

Der Einfluss von Feedback auf die Lehrveranstaltungsevaluation

Dorothea Krampen und Andreas Gold

Feedback begünstigt den akademischen Lernerfolg und ist Teil der pädagogischen Professionalität. Feedback kann maßgeblich zum Gelingen qualitativvoller Lehre beitragen, weshalb davon auszugehen ist, dass sich eine gezielte Anwendung auch in der Bewertung von Lehrveranstaltungen durch die Lernenden niederschlägt. Im Rahmen eines Lehrforschungsprojekts wird untersucht, wie sich eine systematische Variation des Feedbacks der Lehrperson an Studierende auf die Ergebnisse der standardisierten Lehrveranstaltungsevaluation auswirkt. Dazu werden zwei Stichproben von Lehramtsstudierenden aus deckungsgleichen Veranstaltungen betrachtet. Die Bewertung der Lehrveranstaltung in der Experimentalgruppe, die individualisiertes Feedback erhält, fällt signifikant positiver aus als jene in der Kontrollgruppe, die kein individualisiertes Feedback erhält.

Feedback umfasst „Information(en) [...] über Aspekte der eigenen Leistung oder des eigenen Verstehens“ (Hattie, 2013, S. 206). Damit gehört es zum „professionellen Handlungsrepertoire von Lehrenden“ (Hischier, 2016, S. 1) und meint in erster Linie, dass Lehrende Rückmeldungen an Lernende geben, um Lernprozesse zu unterstützen (Buhren, 2015). Als Instrument zur formativen Evaluation während des Lernprozesses (Fuchs & Fuchs, 1986) kann Feedback eine Brücke vom Ist- zum Soll-Zustand schlagen und folglich die Wahrscheinlichkeit der (Lern-)Zielerreichung erhöhen (Hattie, 2013; Timperley, 2013). Damit Feedback wirksam ist, sollte es individuell, konkret, aufgaben- und nicht rein auf Eigenschaften der Person bezogen sein sowie unmittelbar, also in zeitlicher Nähe zum Unterrichtsgeschehen, stattfinden (z. B. Bastian, Combe & Langer, 2016). Feedback in diesem Sinne ist einer der mächtigsten Einflussfaktoren auf den akademischen Lernerfolg und kann zum Gelingen qualitativvoller Lehre auf allen Ebenen des Bildungsbereichs beitragen (Hattie, 2013, S. 206 ff.; Timperley, 2013). Es ist davon auszugehen, dass sich die Anwendung von Feedback auch in der Bewertung von Lehrveranstaltungen durch Lernende widerspiegelt. Meyer (2004) beschreibt nicht zuletzt einen Zusammenhang zwischen Feedback und Lernklima. Nach Hattie (2013, S. 206) besteht die Kunst allerdings darin, Feedback geben und empfangen zu können. Feedback, das Lernende Lehrenden geben, ist genauso wichtig wie Feedback, das Lehrende Lernenden geben.

Fragestellung und Hypothesen

Im Rahmen eines Lehrforschungsprojekts soll folgende Fragestellung untersucht werden:

Wie wirkt sich eine planmäßige Variation des Feedbacks der Lehrperson an die Studierenden auf die Ergebnisse der Lehrveranstaltungsevaluation aus?

Es wird erwartet, dass die Bewertung der Lehrveranstaltung bei Anwendung von individualisiertem Feedback positiver ausfällt als ohne.

Methode

Das Lehrforschungsprojekt wird im Rahmen der digitalen Lehrveranstaltung „*BW-D/Sb1: Mathematische Kompetenzen: Entwicklung, Diagnostik und Förderung*“ an der Goethe-Universität Frankfurt am Main umgesetzt, die im Wintersemester 2020/2021 und im Sommersemester 2021 gleichermaßen über die Lernplattform OLAT angeboten wird. Es handelt sich um ein Blockseminar für Studierende aller Lehramtsstudiengänge im Studienanteil Bildungswissenschaften. Inhalte sind die Entwicklung mathematischer Kompetenzen von den Vorläuferfertigkeiten im Kindergarten über die Grundschule bis hin zur Oberstufe, die Diagnostik und die Förderung mathematischer Kompetenzen sowie konkrete Unterrichtsmodelle für das Schulfach Mathematik. Bei der Veranstaltung, die sich über drei aufeinanderfolgende Tage erstreckt, wird pro Tag eine andere Aufgabe in Einzelarbeit abgefasst (Reflexion zum Text, Fallbeispiel, Essay) und eingereicht. Jede*r Teilnehmer*in erhält am Ende eines Tages ein kurzes schriftliches Feedback dazu von der Lehrperson über das Bewertungswerkzeug in OLAT. Es wird also insgesamt dreimal Feedback gegeben (formativ). Abhängige Variable ist das Ergebnis der standardisierten Lehrveranstaltungsevaluation. In dem Fragebogen (https://www.uni-frankfurt.de/87159694/Informationen_f%C3%BCr_Lehrende) wird neben dem Basisteil routinemäßig der Ergänzungsteil zur digitalen Lehre eingesetzt. Antworten erfolgen auf einer Skala von 1 (*stimme überhaupt nicht zu*) bis 6 (*stimme voll und ganz zu*) durch die Studierenden. Offene Antworten, die ebenfalls möglich sind, werden nicht ausgewertet. Die Lehrveranstaltungsevaluation wird elektronisch am letzten Veranstaltungstag durchgeführt. Die Teilnahme daran ist freiwillig.

Die unabhängige Variable „Feedback“ wird in zwei Stufen variiert: nicht-individualisiertes Feedback versus individualisiertes Feedback. Beim nicht-individualisierten Feedback wird lediglich eine Kennzeichnung der Aufgabenbearbeitung als „bestanden“ oder „nicht bestanden“ vorgenommen sowie eine allgemeine Tagesrückmeldung als Mitteilung

verfasst.¹⁸ Beim individualisierten Feedback werden zusätzlich zur Kennzeichnung als „bestanden“ oder „nicht bestanden“ etwa drei bis fünf Sätze mit konkretem, lernrelevantem Bezug zur eigenen Aufgabenbearbeitung als schriftlicher Kommentar beigefügt.¹⁹

Die Experimentalgruppe (individualisiertes Feedback) bilden die Teilnehmer*innen der Lehrveranstaltung aus dem Wintersemester 2020/2021, die Kontrollgruppe (nicht-individualisiertes Feedback) die Teilnehmer*innen aus dem Sommersemester 2021.

Eine „echte“ Kontrollgruppe, die gar kein Feedback erhält, ist aus didaktischen Gründen nicht angeraten. Es handelt sich somit um ein quasiexperimentelles Nur-Nachtest-Design.

Ergebnisse

Zunächst werden die Stichproben beschrieben, danach deskriptiv- und inferenzstatistische Ergebnisse präsentiert. Da den Dozent*innen die Auswertung der Lehrveranstaltungsevaluation als Zusammenschau übermittelt wird, die keinerlei Zuordnung von Werten zu (anonymisierten) Personen erlaubt, können auf dieser Grundlage nur ausgewählte statistische Analysen durchgeführt werden.

Stichprobenbeschreibung

An der Lehrveranstaltung im Wintersemester 2020/2021 haben insgesamt $N = 31$ Personen teilgenommen. Weil nicht alle den Lehrveranstaltungsevaluationsfragebogen ausgefüllt haben, beläuft sich die Experimentalgruppe auf $n_E = 20$ Lehramtsstudierende. Davon sind 90 Prozent weiblich und 10 Prozent männlich. Zwischen dem dritten und dem siebten Fachsemester befinden sich 85 Prozent der Personen, 15 Prozent befinden sich im neunten oder einem höheren Fachsemester.

Im Sommersemester 2021 haben insgesamt $N = 38$ Personen an der Lehrveranstaltung teilgenommen. Die Kontrollgruppe bilden $n_K = 21$ Lehramtsstudierende, die den Lehrveranstaltungsevaluationsfragebogen ausgefüllt haben. 81 Prozent davon sind weiblich und 19

¹⁸ Beispiel für eine allgemeine Tagesrückmeldung: „Sehr geehrte Teilnehmer*innen der Veranstaltung ‚BW-D/Sb1/Mathematische Kompetenzen: Entwicklung, Diagnostik und Förderung‘ im Sommersemester 2021, vielen Dank für ihre gelungenen Fallbeispiele zum heutigen Thema *Schwache Rechner*innen*, die ich mit Freude gelesen habe! Das Vorgehen zur klinischen Diagnose einer Rechenstörung ist hier noch einmal übersichtlich dargestellt: <https://cfdn.aerzteblatt.de/bilder/2019/02/img139597106.gif>. Morgen wird dann das Thema ‚Förderung‘ im Vordergrund stehen. Bis dahin und herzliche Grüße Dorothea Krampen“

¹⁹ Beispiel für ein individualisiertes Feedback: „Sehr geehrte*r XXX, vielen Dank für die Einreichung Ihrer differenzierten Reflexion. Weiter so! Der von Ihnen angesprochene ‚Knick‘ in Abb. 10 sollte nicht überinterpretiert werden. Das fällt unter Zufallsschwankungen und die Differenz zwischen den Messzeitpunkten ist tatsächlich gering. Vielleicht lässt er sich auch als eine Art Plateau interpretieren, nämlich dass z. B. das Konventions- und Regelwissen irgendwann nicht mehr (weiter) ansteigt. Herzliche Grüße Dorothea Krampen“

Prozent männlich. Zwischen dem ersten und dem vierten Fachsemester befinden sich 81 Prozent der Studierenden, jeweils 9.5 Prozent befinden sich im achten beziehungsweise dem neunten oder einem höheren Fachsemester.

Ein Chi²-Homogenitätstest (z. B. Weiß, 2019) zeigt, dass die relativen Häufigkeiten für das Geschlecht in den beiden unabhängigen Stichproben nicht gleich verteilt sind ($\chi^2 = 8.889(1), p = .003$). Dies gilt es bei der Interpretation der Ergebnisse zu berücksichtigen.

Deskriptivstatistik

Als Erstes interessiert ein deskriptivstatistischer Überblick zu den Globalwerten (siehe Tabelle 1). Der Globalwert des Basisteils des Lehrveranstaltungsevaluationsfragebogens ergibt sich als Mittelwert (*MW*) aus den Items 1.1–1.7. Der Globalwert des Ergänzungsteils zur digitalen Lehre setzt sich als Mittelwert aus den Items 2.1–2.8 zusammen. Zusätzlich wird für den Ergänzungsteil der Mittelwert des Items 2.9 als zusammenfassendes Urteil zur Zufriedenheit mit der digitalen Lehre berichtet.

Tab. 1: Deskriptivstatistische Kennwerte der Globalwerte

Stichprobe	Globalwerte		
	Basisteil (Items 1.1–1.7)	Digitale Lehre (Items 2.1–2.8)	Digitale Lehre: Zufriedenheit (Item 2.9)
	<i>MW</i> (<i>SD</i>)	<i>MW</i> (<i>SD</i>)	<i>MW</i> (<i>SD</i>)
EG	5.7 (0.5)	5.8 (0.4)	5.7 (0.7)
KG	4.9 (1.0)	5.6 (0.7)	5.4 (0.8)
<i>MW</i> -Differenz (EG–KG)	0.8	0.2	0.3

Anm.: EG = Experimentalgruppe, KG = Kontrollgruppe, *MW* = Mittelwert, *SD* = Standardabweichung

Es lässt sich erkennen, dass im Mittel alle Globalwerte der Kontrollgruppe niedriger liegen als die Werte der Experimentalgruppe. Die größte Differenz zwischen den Gruppen findet sich mit 0.8 – und damit fast einer ganzen Antwortkategorie – für den Globalwert des Basisteils. Die Streuung der Antworten ist für diesen Globalwert in der Kontrollgruppe deutlich höher als in der Experimentalgruppe. Die sonstigen Standardabweichungen (*SD*) sind unauffällig.

Ein differenzierteres Bild bietet die Betrachtung von Einzelitems (siehe Tabelle 2). Im Folgenden werden die Items, aus denen sich der Globalwert des Basisteils zusammensetzt (Items 1.1–1.7), genauer dargestellt, da sich ihr Inhalt auf Aspekte bezieht, die für den Lernerfolg besonders relevant sind.^{20,21}

Tab. 2: Deskriptivstatistische Kennwerte relevanter Einzelitems

Item	Stichprobe		
	EG <i>MW (SD)</i>	KG <i>MW (SD)</i>	<i>MW-Differenz (EG–KG)</i>
1.1 Der Besuch der Veranstaltung führt zu einem spürbaren Wissenszuwachs.	5.6 (0.6)	5.1 (0.8)	0.5
1.2 Inhalte werden anschaulich vermittelt.	5.8 (0.6)	4.7 (1.1)	1.1
1.3 In der Veranstaltung werden auch schwierige Inhalte verständlich erklärt.	5.4 (0.9)	4.4 (1.0)	1.0
1.4 Die Relevanz der behandelten Themen wird deutlich.	5.9 (0.3)	5.2 (0.9)	0.7
1.5 Der/die Lehrende ist in der Lage, strukturiert zu erklären.	5.7 (0.5)	4.8 (1.5)	0.9
1.6 Der/die Lehrende gibt hilfreiches Feedback auf die Beiträge der Studierenden.	5.9 (0.5)	4.9 (0.8)	1.0
1.7 Der/die Lehrende achtet darauf, eine wertschätzende Lehr-/Lernatmosphäre herzustellen.	6.0 (0.0)	4.9 (0.7)	1.1

Anm.: EG = Experimentalgruppe, KG = Kontrollgruppe, *MW* = Mittelwert, *SD* = Standardabweichung

Tabelle 2 zeigt, dass die Bewertungen in der Experimentalgruppe für alle interessierenden Einzelitems positiver ausfallen als in der Kontrollgruppe. Die größten Differenzen zwischen den Gruppen von 1.0 oder mehr finden sich – in absteigender Reihenfolge – bei Items 1.2 (*Inhalte werden anschaulich vermittelt.*) und 1.7 (*Der/die Lehrende achtet darauf, eine wertschätzende Lehr-/Lernatmosphäre herzustellen.*) mit jeweils 1.1 sowie bei Items 1.3 (*In der Veranstaltung werden auch schwierige Inhalte verständlich erklärt.*) und 1.6 (*Der/die Lehrende gibt hilfreiches Feedback auf die Beiträge der Studierenden.*) mit jeweils 1.0. Eine relativ hohe Streuung der Antworten lässt sich bei Item 1.5 (*Der/die Lehrende ist in*

²⁰ In Abgrenzung zu den Items des Ergänzungsteils zur digitalen Lehre, die vornehmlich Aspekte der technischen Umsetzung der Lehrveranstaltung erfassen

²¹ Item 1.8 (*Meine Mitstudierenden tragen zu einer konstruktiven Lernatmosphäre bei.*) geht nicht in den Globalwert ein, weshalb es hier nicht aufgeführt, der Vollständigkeit halber jedoch erwähnt wird

der Lage, strukturiert zu erklären.) in der Kontrollgruppe erkennen. Die zumeist niedrigeren Standardabweichungen in der Experimentalgruppe können weithin durch Deckeneffekte erklärt werden.

Inferenzstatistik

Die Experimentalgruppe weist im Vergleich zur Kontrollgruppe einen hoch signifikant positiveren Globalwert des Basisteils der Lehrveranstaltungsevaluation auf ($t(39) = 3.214$, $p = .001$).²² Der Globalwert des Ergänzungsteils zur digitalen Lehre ($t(39) = 1.116$, $p = .136$) sowie die globale Zufriedenheit mit der digitalen Lehre ($t(39) = 1.275$, $p = .105$) unterscheiden sich dagegen nicht überzufällig zwischen den Gruppen.

Bezogen auf die relevanten Einzelitems lassen sich durchgängig signifikante, oftmals hochsignifikante Unterschiede feststellen (siehe Tabelle 3). Die Antworten der Experimentalgruppe fallen im Mittel jeweils deutlich positiver aus als die der Kontrollgruppe.

Tab. 3: Inferenzstatistische Ergebnisse relevanter Einzelitems

Item	<i>t</i> -Test (einseitig)
1.1	$t(39) = 2.255, p = .015^*$
1.2	$t(38) = 3.926, p < .001^{***}$
1.3	$t(33) = 3.103, p = .002^{**}$
1.4	$t(39) = 3.306, p = .001^{***}$
1.5	$t(19) = 1.882, p = .038^*$
1.6	$t(32) = 4.490, p < .001^{***}$
1.7	$t(28) = 6.727, p < .001^{***}$

Anm.: *: $p \leq .05$, **: $p \leq .01$, ***: $p \leq .001$

Diskussion

Das Lehrforschungsprojekt hat untersucht, wie sich eine systematische Variation des Feedbacks der Lehrperson an Studierende auf die Ergebnisse der standardisierten Lehrveranstaltungsevaluation auswirkt. Es lässt sich festhalten, dass die Bewertung der Lehrveranstaltung in der Experimentalgruppe, die individualisiertes Feedback erhält,

²² Gemäß den gerichteten Hypothesen wird hier und im Folgenden jeweils der einseitige *p*-Wert angegeben

verglichen mit der Kontrollgruppe, die kein individualisiertes Feedback erhält, hinsichtlich lernerfolgsrelevanter Aspekte signifikant positiver ausfällt. Die Ergebnisse sind erwartungsgemäß. Unterschiede in den Bewertungen zwischen den Gruppen zeigen sich insbesondere hinsichtlich der inhaltlichen Basisdimensionen „Verständnis“ (Items 1.2 und 1.3), „Lernklima“ (Item 1.7) sowie „Feedback“ (Item 1.6) (siehe Tabellen 2 und 3). Eine planmäßige Anwendung von Feedback wirkt sich damit offenbar direkt oder indirekt auf die genannten Dimensionen aus (z. B. Meyer, 2004), die wiederum direkt oder indirekt den Lernerfolg beeinflussen können. Genaue Zusammenhänge müssen weiter erforscht werden. Desgleichen, ob und unter welchen Bedingungen sich generalisierende Effekte auf die Bewertung nicht unmittelbar lernerfolgsrelevanter Aspekte beobachten lassen, wie sie in der vorliegenden Arbeit ausbleiben (Digitale Lehre: Globalwert und Zufriedenheit, siehe Tabelle 1). Fest steht jedoch, dass Feedback ein Indikator für Unterrichtsqualität sein kann und auf diese Weise gezielt zur Verbesserung der eigenen Lehre herangezogen werden kann.

Dass sich Feedback lohnt, ist vielfach belegt (zusammenfassend Hattie, 2013; Timperley, 2013). Neu hingegen die Erkenntnis, dass sich der Einsatz von Feedback auch in den Ergebnissen der Lehrveranstaltungsevaluation niederschlägt. Dies spricht nicht zuletzt dafür, dass das Instrument geeignet scheint, Veränderungen in der Lehre adäquat abzubilden und im Sinne Hatties *sichtbar* zu machen. Damit kann es im Rahmen der Personal- und Organisationsentwicklung gewinnbringend genutzt werden.

Feedback ist aufwendig. Es beansprucht – gut gemacht (z. B. Bastian et al., 2016) – sowohl personale, zeitliche als auch materielle Ressourcen. Angesichts der erhöhten Gefahr von (beruflicher) Überlastung und Erschöpfung gerade bei Lehrenden (z. B. vbw, 2014), gilt es, dies im Sinne eines angemessenen Kosten-Nutzen-Verhältnisses bei der eigenen Ressourcenplanung zu berücksichtigen.

In dem Lehrforschungsprojekt wurden ausschließlich Lehramtsstudierende betrachtet. Es ist zwar plausibel anzunehmen, dass die Ergebnisse auf andere Studiengänge übertragbar sind, da dem Gewähren und Erfahren von Feedback in vielen Lehr-Lern-Zusammenhängen eine wesentliche Bedeutung zukommt, systematische Untersuchungen sind dennoch von Nöten.

Zum Zeitpunkt der Versuchsplanung war von stark unterschiedlichen Gruppenzusammensetzungen nicht auszugehen, da beide Stichproben (Experimental- und Kontrollgruppe) derselben Population entstammen. Nichtsdestotrotz konnten überzufällige Unterschiede hinsichtlich der Geschlechterverteilung festgestellt werden. Da den Dozent*innen die Auswertung der Lehrveranstaltungsevaluation in einem aufbereiteten, aggregierten

Datenformat zur Verfügung gestellt wird, konnten nicht immer alle zweckmäßigen Berechnungen durchgeführt werden. Unter anderem konnte der Einfluss des Geschlechts nicht post-hoc statistisch kontrolliert werden. Kohorteneffekte sind nicht auszuschließen. Die Lieferung von Rohdatensätzen zur Sekundäranalyse der eigenen Daten aus der Lehrveranstaltungsevaluation wäre daher auf Nachfrage wünschenswert.

Das individualisierte Feedback wurde so gestaltet, dass es die allgemeinen Wirksamkeitskriterien – individuell, konkret, aufgabenbezogen und zeitnah – erfüllt, es wurde aber hinsichtlich dieser Kriterien nicht bewusst variiert. Dies war auch nicht Ziel der Arbeit. Zweifelsohne handelt es sich dabei jedoch um ein lohnendes Unterfangen für künftige Untersuchungen, um tiefere Einsichten zu „gutem“ Feedback zu gewinnen. Zudem kann es einen Unterschied machen, wer das Feedback gibt (Lehrperson, studentische Hilfskraft, Peer). Eine Ergänzung des standardisierten Lehrveranstaltungsevaluationsfragebogens um weitere interessierende Teile beziehungsweise Konstrukte (https://www.uni-frankfurt.de/87159694/Informationen_f%C3%BCr_Lehrende) sollte ebenso erwogen werden.

Einleitend wurde erwähnt, dass die eigentliche Kunst darin besteht, Feedback nicht nur geben, sondern auch empfangen zu können (Hattie, 2013, S. 206). Gewissermaßen eine Doppelsieg-Strategie, aus der alle Beteiligten einen Nutzen erzielen können. Eine Anwendung von Feedback als wichtiger Baustein qualitätvoller Lehre gepaart mit einer standardisierten Evaluation als Rückmeldung an die Lehrenden, wie in dieser Untersuchung realisiert, kann mithin dazu beitragen, die eigene Lehre kontinuierlich zu überprüfen und gegebenenfalls zu modifizieren.

Literatur

- Bastian, J., Combe, A. & Langer, R. (2016). *Feedback-Methoden: Erprobte Konzepte, evaluierte Erfahrungen* (4. Aufl.). Weinheim: Beltz.
- Buhren, C. G. (Hrsg.). (2015). *Handbuch Feedback in der Schule*. Weinheim: Beltz.
- Fuchs, L. S. & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199-208. doi: 10.1177/001440298605300301
- Hattie, J. (2013). *Lernen sichtbar machen*. Baltmannsweiler: Schneider.
- Hischier, D. S. (2016). *Feedback – der Faktor unter der Lupe*. Verfügbar unter: https://www.lernensichtbarmachen.ch/wp-content/uploads/2016/09/Feedback_Faktor-unter-der-Lupe_Final.pdf (abgerufen am 17.06.2022).
- Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelsen Scriptor.

- Timperley, H. (2013). Feedback. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 402-404). New York: Routledge/Taylor & Francis Group.
- vbw – Vereinigung der Bayerischen Wirtschaft e. V. (Hrsg.). (2014). *Psychische Belastungen und Burnout beim Bildungspersonal. Empfehlungen zur Kompetenz- und Organisationsentwicklung. Gutachten*. Münster: Waxmann.
- Weiß, Chr. (2019). Tests zum Vergleich von Häufigkeiten. In Chr. Weiß, *Basiswissen Medizinische Statistik* (7. Aufl., S. 155-168). Berlin: Springer. doi: 10.1007/978-3-662-56588-9_11