Ruprecht-Karls-Universität Heidelberg
Medizinische Fakultät Mannheim
Dissertations-Kurzfassung

Prüfungsqualität des mündlich-praktischen Teils des zweiten Abschnitts der Ärztlichen Prüfung, gemessen anhand der Interrater-Reliabilität und Durchführungsobjektivität

Autor: Elisabeth Narciß
Institut / Klinik: Universitäts-HNO-Klinik
Doktorvater: Professor Dr. med. Boris A. Stuck

Fragestellung: Nach der im Juni 2002 geänderten Approbationsordnung für Ärzte (ÄAppO) besteht die ärztliche Prüfung M2 seit Herbst 2006 neben dem M1 (dem früheren Physikum) nur noch aus einem dreitägigen schriftlichen und einem auf zwei Prüfungstage erweiterten mündlich-praktischen Teil nach Abschluss des Praktischen Jahres. Außer der Verdoppelung der Prüfungszeit sind hier verpflichtend praktische Prüfungsaufgaben sowie strukturierte Prüfungsfragen, die anhand von Fallvignetten das klinische Anwendungswissen der Prüflinge zum Gegenstand haben, eingeführt worden. Dem mündlich-praktischen Prüfungsteil kommt eine besondere Bedeutung sowohl für die Prüfer wie auch die Prüflinge zu. Ob sich durch die Umstrukturierung und die Erweiterung der mündlich-praktischen Prüfung und dem sich hieraus ergebenden gestiegenen personellen und organisatorischen Aufwand auch eine Verbesserung der Qualitätsmerkmale der mündlichen Prüfung ergibt, ist jedoch unklar. Deshalb standen drei Fragenkomplexe im Mittelpunkt dieser Arbeit: Zu einen war die Frage, ob alle Prüflinge dieselben Prüfungselemente absolvierten (Durchführungsobjektivität). Zum anderen waren die Wiederholgenauigkeit und Präzision der Prüferbewertungen zu klären, also die Frage, ob ein Prüfling in einer anderen Prüfungskommission mit anderen Prüfern dieselbe Note erhalten hätte (Reliabilität). Da die Reliabilität bei mündlichen Prüfungen mit mehreren voneinander unabhängig wertenden Prüfern v.a. von deren Übereinstimmung abhängt (Interrater-Reliabilität), fokussierte sich die Arbeit auf die Interrater-Reliabilität der Prüfungsergebnisse an beiden Prüfungstagen sowie die für die einzelnen Prüfungsteile (strukturierte Fragen, freie Fragen, Praxisaufgaben) vergebenen Noten.
Methodik: Die im Jahr 2008 von den Prüfern zurück gesandten standardisierten Prüfungsprotokolle der ersten repräsentativen Prüfungskohorte wurden ausgewertet. Neben den Noten für die einzelnen Prüfungsleistungen in den jeweiligen Fächern an den beiden Tagen wurde die Art der Prüfungsfragen (strukturierte versus freie Prüfungsfragen, Praxisaufgaben) erfasst. Für die einzelnen Prüfungsleistungen wurde eine bestimmte Form des Intraklassenkoeffizienten - der ICC1 unjustiert, einfaktoriell berechnet, der ein Maß für die Interrater-Reliabilität darstellt. Er kann Werte von 0 bis 1 annehmen. Werte ab 0,8-0,9 werden in der Literatur als Qualitätsmarge einer reliablen Prüfung angesehen.
Ergebnisse: Insgesamt konnten 352 Prüfungsprotokolle ausgewertet werden, dies entspricht einem Rücklauf von 42,7%. 89%-95% der Prüflinge erhielten wie gefordert an Tag 1 strukturierte Fragen, 79%-84% wurden an Tag 2 mit Praxisaufgaben geprüft. Damit ist die Durchführungsobjektivität bereits relativ hoch. Die Auswertung ergab eine Interrater-Reliabilität von 0,68 am ersten Prüfungstag und von 0,63 am zweiten Tag, die Interrater-Reliabilität für beide Prüfungstage zusammen genommen war 0,75. Die niedrigere Reliabilität am zweiten Tag erklärt sich durch den Einfluss der Praxisaufgaben, deren Interrater-Reliabilität bei isolierter Betrachtung lediglich bei 0,45 liegt. Überraschend war, dass die strukturierten Fragen nicht wie erwartet besser als die freien Fragen in Bezug auf den ICC1 abschnitten.
Schlussfolgerung: Die Reliabilität der mündlich-praktischen Prüfung insgesamt erreicht den in der prüfungstheoretischen Literatur geforderten Wert von 0,9 noch nicht. Die klare Definition eines Erwartungshorizontes und die Standardisierung der Durchführung der praktischen Aufgaben könnte hier die Interrater-Reliabilität erhöhen. Notwendig sind weiterhin gezieltes Prüfertraining und ein Review der Prüfungsfragen sowie das Sammeln von guten Prüfungsfragen in einer Datenbank. In Bezug auf die inhaltliche Validität der mündlich-praktischen Prüfung besteht weiterer Forschungsbedarf.