

La evaluación como aprendizaje de habilidades cognitivas y metacognitivas. Efectos de un sistema de evaluación multimétodo

Juan J. DI BERNARDO, Viviana de los Ángeles NAVARRO

Objetivo. Analizar y categorizar las actividades de procesamiento cognitivo (APC) que desarrollan los estudiantes en un sistema de evaluación longitudinal multimétodo (ELM).

Material y métodos. Estudio cuali-cuantitativo en el que se valoró el alcance cognitivo de todos los instrumentos de evaluación a partir de las APC que aplican los estudiantes en las preguntas y consignas formuladas, categorizadas según la taxonomía de Bloom: recordar, comprender, aplicar, analizar, evaluar y crear. En cada instrumento se calculó la proporción de ítems que abarcaba cada nivel taxonómico.

Resultados. En la ELM, los estudiantes realizan como APC: recordar (mecanismo fisiopatológico, etiología probable, entidad prevalente, objetivos terapéuticos, tiempo de tratamiento y método diagnóstico); identificar (factores de riesgo, problemas, datos anormales, efectos adversos y fármacos de elección); interpretar (signos/síntomas, laboratorio, imagen y estudio); reconocer (patrones, síndromes, gravedad y urgencia); comparar (estudios, tratamientos y diagnósticos diferenciales); analizar (jerarquizar problemas, plantear hipótesis, priorizar intervenciones, inferir resultados, y reflexionar sobre su desempeño y aprendizaje); decidir (tratamiento, medida de urgencia, conducta correcta y suspensión del tratamiento); justificar (estudios, tratamientos, conductas y preguntas orientadoras); planificar (secuencias diagnósticas, estudios para evaluar la evolución y plan de inmunización); sintetizar (integrar manifestaciones, resumen clínico y ficha bibliográfica); y valorar (razonamientos, compromiso relevante, pronóstico probable, área para internar, validez externa del artículo y fundamentar comentarios). Estas APC abarcan todos los niveles taxonómicos, con representaciones similares, en la nube de palabras, de los niveles cognitivos de orden inferior y superior.

Conclusiones. La combinación de diferentes instrumentos de evaluación permite ampliar el alcance cognitivo del proceso de evaluación y constituye un espacio muy apropiado para que los estudiantes puedan aplicar y desarrollar habilidades cognitivas y metacognitivas.

Palabras clave. Alcance cognitivo. Bloom. Competencias. Educación médica. Impacto educativo. Razonamiento.

Assessment as learning of cognitive and metacognitive skills. Effects of a multi-method evaluation system

Objective. Analyze and categorize the cognitive processing activities (APC) that students develop in a multi-method longitudinal assessment system (ELM).

Material and methods. A qualitative-quantitative study where the cognitive scope of all the evaluation instruments was assessed, based on the APC applied by the students in the questions and requests formulated, categorized according to Bloom's taxonomy: remember, understand, apply, analyze, evaluate and create. For each instrument, the proportion of items corresponding to each taxonomic level was calculated.

Results. In the ELM, the students carried out these APC: remember (physiopathological mechanism, probable etiology, prevalent entity, therapeutic objectives, treatment time, diagnostic method); identify (risk factors, problems, abnormal data, adverse effects, drugs of choice); interpret (signs/symptoms, laboratory, image, study); recognize (patterns, syndromes, severity, urgency); compare (studies, treatments, differential diagnoses); analyze (prioritize problems, propose hypotheses, prioritize interventions, infer results, reflect on their performance and learning); decide (treatment, emergency measure, proper conduct, suspend treatment); justify (studies, treatments, behaviors, guiding questions); planning (diagnostic sequences, studies to evaluate evolution, immunization plan); synthesize (integrate manifestations, clinical summary, bibliographic record); evaluate (reasonings, relevant commitment, probable prognosis, hospitalization area, external validity of the article, justify comments). These APC cover all taxonomic levels, with similar representations in the word cloud of the lower-order and higher-order cognitive levels.

Conclusions. The combination of different evaluation instruments allows expanding the cognitive scope of the evaluation process, constituting a very appropriate space for students to apply and develop cognitive and metacognitive skills.

Key words. Bloom. Cognitive scope. Competencies. Educational impact. Medical education. Reasoning.

Facultad de Medicina de la Universidad Nacional del Nordeste. Corrientes, Argentina.

Correspondencia:

Dr. Juan José Di Bernardo.
Departamento de Educación Médica. Facultad de Medicina. Universidad Nacional del Nordeste. Moreno, 1240. 3400 Corrientes, Argentina.

E-mail:

jjdibernardo@med.unne.edu.ar

Recibido:

21.04.23.

Aceptado:

23.06.23.

Conflicto de intereses:

No declarado.

Competing interests:

None declared.

Cómo citar este artículo:

Di Bernardo JJ, Navarro VA. La evaluación como aprendizaje de habilidades cognitivas y metacognitivas. Efectos de un sistema de evaluación multimétodo. FEM 2023; 26: 167-73. doi: 10.33588/fem.264.1294.

© 2023 FEM



Artículo *open access* bajo la licencia CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

ISSN (ed. digital): 2014-9840

Introducción

La evaluación y el aprendizaje están fuertemente vinculados en el campo de la educación médica. Numerosos trabajos [1-8] han demostrado que la evaluación tiene un efecto positivo en el aprendizaje de los estudiantes, pero, al mismo tiempo, han descrito que puede promover el desarrollo de estrategias de aprendizaje que son menos deseables. Estos efectos de la evaluación no se reflejan solamente en la cantidad y/o calidad del aprendizaje, sino también en los cambios de aptitud de los estudiantes y en el desarrollo de habilidades para el aprendizaje autodirigido [9].

Los efectos de la evaluación sobre el aprendizaje han sido diferenciados en tres momentos: a) antes de la evaluación, lo que se refleja en el comportamiento y los hábitos de estudio que aplican los estudiantes cuando se preparan para un evento de evaluación; b) durante la evaluación, condicionados por las características de la prueba y evidenciados por los resultados de ésta y/o el contenido de los portafolios; y c) después de la evaluación, manifestados durante la retroalimentación y en los logros de aprendizaje [10,11].

La vinculación entre evaluación y aprendizaje puede abordarse desde tres perspectivas diferentes según su propósito: la evaluación del aprendizaje, la evaluación para el aprendizaje y la evaluación como aprendizaje [12]. La primera, relacionada al término 'evaluación sumativa', está dirigida a verificar si los estudiantes lograron los objetivos de aprendizaje esperados o alcanzaron determinados estándares educativos, y tomar decisiones correctas sobre ellos (ingreso, promoción, graduación, etc.). La evaluación para el aprendizaje, denominada 'evaluación formativa', está centrada en los efectos de aprendizaje que tiene la evaluación y permite guiar el aprendizaje de los estudiantes y hacerlo más eficaz (proporcionar retroalimentación, promover estrategias de aprendizaje profundo y apoyar el aprendizaje continuo). La evaluación como aprendizaje abarca las dos perspectivas anteriores integradas al proceso formativo, de tal manera que los programas de evaluación se planifican, implementan, gestionan y evalúan a lo largo de todo el curso para conducir el aprendizaje de los estudiantes.

Además del momento y del propósito de la evaluación, otro aspecto importante para el aprendizaje es el diseño de las evaluaciones, pues la formación de los profesionales de la salud demanda múltiples habilidades cognitivas, psicomotoras, actitudinales y relacionales que los métodos únicos de evaluación no pueden capturar ni desarrollar en su

totalidad, y requieren, por lo tanto, la aplicación de múltiples evaluaciones complementarias, ensambladas e integradas a un 'sistema de evaluación' para lograr los diferentes propósitos y para abordar las múltiples necesidades de los interesados [13].

A partir de estas consideraciones, y contando la asignatura Medicina III con un sistema de evaluación longitudinal multimétodo (ELM) adaptado de Epstein (2007) [14], se planteó como objetivo de esta investigación analizar y categorizar las actividades de procesamiento cognitivo que desarrollan los estudiantes en la ELM.

Material y métodos

Para esta investigación se adoptó un enfoque metodológico mixto (cuali-cuantitativo) de complementación [15], con amplio predominio de la perspectiva cualitativa. El estudio tiene un diseño transversal, y alcance exploratorio y descriptivo.

Para su realización se revisaron todas las evaluaciones que integran el sistema de ELM que se aplica durante el estudio de Medicina III (Fig. 1) y se analizó, de cada instrumento, su diseño y contenido, las tablas de especificaciones de elaboración, las guías e instrucciones para los estudiantes, y los criterios de calificación y rúbricas utilizadas.

Analizando los verbos o frases verbales utilizados en las preguntas y consignas formuladas en las pruebas y guías, se consideró el alcance cognitivo de los procesos mentales que los estudiantes debieron realizar para responder las preguntas o cumplir las tareas solicitadas en las evaluaciones.

Las actividades de procesamiento cognitivo (APC) se categorizaron aplicando la nueva taxonomía de Bloom [16,17], que abarca, para la 'dimensión del proceso cognitivo', seis categorías principales de niveles crecientes: recordar (N-1); comprender (N-2); aplicar (N-3); analizar (N-4); evaluar (N-5); y crear (N-6). Las tres primeras categorías se consideran habilidades de pensamiento de orden inferior, y las tres últimas, habilidades de pensamiento de orden superior.

En el nivel N-1 se agruparon las APC dirigidas principalmente a recuperar, reconocer y recordar conocimientos relevantes de la memoria a largo plazo. En N-2 se incluyeron las APC que permiten construir significado a partir de mensajes escritos o gráficos o actividades, como interpretar, ejemplificar, clasificar, resumir, inferir, comparar o explicar. En N-3 se integraron las APC que requieren la utilización, aplicación, ejecución o implementación de una acción o procedimiento específico. En N-4 se

agruparon las APC que conducen a deconstruir materiales o conceptos en sus partes constituyentes y determinar cómo se relacionan entre sí y con la estructura o propósito general a través de la diferenciación, la comparación, la organización, la atribución y la integración. En N-5 se incluyeron las APC que permiten emitir juicios basados en criterios o estándares a través de la verificación y la crítica. En N-6 se agruparon las APC enfocadas en juntar o integrar elementos para formar un todo coherente o funcional o reorganizar elementos en un nuevo patrón o estructura a través de la generación, planificación o producción.

Estas habilidades del dominio cognitivo se articulan (en la nueva taxonomía de Bloom) con la 'dimensión del conocimiento', que comprende cuatro categorías: conocimiento fáctico, conocimiento conceptual, conocimiento procedimental y conocimiento metacognitivo [16]. Esta última categoría abarca el conocimiento estratégico para aprender y pensar; el conocimiento de las tareas cognitivas y sus contextos; y el autoconocimiento de las propias fortalezas y debilidades para el aprendizaje [18].

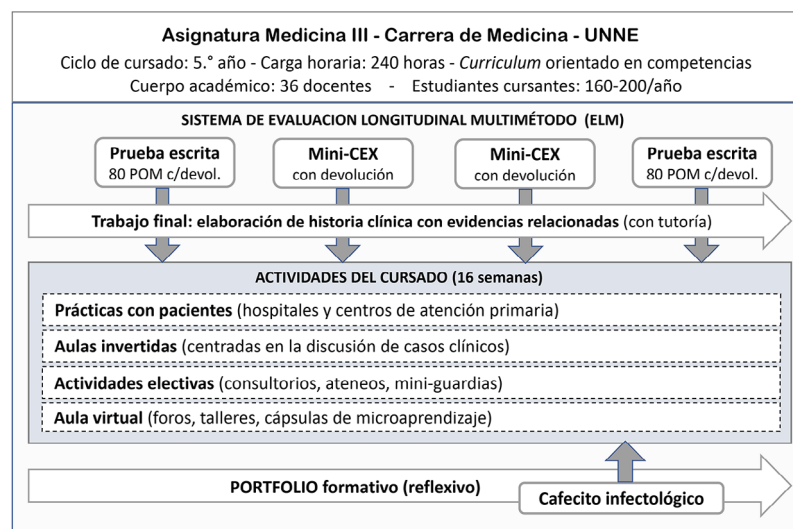
Los datos cualitativos se presentan detallando los verbos y sintagmas enunciados en las preguntas y proposiciones que integran las diferentes evaluaciones, y se sintetizan en tablas y gráficos en nube de palabras. Como datos cuantitativos se presentan, para cada instrumento de evaluación, la proporción de ítems que abarcaba cada nivel cognitivo de la taxonomía de Bloom.

Resultados

Analizando las APC realizadas por los estudiantes en cada uno de los instrumentos que integran el sistema de ELM de Medicina III, se observaron diferentes acciones, operaciones y procesos mentales relacionados con las características y formatos de las herramientas de evaluación utilizadas.

En los exámenes con preguntas de opción múltiple, se observó que todas incluyeron como enunciado una breve viñeta clínica con los datos (clínicos y complementarios) esenciales para elegir la mejor respuesta. Debajo de cada enunciado se plantearon dos o más preguntas con tres opciones de respuesta (una sola correcta), donde los estudiantes tuvieron que realizar las siguientes APC: recordar (etiología más probable, entidad de mayor prevalencia, objetivos terapéuticos, tiempo de tratamiento y método para confirmar un diagnóstico); identificar (datos anormales, estudio más sensible, efecto adverso, antibiótico o fármaco de elección); inter-

Figura 1. Contexto curricular de la asignatura Medicina III. Sistema de evaluación longitudinal multimétodo integrado a las actividades del curso.



pretar (signo o síntoma, resultado de laboratorio, imagen o estudio); reconocer (patrones o síndromes, signos de gravedad, determinación que falta y resultado esperado); comparar (estudios o tratamientos indicados); tomar decisiones (hipótesis diagnóstica más probable, tratamiento de elección o más efectivo, medida de urgencia recomendada, conducta más correcta y suspensión del tratamiento); justificar (preguntas para orientar el diagnóstico y elementos que fundamentan una conducta); planificar (secuencia de pruebas diagnósticas, estudios para evaluar la evolución y plan de inmunización); sintetizar (diagnóstico que explica la mayoría de las manifestaciones); y valorar (razonamiento más apropiado, compromiso más relevante, pronóstico más probable y lugar más apropiado para internar). Por ende, el alcance cognitivo de las preguntas de opción múltiple abarca principalmente los niveles taxonómicos N-1 a N-4.

Los *mini clinical evaluation exercises* (MiniCEX) se aplicaron a los estudiantes en dos momentos diferentes de los trabajos prácticos, durante y después de entrevistar a un paciente real. Se examinaron las habilidades para conducir la entrevista clínica y el examen físico de los pacientes, y se evaluó la capacidad para identificar los problemas activos, interpretar las manifestaciones integrándolas en síndromes o asociaciones patogénicas, proponer hipótesis diagnósticas fundadas, plantear y justificar la secuencia de estudios diagnósticos, y establecer un

Tabla. Principales actividades de procesamiento cognitivo y proporciones de niveles cognitivos abarcados en cada instrumento.

	Principales actividades de procesamiento cognitivo	Niveles cognitivos (proporciones)					
		N-1	N-2	N-3	N-4	N-5	N-6
POM	Recordar, identificar, interpretar, reconocer, ordenar, comparar, tomar decisiones, justificar, planificar, sintetizar y valorar	0,2	0,2	0,36	0,16	0,08	
Mini-CEX	Identificar, interpretar e integrar manifestaciones, proponer y fundamentar hipótesis, y plantear y justificar estudios, tratamientos y plan de acción	0,1	0,15	0,35	0,2	0,1	0,1
Cafecito infectológico	Discutir casos, identificar y jerarquizar problemas, interpretar imágenes/estudios, proponer y justificar diagnósticos y tratamientos, y aportar evidencias	0,1	0,25	0,3	0,2	0,15	
Trabajo final	Redactar historia clínica, identificar y jerarquizar problemas, elaborar una discusión, fundamentar hipótesis, planificar y justificar estudios y tratamientos, y analizar evidencias	0,1	0,2	0,28	0,18	0,12	0,12

Mini-CEX: *mini clinical evaluation exercises*; POM: preguntas de opción múltiple.

plan terapéutico adecuado y pautas para la educación del paciente, priorizando los estudios y estrategias de intervención, en relación con el contexto, el análisis coste/beneficio y la oportunidad, con lo cual el alcance cognitivo de los MiniCEX abarcó principalmente los niveles taxonómicos N-2 a N-6.

En el 'cafecito infectológico', los estudiantes discutieron casos clínicos de infectología, para lo cual tuvieron que identificar y jerarquizar problemas, describir e interpretar imágenes radiológicas, proponer y justificar diagnósticos diferenciales y alternativas terapéuticas, y aportar evidencias científicas para fundamentar sus comentarios. En consecuencia, el alcance cognitivo de esta actividad abarcó principalmente los niveles N-2 a N-5 de la taxonomía de Bloom.

Para el trabajo final, los estudiantes tuvieron que realizar las siguientes tareas: redactar con claridad y precisión una base de datos para la historia clínica; identificar y jerarquizar los problemas que presenta el paciente; elaborar, con los datos disponibles, una discusión diagnóstica coherente; formular y fundamentar las hipótesis diagnósticas; diseñar y justificar un plan de estudios pertinente; establecer un plan terapéutico adecuado; y confeccionar un resumen claro y preciso de la historia clínica. Además, para el trabajo bibliográfico debían seleccionar un problema del paciente; convertir el problema en una pregunta estructurada (formato PICO); realizar la búsqueda bibliográfica; confeccionar la ficha bibliográfica; y evaluar la validez externa del artículo, con lo que el alcance cognitivo

de este ejercicio abarcó cada uno de los seis niveles taxonómicos.

En el portafolio, los estudiantes registraron sus experiencias de aprendizaje a través de la escritura reflexiva y aportaron evidencias de lo aprendido sobre los siguientes aspectos: a) la evidencia y reflexión de su aprendizaje en un ateneo clínico (de su elección); b) el relato de una situación vinculada al profesionalismo (médico o estudiantil) o a la seguridad del paciente (positiva o negativa), observada durante el año, con una reflexión de lo aprendido; y c) una reflexión global sobre lo que aportó el portafolio a su aprendizaje, por lo que con esta herramienta desarrollaron principalmente habilidades metacognitivas.

A partir de la categorización taxonómica de las preguntas y consignas, que se realizó en cada una de las evaluaciones, se calcularon las proporciones de los ítems correspondientes a cada nivel cognitivo incluido en los diferentes instrumentos de evaluación (Tabla).

Analizando globalmente las APC realizadas por los estudiantes en todos los instrumentos e instancias de evaluación, a lo largo del estudio de la asignatura, se observó en la nube de palabras (Fig. 2) que la ELM abarcó todos los niveles taxonómicos de Bloom, con representaciones similares de los niveles cognitivos de orden inferior y de orden superior. Las APC de orden inferior más utilizadas fueron: identificar, interpretar, aplicar, resolver y jerarquizar; mientras que en las de orden superior, las operaciones más frecuentes fueron: valorar, planificar, justificar, decidir y fundamentar.

Discusión

El sistema de ELM que aplica Medicina III integra instrumentos de evaluación que individualmente tienen diferentes alcances cognitivos, pero en conjunto abarcan todos los niveles taxonómicos de orden inferior y superior. De esta manera, el sistema de evaluación constituye un espacio muy completo y favorable para el ejercicio y aprendizaje de las habilidades cognitivas [19] esenciales para la formación profesional.

Si bien las herramientas de evaluación difieren en su formato y estructura, tienen en común que las preguntas y consignas planteadas en ellas son 'ricas en contexto', pues están enfocadas en casos clínicos y problemas de pacientes tomados de la realidad, que los estudiantes deben interpretar y procesar para poder responder o resolver cada ítem de las pruebas. Este diseño hace que las tareas de evaluación sean auténticas, significativas y atractivas, lo que constituye un poderoso estímulo para el aprendizaje [20,21].

Las APC que los estudiantes realizan al resolver los casos y problemas de las evaluaciones son operaciones mentales muy valiosas para aprender y muy adecuadas al nivel de formación alcanzado en quinto año de la carrera, ya que pueden aplicar construcciones cognitivas desarrolladas durante su formación, como un fragmento (*chunk*), que es una fracción de información vinculada a un diagnóstico o tratamiento; un esquema (*schema*), que comprende fragmentos organizados; y un guion (*script*), que es el prototipo de una enfermedad [22,23].

Estas construcciones cognitivas están relacionadas en la perspectiva asociacionista de la memoria, es decir, piezas separadas de información que pueden interrelacionarse debido a su recuperación conjunta y repetida en cada evaluación, y de esa forma se accede a ellas como una sola unidad (o bloque) en la memoria de trabajo, lo que reduce la carga cognitiva y maximiza la capacidad de procesamiento durante una evaluación. En consecuencia, aplicando estos 'bloques de construcción', los estudiantes pueden formar y reconocer patrones a través de la asociación; mejorar la velocidad y precisión del procesamiento cognitivo; y desarrollar con el tiempo habilidades más complejas de razonamiento [22,23].

Por estas razones, las APC que realizan los estudiantes a lo largo de la ELM pueden enmarcarse en la teoría del procesamiento dual, que considera dos tipos o sistemas de razonamiento: el tipo 1, que es automático, subconsciente, más rápido y de menor esfuerzo, se basa en las experiencias almacenadas

Figura 2. Nube de palabras construida con las actividades de procesamiento cognitivo abarcadas en el sistema de evaluación longitudinal multimétodo.



en la memoria a largo plazo (patrones o esquemas construidos); y el tipo 2, que es analítico, más lento y de mayor esfuerzo, combina funciones ejecutivas intensivas en recursos, incluida la memoria de trabajo y la atención [24,25].

Con ese enfoque, Bowen (2006) [26] recomienda que los estudiantes comiencen a reconocer patrones y a desarrollar *scripts* de enfermedad, porque tanto el razonamiento automático como el analítico son efectivos y se usan simultáneamente en forma interactiva. Sin embargo, la aplicación del reconocimiento de patrones puede tener algunos riesgos para los estudiantes, pues, al ser un procesamiento rápido y automático, es propenso a sesgos, como el cierre prematuro del razonamiento [27].

Una observación relevante en el sistema de ELM es la riqueza de tareas de evaluación que demandan APC de orden superior (analizar, evaluar y crear), pues, cuando los estudiantes perciben, a través de señales inferidas de la tarea, que las preguntas y consignas están dirigidas a evaluar niveles cognitivos altos, suelen adoptar enfoques de aprendizaje profundo [11] (alta motivación intrínseca para aprender); por el contrario, cuando los estudiantes intrínsecamente motivados perciben que una evaluación sólo requiere memorizar y recordar, adoptan, por lo general, enfoques de aprendizaje superficial (mínimo esfuerzo y poco compromiso) [6,11].

Respecto a las habilidades metacognitivas, la inclusión del portafolio en el sistema de ELM posibilita a los estudiantes reflexionar sobre sus experiencias de aprendizaje y sus habilidades cognitivas, lo que implica realizar un proceso metacognitivo complejo [28] que les permite aprender activamente de sus propias experiencias, y evaluar sus conocimientos

tos, habilidades de razonamiento y competencias [29,30]. El desarrollo en los estudiantes de la capacidad para reflexionar es esencial para su formación profesional y para trabajar en entornos complejos [31].

Conclusiones

La combinación y la integración de diferentes instrumentos de evaluación, en un sistema de evaluación multimétodo, además de brindar información más amplia y completa sobre el desarrollo de las competencias, también permiten diversificar las tareas de evaluación, y enriquecerlas con elementos auténticos, significativos y atractivos, que amplían el alcance cognitivo del proceso de evaluación, constituyendo así un espacio muy apropiado y valioso para que los estudiantes puedan aplicar y desarrollar habilidades cognitivas y metacognitivas que son fundamentales para el aprendizaje y la formación profesional.

La categorización de las tareas de evaluación en distintos niveles cognitivos permite definir con mayor precisión los objetivos de la evaluación en concordancia con los objetivos curriculares; determinar la amplitud y profundidad cognitiva de las tareas solicitadas en cada evaluación en relación con los resultados de aprendizaje esperados; comparar el alcance cognitivo de las preguntas o consignas entre diferentes instrumentos de evaluación; y utilizar un lenguaje común que facilita la comunicación entre los docentes evaluadores.

Otro beneficio que puede ofrecer la amplitud de niveles cognitivos abarcados en el sistema de ELM es su utilidad para el 'aprendizaje mejorado mediante pruebas' (*test-enhanced learning*) [32], fenómeno que tiene su explicación en que la recuperación activa de información en una evaluación (a través del procesamiento cognitivo) permite la creación de redes organizadas de conocimiento que favorecen la retención a largo plazo y promueven la recuperación de información en contextos donde se necesita [32-34].

Bibliografía

- Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983; 17: 165-71.
- McManus IC, Richards P, Winder BC, Sproston KA. Clinical experience, performance in final examinations, and learning style in medical students: prospective study. *BMJ* 1998; 316: 345-50.
- McLachlan JC. The relationship between assessment and learning. *Med Educ* 2006; 40: 716-7.
- Segers M, Dochy FJRC. Introduction enhancing student learning through assessment: Alignment between levels of assessment and different effects on learning. *Stud Educ Evaluation* 2006; 32: 171-9.
- Cilliers FJ, Schuwirth LWT, Adendorff HJ, Herman N, van der Vleuten CPM. The mechanism of impact of summative assessment on medical students' learning. *Adv Health Sci Educ Theory Pract* 2010; 15: 695-715.
- Al-Kadri HM, Al-Moamary MS, Roberts C, van der Vleuten CPM. Exploring assessment factors contributing to students' study strategies: literature review. *Med Teach* 2012; 34 (Suppl 1): S42-50.
- Cobb KA, Brown G, Jaarsma DA, Hammond RA. The educational impact of assessment: A comparison of DOPS and MCQs. *Med Teach* 2013; 35: 1598-607.
- Heeneman S, Oudkerk Pool A, Schuwirth LW, van der Vleuten CP, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ* 2015; 49: 487-98.
- Monroe KS. The relationship between assessment methods and self-directed learning readiness in medical education. *Int J Med Educ* 2016; 7: 75-80.
- Cilliers FJ, Schuwirth LW, Herman N, Adendorff HJ, van der Vleuten CP. A model of the pre-assessment learning effects of summative assessment in medical education. *Adv Health Sci Educ Theory Pract* 2012; 17: 39-53.
- Cilliers FJ, Schuwirth LWT, van der Vleuten CPM. Modelling the pre-assessment learning effects of assessment: evidence in the validity chain. *Med Educ* 2012; 46: 1087-98.
- van der Vleuten C, Sluijsmans D, Joosten-ten Brinke D. Competence assessment as learner support in education. In Mulder M, ed. *competence-based vocational and professional education, technical and vocational education and training: issues, concerns and prospects*. Cham, Switzerland: Springer; 2017. p. 607-30.
- Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, et al. 2018 Consensus framework for good assessment. *Med Teach* 2018; 40: 1102-9.
- Epstein RM. Assessment in medical education. *N Eng J Med* 2007; 356: 387-96.
- Bericat-Alastuey E. La integración de los métodos cuantitativo y cualitativo en la investigación social. Significado y medida. Barcelona: Editorial Ariel; 1998.
- Krathwohl DR. A revision of Bloom's taxonomy: an overview. *Theory Pract* 2002; 41: 212-8.
- Jensen JL, McDaniel MA, Woodard SM, Kummer TA. Teaching to the test... or testing to teach: exams requiring higher order thinking skills encourage greater conceptual understanding. *Educ Psychol Rev* 2014; 26: 307-29.
- Pintrich PR. The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Pract* 2002; 41: 219-25.
- Cilliers FJ, Schuwirth LWT, Adendorff HJ, Herman N, van der Vleuten CPM. The mechanism of impact of summative assessment on medical students' learning. *Adv Health Sci Educ Theory Pract* 2010; 15: 695-715.
- Boud D, Falchikov N. Aligning assessment with long-term learning. *Eval High Educ* 2006; 31: 399-413.
- Preston R, Gratani M, Owens K, Roche P, Zimanyi M, Malau-Aduli B. Exploring the impact of assessment on medical students' learning. *Eval High Educ* 2020; 45: 109-24.
- de Bruin ABH, Sibbald M, Monteiro S. The science of learning. In Swanwick T, Forrest K, O'Brien BC, eds. *Understanding medical education. Evidence, theory and practice*. 3 ed. Oxford, UK; John Wiley & Sons; 2019. p. 23-36.
- Charlin B, Boshuizen H, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ* 2007; 41: 1178-84.
- Young J Q, Van Merriënboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE Guide No. 86. *Med Teach* 2014; 36: 371-84.
- Evans JS, Stanovich KE. Dual-process theories of higher cognition: advancing the debate. *Perspect Psychol Sci* 2013; 8: 223-41.

26. Bowen JL Educational strategies to promote clinical diagnostic reasoning. *N Eng J Med* 2006; 355: 2217-25.
27. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Qual Saf* 2013; 22: 58-64.
28. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach* 2009; 31: 685-95.
29. Van Tartwijk J, Driessen EW. Portfolios for assessment and learning. AMEE Guide No. 45. *Med Teach* 2009; 31: 790-801.
30. Moniz T, Arntfield S, Miller K, Lingard L, Watling C, Regehr G. Considerations in the use of reflective writing for student assessment: issues of reliability and validity. *Med Educ* 2015; 49: 901-8.
31. Hoffman LA, Shew RL, Vu TR, Brokaw JJ, Frankel RM. Is reflective ability associated with professionalism lapses during medical school? *Acad Med* 2016; 91: 853-7.
32. Larsen DP, Butler AC, Roediger III HL. Test-enhanced learning in medical education. *Med Educ* 2008; 2: 959-66.
33. Green ML, Moeller JJ, Spak JM. Test-enhanced learning in health professions education: a systematic review: BEME Guide No. 48. *Med Teach* 2018; 40: 337-50.
34. Pan SC, Rickard TC. Transfer of test-enhanced learning: meta-analytic review and synthesis. *Psychol Bull* 2018; 144: 710-56.