

# Auswirkungen eines Anreizsystems für differenzierende Prüfungen

V. Fischer<sup>1\*</sup>, H. Müller<sup>1</sup>, M. Krohn<sup>2</sup> & I. Just<sup>3</sup>

1 := Studiendekanat Medizin, Bereich Evaluation & Kapazität; 2:= Bereich Studium & Prüfungen; 3 := Studiendekanat Medizin  
OE 9135

Medizinische Hochschule Hannover, Carl-Neuberg-Str.1, D-30625 Hannover \*fischer.volkhard@mh-hannover.de

## Fragestellung

Können die zentralen LOM-Lehre-Parameter Qualitätsverbesserungen seit ihrer Einführung in 2010 dokumentieren?

Wenn ja, was könnte diese Veränderungen ausgelöst haben?

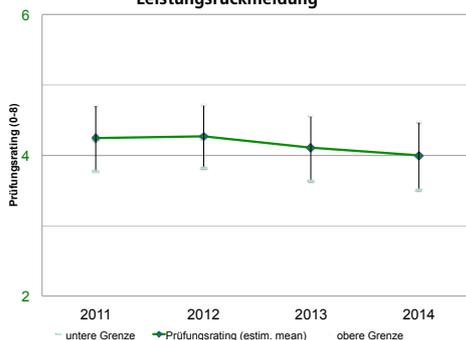
## Methode

- Jedes Modul wird unmittelbar nach der Prüfung evaluiert. Der Fragebogen erfasst neben der globalen Modulbewertung je eine Frage nach in welchen Dimensionen ein Studierender eine Stärke bzw. Verbesserungsmöglichkeiten sieht. Die Dimensionen sind die Lehrkräfte, die Inhalte der Lehrveranstaltungen, Lehr- und Lernmaterialien, der Patientenbezug, Prüfungen und die Organisation der Lehrveranstaltungen. Außerdem fünf weitere Fragen zu besonderen Aspekten.
- Das als Alternative zu Anpassungstests vorgestellte Ratingsystem (Fischer, Müller, Just, 2013) soll differenzierte Leistungsrückmeldungen abbilden. Hier betrachten wir die korrelativen Zusammenhänge zwischen diesem Parameter und den Evaluationsergebnissen.
- Die globale Modulevaluation und das Prüfungsrating wurden als abhängige Variablen in separaten Varianzanalysen auf Qualitätsverbesserungen in den letzten vier Jahren hin betrachtet.
- Auf eine getrennte Darstellung von Ergebnissen für verschiedene Prüfungsarten (mündlich, MCQ, SAQ, Portfolio, OSCE, etc.) wurde hier bewusst verzichtet.

## Ergebnisse 2

- Die ANOVA bezüglich des Ratings der Prüfungsqualität bestätigt das allgemeine Modell ( $F_{(94, 558)} = 1.619$ ;  $\alpha = .001$ ).
- Über die betrachteten Studienjahre hinweg läßt sich kein Haupteffekt nachweisen. Abbildung 2 zeigt die geschätzten Mittelwerte mit ihren oberen und unteren Schranken.
- Aber für die Tertiale läßt sich ein signifikanter Haupteffekt nachweisen ( $F_{(2, 558)} = 3.187$ ;  $\alpha = .042$ ), ebenso für die nach Perzentilen geclusterte Anzahl an Präsenzstunden des Moduls ( $F_{(7, 558)} = 10.024$ ;  $\alpha < .001$ ).
- Für die Interaktionen der UV lassen sich auch in dieser ANOVA keine signifikanten Effekte nachweisen.

Abb. 2:  
Ausmaß einer differenzierten  
Leistungsrückmeldung



## Diskussion

- Die gestiegene positive Bewertung der Module ist erfreulich, während der ausgebliebene Effekt für die Prüfungsqualität überrascht. Eine Interpretation beobachtbarer Unterschiede zwischen verschiedenen Prüfungsformen, was die Differenziertheit ihrer Leistungsrückmeldungen angeht, steht noch aus.
- Das Anreizsystem scheint außer hochschulinternen Diskussionen keine messbaren Veränderungen ausgelöst zu haben.
- Anscheinend verbinden Studierende etwas deutlich Anderes mit positiven Prüfungen als eine differenzierte Leistungsrückmeldung. Aus Diskussionen mit Lehrkräften wissen wir, dass auch diese in Teilen eines solches Rating als Indikator für schwerere Prüfungen interpretieren. Dazu passt, dass Studierende und Lehrkräfte Prüfungen mit weniger Fragen zu bevorzugen scheinen, ebenso wie eine Präferenz für reine Bestehensprüfungen, unabhängig von den gesetzlichen Bestimmungen.
- Die aktuelle Gremiendiskussion an der MHH dreht sich primär nicht um die Verbesserung der Prüfungsqualität, sondern um die Verbesserung der Finanzierungsbedingungen für Lehre durch die Abteilungen.

## Einleitung

Das Hannibal-Curriculum umfasst 46 Pflichtmodule mit mindestens einer summativen Prüfung. In den fünf Studienjahren haben Studierende insgesamt 72 Prüfungen erfolgreich zu bestehen. Aufgrund von Rotationsplänen führt die MHH pro Kalender-jahr ca. 165 Prüfungen durch.

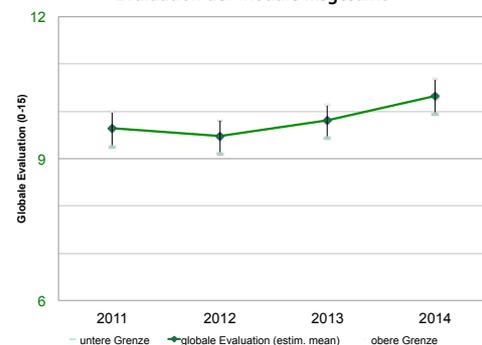
Seit 2010 versucht die Medizinische Hochschule Hannover (MHH) durch finanzielle Anreize für die besten Module des Modellstudiengangs Medizin die Lehre zu verbessern. Zentrale Parameter dieses Systems einer leistungsbezogenen Mittelvergabe in der Lehre (LOM-Lehre) sind die globale Bewertung der Module in der studentischen Evaluation und das Ausmaß, in dem die Modulprüfungen differenzierte Leistungsrückmeldungen ermöglichen.

## Ergebnisse 1

- Die ANOVA bezüglich der globalen Modulbewertung bestätigt das allgemeine Modell ( $F_{(92, 495)} = 2.105$ ;  $\alpha < .001$ ).
- Der Haupteffekt der Studienjahre ist signifikant ( $F_{(3, 495)} = 3.167$ ;  $\alpha = .024$ ). Abbildung 1 zeigt die geschätzten Mittelwerte mit ihren oberen und unteren Schranken.
- Ebenfalls signifikant ist der Haupteffekt für die nach Perzentilen geclusterte Anzahl an Präsenzstunden des Moduls ( $F_{(7, 495)} = 17.632$ ;  $\alpha < .001$ ).
- Für die Tertiale und die Interaktionen der UV lassen sich keine signifikanten Effekte nachweisen.

Abb. 1:

Evaluation der Module insgesamt



## Ergebnisse 3

- Das Prüfungsrating korreliert, wenig überraschend, positiv mit den geclusterten Präsenzstunden ( $\rho = .252$ ), der Anzahl der Prüfungsteilnehmer ( $\rho = .206$ ) und den in einer Klausur erreichbaren Punkten ( $\rho = .270$ ).
- Die globale Modulevaluation korreliert negativ mit dem Prüfungsrating ( $\rho = -.248$ ) und positiv mit den geclusterten Präsenzstunden ( $\rho = .167$ ).
- Die globale Modulevaluation korreliert positiv mit dem Prozentsatz, indem die Prüfungen ( $\rho = .395$ ), die Lehrkräfte ( $\rho = .827$ ), die Veranstaltungsinhalte ( $\rho = .742$ ), die Lehr-Lernmaterialien ( $\rho = .629$ ), der Patientenbezug ( $\rho = .337$ ) und die Modulorganisation ( $\rho = .602$ ) als Stärke des Moduls genannt werden. Die Korrelationen mit den Prozentsatz, in dem diese Dimensionen als verbesserungswürdig bezeichnet werden, sind in ähnlichem Ausmaß negativ signifikant.

## Quellen