Lehre und Prüfung optimieren durch Analyse von Qualitätsindikatoren bei eKlausuren

Dr. Julia Steinhoff-Wagner
Landwirtschaftliche Fakultät
Institut für Tierwissenschaften
Präventives Gesundheitsmanagement

Veranstaltung: Tiergesundheit und Verfahrenstechnik (BA-A-3/07)

Zielgruppe: Bachelor-Studierende im 5. Semester

Gruppengröße: ca. 90 Teilnehmer

Zunehmend mehr Klausuren werden von einer papierbasierten Version auf eine rechnergestützte eKlausur umgestellt. Dabei verändern sich die Aufgaben und Möglichkeiten für die Lehrenden gravierend. Welche Qualitätsindikatoren in Bezug auf eKlausur-Daten geeignet sind, um sowohl die Lehre im Vorfeld zu verbessern, als auch die Prüfung selbst zu reflektieren, wurde aus drei Perspektiven beleuchtet: 1. Literatursuche nach Qualitätsindikatoren, 2. Empirische Untersuchung bei Bonner Lehrenden und 3. Berechnung von Qualitätsindikatoren anhand eines Beispieldatensatzes. Es stehen Lehrenden sehr aussagekräftige Qualitätsindikatoren für eKlausuraufgaben zur Verfügung, wohingegen die Aussagekraft der gesamten Klausur sehr abhängig von der Kohorte ist. Rückschlüsse zur Lehre aus den eKlausur-Daten abzuleiten, ist sehr zeitintensiv und basiert stark auf Annahmen und Erfahrungen.

Methodische Vorgehensweise

Ausgewählte Ergebnisse

Tabelle 1: Übersicht über Klausur- und Aufgaben-spezifische Qualitätsindikatoren aus der Literatur-recherche eingegrenzt durch Anwendung im Beispieldatensatz und Nennung anderer Lehrender aus der empirischen Untersuchung

Qualitäts-indikator	Definition/ Berechnung	Grenzwerte/ Zielgrößen	Bewertung aus Literaturquellen	Handlungsempfehlungen	
		Klausur-spezifisch		Prüfung	Lehre
Objektivität		Reduzierung von Personen und Situationseinflüssen (Zeit, Ort, …)	Subjektivität in der Bewertung bei handschriftlichen und mündlichen Prüfungen [2, 7, 10, 12], Änderungen bei der Auswahl von Antworten bei eKlausuren nicht nachvollziehbar [10], Zeit automatisiert [10], Abschreiben schwieriger [10], Übertragungsfehler geringer [10]	eKlausur beste Prüfungsform wegen hoher Standardisierung	
Reliabilität	Zuverlässigkeit: getestete Note entspricht der tatsächlichen Leistung; Korrelation zwischen Test 1 und Test 2	Korrelation x --> 1 ohne Messfehler [2], Cronbachs alpha > 0.8 minimum, >0.9 besser [2, 4]	Wiederholbarkeit der Prüfungsergebnisse lässt sich schwer bestimmen durch Lerneffekte [4, 6, 7], Zuverlässigkeit steigt mit Anzahl der Aufgaben und Prüfungslänge [6, 7]	Viele Aufgaben verwenden, entsprechend lange Prüfungen durchführen	Lernziele genau definieren
Überein-stimmungs-validität	Korrelation zwischen Pretest/Prognose und Prüfung	>0.9 sehr gut [7]	Übungseffekte überlagern Ergebnis [6, 12], PC-Affinität selektiert bestimmte Studierende [3, 10, 12]	Verschiedene Prüfungsformen beibehalten	Pretests regelmäßig zur Vorbereitung in die Lehre einbinden
Inhalts-validität	Repräsentativität der Inhalte	Steigt mit Übereinstimmung von Inhalt aus Vorlesung mit Prüfungen	Kontrolle der Lernzielorientierung [7]	Regelmäßig Inhalte der Vorlesungen mit denen in der Prüfung abgleichen	
Notenver-teilung	Mittelwert (Durchschnittsnote), Median, Spannweite, Standardabweichung, Variationskoeffizient, relative Häufigkeit	Mittelwert und Median: 2-3; Spannweite = schlechteste Note - beste Note: max. 4; Standardabweichung: 1-1.5; Varitionskoeffizient: 0.4-0.6	Verteilung kohortenabhängig [5, 7], Masternoten besser als Bachelornoten [5, 7, 8], vorgegebene Verteilung ist nicht geeignet [13], Angabe der Kennzahlen schafft Transparenz [1, 8], in einem hoch-standardisierten Prüfungs- und Bewertungsverfahren sind die Noten schlechter [1], durch größere Punkteintervalle bei Alles-oder-Nichts Aufgaben wird die Verteilung gestaucht [13]	Verteilung prüfen, aber nicht Noten danach ausrichten, möglichst hohe Gesamtpunktzahl wählen, bei geringen Punktzahlen für Alles-oder-Nichts Aufgaben	Allgemeine Statistik zur Klausur den Studierenden zugängig machen
	Grenzen: Sockelbetrag und Durchfallquote		Empfohlen sind 20-25% Zusatzpunkte [7], Berücksichtigung von 8% Schusselfehlern [13], Einbeziehen der Ratewahrscheinlichkeit [13]	Hinweis an die Studierenden auch bei Nichtwissen eine Antwort zu wählen	
Klausur-Peer Review		Ausschluss inhaltlicher und formaler Fehler	Wichtigste Aufgabe des Zweitprüfers [2, 11], sehr zeitintensiv [2, 6, 11]	Sorgfältige Vorbereitung der Klausur in Vorfeld durch mehrere Personen, Dokumentation der Änderungen	
Bearbeitungs-statistik	Bearbeitungszeit, Aufgabenreihenfolge		Reihenfolge und Positionseffekte der Aufgaben innerhalb der Klausur [2], Bearbeitungszeit mit Prüfungslänge vergleichen, jedoch abhängig von den gespeicherten Daten	Anpassung der Anzahl Aufgaben, beeinflusst Auswahl der Aufgaben und eventuell deren Formulierung	Rückschlüsse auf Lehre möglich
		Aufgaben-spezifisch			
Empirischer Schwierig-keitsgrad bzw. Aufgaben-schwierigkeit	Relative Häufigkeit richtiger Antworten	0.4-0.9 [11]; 0.2-0.8 [13]; 0.4-0.8 [4]	Aufgabenschwierigkeit ist abhängig von der Kohorte [6, 7, 11], zu leichte Aufgaben sind unökonomisch, aber eventuell sinnvoll [4], zu viele schwere Aufgaben können demotivierend sein [4], normalverteilte Schwierigkeiten über die Klausur [7, 13], Zeitknappheit führt zu Problemen bei der Interpretation [4], Begriff Schwierigkeit ungünstig, weil unabhängig von Kompetenzen [6]	Einstufung als leichte oder schwere Aufgabe alleine noch kein Grund für Änderungen	Schwere Inhalte in der Lehre als Beispielaufgaben einbauen
	Differenz: Erwarteter Schwierigkeitsgrad zum tatsächlichen Schwierigkeitsgrad	0: gut, groß: Fehleinschätzung [4]		Hinweise zur Überarbeitung der Aufgabe	Hinweise zur Überarbeitung der Lehre
Trenn-schärfe	Korrelation zwischen Aufgabenpunkten und Gesamtpunktzahl (Gesamtnote); Lösungswahr-scheinlichkeit in Abhängigkeit der Gesamtleistung	< 0.3 gering; 0.3-0.5 mittel; >0.5 hoch [6]; > 0.3 gut, 0.2-0.2 akzeptabel, 0.1-0.2 schlecht, <0.1 sehr schlecht [4], Achtung Korrelation mit der Gesamtnote ist negativ [13]	Bei geringer Aufgabenzahl gibt es einen starken Einfluss der einzelnen Aufgabe auf das Gesamtergebnis [4], Abbrecher haben ganz niedrige Punkte, gehören aber nicht zwangsläufig zu den schwachen Studierenden [4] Alles-oder-Nichts Aufgaben haben eine sehr schlechte Trennschärfe [7, 11, 13], falls geringe oder umgekehrte Korrelation zur erwarteten: gute haben Probleme, wo schwache keine sehen [7, 8]	Bei wenig Aufgaben (<20) Aufgabenpunkte subtrahieren von Gesamtpunktzahl; Test auf Ausreißer und diese eliminieren; möglichst Teilpunkte vergeben, Überarbeitung der Aufgabe	
Distraktoren-analyse	Häufigkeit der gewählten falschen Antwortoptionen	0: unrealistische Antwort; 0.05 - 0.2 gute Distrktoren [13]	Hohe Bedeutung guter Distraktoren [4, 6, 7, 8, 11, 12], je mehr Distraktoren, desto geringer die Ratewahrschein-lichkeit, aber desto höher die Kontexthinweise [11], über 3 Distraktoren offentsichtlich falsche nehmen zu, Lesezeit verlängert, unabh. von der Trennschärfe [6]	Überarbeitung besonders häufig oder selten gewählten Distraktoren	Missverständnisse ausräumen bei häufig gewählten Distraktoren

Fazit

eKlausuren stellen eine besonders objektive Prüfungsform dar. Auf die rechnergestützte Prüfungsform stellen überwiegend Lehrende mit viel Lehrerfahrung um. Obwohl Korrekturzeit eingespart werden kann, kostet die Erstellung und Kontrolle (Peer Review) der eKlausur viel Zeit und Koordinationsaufwand. Der digitale Ergebnis-datensatz bietet viele Möglichkeiten zur kontinuierlichen Verbesserung von Lehre und Prüfungen (siehe Tabelle 1). Um dieses Potential jedoch ausschöpfen zu können, sollten die Qualitätsindikatoren zentral berechnet und zur Verfügung gestellt werden mit gleichzeitigem Schulungsangebot zur Interpretation der Analyseergebnisse für die Lehrenden.

Was kann/ muss getan werden?

- Automatisierte Auswertung ergänzen um konkretere Bearbeitungsstatistik und teil-nehmerspezifische Auswertungen.
- Möglichkeiten zur Eingabe der Ratewahrschein-lichkeit und eines erwarteten Schwierigkeits-grades schaffen, die mit dem tatsächlichen Schwierigkeitsgrad abgeglichen werden kön-nen.
- Schulungsangebote und individuelle Beratung erhalten bzw. ausbauen.
- Personelle Kontinuität sicherstellen, weil sonst die investierte Zeit für Einarbeitung und Erstellung nicht gerechtfertigt ist.
- Anpassung der Prüfungsordnungen auf tech-nisch machbare eKlausuren.
- Lernzielorientierte Lehre fördern inklusive Aufklärung, was Kompetenzen bedeuten.

Quellenangaben

[1] Achen, A.C., Courant P.N.: What Are Grades Made Of? J. Econ Perspect. (2009).

[2] Epping C.: E-Prüfungen: Gestaltungsempfehlungen für elektronische Prüfungen an Hochschulen. Hamburg, Diplomica Verlag GmbH (2012).

[3] Huber, L.: 'Kompetenzen' prüfen?. In: Dany, S., Szczyrba, B., Wildt, J.: Blickpunkt Hochschuldidaktik Band 118. Arbeitsgemeinschaft für Hochschuldidaktik. Bertelsmann Verlag GmbH & Co KG Bielefeld (2008).

[4] Möltner, A., Schellberg, D., Jünger, J.: Grundlegende quantitative Analysen medizinscher Prüfungen. GMS Zeitschrift für medizinische Ausbildung, 23(3):Doc53 (2006).

[5] Müller-Benedict V., Grözinger G.: Noten an Deutschlands Hochschulen. Springer Fachmedien Wiesbaden GmbH (2017).

[6] Nouns, Z.M. Brauns, K.: Das Prinzip des Progress-Testing. In: Dany, S., Szczyrba, B., Wildt, J.: Blickpunkt Hochschuldidaktik Band 118. Arbeitsgemeinschaft für Hochschuldidaktik. Bertelsmann Verlag GmbH & Co KG Bielefeld (2008).

[7] Roloff, S.: Schriftliche Prüfungen stellen, auswerten- metho-disch, effektiv, objektiv. Hochschule Offenburg Karlsruhe (2012).

[8] Scheffer, S., Fröhmel, A., Georg, W.: Prüfungen praktischer Fertigkeiten: Performance-based Testing in der Medizin. In: Dany, S., Szczyrba, B., Wildt, J.: Blickpunkt Hochschuldidaktik Band 118. Arbeitsgemeinschaft für Hochschuldidaktik. Bertelsmann Verlag GmbH & Co KG Bielefeld (2008).

[9] Spitta, T., Mordau, J.: Entwicklung und Ergebnisse eines allgemeingültigen Fachkonzeptes für die Prüfungsverwaltung an Hochschulen. Universität Bielefeld, Fakultät für Wirtschafts-wissenschaften (1995).

[10] Vogt, M., Schneider, S.: E-Klausuren an Hochschulen. 1. Auflage. Koordinationsstelle Multimedia, JLU Gießen (2009).

[11] Waldschmidt-Dietz, F.: E-Klausuren und Scanner-Klausuren. Ein Leitfaden zum Einsatz von der JLU Gießen (2015).

[12] Wannemacher, K.: Computergestützte Prüfungsverfahren. In: Breitner, M., Bruns, B., Lehner F.: Neue Trends in E-Learning: Aspekte der Betriebswirtschaftslehre und Informatik. Physika-Verlag Heidelberg (2007).

[13] Wollersheim, H.-W.: Skript Prüfungsdidaktik. (2017)

Abschlussveranstaltung der Projektphase WS 2017/18 - 06. Februar 2018
eTeaching-Zertifikat der Rheinischen Friedrich-Wilhelms-Universität Bonn