

Methodenbericht – Kalibrierungsstudie zum Projekt KAT-HS

Gegenstand Studie war die Kalibrierung eines Itempools zur Entwicklung einer computerbasierten, kriteriumsorientierten adaptiven Hochschulklausur zur Vorlesung „Statistik“ im Bachelor-Studiengang Erziehungswissenschaft an der FSU Jena. Primäres Ziel dabei war die Itemparameterschätzung (Rasch-Modell) in einem prototypischen Setting. Da die zu erwartende Zahl an Klausurteilnehmern am Standort Jena für die anzuwendenden statistischen Methoden etwas zu klein war, wurde Kontakt mit weiteren Dozentinnen und Dozenten des Fachs Statistik anderer Universitäten (Ruhr-Universität Bochum, Universität Ulm) aufgenommen, um die Stichprobengröße zu vergrößern. Die zwei zusätzlichen Standorte wurden gewählt, da ihre Lehre sich an der gleichen Grundlagenliteratur orientierte und die Vorlesungen vergleichbare Inhalte abdeckte.

Stichprobe

Die Stichproben umfassten N=69 Erstsemester des Fachs Erziehungswissenschaften an der FSU Jena und N=27 an der RU Bochum bzw. N=43 an der Universität Ulm Erstsemester des Fachs Psychologie. Weitere demographische Angaben wurden nicht erhoben.

Computerisierte Klausur

Die Klausur in Jena wurde dabei als benotete Präsenzklausur durchgeführt. Die Prüfung wurde in den PC-Pools der Universität durchgeführt und über den SaveExamBrowser (ETH Zürich, 2018) und ein eigens entwickeltes Softwarepaket (KAT-HS App, Fink, Spoden, Naumann & Frey, 2019) ausgeliefert. Die Klausuren in Ulm und Bochum wurden als unbenotete Online-Übungsklausur durchgeführt. Hierfür wurde ebenfalls das eigens entwickelte Softwarepaket genutzt. Damit der Lernstand der Studierenden möglichst hoch ist, wurde die Übungsklausur in Ulm und Bochum erst jeweils eine Woche vor dem eigentlichen Klausurtermin freigeschaltet. Um einen Anreiz für die Teilnahme zu schaffen erhielten die Teilnehmerinnen und Teilnehmer der Übungsklausur nach Beendigung Rückmeldung darüber, in welchen Teilbereichen der Klausur Sie noch Lernbedarf aufweisen. Allen Studierenden wurden ein Taschenrechner-Programm und eine Formelsammlung zur Verfügung gestellt. Insgesamt kamen 45 dichotom kodierte Items zum Einsatz (siehe Dokument Codebuch_calibration_200305.docx). Jede Klausur bestand aus sechs Itemclustern zu je fünf Items (= Testlänge von 30 Items), die in einem Balanced Latin Square Design über Clusterpositionen und –reihenfolgen ausbalanciert wurden. Daraus ergaben sich pro Standort sechs Testversionen (siehe Dokument Testdesign_calibration_200818.xlsx) die gleichmäßig über alle Testpersonen verteilt wurden. Die einzelnen Cluster wurden so zusammengesetzt, dass sie ungefähr die gleiche geschätzte Bearbeitungszeit aufwiesen, sowie ein möglichst breites Spektrum des Inhaltsbereiches der Vorlesung abdeckten. Um Präsenz- und Onlineklausur statistisch miteinander verbinden und Differential Item Functioning Analysen durchführen zu können, wurde eine überschneidende Itemmenge von 15 Items genutzt. Die restlichen 15 Items unterschieden sich zwischen Präsenz- und Onlineklausur. Die Bearbeitungszeit der Klausur betrug 90 Minuten. Um zur nächsten Aufgabe zu gelangen, mussten die Studentinnen und Studenten das vorhergehende Item beantworten (forced

choice). Itemreview war nicht erlaubt. Dieses Vorgehen führte dazu, dass alle Studierenden die Klausur vollständig bearbeiteten. Fehlende Werte ergaben sich ausschließlich aus dem Testheftdesign.

Literatur

ETH Zürich, Lehrentwicklung und -technologie (2018). Safe Exam Browser (Version 2.2) [Computer Software]. Zürich: ETH Zürich, Lehrentwicklung und -technologie (LET).

Fink, A., Spoden, C., Naumann, P. & Frey, A. (2019). KAT-HS APP (Beta-Version) [Computer Software]. Frankfurt am Main: Goethe-Universität Frankfurt. Verfügbar unter: <https://kat-hs.uni-frankfurt.de/materialien/software/>