

## Steigerung von Interrater-Reliabilitäten im Münsteraner Studierfähigkeitstest

Hertel-Waszak A, Schönfeld E, Brouwer B, Ahrens H, Kurpat R, Geldmacher T und Marschall B

Institut für Ausbildung und Studienangelegenheiten, Medizinische Fakultät, Westfälische Wilhelms-Universität Münster

### Hintergrund

Im Münsteraner Studierfähigkeitstest kommt neben anderen Verfahren ein Multipler Mini Aktionstest (MMA) zum Einsatz. Der Prozess der Entwicklung eines gemeinsamen Bewertungsstandards in der Jurorenschulung erfuhr zum Wintersemester 2016 hin eine Verbesserung. So wurde das Verhalten der Simulationsbewerberinnen und -bewerber in den praktischen Probedurchläufen der Schulung bis einschließlich Sommersemester 2016 dem Zufall überlassen. Zum Wintersemester 2016 dagegen wurden die Simulationsbewerberinnen und -bewerber auf konkretes und voneinander unterschiedliches Verhalten in den spezifischen Stationen geschult. Auf diese Weise sollten die Jurorinnen und Juroren auf eine große Bandbreite an möglichem Verhalten vorbereitet und Beobachtungsfehler wie Antworttendenzen und logische Fehler [1] weiter gemildert werden. Diese Weiterentwicklung der Schulung sollte mit Reliabilitätssteigerungen einhergehen.

### Methode

Beim MMA für die Zulassung zum Studium der Humanmedizin durchlaufen 160 Bewerberinnen und Bewerber zehn Stationen, die von jeweils zwei Jurorinnen und Juroren beobachtet und beurteilt werden. Dementsprechend besuchen je acht Jurorinnen und Juroren eine moderierte Schulung, die bei der Beurteilung der Bewerberinnen und Bewerber je vier Paare bilden. Im MMA zum Wintersemester 2016 studierten die Simulationsbewerberinnen und -bewerber der praktischen Durchläufe erstmalig anhand von Skripten spezifische Rollen ein. Für jede Station wurden vier Rollen geschrieben: gut, mittelgut, gut in einer Bewertungsdimension und schlecht in der anderen und umgekehrt. Da die Stationen einmalig von einem MMA zum nächsten inhaltlich gleich ausfielen, war ein direkter Vergleich der Stationen möglich.

### Ergebnisse

Als Maß für die Reliabilität wurde die Interrater-Reliabilität jeder Station in Form von Intraklassen-Korrelationen (ICCs) herangezogen. Gemäß dem Design des MMA wurden ICCs einfach, zufällig ermittelt. Wie Abbildung 1 zu entnehmen ist, waren bei fünf von neun Stationen Steigerungen der ICCs vom Sommer- zum Wintersemester 2016 zu verzeichnen. Cohen's q als Maß für die Unterschiedlichkeit zwischen Reliabilitäten [2] rangierte von .2 bis .63. Für eine Station ergab sich ein Effekt in umgekehrter Richtung (Cohen's q) und bei drei Stationen gab es keine Veränderung.

#### Vergleich Intraklassenkorrelationen (ICCs) Wintersemester 2016 und Sommersemester 2016

MMA-Stationen	ICC WS 2016	ICC SS 2016	Cohens q	Effekt
Station 1	.72	.54	.289	klein (WS > SS)
Station 2	.76	.35	.631	groß (WS > SS)
Station 3	.60	.57	.046	---
Station 4	.50	.55	.069	---
Station 5	.47	.62	.215	klein (WS < SS)
Station 6	.76	.64	.238	klein (WS > SS)
Station 7	.90	.79	.401	mittel (WS > SS)
Station 8	.76	.66	.203	klein (WS > SS)
Station 9	.77	.61	.311	mittel (WS > SS)
Station 10	.56	.55	.014	---

Abbildung 1: Cohen's q < .1: kein Effekt; .1 bis .3: kleiner Effekt; .3 bis .5: mittlerer Effekt; > .5: großer Effekt



Abbildung 2: Jeweils zwei Jurorinnen/Juroren beobachten dieselbe Station. Jede der zehn Stationen existiert vier Mal.

### Diskussion/Schlussfolgerung

Im MMA zum Sommersemester 2016 wurden überwiegend bereits zufriedenstellende Interrater-Reliabilitäten erzielt. Als systematische Änderung zum Wintersemester 2016 wurden geschulte Simulationsbewerberinnen und -bewerber in den Jurorenschulungen eingesetzt. Die Interrater-Reliabilitäten der meisten Stationen verzeichneten vom Sommer- zum Wintersemester eine nochmalige Steigerung. Auch wenn der Einfluss konfundierender Variablen (z.B. Wechsel von Stationspatinnen und -paten, veränderte Zusammensetzung der Jurorenschaft) nicht auszuschließen ist, könnten diese Effekte auf die aktive Optimierung der Schulung der Jurorinnen und Juroren mittels geschulter Simulationsbewerberinnen und -bewerber zurückgehen.

### Literatur:

[1] Schmidt-Atzert L, Amelang, M. Psychologische Diagnostik. 5. Auflage. Heidelberg: Springer; 2012.

[2] Lenhard W, Lenhard A. Berechnung von Effektstärken. Dettelbach: Psychometrika: DOI: 10.13140/RG.2.1.3478.4245; 2016 [zitiert Sept 2017]. Verfügbar unter:

<https://www.psychometrika.de/effektstaerke.html>.