

DeGEval Jahrestagung
Dr. Elisa Knödler
Prof. Dr. Moritz Heene
Ludwig-Maximilians-Universität München
17.09.2020

Lehrveranstaltungsevaluation mit einem verhaltensbasierten und psychometrisch soliden Messinventar – ein Beitrag zur Qualitätssicherung in Lehre und Studium? Eine Mehrebenenanalyse.





- (1) Konstruktionsbasis des verhaltensbasierten Messinventars
- (2) Eine Analyse über die Semester hinweg
- (3) Ausblick

Evaluation
verbessert die
Qualität der
Lehre nicht.
(Berthold, 2002)

Evaluation hat
keinen **Nutzen**.
(Döring, 2002)

Evaluation ist mit
einem viel zu hohen
Aufwand verbunden.
(Hanft, 2004)



<https://de.fotolia.com/id/41357852>



- Fehlende Theorie bei der Konstruktion (z.B. Koch, 2004; Marsh, 1984)
- Fehlende Testung und Validierung (z.B. Covert & Mason, 2016; Rindermann, 2009)



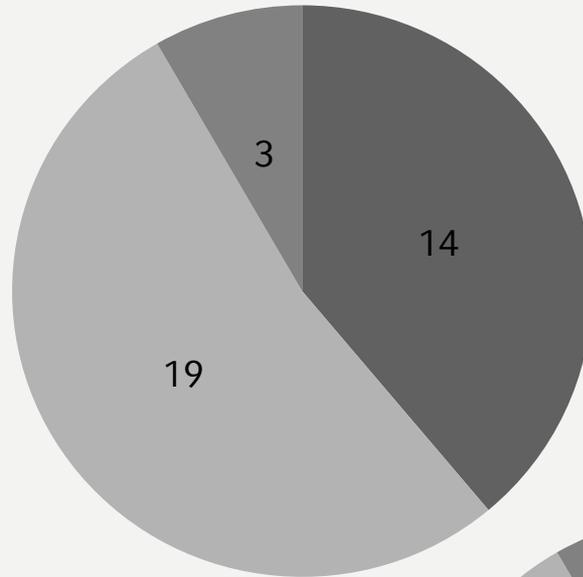
Konstruktionsbasis

- Theorien
 - Organisations- und Evaluationstheorien
 - Theorien zu Qualität, Qualitätsmanagement und -sicherung im Hochschulbereich
 - Methodische Aspekte aus der LVE-Forschung
- Zwei Vorstudien
 - Leitbild guter Lehre zur Faktorenkonstruktion
 - CITs zur Itemgenerierung

→ Pilotiert und Validiert

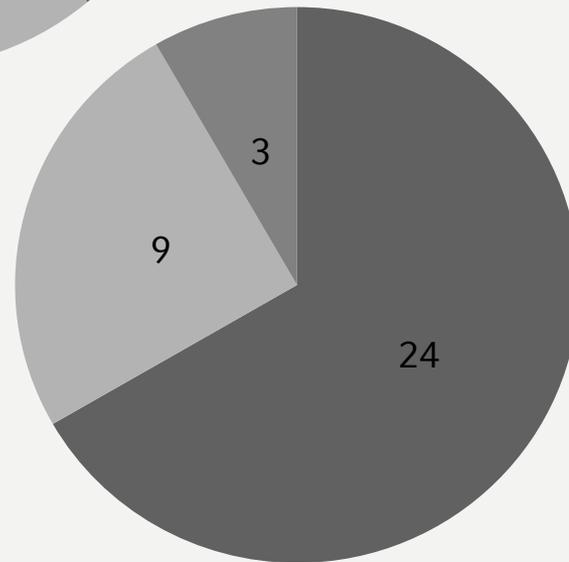


Vorlesungen:



- Verhaltensbasierte Items aus den CITs
- Modifizierte Items elaborierter Instrumente
- Modifizierte Items aus dem alten Bogen

Seminare:



- Verhaltensbasierte Items aus den CITs
- Modifizierte Items elaborierter Instrumente
- Modifizierte Items aus dem alten Bogen



Einsatz in der Fakultät

- LVE in jedem Semester
- Erhebungszeitpunkt ca. 3 - 4 Wochen vor Semesterende
- Rückmeldung während des Semesters (MW, MD, SD, Faktorwerte, Vergleichslinien)

Fakultätsübergreifender Einsatz

- Sommersemester 16: 3 Fakultäten
 - 32 Vorlesungen (N = 1064) und 68 Seminare (N = 966)
- Wintersemester 17/18: 3 Fakultäten
 - 52 Vorlesungen (N = 1786) und 63 Seminare (N = 846)



Einsatz in der Fakultät – Berechnung über mehrere Semester

Sommersemester 2017 bis Wintersemester 2018/2019

- Seminarbogen $N = 2721$
- Vorlesungsbogen $N = 2930$

Die Daten werden mithilfe von gemischten Modellen analysiert

Cumulative Link Mixed Models (CLMM)

- Zielgröße wird ordinalskaliert betrachtet
- Kumulierte Logits sind proportional über alle Kategorien hinweg (die Chancenverhältnisse sind unabhängig von der betrachteten Schwelle, hier Note, und sind nur proportional zum Unterschied von x zu $(x+1)$).



Feste Effekte

- Geschlecht
- Alter
- Pflicht-/Wahlveranstaltung
- Interesse an der behandelten Thematik zu Beginn
- Faktoren
 - Lernendenzentrierung
 - Nachvollziehbarkeit
 - Didaktischer Aufbau (VL)/Didaktik und Engagement (S)

Zufällige Effekte

- Random Intercept pro Person Studierender
- Random Intercept pro Lehrenden
- Random Intercept der einzelnen Veranstaltungen (innerhalb eines Dozierenden)



Vorlesungen

coefficients	estimate	std. error	z value	P value
Sympathie	0.82	0.12	6.92	0.000***
Geschlecht (weiblich)	-0.17	0.23	-0.77	0.444
Alter	0.26	0.16	1.67	0.095 .
Pflicht	-0.14	0.24	-0.60	0.548
Interesse	0.21	0.07	2.98	0.003 **
Didaktischer Aufbau	2.22	0.23	9.82	0.000 ***
Lernendenzentrierung	0.24	0.12	2.02	0.043 *
Nachvollziehbarkeit	0.22	0.29	0.74	0.460
MW Sympathie Doz.	0.42	0.23	1.83	0.067 .



Vorlesungen

Keinen signifikanten (Signifikanzniveau $\alpha = 0.05$) Einfluss auf die Bewertung der Lehrenden bzw. einen Unterschied zwischen den betrachteten Gruppen haben:

- Geschlecht, Alter, Pflichtveranstaltung und „Sympathie auf Lehrendenebene“ sowie Faktor „Nachvollziehbarkeit“

Einen *signifikanten* Einfluss haben:

- Sympathie, Faktoren „Lernendenzentrierung“ und „Didaktische Aufbau“ sowie das Interesse zu Beginn der Veranstaltung



Gruppe	p-Werte
Studierende	0.05
Dozierende	<. 0.001

- Der zufällige Effekt „Studierender“ ist *nicht signifikant*
 - keinen Nachweis für eine individuelle Tendenz eines Studierenden beim Bewerten verschiedener Seminare, wenn alle Effekte der Einflussgrößen einbezogen wurden.
- Der zufällige Effekt „Dozierender“ ist *signifikant*
 - Weiterhin statistisch substanziell verbleibende Variabilität zwischen den Lehrenden erklärt wird, nachdem die Effekte der Einflussgrößen einbezogen wurden.



Seminare

coefficients	estimate	std. error	z value	P value
Sympathie	0.98	0.09	10.39	0.000 ***
Geschlecht (weiblich)	-0.10	0.18	-0.07	0.571
Alter	0.02	0.07	0.34	0.738
Pflicht	0.22	0.22	1.00	0.318
Interesse	0.31	0.05	5.83	0.000 ***
Didaktik und Engagement	0.97	0.11	8.49	0.000 ***
Lernendenzentrierung	0.22	0.09	2.52	0.012 *
Nachvollziehbarkeit	0.81	0.14	5.61	0.000 ***
MW Sympathie Doz.	0.41	0.17	2.42	0.015 *



Seminare

Keine Unterstützung der Hypothese zum Einfluss auf die Bewertung der Lehrenden bzw. einen Unterschied zwischen den betrachteten Gruppen haben:

- Alter, Pflichtveranstaltung und „Sympathie auf Lehrendenebene“

Einen *signifikanten* Einfluss haben:

- **Geschlecht**, Sympathie, Faktoren „Lernendenzentrierung“, „**Nachvollziehbarkeit**“ und „Didaktik und Engagement “ sowie das Interesse zu Beginn der Veranstaltung



Gruppe	p-Wert
Studierende	<. 0.001
Dozierender:Veranstaltung	<. 0.001
Dozierende	<. 0.001

- Alle drei zufällige Effekte sind signifikant!
- „Studierender“ → Nachweis für individuelle Tendenz der Studierenden beim Bewerten mehrerer Veranstaltungen
- „Veranstaltungen innerhalb eines Dozierenden“ → erklärt die Variabilität zwischen den Veranstaltungen innerhalb eines Lehrenden
- „Dozierender“ → erhebliche Variabilität zwischen den Lehrenden

	Vorlesungen	Seminare
Geschlecht	-	✓
Alter	-	-
Pflicht	-	-
Interesse zu Beginn	✓	✓
Didaktik	✓	✓
Lernendenzentrierung	-	✓
Nachvollziehbarkeit	-	✓
Studierender	-	✓
Lehrender	✓	✓
Lehrender:Veranstaltung	n.a.	✓

Signifikanzniveau: $\alpha = .05$



Es gibt Unterschiede zwischen den Lehrenden, was bedeutet, dass manche Lehrende unabhängig vom Inhalt tendenziell besser oder schlechter bewertet werden.

- Eine Analyse mit der Kontrollvariable „Sympathie“ („*Mir war das Auftreten des Dozierenden sympathisch.*“) könnte Aufschluss darüber geben, ob die gesamte Bewertung von dieser Tatsache beeinflusst wird und in welchem Maße.

Im Semindatensatz gibt es signifikante Unterschiede in der Bewertung eines Lehrenden in unterschiedlichen Veranstaltungen.

- Ist die Bewertung des Lehrenden vom Fach abhängig? (z.B. Statistikkurse)

Das Interesse eines Studierenden zu Beginn einer Veranstaltung hat bei beiden Lehrformaten einen signifikanten Einfluss. Je größer das Interesse zu Beginn desto besser fällt im Durchschnitt die Bewertung aus.

- Wie lässt sich das Interesse zu Beginn beeinflussen?

**VIELEN DANK FÜR IHRE
AUFMERKSAMKEIT!**



- Berthold, C. (2002). Von der Evaluation zur strategischen Hochschulentwicklung – 16 Thesen. In: Reil, T./Winter, M. (Hg.): Qualitätssicherung an Hochschulen: Theorie und Praxis. Bielefeld, S. 160-165.
- Covert, R. W. & Mason, E. J. (2016). Factorial Validity of a Student Evaluation of Teaching Instrument. *Educational and Psychological Measurement*, 34(4), 903–905.
- Döring, N. (2002). Für Evaluation und gegen Evaluitis. Warum und wie Lehrevaluation an deutschen Hochschulen verbessert werden sollte. In B. Berendt, H.-P. Voss & J. Wildt (Hrsg.), *Raabe - nachschlagen, finden. Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten* (2nd ed.). Stuttgart: Raabe.
- Koch, E. (2004). *Gute Hochschullehre: Theoriebezogene Herleitung und empirische Erfassung relevanter Lehraspekte*. Zugl.: Köln, Univ., Diss., 2003. *Schriften zur pädagogischen Psychologie: Vol. 11*. Hamburg: Kovač.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707–754.
- Rindermann, H. (2009). *Lehrevaluation: Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierten Unterrichts* (2. Aufl). *Psychologie: Bd. 42*. Landau: Empirische Pädagogik e.V.