

NUEVOS INSTRUMENTOS DE EVALUACIÓN OBJETIVA DE COMPETENCIAS TÉCNICAS EN CIRUGÍA

NEW TOOLS FOR OBJECTIVE ASSESSMENT OF TECHNICAL SKILLS IN SURGERY.

Andrés C. Kasparian, Ricardo Chércoles

INTRODUCCIÓN

La adquisición de habilidades quirúrgicas constituye un factor central en la formación de todo cirujano. No hay duda que la destreza técnica es considerada por el público en general y por los propios médicos en formación como uno de los factores más importantes de la lista de cualidades que debería poseer todo cirujano.

Una encuesta publicada en el año 2001 realizada con cirujanos del Reino Unido que alcanzaron el grado de “maestro”, reveló que más del 80% de los encuestados identificaron la excelencia en las habilidades psicomotoras como el factor más determinante de una carrera quirúrgica exitosa (6, 7). Otro reporte publicado en el mismo año proveniente de centros europeos y norteamericanos, basado en encuestas y reuniones de consenso con cirujanos “maestros” de diferentes especialidades, consideró a la destreza operatoria como el atributo de mayor importancia luego del juicio clínico (8, 35).

Sin embargo, dentro de todos los requisitos incluidos en la formación de un cirujano, la evaluación de las habilidades técnicas constituye el factor más débil y menos desarrollado. En la actualidad las herramientas de evaluación de los programas de residencias quirúrgicas hacen fuertemente hincapié en la adquisición de conocimientos y en la aplicación de los mismos en forma criteriosa, en tanto que los recursos para evaluar las competencias técnicas se caracterizan por presentar subjetividad, y falta de confiabilidad y validez (11, 29).

Este hecho guarda relación con la vigencia hasta la actualidad de un modelo clásico de formación en cirugía introducido por Sir William Halsted en el año 1898 en el John Hopkins Hospital de los Estados Unidos de Norteamérica. Dicho modelo, diseñado a semejanza del sistema alemán, estaba basado en una relación maestro-aprendiz y en

la adquisición de responsabilidades graduadas y progresivas. El entrenamiento estaba regido por el concepto de “ver y hacer”, es decir, el aprendiz trabajaba a la par de su maestro para asegurar la progresión de sus habilidades por imitación. La solvencia del aprendiz estaba determinada por el maestro de acuerdo a su desempeño. Esta evaluación estaba basada muchas veces en rasgos de la personalidad más que en criterios objetivos (16,24).

Sin embargo, los avances en las teorías de la educación, y las presiones crecientes en la práctica profesional han llevado a cuestionar seriamente la confianza en este sistema para la adquisición de habilidades técnicas. La creciente complejidad de los casos hospitalarios, y el mayor énfasis en mitigar el error médico por los problemas médico legales, ha modificado de manera significativa la posibilidad de los instructores para formar a los residentes. La limitación reglamentaria de la carga horaria para los residentes, y el interés en optimizar el uso del quirófano han determinado una menor exposición de los médicos en formación a los procedimientos con pacientes en la sala de operaciones.

Surge por estos motivos la necesidad de un nuevo modelo de educación en cirugía basado en un programa estructurado de formación, en el cual el entrenamiento de las técnicas operatorias comienza fuera del quirófano, en un laboratorio, y la adquisición y evaluación de competencias se realiza con instrumentos que pretenden ser lo más objetivos posibles (1,16, 24, 28).

COMPETENCIAS Y SU EVALUACIÓN

Kane (19) definió la competencia como: “... el grado de utilización de los conocimientos, las habilidades y el buen juicio asociados a la profesión, en todas las situaciones que se pueden confrontar en el ejercicio de la práctica profesio-

nal” es decir que la competencia representa la capacidad de un profesional de utilizar su buen juicio así como también sus conocimientos, habilidades, y actitudes para solucionar problemas complejos que se presentan en el ámbito de su actividad.

El concepto de competencia es una construcción conceptual compleja que se puede abordar de manera parcial a través de los diferentes aspectos mesurables que la constituyen: conocimientos, habilidades técnicas, actitudes, juicio, capacidad resolutoria frente a nuevos problemas, capacidad de comunicación, etc.

Hay muchos motivos por los cuales hay que evaluar las competencias tanto en estudiantes como en profesionales. Se pueden mencionar: 1) la evaluación como requisito integrante del proceso educativo 2) para certificar la competencia (evaluación sumativa) 3) para retroalimentar el proceso educativo (evaluación formativa) 4) para evaluar los propios programas de formación 5) para confirmar los valores y los estándares de la profesión y 6) por la responsabilidad e imputabilidad frente a la sociedad.

Miller (23) ha desarrollado un modelo de competencia profesional representado por una pirámide compuesta de varios niveles. En la base se encuentran los conocimientos (saber) sobre los que se apoya la competencia (saber como). A un nivel superior se encuentra el desempeño (mostrar como) y finalmente la acción en una práctica real (el hacer).

En 1996 Van Der Vleuten (34) propone un modelo teórico conceptual para definir la utilidad de una herramienta de evaluación de competencias profesionales que correlaciona distintas propiedades y que puede expresarse con la siguiente ecuación:

U: utilidad del instrumento; F: fidelidad o confiabilidad; V: validez; E: impacto educativo; A: aceptabilidad; C: costo; (p) contribución parcial. Modificado de Van Der Vleuten C. 1996. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1: 41-67

La confiabilidad puede definirse como la expresión cuantitativa de la reproducibilidad con la que un instrumento mide el mismo atributo. Hace re-

ferencia a la precisión del test. Es decir, en que medida el test genera los mismos resultados si se repite en dos sucesivas oportunidades bajo las mismas circunstancias, o sea, sin la modificación propia del aprendizaje entre las tomas (15, 30). Se representa con el coeficiente alfa de Cronbach. Cuando las evaluaciones son llevadas a cabo por más de un observador se aplica también el concepto de confiabilidad inter-evaluadores, que mide el grado de concordancia entre dos o más evaluadores que utilizan el mismo instrumento (25).

La validez es un concepto complejo que hace referencia a la capacidad del instrumento de evaluación de medir o no lo que se propone como objetivo. Puede desglosarse en distintos subconceptos: validez de constructo hace referencia a la medida en la cual el examen puede diferenciar distintos niveles de experiencia. Validez de contenido refiere si el objeto o campo de conocimiento que pretende ser evaluado es medido por el instrumento. Por ejemplo: si se pretende evaluar aspectos técnicos y se terminan evaluando conocimientos, el instrumento de evaluación carece de validez de contenido. Validez concurrente es el grado en el cual los resultados de la evaluación se correlacionan con el gold Standard para ese campo del conocimiento. Validez de superficie es el grado en el cual el instrumento tiene el aspecto de situaciones reales. Validez predictiva la habilidad del examen de predecir el desempeño a futuro (14, 23, 29, 31, 33).

Reznick (27) en el año 1993 reportaba cinco métodos disponibles para la evaluación de las competencias técnicas, con grados variables de confiabilidad y validez:

Los exámenes de certificación administrados por las instituciones formadoras o reguladoras al final de un período lectivo o luego de cumplir con el programa de formación, basan su evaluación en los conocimientos y las habilidades clínicas, y no evalúan la capacidad técnica del candidato (30).

Los registros de procedimientos (cuadernos de bitácora) efectuados a lo largo del proceso de formación son obligatorios en los programas de residencia de algunos lugares del mundo como el Reino Unido, y son requeridos al momento de las evaluaciones de fin de año o final de residencia, entrevistas de trabajo, etc. Sin embargo ha

sido demostrado que es un registro meramente indicativo de la ejecución de los procedimientos y no un reflejo de la habilidad operatoria, es decir carecen de validez de constructo.

El registro del tiempo empleado para la ejecución de los procedimientos tampoco evalúa la calidad de la técnica, y es un parámetro no confiable durante procedimientos reales debido a la influencia de numerosas circunstancias que pueden no ser previstas, y que tendrían poca relación con el desempeño del cirujano.

La observación directa tal como se realiza actualmente en el quirófano en casi todas las residencias de cirugía, es global y no está basada en criterios específicos, por lo tanto no resulta confiable. Como está influenciada por la subjetividad del observador tiene una pobre confiabilidad de reevaluación e interobservador. Además los procedimientos quirúrgicos son únicos y efímeros, por lo tanto no hay posibilidad de revisar el desempeño del candidato en caso de dudas, y esto aumenta aún más la subjetividad y la falta de confiabilidad.

Los datos de morbilidad y mortalidad con frecuencia son utilizados como marcadores del desempeño operatorio, sin embargo esta información está influenciada en gran medida por factores dependientes del paciente: variabilidad biológica, disparidad en la presentación de las enfermedades, etc. y por otros factores no relacionados con los pacientes: soporte institucional, acceso a los recursos, etc. Por lo tanto estas cifras carecen de confiabilidad, y no reflejan verdaderamente la competencia técnica.

Como consecuencia de los problemas de confiabilidad y subjetividad de los métodos mencionados previamente, se han propuesto y desarrollado métodos más objetivos de evaluación de habilidades técnicas, entre los que se pueden mencionar:

Las listas de cotejo y evaluaciones globales: la enumeración de criterios definidos para cotejar la ejecución o no de algún procedimiento hace que el proceso de evaluación sea más objetivo, válido y confiable. Se transformarían a los evaluadores en observadores más que en interpretadores de la conducta, tendiendo a minimizar de esta forma la subjetividad en el proceso de evaluación (26).

La amplia aceptación del examen clínico objetivo estructurado (ECO) llevó a un gru-

po de Toronto a desarrollar un concepto similar aplicado a la evaluación de habilidades técnicas (OSATS), que consiste en 6 estaciones donde los candidatos desarrollan procedimientos quirúrgicos estandarizados en modelos animales o inanimados, durante un período limitado de tiempo. El desempeño es evaluado con una lista de cotejo específica de la consigna a desarrollar, y una evaluación del desempeño global que consiste en 7 componentes genéricos de destreza operatoria que son cuantificados con una escala de 5 puntos de Likert, con descriptores explícitos para los puntos extremos y el medio (22, 29). El inconveniente de este tipo de evaluaciones es la inversión en recursos humanos y el tiempo que debe dedicarse en el análisis de las grabaciones u observación directa (26).

Los sistemas de análisis de destreza manual como el ICSAD (Imperial Collage Surgical Assessment Device) es un sistema de rastreo, disponible comercialmente, que consiste en un generador de campo electromagnético y dos sensores que son colocados en las manos del cirujano. Hay un procesador de datos que utiliza un software para medir el número y velocidad de movimientos, la distancia recorrida por las manos y el tiempo para ejecutar la tarea. Diversos estudios han demostrado la validez de constructo de este sistema (9).

Los sistemas de realidad virtual pueden definirse como un conjunto de tecnologías que permiten a un individuo interactuar eficientemente con bases de datos computarizadas en tres dimensiones y en tiempo real, usando sus órganos de los sentidos naturales y habilidades. Se puede mencionar el MIST-VR (Minimally Invasive Surgical Trainer – Virtual Reality) reproduce una serie de imágenes tridimensionales que pueden ser manipuladas, fue ideado como entrenamiento para las colecistectomías laparoscópicas. Los datos que surgen del uso de este sistema se pueden medir y evaluar, y MIST-VR ha sido validado ampliamente para la evaluación de habilidades básicas en cirugía laparoscópica (25).

El análisis de los productos finales: los resultados de la cirugía y la posterior evolución del paciente no se pueden atribuir a cuestiones técnicas exclusivamente. De hecho las consecuencias de una mala técnica pueden ponerse en evidencia muchos años después. Szalay y col. (32) propu-

sieron la evaluación del producto final de los modelos inanimados que constituyen las diferentes estaciones del OSATS. Sus estudios demostraron validez de constructo y una muy buena correlación con la evaluación del OSATS en vivo, superando de esta forma los inconvenientes derivados de las observaciones en vivo y las grabaciones. Datta et al. analizaron la tasa de filtrado y el área de sección transversal de modelos de anastomosis vasculares (10). Hanna et al. estudiaron la calidad de los nudos hechos por vía laparoscópica usando un tensiómetro y desarrollaron una escala de calidad y confiabilidad (17). En resumen, el conocimiento actual sobre el tema pone en evidencia falencias de peso en los métodos de evaluación de las habilidades técnicas vigentes en la actualidad. Los métodos tradicionales vigentes en la mayoría de los centros formadores de nuestro medio carecen de validez y confiabilidad en relación a la destreza quirúrgica. La formación quirúrgica dentro del marco de un programa formal estructurado de residencias médicas requiere una retroalimentación objetiva constante de la adquisición de habilidades técnicas para corrección y reorientación, y para ello precisa de instrumentos de evaluación objetivos. Todas las profesiones en general, relacionadas o no con la salud, implican la ejecución de procedimientos que requieren la adquisición de destrezas técnicas de variada complejidad. La formación de profesionales de ámbitos tan disímiles como son un piloto de avión y un cirujano, en apariencia sin ningún punto de contacto, tienen en común el hecho que ambas requieren el desarrollo de habilidades técnicas y que ambas deben enfrentar condiciones en las cuales, eventualmente, se pone en riesgo la vida. Muchos aspectos relacionados con el entrenamiento y la evaluación de los pilotos de avión podrían ser aplicables a la formación de los cirujanos. (24) Existen métodos validados de evaluación objetiva que podrían comenzar a utilizarse para evaluar las habilidades quirúrgicas con fines formativos y de certificación. Aunque se requieren futuras investigaciones antes de ser utilizados para evaluaciones sumativas y de revalidación (2, 3, 5, 12, 13, 20, 21).

BIBLIOGRAFÍA

1. ACGME. Outcome Project. <http://www.acgme.org/Outcome/>

2. Anderson CI, Jentz AB, Harkema JM, Rao Kareti L, Apelgren KN, Slomski CA. 2005. Assessing the competencies in general surgery residency training. *Am J Surg* 189: 288-292
3. Bann S, Khan M, Datta V, Darzi A. 2005. Surgical skill is predicted by the ability to detect errors. *Am J Surg*. 189: 412-415.
4. Beard JD. 2007. Assessment of Surgical Competence. *BJS* 94: 1315-1316
5. Boehler ML, Schwind CJ, Rogers DA, Ketchum J, O'Sullivan E, Mayforth R, Quin J, Wohltman C, Johnson C, Williams RG, Dunnington G. 2007. A Theory-Based Curriculum for Enhancing Surgical Skillfulness. *J Am Coll Surg*. 205: 492-497.
6. Cosman P, Hemli JM, Ellis AM, Hugh TJ. 2007. Learning the Surgical Craft: A Review of Skills Training Options. *ANZ J. Surg*. 77: 838-845.
7. Cuschieri A. 2001. Master surgeons and surgical proficiency. En *Metrics: Objective Assessment of Surgical Skills – Developing Quantitative Measurements through Surgical Simulation*. Scottsdale, Arizona: US Army Medical Research Command, Telemedicine and Advances Technology Research Center (TATRC)
8. Cuschieri A, Francis N, Crosby J, Hanna GB. 2001. What do Master Surgeons Think of Surgical Competence and Revalidation. *Am J Surg* 182:110-116.
9. Datta V, Chang A, Mackay S, Darzi A. 2001. The use of electromagnetic motion tracking analysis to objectively measure open surgical skills in the laboratory based model. *J Am Coll Surg*. 193: 479-485.
10. Datta V, Mandalia M, Mackay S, Chang A, Chesire N, Darzi A. 2002. Relationship between skill and outcome in the laboratory based model. *Surg*. 131: 318-323.
11. Darzi A, Datta V, Mackay S. 2001. The Challenge of Objective Assessment of Surgical Skill. *Am J Surg* 181: 484-486.
12. Dath D, Regehr G, Birch D, Schlachta C, Poulin E, Mamazza J, Reznick R, MacRae H. 2004. Toward Reliable Operative Assessment. *Surg Endosc* 18: 1800-1804.
13. Doyle JD, Webber EM, Sidhu RS. 2007. A universal global rating scale for the evaluation of technical skills in the operating room. *Am J Surg*. 193: 551-555.

14. Goff BA, Lentz GM, Lee D, Fenner D, Morris J, Mandel LS. 2001. Development of a Bench Station Objective Structured Assessment of Technical Skills. *Obstet Gynecol.* 98: 412-416.
15. Greenberg JA, Irani JL, Greenberg CC, Blanco MA, Lipsitz S, Ashley SW, Breen EM, Hafler JP. 2007. The ACGME competencies in the operating room. *Surgery.* 142: 180-184.
16. Hamford JM, Hall JC. 2000. Acquiring Surgical Skills. *BJS* 87: 28 – 37.
17. Hanna GB, Frank TG, Cuschieri A. 1997. Objective assessment of endoscopic knot quality. *Am J Surg.* 174: 410-413.
18. Intercollegiate Surgical Curriculum Programme. <http://www.iscp.ac.uk>.
19. Kane MT. 1992. The assessment of clinical competence. *Eval Health Prof.* 15(2): 163-182.
20. Larson JL, Williams RG, Ketchum J, Boehler M, Dunnington GL. 2005. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. *Surgery.* 138: 640-649.
21. MacRae H, Regehr G, Leadbetter W, Reznick R. 2000. A Comprehensive Examination for Senior Surgical Residents. *Am J Surg.* 179: 190-193.
22. Martin JA, Regehr G, Reznick R, MacRae H, Mumaghan J, Hutchinson G. 1997. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 84: 273-278.
23. Miller G. 1990. The assessment of clinical skills/competence/performance. *Academic Medicine.* 65: S63-S67.
24. Moen MD, Moen RO, Moen O.L. 2007. Teaching and Evaluating Technical Skills: Applying the Aviation Model to Gynecologic Surgery. *J Pelvic Med Surg.* 13: 107-112
25. Moorthy K, Munz Y, Sarker SK, Darzi A. 2003. Objective assessment of technical skills in surgery. *BMJ.* 327: 1032-1037
26. Regehr G, MacRae H, Reznick R, Szalay D. 1998. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE format examination. *Acad Med.* 73: 993-997.
27. Reznick R. 1993. Teaching and testing surgical skills. *Am J surg.* 165: 358-361.
28. Reznick R, MacRae H. 2006. Teaching Surgical Skills – Changes in the Wind. *N Engl J Med* 355: 2664-2669.
29. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. 1997. Testing Technical Skill via an Innovative “Bench Station” Examination. *Am J Surg.* 173: 226–230.
30. Scott DJ, Valentine RJ, Bergen PC, Rege RV, Laycock R, Tesfay ST. 2000. Evaluating surgical competency with the American Board of Surgery in-training examination, skill testing, and intraoperative assessment. *Surgery.* 128: 613-622.
31. Smee S. 2003. ABC of learning and teaching in medicine. Skill based assessment. *BMJ.* 326: 703-706.
32. Szalay D, MacRae H, Regehr G, Reznick R. 2000. Using Operative Outcome to Assess Technical Skill. *Am J Surg* 180: 234-237.
33. Van Der Vleuten C, Schuwirth L. 2005. Assessing professional competence: from methods to programmes. *Medical Education.* 39: 309-317.
34. Van Der Vleuten C. 1996. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1: 41-67.
35. Yule S, Flin R, Paterson-Brown S, Maran N, 2006. Non-technical Skills for surgeons in the operating room: A review of the literature. *Surgery.* 139: 140-149.