



Sobre la confiabilidad de un examen clínico objetivo estructurado

On the reliability of an objective structured clinical examination

Jimmie Leppink*

Apreciada Editora:

Para empezar, querría darle las gracias por la existencia y el mantenimiento de esta gran revista, que cada cuatro meses presenta – con acceso abierto para todos – una variedad de artículos de alta calidad y sobre temas muy importantes, en un idioma que da acceso a más de 500 millones de personas en el mundo. Con cada número, es un placer leer los contenidos y reflexionar sobre ellos.

Los artículos en el último número (2022, Número 1) ayudan a recordar la interdependencia de la enseñanza, del aprendizaje y de la evaluación. Entre otras cosas, si están bien diseñadas, las actividades de evaluación pueden ser formas de enseñar y fomentar el aprendizaje. Según el tipo de la actividad y los objetivos de aprendizaje evaluados, se puede evaluar el rendimiento indi-

vidual,¹ del equipo y de las personas que forman parte del equipo² o una combinación de ambos.

En cualquier caso, las actividades de evaluación requieren recursos de un sistema sanitario y/o universitario que tiene que servir prioridades locales dentro de límites relativamente estrechos y, por lo tanto, una pregunta clave es *cuántos recursos se necesitan* para llevar a cabo una actividad de evaluación de una calidad que nos permita tomar decisiones importantes sobre el desarrollo de nuestros estudiantes, residentes o profesionales. El artículo de Machuca-Contreras y compañeros¹ en el primer número de 2022 en esta revista presenta un gran trabajo sobre esta pregunta clave. Su estudio investiga cuestiones de la validez y la confiabilidad, lo que ayuda a investigadores y formadores a comprender cuántas estaciones suelen requerirse en el tipo de contexto dado. Además, los autores presentan varias

Tabla 1: Correlaciones lineales entre valoraciones de tres criterios (a, b, c) en tres estaciones (1, 2, 3): los números marcados en amarillo son correlaciones dentro de la misma estación.

	s1a	s1b	s1c	s2a	s2b	s2c	s3a	s3b
s1a	—							
s1b	0.630	—						
s1c	0.590	0.582	—					
s2a	0.190	0.362	0.080	—				
s2b	0.299	0.356	0.121	0.548	—			
s2c	0.214	0.421	0.066	0.579	0.588	—		
s3a	0.192	0.389	0.156	0.391	0.271	0.032	—	
s3b	0.268	0.270	0.400	0.218	0.188	0.014	0.521	—
s3c	0.295	0.369	0.205	0.240	0.380	0.148	0.600	0.590

* Hospital Virtual
Valdecilla, Santander,
Cantabria, España.

doi: 10.35366/107393

Citar como: Leppink J. Sobre la confiabilidad de un examen clínico objetivo estructurado. Rev Latinoam Simul Clin. 2022; 4 (2): 76-78. <https://dx.doi.org/10.35366/107393>



herramientas para investigar dichas cuestiones e informan de manera muy transparente sobre las dificultades que suelen encontrarse en este tipo de estudio, como criterios que no funcionan como lo esperado o métodos estadísticos que no dan soluciones como nos gustaría.

Esto me lleva a una idea que forma parte de la convención en la práctica internacional de este tipo de estudios y unas posibles soluciones para poder manejar algunas de las dificultades encontradas en este tipo de situación en todo el mundo. La idea sobre la que me gustaría reflexionar es que para un estudio de este tipo se necesita una mues-

tra grande – por ejemplo 100 estudiantes – para que se pueda utilizar análisis factorial y alfa de Cronbach para investigar los criterios y determinar cuántas estaciones se necesitan para llegar a una confiabilidad suficientemente alta (por ejemplo 0.7 o 0.8). Aunque este enfoque sigue siendo parte de la convención internacional, tiende a resultar en una subestimación de la confiabilidad y, en consecuencia, una sobreestimación de los recursos requeridos, porque tanto análisis factorial como alfa de Cronbach requieren suposiciones sobre las variables de interés que en la práctica no necesariamente son adecuadas.^{3,4} Sobre todo, alfa de Cronbach, un coeficiente introducido en 1951,⁵ es muy restrictivo y, por lo tanto, muchos autores – incluso el mismo autor Cronbach⁶ – han recomendado considerar alternativas disponibles también, como por ejemplo el omega de McDonald,⁷ y presentar unas de estas alternativas o una alternativa en combinación con alfa de Cronbach. Por ejemplo, se puede presentar alfa y omega, y si los datos no desvían mucho de las suposiciones restrictivas del alfa de Cronbach, alfa y omega suelen dar más o menos el mismo resultado.³ Además, es recomendable utilizar múltiples criterios por estación en un examen clínico objetivo estructurado (EEOE), como por ejemplo en el estudio de Machuca-Contreras y compañeros,¹ que requieren una valoración categórica o cuantitativa. Presento un ejemplo simulado de un ECOE de tres estaciones (s1, s2, s3) con tres criterios de evaluación (a, b, c) valorados en una escala de 1 (mínimo) a 5 (máximo) en cada estación, de un grupo de 20 estudiantes. No son datos reales, pero los datos de esta simulación se parecen a las características de ECOE que yo solía dirigir en un puesto de trabajo anterior. El software utilizado para el análisis de datos es el paquete de Fuente abierta (*Open Source*) JASP versión 0.16.2.⁸ La

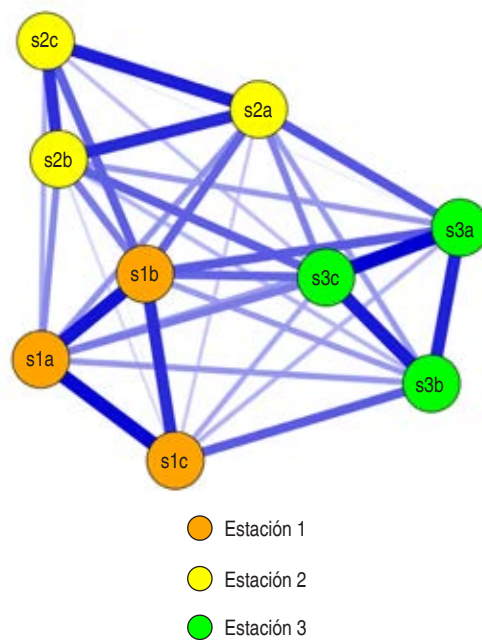


Figura 1: Red de correlaciones lineales entre valoraciones de tres criterios en tres estaciones: líneas más gordas representan correlaciones más altas.

Tabla 2: La confiabilidad en función de los criterios incluidos.		
Criterios	Omega de McDonald	Alfa de Cronbach
s1a, s2a y s3a	0.535	0.503
s1b, s2b y s3b	0.573	0.530
s1c, s2c y s3c	0.334	0.306
las sumas de s1, s2 y s3	0.640	0.636
1a, 1b y 1c	0.817	0.797
2a, 2b y 2c	0.800	0.800
3a, 3b y 3c	0.796	0.794
Todos los 9 criterios	0.816	0.812

Tabla 1 muestra las correlaciones entre los nueve criterios valorados en la serie de tres estaciones.

La *Figura 1* presenta una visualización de las correlaciones reportadas en la *Tabla 1* en la forma de análisis de red, un método que consta de una alternativa más flexible que el análisis factorial para este tipo de datos.⁴

El patrón en estos datos es común en ECOE con múltiples criterios por estación: hay una tendencia hacia una intercorrelación positiva y sobre todo entre criterios de la misma estación. La *Tabla 2* presenta el omega de McDonald y el alfa de Cronbach para distintas combinaciones de criterios.

La *Tabla 2* indica que es posible llegar a una confiabilidad de 0.8 con una estación con tres criterios, pero no se llegaría ni a 0.6 con tres estaciones con un criterio valorado por estación y tampoco se quedaría por debajo de 0.7 si se utilizase la suma de los tres criterios de cada estación. Esto es importante porque en no pocos sitios todavía se utiliza ECOE con un criterio por estación o se estima la confiabilidad utilizando las sumas de criterios por estación en vez de los criterios como tal, y una consecuencia de esta práctica son las estimaciones de confiabilidad demasiado pesimistas resultando en estimaciones de recursos necesarios demasiado altas.^{3,4} Especialmente donde se trabaja con un criterio por estación suele haber diferencias entre omega y alfa con omega normalmente indicando un valor más alto y realista.

Otra posible consecuencia del uso del alfa de Cronbach con datos que no cumplen los requisitos de este coeficiente es que se termina descartando criterios como si no tuvieran suficiente calidad, aunque sí funcionarían bien si se utilizaran coeficientes como omega u otros métodos que no requieren suposiciones tan restrictivas, resultando en la exclusión de criterios – y quizás estaciones – que no deberían ser excluidos.

Dicho esto, todos los métodos tienen sus ventajas y desventajas, y tienen en común que no son

perfectos. Por lo tanto, es aconsejable analizar los datos desde suposiciones distintas (más y menos restrictivas) y reportar los resultados obtenidos con métodos distintos, como por ejemplo omega y alfa cuando se trata de cuestiones de confiabilidad, o – si el tamaño de la muestra permite el uso de análisis factorial – utilizando tanto análisis factorial como algún tipo de análisis de red como se presenta en la *Figura 1*.

REFERENCIAS

1. Machuca-Contreras F, Maldonado-Holtheuer M, Villanueva-Quezada C. Diseño y evaluación psicométrica de un examen clínico objetivo estructurado de primeros auxilios. *Rev Latinoam Simul Clin.* 2022; 1: 3-10. Disponible en: <https://doi.org/10.35366/104949>
2. García-Méndez JA, Díaz-Agea JL, Leal-Costa C, Jiménez-Rodríguez D, Rojo-Rojo A, Fenzi G, Pujalte-Jesús MJ. Simulación clínica 3.0: El futuro de la simulación: el factor grupal. Disponible en: <https://doi.org/10.35366/104953>
3. Leppink J. How we underestimate reliability and overestimate resources needed: revisiting our psychometric practice. *Health Prof Educ.* 2019; 5 (2): 91-92. <https://doi.org/10.1016/j.hpe.2019.05.003>
4. Leppink J. The art of modelling the learning process: uniting educational research and practice. Cham, Switzerland: Springer; 2020. Available in: <https://doi.org/10.1007/978-3-030-43082-5>
5. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334. Available in: <https://doi.org/10.1007/BF02310555>
6. Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. *Educ Psychol Meas.* 2004; 64: 391-418. Available in: <https://doi.org/10.1177/0013164404266386>
7. McDonald RP. Test theory: a unified Treatment, Mahwah, N.J.: L. Erlbaum Associates, 1999.
8. JASP Team JASP (Version 0.16.2) [Computer software]. (2022). Available in: <https://jasp-stats.org/>

Correspondencia:

Dr. Jimmie Leppink

E-mail: jleppink@hvvaldecilla.es

www.medigraphic.org.mx