

Prüfung kommunikativer Kompetenzen im CoMeD-OSCE: Sekundäranalyse zum Einfluss von Ratermerkmalen auf die Interrater-Reliabilität

Pentzek, M.¹; Immecke, J.¹; Rotthoff, T.²; in der Schmitt, J.¹; Karger, A.³; Raski, B.³; Mortsiefer, A.¹

¹ Universität Düsseldorf, Medizinische Fakultät, Institut für Allgemeinmedizin

² Universität Düsseldorf, Medizinische Fakultät, Studiendekanat und Klinik für Endokrinologie, Diabetologie und Rheumatologie

³ Universitätsklinikum Düsseldorf, Klinisches Institut für Psychosomatische Medizin und Psychotherapie

Einleitung

Im Rahmen des Düsseldorfer CoMeD-Projekts wird die kommunikative Leistung von Studierenden in 4 OSCE-Stationen mit Schauspielpatienten geprüft. Die Bewertung erfolgt mit dem Berliner Global Rating (BGR), das aus den 4 Items „Empathie“, „Struktur“, „Verbaler Ausdruck“, „Nonverbaler Ausdruck“ besteht.

Studienfrage

Ist die Beurteilungsübereinstimmung zweier Rater abhängig von der Ähnlichkeit dieser beiden Rater?

Methoden

In 169 OSCE-Situationen (von 113 Studierenden) wurde die Leistung von jeweils 2 Ratern beurteilt (aus einem Pool von insgesamt 19 Ratern). Es fanden sich 16 Raterpaare mit unterschiedlicher Ähnlichkeit. Als Kriterien der Rater-Ähnlichkeit wurden 4 Merkmale ausgewählt: Geschlecht (gleich vs. gemischt); Erfahrung als OSCE-Rater; Teilnahme an der vorhergehenden Raterschulung (keiner der Rater vs. beide vs. unterschiedlich); Tätigkeit in der Patientenversorgung. Die Abweichungen in den BGR-Scores der Rater wurden verglichen und Intraklassenkorrelationen (ICC) berechnet.

Ergebnisse

Die Abweichungen der Rater unterscheiden sich nicht zwischen Paaren mit wenig, viel oder unterschiedlicher Rater-Erfahrung; auch der Besuch der vorhergehenden Raterschulung zeigt keinen Effekt (jeweils Kruskal-Wallis H-Test, $p > 0,10$). Gleichgeschlechtliche Raterpaare stimmen besser überein als gemischtgeschlechtliche (Mann-Whitney U-Test, $p = 0,001$; ICC 0,76 vs. 0,41). Paare, bei denen beide Rater in der Patientenversorgung tätig sind, stimmen besser überein als Paare, bei denen nur einer oder keiner der Rater in der Patientenversorgung tätig ist (H-Test, $p < 0,001$; ICC 0,73 vs. 0,56 vs. 0,47).

Schlussfolgerungen

Diese explorative Analyse weist darauf hin, dass die Reliabilität eines Kommunikations-OSCE von den Merkmalen der eingesetzten Rater abhängt. Sowohl das Geschlecht als auch die klinische Tätigkeit scheinen die Bewertung zu beeinflussen. Da die Teilnahme an vorhergehenden Rater-Schulungen keinen Effekt auf die Rater-Übereinstimmung zeigte, ist dies Anlass zur Überprüfung des Schulungskonzepts. Mehr Aufschluss könnten größere Studien mit multivariaten Analysen und einer Cluster-Auswertung der Raterpaare geben.