durch die Vorabmessung klar wird, welche Aspekte des Treatments als besonders relevant angesehen werden.



Abbildung 8: Versuchsplan mit Vor- und Nachtest sowie Kontrollgruppe.

Ein Lösungsansatz bietet der Solomon-4-Gruppenplan [5]. Er berücksichtigt nicht nur die Hauptfaktoren, sondern kontrolliert auch mögliche Auswirkungen der Messungen selbst, wie in Abbildung 9 dargestellt.

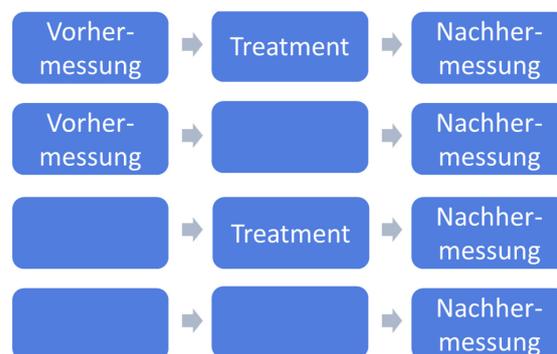


Abbildung 9: Solomon-4-Gruppenplan mit doppelter Kontrollgruppe (Treatment und Messung) [5].

Probleme bei der Umsetzung von Forschungsprojekten

Bei vielen universitären Kursen melden sich die Studierenden selbst an, basierend auf ihren Interessen, ihren Anforderungen oder ihrem Studienverlauf. Es wäre nicht praktikabel, sie zufällig einem bestimmten Kurs oder einer Lehrperson zuzuweisen.

Das Randomisieren von Studierenden könnte als unfair empfunden werden, insbesondere wenn ein Kurs oder eine Lehrmethode als überlegen oder begehrter angesehen wird. Studierende könnten sich benachteiligt fühlen, wenn sie zufällig einem weniger bevorzugten Kurs zugewiesen werden (insbesondere, wenn es sich um die Kontrollgruppe handelt).

Die Implementierung einer Randomisierung in universitären Kursen würde einen erheblichen organisatorischen Aufwand erfordern. Es müssten Systeme eingerichtet werden, um sicherzustellen, dass die Zuteilung der Studierenden korrekt und wirklich zufällig erfolgt.

Angesichts all dieser Überlegungen kann die Durchführung einer Randomisierung in universitären Kursen so aufwendig und komplex sein, dass sie den potenziellen Nutzen übersteigt, insbesondere wenn das Hauptziel der Evaluation ist, rückmeldungsorientiertes Feedback zur Verbesserung der Lehre zu sammeln und nicht unbedingt kausale Zusammenhänge herzustellen.

Ausweg aus dem Dilemma

Eine einfache Möglichkeit, die genannten Probleme zu adressieren, besteht darin, Wirkungsanalysen mittels Regressions- oder Pfadanalysen durchzuführen. Auch wenn diese Ansätze keine Kausalitäten im strengen Sinne nachweisen, ermöglichen sie es, Unterschiede in den abhängigen Variablen durch die Einflüsse von unabhängigen, Stör- und Drittvariablen zu erläutern. Deshalb ist es von Vorteil, zusätzlich zur abhängigen Variable ein detailliertes Verständnis potenzieller Stör- und Drittvariablen zu haben und diese ebenfalls zu erfassen. Bei den Probandenmerkmalen könnten beispielsweise Unterschiede in kognitiven Fähigkeiten [6], in der mentalen kognitiven Belastung [7] oder im themenspezifischen Vorwissen die Varianz der abhängigen Variable beeinflussen. Diese Faktoren könnten zusätzlich zu den durch die unabhängige Variable verursachten Unterschieden beitragen.

5. Fazit und Ausblick

Die in den letzten beiden Kapiteln geführte Diskussion zeigt, wie unterschiedlich die Funktionen bzw. Fragestellungen von Evaluation(-

sforschung) sein können. Daraus ergeben sich unterschiedliche Herangehensweisen in der Entwicklung und Ausgestaltung der Instrumente. Diese Unterschiede determinieren die Interpretation der Ergebnisse und die Konsequenzen für Lehrende. Dies zu beachten und klar zu definieren ist die Basis einer gelingenden Evaluation. Im Workshop „Wie gut ist meine Lehre (wirklich)?“ im Rahmen der Lessons Learned Konferenz 2023 wurde mit den Lehrenden über Evaluation(-sforschung), ihre unterschiedlichen Funktionen und über die Entwicklung und Gestaltung diskutiert. In einem kollaborativen Prozess entstand ein Evaluationsbaustein, der von Lehrenden genutzt werden kann (Beschreibung im Abschnitt 3). Außerdem wurden unterschiedliche Forschungsansätze hinsichtlich ihrer Praktikabilität erörtert. Das gemeinsame Fazit war, weniger Unterschiedshypothesen zu prüfen, da diese in der Regel mit komplexen und aufwändigen Designs verbunden sind. Der Schwerpunkt wurde stattdessen auf Wirkungsanalysen verlagert, um Variationen in einer abhängigen Variable besser erklären zu können.

Zur Lessons Learned Konferenz 2024 wird es dann um die gemeinsame Auswertung, Bewertung und Diskussion der Ergebnisse gehen. Unsere Vision einer gemeinsamen Evaluationsentwicklung in der Lessons Learned Community wird somit weiter vorgebracht.

Literatur

- [1] <https://tu-dresden.de/zqa/qualitaetsanalyse/lehreveranstaltungsevaluation>
- [2] Musterfragebogen Vorlesung des Zentrums für Qualitätsanalyse der TU Dresden. https://tu-dresden.de/zqa/ressourcen/dateien/qa/lve/LVE_Vorlesung_virtuell.pdf?lang=de
- [3] Deci, Edward L.; Ryan, Richard M. Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik Zeitschrift für Pädagogik 39 (1993) 2, S. 223-238
- [4] <https://bildungsportal.sachsen.de/umfragen/>
- [5] Döring, Nicola; Bortz, Jürgen. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (2016). Berlin, Heidelberg: Springer.
- [6] Heller, Kurt A.; Perleth, Christoph. Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (2000). Göttingen: Hogrefe.
- [7] Leppink, J., Paas, F., Van der Vleuten, C.P.M. et al. Development of an instrument for measuring different types of cognitive load. Behav Res 45, 1058-1072 (2013).