



ORIGINAL

Correlación de competencias evaluadas en una evaluación clínica objetiva estructurada, un test de opción múltiple y un examen de casos clínicos por ordenador en estudiantes de último curso de Medicina



Emilio Cervera-Barba^{a,*}, Sophia Denizon-Arranz^a, Raúl Castañeda-Vozmediano^a, José Manuel Blanco-Canseco^{a,b}, M. Isabel Gámez-Cabero^{a,b} e Irene Salinas-Gabiña^{a,c}

^a Comité de Evaluación Clínica Objetiva Estructurada (ECOE), Facultad de Medicina, Universidad Francisco de Vitoria, Madrid, España

^b Centro de Salud Valle de la Oliva, Madrid, España

^c Servicio de Medicina Intensiva, Hospital Universitario del Henares, Madrid, España

Recibido el 25 de junio de 2025; aceptado el 1 de agosto de 2025

Disponible en Internet el xxxx

PALABRAS CLAVE

Evaluación Clínica Objetiva Estructurada (ECOE);
Prueba de casos clínicos computarizados;
Casos clínicos simulados computarizados;
Pacientes virtuales;
Estudiantes de Medicina;
Evaluación de la competencia

Resumen

Introducción: los alumnos del último curso de Medicina deben realizar una evaluación final mediante una ECOE (evaluación clínica objetiva estructurada). En algunas facultades, por adaptación a la pandemia, se incorporó un examen informatizado de casos clínicos (PVCC), que pretendía evaluar competencias similares a la ECOE y que se pondera en la calificación final junto con esta. En 2023 se compararon una ECOE, una PVCC y un examen tipo test de opción múltiple (MCQ, por sus siglas en inglés), con el objetivo de analizar la concordancia de resultados globales y por competencias, bajo la hipótesis de que la PVCC mide un constructo más cercano al MCQ que a la ECOE.

Métodos: participaron 108 estudiantes. Realizaron una ECOE de 21 estaciones, una PVCC con 10 casos clínicos informatizados y un MCQ de 150 preguntas sobre 26 casos clínicos relacionados con las otras 2 pruebas. Se evaluaron 5 competencias comunes: anamnesis, exploración física, juicio clínico, prevención/promoción y profesionalismo. Se analizó la consistencia interna (alfa de Cronbach) y la correlación entre pruebas (coeficiente de Spearman, Rho).

Resultados: las puntuaciones fueron más bajas en el MCQ. La ECOE y la PVCC solo coincidieron en el resultado global, no por competencias. La mayor correlación de puntuaciones se observó entre MCQ y PVCC ($Rho = 0,49$), y la menor, entre ECOE y PVCC ($Rho = 0,27$). Solo se hallaron correlaciones significativas en *juicio clínico* y *exploración física* entre el MCQ y las otras pruebas. La ECOE fue la prueba con mayor consistencia interna ($\alpha = 0,754$).

* Autor para correspondencia.

Correo electrónico: emilio.cervera@ufv.es (E. Cervera-Barba).

Conclusión: las 3 pruebas evalúan dimensiones distintas. La PVCC mide un constructo más próximo al MCQ. Las competencias evaluadas por ECOE y PVCC no se correlacionan.

© 2025 El Autor/Los Autores. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

KEYWORDS

Objective Structured Clinical Examination (OSCE);
Computerised Case-based Test;
Computer-based case simulation;
Virtual patients;
Medicine students;
Competence assessment

Correlation of competencies assessed in an objective structured clinical examination, a multiple-choice question exam and a computer-based cases test in last year medical students

Abstract

Background: Final-year medical students are required to complete a summative assessment through an OSCE (Objective Structured Clinical Examination). In some faculties, as an adaptation during the pandemic, a computer-based Cases Test (CCT) was introduced. This test was intended to assess competencies similar to those evaluated in the OSCE and is weighted as part of the final grade alongside the OSCE. In 2023, we compared an OSCE, a CCT, and a multiple-choice question (MCQ) exam, with the aim of analysing the concordance of overall and competency-specific results. Our hypothesis was that the CCT assesses a construct more closely related to the MCQ than to the OSCE.

Methods: A total of 108 students participated. They completed a 21-station OSCE, a CCT consisting of 10 computer-based clinical cases, and a 150-question MCQ exam based on 26 clinical scenarios aligned with the other two assessments. Five shared competencies were evaluated: history taking, physical examination, clinical reasoning, prevention/promotion, and professionalism. Internal consistency was analysed using Cronbach's alpha, and correlations between assessments were calculated using Spearman's rho coefficient.

Results: Scores were lowest in the MCQ. The OSCE and CCT coincided only in overall scores, not by individual competencies. The strongest correlation was observed between the MCQ and CCT ($\rho = 0.49$), and the weakest between the OSCE and CCT ($\rho = 0.27$). Significant correlations were found only in clinical reasoning and physical examination between the MCQ and the other assessments. The OSCE showed the highest internal consistency ($\alpha = 0.754$).

Conclusion: The three assessments evaluate different dimensions. The CCT measures a construct more closely aligned with the MCQ. Competencies assessed by the OSCE and the CCT do not correlate.

© 2025 The Author(s). Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

El objetivo de la evaluación en educación médica es desarrollar mediciones fiables de la competencia de los evaluados que permitan asegurar un adecuado desempeño posterior, ofrecerles retroalimentación para favorecer su avance en el aprendizaje y obtener información para adecuar los planes formativos. La evaluación sumativa debería, además, ser capaz seleccionar a los evaluados por un nivel mínimo de competencia que garantice a la sociedad una práctica profesional competente¹⁻³. La evaluación de las distintas dimensiones de la competencia es un proceso complejo y requiere la adecuación de las metodologías a lo que se pretende evaluar^{4,5}. Una correcta evaluación debería ser continuada, programada y multidimensional para favorecer el conocimiento de la realidad competencial del evaluado^{5,6}.

En el grado de Medicina la principal herramienta de evaluación sumativa son los tests de opción múltiple (MCQ, por sus siglas en inglés) tipo A de única mejor respuesta, que tienen elevada validez y fiabilidad si están bien contruidos^{7,8}.

Los alumnos realizan exámenes durante todos los años de carrera para entrenar la misma metodología evaluativa que se encontrarán al finalizar el grado, en el proceso selectivo de acceso a la formación sanitaria especializada.

La ECOE (Evaluación Clínica Objetiva Estructurada) es una metodología evaluativa creada hace 50 años⁹ y también ha demostrado su validez y fiabilidad para la evaluación de la competencia, especialmente en las profesiones sanitarias^{10,11}. En la ECOE, el alumno se enfrenta en un ambiente simulado a situaciones clínicas representadas por pacientes simulados, maniquíes o simuladores robóticos. Cada una de las situaciones se denomina *estación* y los alumnos se exponen a un número variable de estaciones donde son evaluados por pacientes estandarizados o evaluadores en su desempeño en distintas competencias profesionales. Las facultades de medicina españolas tienen la obligación de realizar una ECOE al final del grado, como requisito de aptitud para alcanzar el egreso y la concesión del nivel MECES-3¹².

A consecuencia de la pandemia COVID-19, el Comité ECOE interfacultativo (CI-ECOE) de la Conferencia Nacional de

Decanas y Decanos de Facultades de Medicina de España (CNDFME) creó un examen de casos clínicos por ordenador (*Computer-based Cases Test*, CCT) que sustituyó a la ECOE durante el confinamiento¹³. Aunque entonces se denominó ECOE-CCS y actualmente recibe el nombre de Prueba Virtual Clínica Complementaria (PVCC), la metodología evaluativa es la de una prueba CCT, descrita por Schuwirth et al.¹⁴. Pasada la pandemia se mantuvo su organización y los resultados de esta prueba se ponderan con la ECOE en la evaluación sumativa final de cada alumno, asumiendo por consenso pero sin validación previa, que la PVCC puede evaluar, de forma análoga a la ECOE, 5 de las competencias evaluadas en esta: 1) anamnesis, 2) exploración física, 3) juicio clínico, plan de manejo diagnóstico y terapéutico, 4) prevención y promoción de la salud, y 5) aspectos ético-legales y profesionalismo. En un estudio desarrollado en nuestra facultad se comprobó que dicha similitud no parecía confirmarse por la ausencia de correlación de resultados de los alumnos entre ambas pruebas¹⁵.

En este estudio se pretende analizar los resultados de los alumnos en una ECOE, un examen tipo MCQ y la PVCC. El objetivo principal es analizar los resultados globales y por competencias en las 3 pruebas para conocer si existe correlación entre ellas. Como objetivos secundarios, se analizará la consistencia interna de cada prueba y se explorará la opinión de los participantes sobre ellas. La hipótesis de partida es que los resultados en la PVCC, tanto globales como por competencias, se asimilarán más a los obtenidos en el MCQ que en la ECOE, lo que sería un indicio de similitud de los procesos cognitivos desarrollados al enfrentarse a dichas pruebas.

Metodología

Estudio observacional analítico transversal retrospectivo. La población de estudio estuvo compuesta por los alumnos de 6.º curso de Medicina de la Universidad Francisco de Vitoria (N = 112), que se enfrentaron en 2023 a las pruebas obligatorias de evaluación de la competencia final de grado (ECOE y PVCC) y, optativamente, a un examen MCQ.

Las 3 pruebas fueron:

1) Una prueba ECOE de 21 estaciones con pacientes estandarizados, simuladores y evaluadores, pruebas complementarias, y redacción de informes. Se realizaron 2 versiones diferentes (EEOE-1 y ECOE-2) para aminorar el efecto copia entre participantes de distintos turnos. Las versiones eran diferentes en 11 estaciones, con presentación clínica parecida, pero diagnósticos diferentes. Ambas eran similares en la ponderación de competencias, ítems y puntuaciones (EEOE-1: 540 ítems, 1.561 puntos. ECOE-2: 545 ítems, 1.561 puntos). La ECOE se realizó los días 11, 12 y 13 de mayo y tuvo una duración de 3 horas y 50 minutos. En la [tabla 1](#) se presenta la tabla de especificaciones.

2) La PVCC organizada por el CI-EEOE y común a todas las facultades de Medicina, con 10 casos clínicos por ordenador a través de plataformas virtuales (Moodle, Blackboard, Sakai) y un total de 72 cuestiones y 1.000 puntos. Las cuestiones contenían preguntas tipo Cloze y de opción múltiple. Tuvo una duración de 2 horas y se realizó el 5 de mayo. En la [tabla 2](#) se muestran las especificaciones de la prueba.

3) Un examen MCQ tipo A de 150 preguntas (150 puntos) distribuidas en 26 casos clínicos similares a los desarrollados en la ECOE y la PVCC. Las preguntas se crearon por 21 profesores de la facultad con probada experiencia en la elaboración de MCQ. Las preguntas de los casos, de 4 respuestas y una única opción válida, se distribuyeron proporcionalmente al peso de las competencias evaluadas por los ítems de la ECOE y las cuestiones de la PVCC ([tabla 3](#)). Por cada 3 preguntas falladas se restaba una pregunta acertada. El examen se realizó el 10 de mayo y tuvo una duración de 3 horas.

Al finalizar cada prueba, los participantes completaron una encuesta de satisfacción anónima donde se les preguntaba la opinión sobre el tiempo para realizar cada prueba (1=muy escaso a 5=excesivo), el nivel de dificultad global percibido (1-excesivamente fácil a 10-excesivamente difícil), y el grado en que consideraban que cada prueba era capaz de evaluar las competencias generales de un alumno de 6.º de Medicina, y la competencia en el manejo clínico ante pacientes reales (ambas: 1-nada relacionado a 10-totalmente relacionado).

Las variables de estudio fueron: variables de filiación (código identificativo de participante; versión ECOE realizada: ECOE-1 o ECOE-2); variables de resultados de alumnos en las 3 pruebas (puntuación total y por competencias; porcentajes de acierto de cada ítem, pregunta o cuestión, totales y por competencias); resultados de las encuestas de satisfacción de cada prueba.

Solo se compararon los resultados de la ECOE en las 5 competencias citadas más arriba y contrastadas con las otras 2 pruebas. Por ello, finalmente se compararon con la PVCC y MCQ solo 356 y 360 ítems (EEOE-1 y ECOE-2, respectivamente) y un total de 1.051 puntos en ambas versiones.

Los datos se obtuvieron desde las aplicaciones en que se registraron: Moodle para la PVCC, Excel para los resultados obtenidos mediante lectura óptica de hoja de exámenes tipo test, aplicación GEXECO® (Innocan) para la ECOE, y Excel de formulario Google® para las encuestas de satisfacción. Los datos personales de los participantes no eran conocidos por los investigadores del estudio, ya que las bases de datos proporcionadas estaban seudonimizadas desde el origen, asignando un código alfanumérico único por cada estudiante. El equipo investigador se comprometió a trabajar sobre la base unificada de datos seudonimizada y a omitir cualquier acción encaminada a la identificación de los participantes, con el fin de mantener el anonimato.

Análisis estadístico

La normalidad de la distribución de las calificaciones, formuladas tanto como porcentaje directo de aciertos (sobre 100) como puntuaciones obtenidas según las ponderaciones y penalizaciones de cada prueba (sobre 10), se estudió mediante la prueba Anderson-Darling. Las calificaciones finales y las correspondientes a los distintos dominios se mostraron como mediana y rango intercuartílico.

Para estudiar la correlación de calificaciones entre dominios competenciales homónimos, así como en las puntuaciones totales de las 3 pruebas, se estimó el coeficiente de correlación de Spearman (Rho). La fiabilidad de las pruebas como consistencia interna de cada dominio

Tabla 1 Tabla de especificaciones de la evaluación clínica objetiva estructurada (ECOE). Se presentan datos de ECOE-1/ECOE-2 cuando no coinciden

Estación	Caso	Anamnesis		Exploración física		Habilidades Técnicas		Comunicación		Juicio clínico		Prevención		Redacción Informes		Profesionalismo		Total ECOE	
		Ptos	Ítem	Ptos	Ítem	Ptos	Ítem	Ptos	Ítem	Ptos	Ítem	Ptos	Ítem	Ptos	Ítem	Ptos	Ítem	Ptos	Ítem
E-1	Traumatología*	17	8	15	5			20	5	12	4	8	3			6	2	78	27
E-2	Exploración maniquí*			52	20					15	4							67	24
E-3	Neurología			85	28					10	2							95	30
E-4	Reumatología*			8	2	36	12			9	3	4	1			15	5	72	23
E-5	Pediatría	21	8					20	5	20	4	19	4					80	21
E-6	Neurología*					12	2			36	17	8	2			9	3	65	24
E-7	Psiquiatría	30	12	20	5			20	5	20	6					10	2	100	30
E-8	Informe Psiquiatría									16	5			54	22			70	27
E-9	Cardiología*	43	15	10	2			20	5	15	2/6							88	24/28
E-10	Informe Cardiología*									33	10/11			57	25/26			90	35/37
E-11	Metabolismo*	28	12					20	5	14	3	20	4					82	24
E-12	Urgencias			20	4	30	14									12	6	62	24
E-13	Neumología					25	10			23	6							48	16
E-14	Reumatología*	30	11	35	8			20	5	15	6							100	30
E-15	Cirugía					21	16			11	4	12	3			16	2	60	25
E-16	Digestivo*	10	7/6					20	5	20	7	15	6/5					65	25/23
E-17	Informe Digestivo									13	3/4			45	18			58	21/22
E-18	Cardiología*			10	6	30	15			11	3							51	24
E-19	Neumología	38	15	10	3			20	5	13	4							81	27
E-20	Infecciosas	43	22					20	5	4	1	10	5					77	33
E-21	Ginecología	18	6	34	15			20	5									72	26
Total por competencia		278	116/115	299	98	154	69	200	50	310	94/100	96	28/27	156	65/66	68	20	1561	540/545
Porcentaje (%) sobre total ECOE		17,8	21,5/21,1	19,1	18,1/18,0	9,9	12,8/12,7	12,8	9,3/9,2	19,9	17,4/18,3	6,1	5,2/4,9	10	12,0/12,1	4,4	3,7/3,7	100	100

ECOE: evaluación clínica objetiva estructurada; Ptos: puntos.

* Estación con versiones diferentes para ECOE-1 y ECOE-2.

Tabla 2 Tabla de especificaciones de la prueba virtual clínica complementaria (PVCC)

N°	Caso	Anamnesis		Exploración física		Juicio clínico		Prevención		Profesionalismo		Total PVCC	
		Ptos	Cuest.	Ptos	Cuest.	Ptos	Cuest.	Ptos	Cuest.	Ptos	Cuest.	Ptos	Cuest.
C-1	Urgencias	15	1	0	0	20	3	0	0	65	6	100	10
C-2	Digestivo	30	1	30	2	40	3	0	0	0	0	100	6
C-3	Pediatría	30	2	10	1	40	4	20	1	0	0	100	8
C-4	Neurología	20	1	45	3	35	3	0	0	0	0	100	7
C-5	Obstetricia	25	1	10	1	40	3	25	1	0	0	100	6
C-6	Cardiovascular	20	2	0	0	30	3	50	3	0	0	100	8
C-7	Reumatología	30	2	20	1	50	3	0	0	0	0	100	6
C-8	Neumología	25	1	17	1	50	6	8	1	0	0	100	9
C-9	Oncología	0	0	0	0	80	3	15	1	5	1	100	5
C-10	Psiquiatría	20	1	30	1	30	3	0	0	20	2	100	7
Total por competencia		215	12	162	10	415	34	118	7	90	9	1000	72
Porcentaje (%) sobre total PVCC		21,50	16,67	16,20	13,89	41,50	47,22	11,80	9,72	9,00	12,50	100	100

Cuest.: cuestiones; Ptos: puntos; PVCC: prueba virtual clínica complementaria.

competencial, así como de cada prueba, se reportó mediante el índice alfa de Cronbach (α).

También se comparó el promedio de satisfacción reportado por los participantes entre cada par de pruebas mediante la prueba U de Mann Whitney.

Para el análisis estadístico se empleó el software R (versión 4.3.2.). Se asumió como error tipo I un $\alpha = 0,05$.

Resultados

A las pruebas ECOE y PVCC concurren 112 alumnos. 4 no se presentaron al MCQ, por tanto, la muestra final del estudio quedó constituida por 108 participantes.

En el test de normalidad univariante Anderson-Darling, se observó que la puntuación total del MCQ ($p = 0,42$) y la PVCC ($p = 0,62$) se distribuyeron de forma significativamente normal, a diferencia de la observada en la ECOE ($p = 0,017$). En el caso de los porcentajes totales de acierto, las 3 pruebas presentaron normalidad en sus distribuciones (MCQ: $p = 0,276$; PVCC: $p = 0,061$; ECOE: $p = 0,122$), aunque no en alguno de sus dominios, por lo que los promedios se presentarán con medianas.

Los resultados descriptivos de puntuaciones y porcentaje de acierto figuran en la [tabla 4](#). Aplicando las ponderaciones y penalizaciones respectivas de cada prueba y obteniendo cada puntuación sobre 10, los alumnos obtuvieron descriptivamente peores resultados en el MCQ que en el resto de las pruebas. Tratando los resultados como porcentaje de acierto de las 3 pruebas, la calificación del MCQ fue superior descriptivamente a sus puntuaciones anteriores (puesto que no se contempló aquí la penalización del tipo test). Tanto en un caso como en otro, si interpretamos como puntuaciones similares una diferencia de mediana (Δ Mediana) de $\leq 0,5$ puntos y $\leq 5\%$ en el caso de los porcentajes de acierto, las únicas similitudes descriptivas encontradas entre pruebas en las puntuaciones y el porcentaje de aciertos fueron entre la ECOE y la PVCC en la nota total (Δ Mediana Puntuación = 0,01; Δ Mediana % = 3,33) y en la competencia de profesionalismo (Δ Mediana Puntuación = 0,44; Δ Mediana % = 1,36).

Los resultados de las correlaciones entre cada par de pruebas se resumen en la [tabla 5](#). Todas las correlaciones observadas fueron de signo positivo. Se observa que, en las puntuaciones totales, la correlación significativa más alta se obtuvo entre MCQ y PVCC ($Rho = 0,494$; $p < 0,001$), y la más baja, entre ECOE y PVCC ($Rho = 0,269$; $p < 0,001$). Las correlaciones significativas más altas por competencias se obtuvieron entre MCQ y PVCC en juicio clínico ($Rho = 0,399$; $p < 0,001$), y con menor intensidad en otras competencias. Sin embargo, entre las pruebas ECOE y PVCC no se observó ninguna correlación significativamente distinta de cero en 4 de las 5 competencias: anamnesis, prevención y promoción de la salud, aspectos ético-legales y profesionalismo, así como en exploración física.

Similares correlaciones, aunque con menor intensidad, se mantuvieron en el porcentaje de aciertos obtenido en cada prueba. De forma discrepante con lo anteriormente descrito, tan solo se encontró una ausencia de significación estadística en la correlación de juicio clínico entre ECOE y PVCC ($Rho = 0,096$; $p = 0,321$), y la estimación de correlación significativa en profesionalismo entre MCQ y

Tabla 3 Distribución de preguntas del test de opción múltiple (MCQ) en relación con las estaciones, casos clínicos y competencias de la evaluación clínica objetiva estructurada (ECO) y prueba virtual clínica complementaria (PVCC)

Estación ECO	Caso Clínico MCQ	Anamnesis	Exploración Física	Juicio clínico	Prevención	Profesionalismo	Total
E-1	T-26 Traumatología	2	1	1	1	1	6
E-2	T-1 Exploración maniquí	0	5	1	0	0	6
E-3	T-21 Neurología	0	8	1	0	0	9
E-4	T-13 Reumatología	0	1	1	0	1	3
E-5	T-4 Pediatría	2	0	2	2	0	6
E-6	T-12 Neurología	0	0	3	1	1	5
E-7	T-18 Psiquiatría	3	2	2	0	1	8
E-8	T-18 Informe Psiquiatría	0	0	2	0	0	2
E-9	T-8 Cardiología	4	1	1	0	0	6
E-10	T-8 Informe Cardiología	0	0	3	0	0	3
E-11	T-5 Metabolismo	3	0	1	2	0	6
E-12	T-3 Urgencias	0	2	0	0	1	3
E-13	T-2 Neumología	0	0	2	0	0	2
E-14	T-22 Reumatología	3	3	1	0	0	7
E-15	T-14 Cirugía	0	0	1	1	2	4
E-16	T-17 Digestivo	2	0	2	1	0	5
E-17	T-17 Informe Digestivo	0	0	1	0	0	1
E-18	T-9 Cardiología	0	1	1	0	0	2
E-19	T-16 Neumología	4	1	1	0	0	6
E-20	T-23 Infecciosas	4	0	0	1	0	5
E-21	T-19 Ginecología	2	3	0	0	0	5
Caso PVCC Preguntas de estaciones ECOE (n, %)		29 (29%)	28 (28%)	27 (27%)	9 (9%)	7 (7%)	100
C-1	T-3 Urgencias	1	0	1	0	3	5
C-2	T-24 Digestivo	1	1	2	0	0	4
C-3	T-15 Pediatría	1	1	2	1	0	5
C-4	T-21 Neurología	1	2	2	0	0	5
C-5	T-6 Obstetricia	1	1	2	1	0	5
C-6	T-11 Cardiovascular	1	2	2	0	0	5
C-7	T-20 Reumatología	2	1	2	0	0	5
C-8	T-7 Neumología	1	1	3	0	0	5
C-9	T-25 Oncología	0	0	4	1	1	6
C-10	T-10 Psiquiatría	1	2	1	0	1	5
Preguntas de casos clínicos PVCC (n, %)		10 (20%)	11 (22%)	21 (42%)	3 (6%)	5 (10%)	50
Total preguntas por competencia		39	39	48	12	12	150
Porcentaje (%) preguntas sobre total MCQ		26	26	32	8	8	100

ECO: evaluación clínica objetiva estructurada.

C: caso clínico de la PVCC; E: estación de la ECO; MCQ: test de opción múltiple; T: caso clínico del MCQ.

PVCC ($Rho = 0,205$; $p < 0,001$), así como en prevención y promoción de la salud entre ECO y PVCC ($Rho = 0,243$; $p < 0,001$). De cualquier forma, la correlación más alta siguió estando presente entre las puntuaciones totales de MCQ y PVCC ($Rho = 0,329$; $p < 0,001$), y la más baja, entre las de la ECO y PVCC ($Rho = 0,215$; $p < 0,001$).

En cuanto a la consistencia interna de cada prueba, se observó que los valores más elevados del coeficiente alfa de Cronbach en cada competencia siempre fueron superiores para la ECO respecto a MCQ y PVCC (tabla 6). En concreto, los valores más altos de fiabilidad fueron encontrados en: anamnesis ECO ($\alpha = 0,754$), exploración física ECO ($\alpha = 0,715$), juicio clínico ECO ($\alpha = 0,645$) y juicio clínico MCQ ($\alpha = 0,497$).

Por último, un alto porcentaje de participantes respondieron a las encuestas de satisfacción: 82,1% (ECO),

95,5% (PVCC) y 93,5% (MCQ). Ninguna de las medidas de satisfacción se distribuyó de forma significativamente normal ($p < 0,05$). Los alumnos consideraron que el tiempo necesario para realizar las pruebas (escala sobre 5) fue significativamente más escaso en la PVCC ($Mdn = 3$, $RIQ = 2-3$) respecto a la ECO ($Mdn = 3$, $RIQ = 3-4$, $p < 0,001$) y MCQ ($Mdn = 3$, $RIQ = 3-4$, $p < 0,001$), pero igual entre ECO y MCQ ($p = 0,618$), concediendo una dificultad significativamente igual ($p > 0,05$) a las 3 pruebas (ECO y PVCC: $Mdn = 7$, $RIQ = 7-8$; MCQ: $Mdn = 7$, $RIQ = 6-8$). No obstante, opinaron que la ECO fue la mejor prueba ($Mdn = 8$, $RIQ = 7-9$) para medir las competencias generales de un estudiante de 6.º en comparación con PVCC ($Mdn = 7$, $RIQ = 5-8$, $p < 0,001$) y MCQ ($Mdn = 7$, $RIQ = 6-8$, $p < 0,001$), no encontrando diferencias significativas en la comparación entre estas 2 últimas ($p = 0,060$). Algo idéntico a lo anterior

Tabla 4 Puntuaciones y porcentajes de acierto alcanzados en cada prueba, totales y por competencias

Resultados por puntuaciones (Mediana [RIC], sobre 10)			
	MCQ	PVCC	ECOE
Total	4,61 [4,08–5,02]	6,43 [6,18–6,76]	6,42 [6,20–6,71]
Anamnesis	4,19 [3,50–4,70]	7,39 [7,12–7,63]	6,67 [6,13–7,10]
Exploración física	4,87 [4,19–5,56]	6,35 [5,82–6,93]	7,59 [7,16–8,01]
Juicio clínico	4,44 [3,61–5,14]	6,24 [5,67–6,74]	5,61 [5,18–6,11]
Prevención	4,44 [3,33–5,56]	6,38 [5,54–6,81]	5,52 [4,84–6,46]
Profesionalismo	4,44 [3,33–5,56]	5,44 [4,72–6,33]	5,88 [5,00–6,69]
Resultados por porcentajes de acierto (Mediana [RIQ], sobre 100)			
	MCQ	PVCC	ECOE
Total	58,00 [53,67–62,67]	64,33 [61,10–67,04]	67,63 [64,69–70,82]
Anamnesis	56,41 [51,28–60,26]	67,24 [61,90–71,86]	76,65 [73,59–79,44]
Exploración física	61,54 [56,41–66,67]	76,53 [70,92–79,59]	71,33 [65,67–76,67]
Juicio clínico	58,33 [52,08–63,54]	52,56 [48,94–57,00]	65,07 [59,03–68,76]
Prevención	58,33 [50,00–66,67]	53,57 [45,44–59,26]	65,42 [58,10–69,24]
Profesionalismo	58,33 [50,00–66,67]	65,00 [55,00–70,00]	63,64 [56,82–69,70]

ECOE: evaluación clínica objetiva estructurada; MCQ: test de opción multiple; PVCC: prueba virtual clínica complementaria.

sucedio con la opinión sobre la capacidad para medir el manejo clínico ante pacientes reales, que fue significativamente superior en la ECOE (Mdn = 7, RIQ = 7–9) con respecto a la PVCC (Mdn = 7, RIQ = 4–8, $p < 0,001$) y a MCQ (Mdn = 7, RIQ = 5–8, $p < 0,001$), sin diferencias entre estas 2 últimas ($p = 0,104$).

Discusión

Los resultados confirman la hipótesis de partida y los hallazgos del estudio anterior¹⁵.

La literatura muestra que la ECOE es capaz de evaluar el tercer nivel de la Pirámide de Miller («mostrar cómo», en un entorno de simulación), mientras que los tests contextualizados con casos clínicos y los exámenes de casos clínicos computarizados evalúan el segundo nivel (aplicación del conocimiento, «saber cómo»)^{16,17}. Evalúan, por tanto, competencias de distinto nivel y el punto diferencial es la

actuación en una situación clínica simulada, en la que aplica su competencia (conocimientos, habilidades y actitudes)⁴. Por ello, el mejor método para medir la competencia clínica de los evaluados, en una revisión sistemática sobre los métodos de evaluación sumativa, resultó ser la ECOE, y, con menor evidencia, los MCQ y otros formatos de test¹⁸.

En nuestro estudio, los peores resultados en porcentaje de acierto correspondieron al test, en el que los alumnos se entrenan desde el inicio del grado. No obstante, se trataba de un test orientado hacia las 5 competencias evaluadas en las otras 2 pruebas, cuando habitualmente se orienta a conocimientos teóricos o aplicados al manejo de situaciones clínicas, como muestra la mejor consistencia interna del MCQ en la competencia de juicio clínico.

Merece la pena resaltar que la mayor correlación significativa en resultados totales y en la competencia de juicio clínico corresponde al MCQ y la PVCC, probablemente por explorar el mismo constructo. La no correlación completa entre ellas en el resto de competencias se

Tabla 5 Correlaciones en puntuaciones y porcentajes de acierto, totales y por competencias, para cada par de pruebas (coeficiente de correlación de Spearman, Rho)

Correlaciones de puntuaciones						
	Total	Anam	Exp. Física	J. Clínico	Prev	Prof
MCQ-PVCC	0,494*	–0,050	0,248*	0,399*	0,142	0,130
MCQ-ECOE	0,318*	0,045	0,219*	0,271*	0,178	0,108
ECOE-PVCC	0,269*	0,057	0,129	0,220*	0,141	0,016
Correlaciones de porcentajes de acierto						
MCQ-PVCC	0,329*	0,001	0,220*	0,238*	0,162	0,205*
MCQ-ECOE	0,313*	0,035	0,243*	0,301*	0,174	0,136
ECOE-PVCC	0,211*	0,019	0,142	0,095	0,275*	0,179

* $p < 0,05$. Anam: anamnesis; Exp. Física: exploración física; J. Clínico: juicio clínico, plan de manejo diagnóstico y terapéutico; Prev: prevención y promoción de la salud; Prof: aspectos ético-legales y profesionalismo. Se resaltan en **negrita** las correlaciones con significación estadística.

Tabla 6 Consistencia interna de las competencias evaluadas en cada prueba (alfa de Cronbach)

Prueba	Anam	Exp. Física	J. Clínico	Prev	Prof
ECOE	0,754	0,715	0,645	0,239	0,410
PVCC	0,163	0,311	0,393	0,113	0,241
MCQ	0,292	0,266	0,497	0,090	-0,120

Anam: anamnesis; ECOE: evaluación clínica objetiva estructurada. Exp. Física: exploración física; J. Clínico: juicio clínico, plan de manejo diagnóstico y terapéutico; MCQ: test de opción múltiple; Prev: prevención y promoción de la salud; Prof: aspectos ético-legales y profesionalismo; PVCC: prueba virtual clínica complementaria. Se resaltan en **negrita** los mejores resultados en cada competencia.

justificaría por las diferencias inherentes de ambas pruebas, novedosa para los alumnos en el caso de la PVCC, con un tiempo percibido como insuficiente para la PVCC, así como porque en ambas pruebas las competencias evaluadas son las de la ECOE y no son extrapolables a lo que se evalúa con los MCQ y PVCC. Se cree que la diferencia de resultados en las 3 pruebas no es atribuible al nivel de dificultad de cada una, ya que fue percibido por los participantes como similar.

Varios estudios han comparado resultados entre ECOE y MCQ. Al igual que en nuestro estudio, las correlaciones totales son bajas entre ambas pruebas lo que reafirma que evalúan constructos y competencias distintas, más centradas en la competencia al actuar, en el caso de la ECOE, que en la competencia al pensar, en el caso del MCQ¹⁹⁻²². En estudios que comparan por competencias aisladas entre ECOE y MCQ, tampoco la concordancia fue elevada^{23,24}.

Aunque no existe mucha bibliografía que compare ECOE con exámenes de casos clínicos computarizados, en un estudio se comparó una CCT con el desempeño ante pacientes simulados ocultos en una consulta convencional y las correlaciones fueron negativas²⁵. Concluyeron que la CCT debe utilizarse para evaluar la competencia en la toma de decisiones, pero no ante situaciones clínicas.

En nuestro estudio, los alumnos consideraron que la ECOE era la prueba que mejor podía medir la competencia general al final del grado y en el manejo clínico ante pacientes reales. Estos resultados coinciden con la percepción de estudiantes de facultades alemanas, que consideraron que los MCQ solo medían conocimientos y la ECOE otras competencias adicionales²⁶.

Las limitaciones de nuestro estudio son que ha sido realizado con una promoción de nuestra facultad y con las pruebas ECOE y MCQ propias.

Para profundizar en este tema, una perspectiva adicional para futuros estudios podría ser la aplicación de modelos de diagnóstico cognitivo (CDM, por sus siglas en inglés), utilizados especialmente en psicología educativa, que permitirían estudiar la calidad psicométrica de cada prueba y analizar sus diferencias en los atributos y procesos cognitivos específicos evaluados²⁷. Requieren de un análisis estadístico más complejo, pero podrían ayudar a confirmar las diferencias cognitivas que podrían existir entre cada prueba a raíz de los resultados de este trabajo.

En situaciones excepcionales, hay que implementar soluciones excepcionales. En Canadá, durante la pandemia,

se modificó la ECOE para realizar exploraciones físicas sin tocar, lo que aumentó la competencia de juicio clínico, pero esto no afectó a la independencia evaluativa de los test y casos clínicos que, en conjunto, dan acceso a la formación especializada²⁸. Los autores observan que es importante reunir evidencia adicional para respaldar la validez de cualquier inferencia que afecte a las puntuaciones de los exámenes.

Los resultados obtenidos muestran que las 3 pruebas evalúan distintas perspectivas de la competencia, lo que permite disponer de información de la realidad del evaluado desde distintos niveles de evaluación. Nuestros estudios muestran que las competencias evaluadas con la PVCC y la ECOE no se correlacionan y que la PVCC está más próxima a un examen MCQ que a una ECOE y, por ello, no se pueden ponderar como iguales en la evaluación sumativa de la ECOE final de grado. Las facultades de Medicina tenemos la misión y la obligación de determinar, con rigor evaluativo y metodológico, un nivel mínimo de competencia de los egresados que prediga un adecuado desempeño cuando se incorporen a la vida clínica y, hasta el momento, la evidencia nos muestra que es la ECOE, y no la PVCC ni los MCQ, la prueba que mejor predice dicho desempeño.

Consideraciones éticas

Los autores declaran que este estudio se llevó a cabo de acuerdo con los principios éticos que tienen su origen en la Declaración de Helsinki, y siguiendo las buenas prácticas clínicas contempladas en la Ley 14/2007 de Investigación Biomédica. Los datos de los sujetos en investigación fueron sometidos a tratamiento confidencial, de acuerdo con la Ley Orgánica de Protección de Datos 3/2018. La investigación se realizó en el marco del respeto a los derechos y libertades fundamentales, con garantías de confidencialidad en el tratamiento de los datos de carácter personal. El proyecto fue aprobado por el Comité de Ética de la Investigación de la Universidad Francisco de Vitoria, con fecha 23 de mayo de 2024.

En relación con la exención de consentimiento informado de los participantes, deseamos exponer que la aclaración sobre este punto nos fue exigida previamente en el protocolo presentado al Comité de Ética para su aprobación.

El equipo investigador se comprometió individualmente, con documento firmado por manuscrito, a mantener la identidad y datos de los participantes en el anonimato y a omitir cualquier acción encaminada a la identificación de los participantes del estudio. Por todo ello, el Comité de Ética aceptó la exención del consentimiento informado.

Financiación

Los autores declaran que el estudio presentado en este artículo no ha recibido ninguna ayuda económica.

Conflicto de intereses

Los autores declaran que no tienen ningún conflicto de intereses.

Agradecimientos

A Juan José Gil de Arana por su indispensable ayuda en el registro y tratamiento de los datos. A los profesores de la facultad que, junto con los autores de este artículo, colaboraron en la edición del test MCQ: Abellán Martínez J., Carballo González N., Cañas de Paz F., Díaz Miguel V., García Torrejón M.C., Gutiérrez Guisado J., Jiménez de los Galanes Marchán S., León Vázquez F., Llanes González L., López Escobar A., Martínez Porras J.L., Moreno Cuerda V.J., Ruiz Sánchez A., Toribio Díaz E., Touza Fernández A., Villén Villegas, T.

Bibliografía

- Wass V, Van Der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *The Lancet*. 2001;357(9260):945–9. doi:[10.1016/S0140-6736\(00\)04221-5](https://doi.org/10.1016/S0140-6736(00)04221-5).
- Schuwirth LWT, van der Vleuten CPM. General overview of the theories used in assessment: AMEE guide no. 57. *Med Teach*. 2011;33(10):783–97. doi:[10.3109/0142159X.2011.611022](https://doi.org/10.3109/0142159X.2011.611022).
- Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387–96. doi:[10.1056/NEJMr054784](https://doi.org/10.1056/NEJMr054784).
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med J Assoc Am Med Coll*. 1990;65(9 Suppl):S63–7. doi:[10.1097/00001888-199009000-00045](https://doi.org/10.1097/00001888-199009000-00045).
- Lockyer J, Carraccio C, Chan MK, Hart D, Smee S, Touchie C, et al. Core principles of assessment in competency-based medical education. *Med Teach*. 2017;39(6):609–16. doi:[10.1080/0142159X.2017.1315082](https://doi.org/10.1080/0142159X.2017.1315082).
- Boulet JR, Durning SJ. What we measure ... and what we should measure in medical education. *Med Educ*. 2019;53(1):86–94. doi:[10.1111/medu.13652](https://doi.org/10.1111/medu.13652).
- Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. *Med Educ*. 1985;19(3):238–47. doi:[10.1111/j.1365-2923.1985.tb01314.x](https://doi.org/10.1111/j.1365-2923.1985.tb01314.x).
- Naidoo M. The pearls and pitfalls of setting high-quality multiple choice questions for clinical medicine. *South Afr Fam Pract*. 2023;65(1):5726. doi:[10.4102/safp.v65i1.5726](https://doi.org/10.4102/safp.v65i1.5726).
- Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J*. 1975;1(5955):447–51. doi:[10.1136/bmj.1.5955.447](https://doi.org/10.1136/bmj.1.5955.447).
- Hodges B. Validity and the OSCE. *Med Teach*. 2003;25(3):250–4. doi:[10.1080/01421590310001002836](https://doi.org/10.1080/01421590310001002836).
- Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ*. 2011;45(12):1181–9. doi:[10.1111/j.1365-2923.2011.04075.x](https://doi.org/10.1111/j.1365-2923.2011.04075.x).
- García-Estañ López J. Prueba nacional de evaluación de competencias clínicas de la conferencia nacional de decanos de facultades de medicina de España. *Educ Med FEM*. 2013;16 (Supl 3):S59–62.
- García-Seoane JJ, Ramos-Rincón JM, Lara-Muñoz JP, López-Román A, Cardoner-Álvarez N, Lorenzo-Peñuelas A, et al. Cambios en el examen clínico objetivo y estructurado (ECOE) de las facultades de medicina durante la COVID-19. Experiencia de una ECOE de casos-clínicos computarizados simulados (ECOE-CCS) conjunta. *Rev Clínica Esp*. 2021;221(8):456–63. doi:[10.1016/j.rce.2021.01.004](https://doi.org/10.1016/j.rce.2021.01.004).
- Schuwirth LWT, Van Der Vleuten CPM, De Kock CA, Peperkamp AGW, Donkers HJLM. Computerised case-based testing: a modern method to assess clinical decision making. *Med Teach*. 1996;18(4):294–9. doi:[10.3109/01421599609034180](https://doi.org/10.3109/01421599609034180).
- Cervera-Barba E, Blanco-Canseco JM, Fidalgo-Montero P, Gámez-Cabero MI, Hijano-Bandera F, Holgado-Catalán MS, et al. Concordancia competencial entre el examen clínico objetivo estructurado y la prueba de casos clínicos computarizados de medicina: ¿estamos evaluando lo mismo? *Educ Médica*. 2023;24(6):100841. doi:[10.1016/j.edumed.2023.100841](https://doi.org/10.1016/j.edumed.2023.100841).
- Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet Lond Engl*. 2001;357(9260):945–9. doi:[10.1016/S0140-6736\(00\)04221-5](https://doi.org/10.1016/S0140-6736(00)04221-5).
- Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The objective structured clinical examination (OSCE): AMEE guide no. 81. Part II: organisation & administration. *Med Teach*. 2013;35(9):e1447–63. doi:[10.3109/0142159X.2013.818635](https://doi.org/10.3109/0142159X.2013.818635).
- Terry R, Hing W, Orr R, Milne N. Do coursework summative assessments predict clinical performance? A systematic review. *BMC Med Educ*. 2017;17(1):40. doi:[10.1186/s12909-017-0878-3](https://doi.org/10.1186/s12909-017-0878-3).
- Martins IP, Fonseca AC, Pires L, Ferro JM. Undergraduate neurology teaching: comparison of an inpatient versus outpatient clinical setting. *Eur J Neurol*. 2021;28(4):1108–12. doi:[10.1111/ene.14677](https://doi.org/10.1111/ene.14677).
- Chang O, Holbrook AM, Lohit S, Deng J, Xu J, Lee M, et al. Comparability of objective structured clinical examinations (OSCEs) and written tests for assessing medical school students' competencies: a scoping review. *Eval Health Prof*. 2023;46(3):213–24. doi:[10.1177/01632787231165797](https://doi.org/10.1177/01632787231165797).
- McLaughlin K, Vitale G, Coderre S, Violato C, Wright B. Clerkship evaluation—what are we measuring? *Med Teach*. 2009;31(2):e36–9. doi:[10.1080/01421590802334309](https://doi.org/10.1080/01421590802334309).
- Rahayu GR, Suhoyo Y, Nurhidayah R, Hasdianda MA, Dewi SP, Chaniago Y, et al. Large-scale multi-site OSCEs for national competency examination of medical doctors in Indonesia. *Med Teach*. 2016;38(8):801–7. doi:[10.3109/0142159X.2015.1078890](https://doi.org/10.3109/0142159X.2015.1078890).
- Matet A, Fournel L, Gaillard F, Amar L, Arlet JB, Baron S, et al. Impact of integrating objective structured clinical examination into academic student assessment: large-scale experience in a French medical school. *PLoS One*. 2021;16(1):e0245439. doi:[10.1371/journal.pone.0245439](https://doi.org/10.1371/journal.pone.0245439).
- Craig B, Wang X, Sandella J, Tsai TH, Kuo D, Finch C. Examining concurrent validity between COMLEX-USA level 2-cognitive evaluation and COMLEX-USA level 2-performance evaluation. *J Osteopath Med*. 2021;121(8):687–91. doi:[10.1515/jom-2021-0007](https://doi.org/10.1515/jom-2021-0007).
- Schuwirth L, Gorter S, Van der Heijde D, Rethans JJ, Brauer J, Houben H, et al. The role of a computerised case-based testing procedure in practice performance assessment. *Adv Health Sci Educ Theory Pract*. 2005;10(2):145–55. doi:[10.1007/s10459-004-2784-9](https://doi.org/10.1007/s10459-004-2784-9).
- Müller S, Settmacher U, Koch I, Dahmen U. A pilot survey of student perceptions on the benefit of the OSCE and MCQ modalities. *GMS, J Med Educ*. 2018;35(4). doi:[10.3205/zma001197](https://doi.org/10.3205/zma001197).
- de la Torre J, Minchen N. Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicol Educ*. 2014;20(2):89–97. doi:[10.1016/j.pse.2014.11.001](https://doi.org/10.1016/j.pse.2014.11.001).
- Gotzmann A, Boulet J, Zhang Y, McCormick J, Wojcik M, Bartman I, et al. Conducting an objective structured clinical examination under COVID-restricted conditions. *BMC Med Educ*. 2024;24(1):801. doi:[10.1186/s12909-024-05774-8](https://doi.org/10.1186/s12909-024-05774-8).