

## Metodología Docente: ¿Check list o Rating list?

**Resumen:** Se resaltan las indicaciones, ventajas y desventajas así como algunas de sus propiedades psicométricas de las escalas de codificación y las evaluativas.



En la evaluación de una determinada habilidad clínica (tarea, actividad, conducta,?) mediante la observación, bien sea por un tutor clínico o por un paciente simulado o por un no-experto entrenado requiere la elección de un formato de recogida de la información válido y fiable. Habitualmente la elección del tipo de formato recae entre el uso del 'listado de comprobación' o 'codificación' o 'checklist' versus la 'escala evaluativa' o 'rating list' y esta elección requiere de la consideración de algunos importantes aspectos que tienen que ver con las propiedades psicométricas de cada uno de estos dos instrumentos.

Listados de comprobación-codificación-(checklist): Este tipo de formato **Codifica** (Coding), es decir clasifica las diferencias sin medirlas. El observador indica aquí si **una categoría o acción determinada se completa o no** durante la interacción (?se va la las manos?; ?se presenta al paciente? o ?toma la tensión arterial con el manguito adecuado?). El espectro de respuestas es de: SI/NO/NP, o también de ?hecho/no hecho?. Los efectos que el uso de este tipo de formatos tiene en el aprendizaje es que promueven la trivialización (los estudiantes tienden a memorizar los checklists) y la priorización de lo concreto. El uso de esto por parte de los observadores no expertos como por ejemplo un paciente simulado adecuadamente entrenado es muy fiable.

Escala-evaluación-(rating list): Este tipo de formato **Valora** (Rating) es decir, mide la conducta sobre un continuum. El observador aquí **Juzga** (holístico) o cuantifica la calidad de una conducta o habilidad (requiere de su interpretación). Por ejemplo, las habilidades interpersonales pueden ser evaluadas en un ítem con una escala tipo LIKERT que exprese el rango de la calidad de la actuación. Aunque probablemente existe un continuum subyacente que va de lo más holístico a lo más analítico (atomístico), los checklist tienden a captar si una acción tiene lugar o no mientras que la escala evaluativa requiere la interpretación de acciones. Los efectos que el uso de este tipo de formatos evaluativos tiene en el aprendizaje es que promueven un aprendizaje más para la comprensión y la integración del conocimiento y habilidades.

La investigación sobre el uso de checklist vs rating list en el contexto de las pruebas ECOE nos ofrece interesantes resultados: En términos de ?reproducibilidad? de las puntuaciones, las escalas evaluativas globales son tan buenas como los checklist. Un resultado aparentemente paradójico es que las rating list en las ECOE tienen más fiabilidad inter-estación pero una más baja fiabilidad inter-observador que los checklist. Aunque aparentemente una mayor intersubjetividad debería existir en los juicios holísticos, este tipo de juicios parece que capta habilidades que son generalizables de una estación a otra. Los resultados del uso de este tipo de escalas por los PS y observadores cualificados es sin embargo similar. Esto quiere decir que: recoger un amplio número de juicios subjetivos de diferentes evaluadores en diferentes contextos lleva a resultados más fiables.

¿Se deriva de aquí que hay que recomendar el uso de rating frente al coding? La respuesta es NO. El uso de uno u otro formato depende del objetivo de la evaluación: En situaciones donde la pericia (dominar una técnica o habilidad específica; p.e. exploración física, resucitación CP, intubación,?) es importante, los checklists son más apropiados. En niveles iniciales (juniors) del aprendizaje. La actuación holística (ligada al concepto de competencia), generalmente requerida en niveles más seniors de aprendizaje, se

recomienda el uso de escalas.

**Algunos conceptos útiles:**

Precisión (exactitud) (en inglés accuracy) de un sistema de medida, es el grado en el que dicha medida se acerca a la cantidad del valor verdadero. Se relaciona con la ?reproducibilidad?, que es el grado en el que medidas repetidas en las mismas circunstancias dan los mismos resultados (también ?fiabilidad? ?repetibilidad?).

Referencias:

**Van der Vleuten CPM, Swanson DB.** Assessment of clinical skills with standardized patients: state of the art revisited. Training and learning in Medicine 2013;25(S1), S17-S25

**Van Luijk SJ, Van der Vleuten CPM.** A comparison of checklists and rating scales in performance-based testing. In Current developments in assessing clinical competence (pp. 357?82). Montreal: Can-Heal, 1992.

**Cunnington JPW, Neville AJ, Norman GR.** The risk of thoroughness: reliability and validity of global ratings in checklists in an OSCE. Advances in Health Sciences Education 1996;1:227?33.

**Regehr G, MacRae H, Reznick R, Szalay D.** Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Academic Medicine 1998;73:993?7.

Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. Academic Medicine 1999;74:1129?34.