

Reliabilität und Durchführbarkeit einer Ultraschallprüfung für Studierende in der Schweiz

Andri Vetsch^a, Christoph Berendonk^b, Roman Hari^c

^a Berner Institut für Hausarztmedizin (BIHAM), Universität Bern; ^b Institut für Medizinische Lehre (IML), Universität Bern; ^c Berner Institut für Hausarztmedizin (BIHAM), Universität Bern und Redaktor PHC

Im Rahmen des neuen «Basiskurses Sonografie» erhalten interessierte Studierenden die Möglichkeit, bereits während des Studiums eine SGUM-akkreditierte Grundausbildung in praktischer Sonografie zu absolvieren. Das praktische Lernen geschieht dabei ausschliesslich im Peer-Tutoring, weshalb der Erhalt des Zertifikats einer Überprüfung des Lernerfolgs bedarf. Dieser Artikel beschreibt die Entwicklung, Pilotierung und die Reliabilität dieser praktischen Ultraschallprüfung im Format einer «objektiv-strukturierten klinischen Prüfung».

Ausgangslage

Ultraschallkenntnisse werden für angehende Ärztinnen und Ärzte immer wichtiger, und die Ausbildung verschiebt sich entsprechend von der Weiter- in die Ausbildung. Unter der Leitung des Berner Instituts für Hausarztmedizin (BIHAM) hat die Schweizerische Gesellschaft für Ultraschall in Medizin (SGUM) einen neuen «Basiskurs Sonografie» geschaffen, der interessierten Studierenden eine Grundausbildung in praktischer Sonografie ermöglicht [1]. Hierbei agieren speziell ausgebildete Studierende als «Peer-Tutoren» für ihre Kommilitoninnen und Kommilitonen.

Inhaltlich orientiert sich der Basiskurs am Blueprint des «Grundkurs Abdomen», dem ersten von drei Kursen in der Weiterbildung zum Fähigkeitsausweis «Abdomen-Sonografie» der SGUM [2]. Der Kurs ist als «Blended Learning» konzipiert mit einem kostenlos zugänglichen E-Learning-Modul (5 h) und einer praktischen Ausbildung in 16 Stunden «Peer Tutoring», verteilt über mehrere Monate [1]. Da die ganze praktische Ausbildung somit ohne fachärztliche Supervision stattfindet, wurde die Kurs-Akkreditierung durch die SGUM als Äquivalent des «Grundkurs Abdomen» an das Bestehen einer praktischen Abschlussprüfung geknüpft.

Als Prüfungsform wurde das Format einer «objektiv-strukturierten klinischen Prüfung» (*Objective Structured Clinical Examination, OSCE*) im Sinne eines Postenlaufs mit mehreren praktischen Prüfungsposten gewählt. In diesem häufig eingesetzten Instrument zur Überprüfung von klinischen Fähigkeiten werden Teilnehmende an mehreren Posten durch jeweils unterschiedliche Examinatoren beurteilt [3]. Für die Erstellung des Ultraschall-OSCE wurden die Stationen aus einer Studie von Matthias Hofer in Deutschland übernommen [4], an die lokalen Gegebenheiten angepasst und zu einem Prüfungsparcours mit fünf Posten à 5 min zusammengestellt. Dieser Artikel beschreibt die Pilotierung und Auswertung dieser praktischen Ultraschallprüfung für den «Basiskurs Sonografie».

Methodik

Pilot

Die Durchführbarkeit und Reliabilität des neuen OSCE wurden im Rahmen eines Pilotversuchs näher betrachtet. Dieser konnte im Anschluss an einen konventionellen SGUM «Grundkurs Abdomen» im Februar 2019 durchgeführt werden.

Im Vorfeld der Prüfung wurde eine Examinatorenschulung durchgeführt.

Diese diente dem Kennenlernen des Prüfungsformats und der Erhöhung der Bewertungsübereinstimmung zwischen den Prüfern (Interrater-Reliabilität, IR). Die zukünftigen Prüfer bewerteten dabei eine auf Video

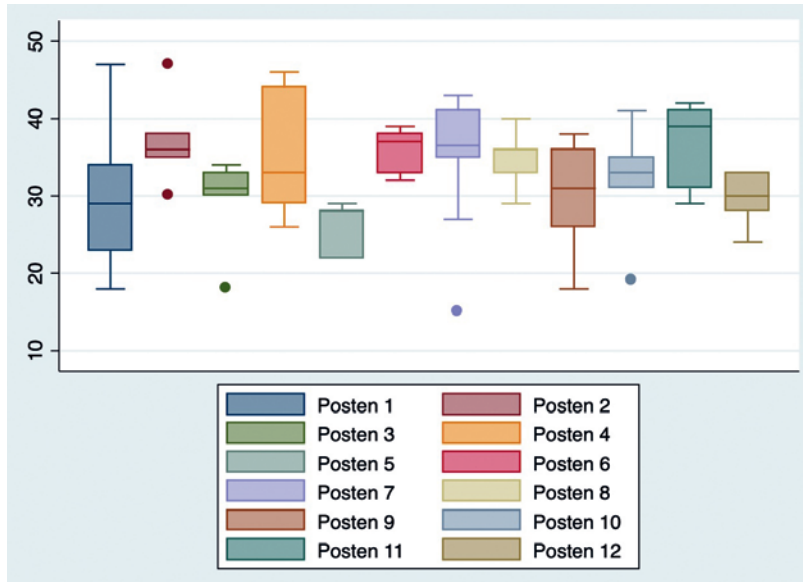


Abbildung 1: Mittlere Punktzahl der Teilnehmenden bei den 12 Posten (0–50 Punkte).

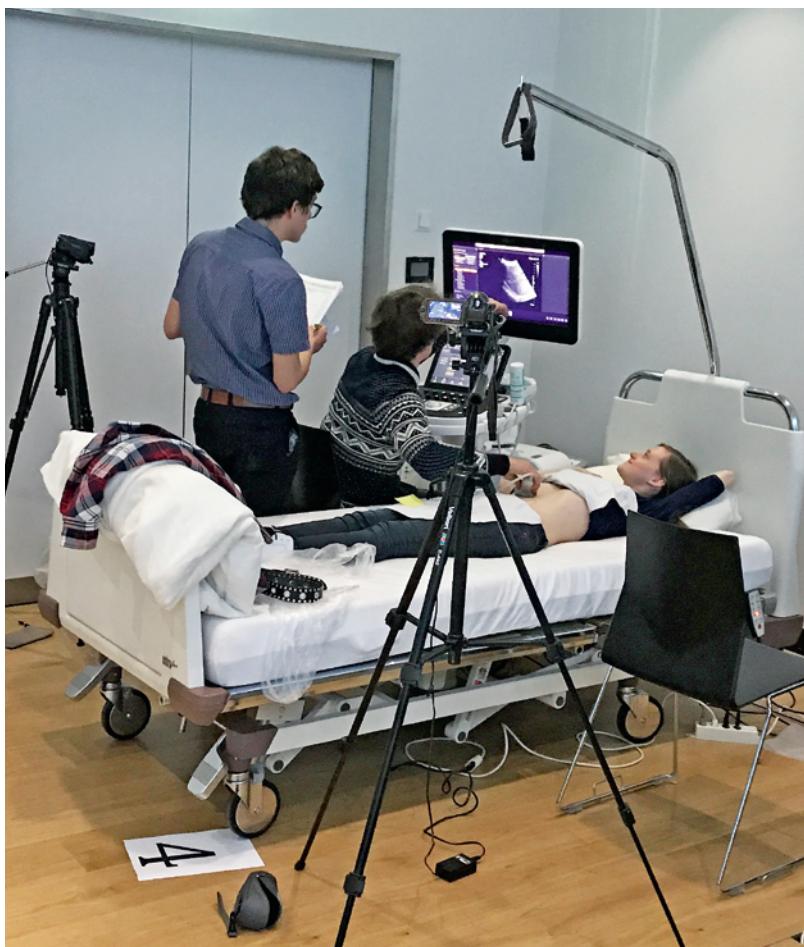


Abbildung 2: Szene aus dem Prüfungsstellenlauf (die Videoaufnahmen dienen später der Examinatorenschulung).

aufgezeichnete Ultraschalluntersuchung von Studierenden und diskutierten anschließend ihre Bewertungen untereinander. Anschließend bewerteten die Examinatoren unabhängig voneinander vier verschiedene Videoszenarien. Die Interrater-Reliabilität wurde mittels Intra-Klassen-Korrelation quantifiziert.

Der Prüfungs-Probelauf fand unmittelbar im Anschluss an einen zweieinhalbtägigen SGUM-Grundkurs Abdomen mit 15 Teilnehmenden statt. Jeder Absolvent bestritt dabei fünf Posten selber und diente daneben als «Modell» für die Prüfung seiner Kommilitonen. Die Auswahl und Reihenfolge der Posten wurde zufällig festgelegt.

Jeder einzelne Prüfungs-Posten (Abb. 2) war identisch gegliedert und hatte die gleichen Unterkapitel: *Schallkopf-Handhabung* und *Patientenführung* (40% der Punkte), *Bilderläuterung* und *Untersuchung* (25% der Punkte) und *Theorie* (20% der Punkte). Dazu kommt eine Bewertungsmöglichkeit für den *Gesamteindruck* (15% der Punkte).

Die zukünftigen Prüfer bewerteten eine aufgezeichnete Ultraschalluntersuchung und diskutierten ihre Bewertungen untereinander.

Die Ergebnisse aller Prüfungsstellen wurden digitalisiert. In der Analyse wurde die Schwierigkeit der einzelnen Posten und die Streubreite der Mittelwerte der Teilnehmenden berechnet. Durch die Berechnung des «Cronbachs Alpha» wurde die interne Konsistenz der Posten evaluiert.

Durchführbarkeit

Neben psychometrischer Aspekte stellten sich auch ganz praktische Fragen an den Pilot: Wie gut durchführbar ist diese neuartige Prüfung? Ist der Prüfungsablauf für die Teilnehmenden verständlich? Wie sicher fühlen sich Examinatoren bei der Bewertung? Durch direkte Befragungen von Prüfungsteilnehmenden und Examinatoren direkt im Anschluss an die Prüfung wurden diese Aspekte thematisiert.

Resultat

Schwierigkeit der Prüfung

15 Studierende absolvierten je fünf Posten à 5 min. Die Bestehensgrenze wurde bei 26/50 Punkten festgelegt – dies durch eine inhaltsbasierte Standardsetzung einer Expertengruppe der SGUM. Von den 15 Studierenden erreichten 14 (93%) die geforderte Mindestpunktzahl (min: 24,0, max: 39,8), der Mittelwert aller Teilnehmenden betrug 32,9 (SD 4,2).

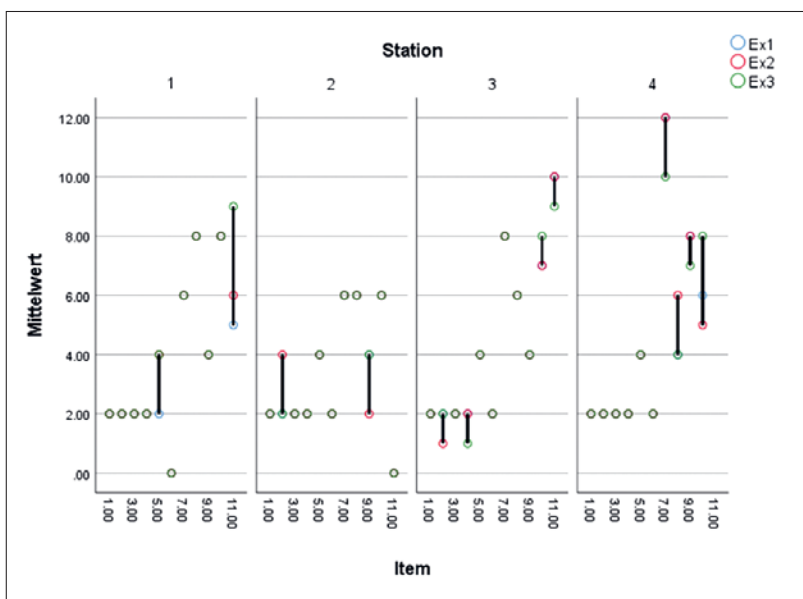


Abbildung 3: Bewertung der Stationen durch drei Examiner/-innen, elf Fragen pro Station, Diskordanzen als schwarze Striche eingezeichnet.

Schwierigkeit der Posten

Soweit aufgrund der tiefen Anzahl der Probanden beurteilbar, war die Schwierigkeit der einzelnen Posten in etwa vergleichbar, mit allerdings einer beträchtlichen Streuung der einzelnen Resultate (Abb. 1). Der Posten 5 (Leberdurchmusterung) wurde von den Examiner/-innen als etwas schwieriger empfunden, die erreichte Durchschnittspunktzahl lag bei diesem Posten mit 25,8 Punkte deutlich unter dem generellen Durchschnitt (32,9 Punkte). Auf eine statistische Varianzanalyse wurde aufgrund der tiefen Zahl der Messpunkte verzichtet.

Reliabilität

Reliabilität der Examinatoren

Zur Überprüfung, wie homogen die Examiner/-innen die Prüfungsleistungen bewerten, wurden vier Prüfungsleistungen unabhängig von drei Examiner/-innen bewertet. In Abbildung 3 sind von links nach rechts Resultate der vier Stationen in chronologischer Abfolge abgebildet, was die Entwicklung der Reliabilität im Verlauf des Trainings aufzeigt. Vielerorts stimmten alle drei Bewertungen ganz überein (nur ein einzelner Punkt), einzelne Abweichungen gab es vor allem im Theorieteil (jeweils Item «11.00»), diese sind als schwarze Striche erkennbar. Insgesamt wurden 43 Items von drei Ratern beurteilt, der Korrelationskoeffizient (ICC für ein zweifach zufälliges Modell) innerhalb einer Klasse erreichte dabei mit 0,95, einen sehr hohen Wert. Die Übereinstim-

mung zwischen den Examiner/-innen erwies sich von Beginn an als sehr gut. Das Klären der Differenzen nach jedem Posten führte zu keiner zusätzlichen Verbesserung der Übereinstimmung.

Einschränkend ist zu bemerken, dass es aufgrund der recht guten Leistungen in den simulierten Posten möglicherweise zu einem gewissen Ceiling-Effekt gekommen ist. Dies wird unterlegt durch die Beobachtung, dass die schwierigste Aufgabe die höchste Streuung aufwies.

Reliabilität des Messinstruments

Neben den Examiner/-innen sind auch Eigenschaften der Prüfungsposten für eine verlässliche und reproduzierbare – also reliable Prüfung wichtig.

Von den drei durchgeführten Rotationen wiesen die zweite und die dritte, mit einem Cronbachs Alpha von 0,66 und 0,72, eine gute bis sehr gute Reliabilität auf. In der ersten Runde war die Kennzahl mit 0,32 allerdings nicht zufriedenstellend. Eine zufällige Auswahl von fünf Posten führte also in dieser (kleinen) Stichprobe nicht zuverlässig zu konsistenten Prüfungen.

Durchführbarkeit

Die Durchführung der Prüfung verlief ohne grössere Zwischenfälle. Die Probanden empfanden die Prüfung als fair und inhaltlich adäquat – wenn auch als etwas hektisch. Zeitknappheit war das dominierende Thema in allen Feedbackrunden, sowohl für den Wechsel der Posten, als auch um die einzelnen Aufgaben zu bewältigen. Bei einzelnen Posten bestand das Problem, dass die Aufgabenstellung bei einigen Probanden aus anatomischen Gründen erschwert war – oder umgekehrt so einfach zu untersuchen, dass gewisse (punkterelevante) Anweisungen an den Patienten zur Optimierung des Bildes gar nicht nötig waren.

Die Probanden empfanden die Prüfung als fair und inhaltlich adäquat – wenn auch als etwas hektisch.

Die Examiner/-innen gaben an, von der Routine durch die Vorbereitung und dann auch im Verlauf der Prüfung sehr profitiert und im Bewertungsvorgang an Sicherheit gewonnen zu haben. Einzelne Unsicherheiten bezüglich Punktevorgabe blieben aber bestehen. Zusammenfassend kann der organisatorische Aspekt im Pilot als erfolgreich bezeichnet werden. Eine sorgfältige Examinatorenschulung mit der Möglichkeit, Fragen zu stellen, wurde als entscheidender Faktor für einen problemlosen Ablauf identifiziert.

Diskussion

Im Pilotversuch zeigte sich der Postenlauf mit fünf strukturierten Prüfungsposten als durchführbar, von adäquater Schwierigkeit mit ordentlichen Werten bezüglich der internen Konsistenz. Die diesbezüglich etwas inkonsistenten Resultate der drei durchgeführten Prüfungspostenläufe (je fünf Posten) mit einem Cronbachs Alpha zwischen 0,32–0,72 ist etwas unklar. Ein möglicher Grund ist die doch sehr tiefe Zahl der Probanden pro Prüfungspostenlauf, wodurch einzelne Ausreisser das Resultat stark beeinflussen können. In der ursprünglichen Arbeit von Hofer hatte sich bei fünf Stationen ein Cronbachs Alpha von 0,69 ergeben. Eine Erhöhung der Postenanzahl von fünf auf sechs Posten pro Prüfung wäre aber zu erwägen.

Die Auswertung der Examinatorenschulung zeigte eine ausgezeichnete Inter-Reliabilität bereits bei der gemeinsamen Bewertung des ersten Fallbeispiels. Dies ist mutmasslich der klaren Strukturierung der Prüfungsposten mit vorgegebenen Bewertungs-Checklisten geschuldet. Zu beachten hier ist sicherlich, dass die Examiner/-innen in unserem Pilot etwa gleich alt waren und dieselbe Examinatoren-Ausbildung zur selben Zeit durchliefen. Eine höhere Diversität der Examiner/-innen bezüglich Alter, Ausbildung und

Erfahrung dürfte hierfür eine gemeinsame Standardsetzung noch wichtiger machen.

Eine mögliche Einschränkung ist die geringe Stichprobenzahl im vorliegenden Pilot. Eine fortgesetzte Evaluation weiterer Prüfungsdurchgänge mit grösseren Stichproben ist hier sicher wünschenswert.

Von der praktischen Seite wurde klar, dass ausreichend Zeit für die Wechsel zwischen den Stationen sowie eine gute Selektion der Probanden (ggf. «Vor-Schallen» durch die Examiner/-innen) wichtig sind.

Abschliessend handelt es sich bei der vorgeschlagenen Prüfung um ein reliables und gut durchführbares Instrument, um eine objektive Beurteilung der praktischen Ultraschallfähigkeiten von Studierenden zu beurteilen und zu vergleichen.

Literatur

- 1 Raschle N, Hari R. Blended Learning Basic Course Sonography – A SGUM Accredited Ultrasound Course Based on Peer-Tutoring. Praxis. 2018;107(23):1255–9.
- 2 SIWF. Fähigkeitsprogramm Abdomen (SGUM) 2009 [Available from: https://www.fmh.ch/files/pdf19/sonographie_d.pdf.
- 3 Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). Medical education. 1979;13(1):41–54.
- 4 Hofer M, Kamper L, Sadlo M, Sievers K, Heussen N. Evaluation of an OSCE assessment tool for abdominal ultrasound courses. Ultraschall in der Medizin (Stuttgart, Germany: 1980). 2011;32(2):184–90.

Korrespondenz:
Dr. med. Roman Hari
Berner Institut für
Hausarztmedizin (BIHAM)
Gesellschaftsstrasse 49
CH-3012 Bern
[roman.hari\[at\]](mailto:roman.hari[at]biham.unibe.ch)
biham.unibe.ch