

# Diseño, aplicación y control de calidad de un examen OSCE de competencias de cirugía en la revalidación práctica del título de médico cirujano

## An objective structured clinical examination OSCE of general surgery for title revalidation of graduates of foreign medical schools

Felipe Celedón<sup>1</sup>, Claudia Morales<sup>2,3</sup>, Adriana López<sup>4</sup>, Katherine Daniel<sup>5</sup>, Nebenka Lombardic<sup>5</sup>, Ernesto Guiraldes<sup>6</sup>

### Resumen

**Introducción:** para revalidar su título en Chile, los médicos graduados en el exterior (MGEs) deben someterse a un examen compuesto de un módulo teórico y otro práctico. Para realizar este último, la Universidad Mayor utiliza la modalidad OSCE. Este formato garantiza apropiados grados de validez, fiabilidad y objetividad si su diseño y aplicación son meticulosos y detallados. Sus resultados deben ser sometidos sistemáticamente a un control de calidad. Este último, un componente crucial de toda evaluación de alta trascendencia, se menciona escasamente en la literatura biomédica nacional. Nuestro objetivo es describir aquí el diseño, administración y análisis psicométrico de un OSCE destinado a evaluar las competencias quirúrgicas generales de (MGEs). **Métodos:** el OSCE fue diseñado por un equipo multidisciplinario. Se crearon diez estaciones que estaban en correspondencia con las competencias quirúrgicas generales esperadas en un graduado. El análisis psicométrico de los resultados incluyó el uso del coeficiente alfa de Cronbach y la correlación de Pearson. **Resultados:** los veinte MGEs aprobaron el examen con puntajes entre 64,8 y 96,0%. El coeficiente de Cronbach global fue 0,67. La correlación entre los puntajes de cada estación con el del OSCE global alcanzó significación estadística en todos los casos. **Conclusiones:** un OSCE que evalúa la competencia quirúrgica general de un médico graduado es factible. Las propiedades psicométricas del examen fueron satisfactorias; sin embargo es aconsejable perfeccionar aún más el proceso de elaboración del examen así como los casos empleados y otros elementos claves.

**Palabras clave:** evaluación clínica objetiva estructurada; competencia clínica; certificación; confiabilidad; validez; cirugía.

### Abstract

**Background:** international medical graduates seeking to practice medicine in Chile must undergo a two-step examination process consisting of 2 modules, the first one assessing theoretical knowledge and the second evaluating practical skills. Our institution uses the objective structured examination (OSCE) modality to assess candidates taking the clinical module. The OSCE format tends to correlate with high degrees of fairness, validity and reliability when it is adequately designed and implemented. A proper psychometric analysis of the OSCE should follow its administration. Quality control of this high-stakes assessment is not customarily performed in Chile. Here, we describe the design, administration and *post hoc* psychometric analysis of an OSCE aimed at evaluating general surgical competencies of graduates of foreign medical schools. **Methods:** a multidisciplinary team carried out the design of the OSCE. Its aim was to ensure suitable levels of validity and objectivity of the assessment. A consensus blueprint was designed to match the expected surgical competencies of a graduate in general medicine. Psychometric analysis of OSCE data included the use of Cronbach's alpha and Pearson correlation coefficient. **Results:** all twenty candidates passed this 10-station examination. Their OSCE scores ranged from 64.8 to 96.0%. Overall Cronbach score for the OSCE was 0.67. Correlation between individual station scores and the total score attained significance in all cases. **Conclusions:** an OSCE evaluating general surgical competence of a medical graduate for licensing purposes is feasible. Its psychometric properties were satisfactory. Further refinement of the blueprinting process and of the OSCE cases and its constituent components is advisable.

**Keywords:** objective structured clinical examination; clinical competence; licensure; international medical graduate (IMG); reliability; surgery.

Fecha de envío: 5 de marzo de 2019 - Fecha de aceptación: 25 de septiembre de 2019

(1) Servicio de Cirugía, Hospital de la Fuerza Aérea de Chile,  
(2) Escuela de Medicina, Facultad de Ciencias, Universidad Mayor, Santiago, Chile  
(3) Facultad de Medicina, Universidad Andrés Bello, Santiago, Chile  
(4) Servicio de Medicina Interna & Departamento de Educación, Hospital de la Fuerza Aérea de Chile.  
(5) Centro de Simulación, Universidad Mayor, Santiago, Chile.  
(6) Comité de Calidad en Docencia, Escuela de Medicina, Facultad de Ciencias, Universidad Mayor, Santiago, Chile.  
Autor de correspondencia: [ernesto.guiraldes@umayor.cl](mailto:ernesto.guiraldes@umayor.cl)



## Introducción

El OSCE es un examen multidimensional que se ha convertido en una herramienta de evaluación tanto sumativa como formativa, ampliamente usada en ciencias de la salud (Harden *et al.*, 1975; Turner & Dankoski, 2008; Khan *et al.*, 2013a; Harden *et al.*, 2015). Su formato permite garantizar criterios evaluativos uniformes, estandarizados y explícitos (Patricio *et al.*, 2013; Harden *et al.*, 2015). Ello a su vez, asegura un alto grado de objetividad y confiabilidad, siempre y cuando se cumplan las condiciones adecuadas de calidad en su diseño y ejecución (Khan *et al.*, 2013a; Khan *et al.*, 2013b; Harden *et al.*, 2015; Daniels & Pugh, 2018). Los atributos mencionados suelen faltar en otras herramientas tradicionalmente usadas para evaluar competencias clínicas (Harden *et al.*, 2015).

En todo proceso evaluativo, muy especialmente cuando implica una toma de decisiones de trascendencia, es imperativo garantizar la equidad y la seguridad de los procedimientos y establecer estándares de calidad y de precisión que lo hagan defendible y aceptable (Turner & Dankoski, 2008; Yeates & Sebok-Syer, 2016).

En Chile, el importante proceso de evaluación de los médicos graduados en el exterior (MGEs) y que postulan a revalidar su título, consta de una sección teórica y otra práctica, que requieren aprobación en forma independiente (Kunakov & Bozzo, 2015). La primera parte, que se rinde a escala nacional es un examen teórico de múltiple elección. La segunda, que requiere haber aprobado la primera sección, se basa en una sesión de evaluación clínica de cada una de las especialidades básicas: medicina interna, cirugía, pediatría, y obstetricia-ginecología (Kunakov & Bozzo, 2015). Su modalidad ha tendido a evolucionar hacia el formato OSCE más recientemente, luego de haberse iniciado como un examen oral no estandarizado, que si bien se verifica en un ambiente de atención médica real, no explicita generalmente unos criterios evaluativos uniformes (Kunakov & Bozzo, 2015).

La Universidad Mayor, que participa en este proceso de reválida, usa la modalidad OSCE para este propósito. Su escuela de medicina tiene experiencia desde el año 2000 en ejecutar el OSCE como herramienta evaluativa sumativa en el pregrado y, desde la pasada década, también como metodología formativa. A partir de 2016, los exámenes OSCE son sometidos a una medición amplia de calidad del instrumento, que incluye aplicación de las pruebas psicométricas pertinentes. Este análisis se ha aplicado también retrospectivamente a los OSCEs anteriores, cuyos resultados también se han incorporado en el ciclo de *assessment*. Ello ha permitido a la Escuela determinar los efectos de los cambios introducidos.

Intuitivamente, en el medio clínico, se conceptualiza la implementación de OSCEs de cirugía destinados a médicos generales, como

un proceso complejo y costoso, principalmente por sus contenidos procedimentales específicos y la logística que ello supone. Sin embargo, las competencias respectivas están descritas en el perfil de ASOFAMECH y dicha guía permite catalogar un amplio abanico de competencias mensurables en un OSCE de este tipo. En la búsqueda de este explícito objetivo es vital considerar que, al igual que en las otras especialidades médicas básicas, éste no es un certamen destinado a evaluar competencias especializadas de residentes o graduados de la disciplina respectiva.

Hay escasa información sobre el aseguramiento de la calidad del examen OSCE en el ámbito de la educación médica en Chile (Triviño *et al.*, 2002; Kunakov & Bozzo, 2018). En este artículo presentamos la experiencia de la Escuela de Medicina de la Universidad Mayor en el diseño preparatorio y evaluación ulterior de calidad, de un examen OSCE destinado a la revalidación del título de MGEs y orientado a evaluar competencias en el ámbito de cirugía, del nivel de la práctica de un médico general en Chile.

## Material y método

Para diseñar el examen, la escuela de medicina constituyó un equipo de 4 integrantes, compuesto por un especialista en cirugía y tutor de internos, una especialista en medicina interna y coordinadora del curso integrado médico-quirúrgico de 4º año y dos académicos *senior* (CM y EG) con experiencia clínica y de gestión y formación profesional en educación médica.

El grupo de trabajo se reunió periódicamente desde seis semanas anteriores al examen y, a partir del perfil de conocimientos "EUNACOM", derivó un conjunto de competencias quirúrgicas medulares a evaluar en un médico general. Para este fin, se propusieron varias estaciones, cada una destinada a enfatizar la evaluación de alguna de las áreas de dominio de la disciplina quirúrgica: cognitiva o procedimental. Entre los requisitos de validez de constructo del examen, se estableció que el conjunto de las estaciones debía cubrir una muestra representativa de temas quirúrgicos y alinearse apropiadamente con las competencias descritas por ASOFAMECH para un médico general. Durante este período, el grupo sostuvo permanente comunicación con miembros del equipo institucional de Cirugía, así como con los integrantes del centro de simulación Clínica de la Universidad Mayor (CSC), con quienes se analizó detalladamente la logística del evento, con énfasis en aspectos específicos de la simulación del ámbito quirúrgico.

Previamente, la Universidad Mayor había realizado otros exámenes OSCE de reválida, en especialidades diferentes. La experiencia allí recabada se utilizó para introducir algunas innovaciones en la organización y estructura del OSCE de cirugía. La mayoría de los

MGEs habían llegado al ciclo de reválida sin conocer el formato OSCE (Morales *et al.*, 2018). Se decidió que el nuevo OSCE constaría nuevamente de 10 estaciones, esta vez con una duración de siete minutos cada una. En la experiencia previa se había observado que cinco minutos no eran en general suficientes para que los candidatos cumplieran a cabalidad todas las tareas demandadas por la estación. En consenso con los profesionales del CSC se determinó que tres estaciones se desarrollarían con paciente simulado; cuatro, con fantasmas, y tres evaluarían toma de decisiones en situaciones clínicas teóricas. Se prestó especial atención al lenguaje empleado en el material escrito, evitando jerga o expresiones excesivamente locales.

De esta manera se originó el “blueprint” del examen. Dos semanas antes del OSCE, la Dirección de la Escuela de Medicina envió a los MGEs inscritos un correo electrónico puntualizándoles que el examen estaba destinado a evaluar competencias del ámbito quirúrgico esperables en un médico general. Se les hizo llegar también sugerencias específicas, basadas en la experiencia de los evaluadores de los OSCES previos rendidos por el grupo de revalidantes y que concordaban en gran parte con la lista de errores comunes cometidos en OSCES, elaborada por el Medical Council of Canada (Medical Council of Canada – Le Conseil Médical du Canada, 2019). Primordiales en las sugerencias enviadas, eran las recomendaciones de leer cuidadosamente las instrucciones en cada estación y no hablar en exceso.

Como evaluadores de estación se seleccionó tanto a docentes experimentados en estas funciones, como a otros más novicios, a los cuales se entrenó en forma dirigida. Se reclutaron pacientes simulados estandarizados (PSE) que satisficieran los requisitos del contexto específico de las estaciones planeadas. Éstos son actores profesionales que forman parte del equipo habitual de PSE de la Universidad Mayor. Oportunamente ellos recibieron los guiones, además de información específica sobre la simulación, de la estación en que les correspondía intervenir.

El examen se llevó a cabo en el CSC y se organizó a lo largo de un solo circuito. La clasificación temática de las estaciones está descrita en la tabla 1. Se fijó como puntaje de aprobación por estación la obtención de, al menos el 60% del puntaje asignado. Para pasar el examen un candidato necesitaba tener un mínimo de puntaje global de 60% y seis estaciones (60%) aprobadas.

Para identificar áreas que requerían ajuste, se practicó un detallado análisis post-examen, que comprendió una revisión por expertos, de los resultados de cada estación mediante criterios cualitativos de calidad recomendados para OSCES (Tavakol & Dennick, 2012), así como pruebas psicométricas *ad-hoc*. Éstas incluyeron la estimación de la consistencia interna de los resultados (“confiabilidad”), con el coeficiente alfa de Cronbach (Cronbach & Shavelson, 2004) y el cálculo de la correlación entre el puntaje de cada estación y el promedio de los puntajes de todas las otras, mediante la prueba de Pearson (Sibert *et al.*, 2001; Roberts *et al.*, 2006). Para la diferencia de dos medias se usó la prueba t de Student. Un valor  $p < 0,05$  se consideró como significativo.

**Resultados**

Las características de las estaciones del OSCE, así como los resultados numéricos obtenidos en cada estación se presentan en la tabla 1.

Veinte MGEs rindieron el examen. Todos lo aprobaron, con puntajes entre 64,8 y 96,0%. El puntaje promedio fue de 85,1 con desviación estándar, 7,2. La mediana fue de 84,3.

El valor del alfa de Cronbach del examen global fue de 0,67 (tabla 2). Asimismo, para cada estación se muestran los resultados que el alfa de Cronbach daría si la estación respectiva se hubiera omitido.

La tabla 2 muestra también la correlación de Pearson entre cada una de las estaciones y el puntaje total (excluido el puntaje de la estación en cuestión). Esta correlación fue estadísticamente significativa para todas las estaciones.

**Tabla 1:** Examen OSCE de cirugía para médicos revalidantes (n = 20): estaciones que lo componen y resultados obtenidos.

Estaciones	1. Emergencia quirúrgica respiratoria	2. Examen de cuello	3. Complicaciones quirúrgicas diabetes	4. Abdomen agudo	5. Patología herniaria y urológica	6. Patología mamaria	7. Manejo inicial heridas	8. Patología proctológica	9. Patología digestiva	10. Patología urológica
Promedio obtenido en la estación (en %)	72,8	76,0	78,5	95,4	80,3	92,0	76,3	91,7	93,0	96,0
Desviación estándar	20,8	21,6	12,5	5,5	17,5	8,3	19,5	14,2	8,7	5,8

**Tabla 2:** Examen OSCE de cirugía para médicos revalidantes (n = 20): Resultados prueba alfa de Cronbach y prueba de Pearson.

	<b>Resultados: Prueba alfa de Cronbach<sup>a</sup></b>	<b>Resultados de la Prueba de Pearson (valor de p)<sup>b</sup></b>
Examen global	0,67	
Estaciones		
1. Emergencia quirúrgica respiratoria	0,70	<0,05
2. Examen cuello	0,56	<0,0005
3. Complicaciones quirúrgicas diabetes	0,63	<0,05
4. Abdomen agudo	0,65	<0,05
5. Patología herniaria y urológica	0,65	<0,05
6. Patología mamaria	0,65	<0,05
7. Manejo inicial heridas	0,67	<0,05
8. Patología proctológica	0,57	<0,0005
9. Patología digestiva	0,63	<0,025
10. Patología urológica	0,63	<0,05

<sup>a</sup> Los resultados de la prueba alfa de Cronbach para cada estación (1 a 10), corresponden a la estimación de la consistencia interna ("confiabilidad") que habría tenido el examen si se hubiera excluido la correspondiente estación.

<sup>b</sup> Estos resultados corresponden a la significación estadística de la correlación que existe entre los resultados de la estación correspondiente y el resultado global del examen (debidamente excluida la estación específica en estudio).

## Discusión

Toda herramienta de evaluación sumativa debe ser sujeta a un análisis de calidad (Pell *et al.*, 2016). Ello es especialmente importante cuando los resultados de la evaluación conllevan decisiones trascendentes (*high stakes*) para la sociedad y para el individuo implicado (van der Vleuten, 2000; Turner & Dankoski, 2008; Brannick *et al.*, 2011). Eventuales fallas en el proceso o en el instrumento evaluativo pudieran conducir al indeseado efecto de aprobar a un candidato incompetente o reprobar a uno idóneo. Por consiguiente, debe verificarse en cada caso que el instrumento de evaluación usado satisface criterios establecidos de calidad. (Brannick *et al.*, 2011, Khan *et al.*, 2013b; Pell *et al.*, 2016). Con ello se está cumpliendo un principio básico de justicia para la colectividad y los evaluados (Van der Vleuten, 2000; Epstein & Hundert, 2002; Turner & Dankoski, 2008).

Las características del formato OSCE que han hecho a este examen popular y eficiente en ciencias de la salud no garantizan automáticamente que en su administración se den siempre un apropiado grado de confiabilidad e inferencias de validez. Estos atributos, esenciales en la evaluación de los aprendizajes, sólo se cumplen si se satisfacen ciertas condiciones en su diseño y ejecución (Roberts *et al.*, 2006; Brannick *et al.*, 2011; Khan *et al.*, 2013a; Khan *et al.*, 2013b; Harden *et al.*, 2015; Daniels & Pugh, 2018). Así, el análisis *post hoc* del examen, no debe limitarse sólo a computar los resultados derivados de las calificaciones de los candidatos (Fuller *et al.*, 2016), sino que muy importantemente, debe incluir las apropiadas pruebas psicométricas que avalan la calidad del instrumento (Khan *et al.*, 2013b; Harden *et al.*, 2015, Fuller *et al.*, 2016; Pell *et al.*, 2016; Daniels & Pugh, 2018).

El examen aquí reportado alcanzó un alfa de Cronbach global de 0,67. Por otra parte, la correlación entre el puntaje promedio en cada estación y el puntaje global del examen fue en cada caso, estadísticamente significativa. Convencionalmente, se considera como deseable un valor de alfa de Cronbach (prueba que estima la consistencia interna de un test) igual o mayor de 0,7, aunque este punto está abierto a discusión (Khan *et al.*, 2013b). La correlación de Pearson contribuye a determinar la homogeneidad de un test y algunos autores la consideran, por ello como un indicador indirecto a favor de su validez de constructo (Sibert *et al.*, 2001; Roberts *et al.*, 2006; Tavakol & Dennick, 2011).

El grado de confiabilidad de un test aumenta conforme exista un mayor número de ítems a evaluar (en el OSCE: estaciones) (Turner & Dankoski, 2008; Khan *et al.*, 2013b). Conjuntamente con estimar un valor alfa global, también es útil computar qué estaciones están alineadas con el resultado general de la prueba y cuales -por el contrario- le restan efectivamente confiabilidad. Esto se explora recalculando el alfa de Cronbach, omitiendo sucesivamente de los resultados cada una de las estaciones. Dado que un mayor número de estaciones tiende a incrementar la confiabilidad del test, se esperaría que la omisión de alguna de ellas debiera reducir el valor del alfa de Cronbach. Si ocurre lo inverso en alguna estación, ello significa que ésta ha sido problemática y está ejerciendo cierta interferencia en la consistencia interna del test, y por ende en su confiabilidad. Las posibles causas de ello han sido descritas (Auewarakul *et al.*, 2005; Fuller *et al.*, 2016; Pell *et al.*, 2016; Daniels & Pugh, 2018). En el examen de Cirugía aquí presentado, el haber excluido la estación 1, habría incrementado la confiabilidad global del test a 0,7. Ello debe motivar una revisión pormenorizada de la estación para identificar los elementos perfectibles, para futuro uso (Daniels & Pugh, 2018).

Las cifras del análisis psicométrico de calidad del OSCE aquí comentado, pueden considerarse como satisfactorias al tratarse

de un examen estructurado que se aplicó por primera vez a una población de examinandos cuyo perfil profesional era relativamente inédito y que constituía para la institución una cohorte nueva. A diferencia de nuestros internos de medicina, procedentes de una cultura uniforme, y que configuran una cohorte conocida que ha seguido un currículo acotado, los revalidantes llegan con diferencias en sus historiales formativos, experiencias profesionales y culturas de origen. Además, en su mayoría ellos no se habían enfrentado al formato OSCE antes de iniciar su ciclo de reválida (Morales *et al.*, 2018). Los efectos de esta circunstancia fueron más evidentes al comienzo del ciclo, pero al adoptar la institución una actitud proactiva, difundiendo instrucciones ya adoptadas en otros países y enfocadas a revalidantes novatos, el impacto adverso de esta falta de experiencia se vio atenuado.

Ejecutar un OSCE nuevo, para examinandos cuyo perfil es relativamente desconocido, hace relativamente incierta la obtención de un alto grado de calidad psicométrica del examen. Ello debe ser considerado al analizar los resultados psicométricos del nuestro. Similares experiencias han sido reportadas por grupos académicos que inician la práctica del OSCE en una población de estudiantes o graduados no familiarizados con este examen estructurado (Hilliart & Tallett, 1998; Sibert *et al.*, 2001; Al-Naami 2008). En general, un índice de confiabilidad algo menor a la cifra "emblemática" de 0,7 no es en absoluto inusual en la práctica del OSCE (Grand'Maison *et al.*, 1992; Walkers *et al.*, 2005; Roberts *et al.*, 2006; Turner & Dankoski 2008; Behrens *et al.*, 2018), inclusive en instituciones experimentadas, como ha sido reportado en revisiones sistemáticas (Turner & Dankoski 2008; Brannick *et al.*, 2011). Un mayor número de estaciones puede, por otra parte, incrementar la confiabilidad del test (Sibert *et al.*, 2001; Roberts *et al.*, 2006, Turner & Dankoski 2008; Khan *et al.*, 2013b). Los OSCEs con menos de diez estaciones tienden a exhibir unos parámetros de confiabilidad menores que aquellos con un mayor número (Brannick *et al.*, 2011).

Los veinte MGEs aprobaron el examen al haber superado el umbral del 60% de rendimiento. Se puede estimar que este resultado refleja también que el *blueprint* del examen avaló una apropiada congruencia entre el instrumento de evaluación diseñado y los perfiles declarados por ASOFAMECH respecto a las competencias esperables para un médico general (coherentes con aquellas propias de un egresado), en el área de cirugía general. Ésta es una garantía primaria para acrecentar la validez de inferencias de los resultados del examen (van der Vleuten, 2000; Roberts *et al.*, 2006; Khan *et al.*, 2013a; Khan *et al.*, 2013b). También colabora a reforzarla, la adjudicación de un tiempo realista para cumplir con las tareas

requeridas por cada estación (Khan *et al.*, 2013). Las competencias aquí requeridas son afines a las aplicadas por los médicos generales en el cumplimiento de los programas de salud que integran la oferta de prestaciones de un centro de atención primaria.

En un examen de alta trascendencia, (*high stakes*) sus resultados deben estar sólidamente respaldados por un aseguramiento de la calidad del instrumento usado. El análisis psicométrico aquí empleado avaló dicho instrumento, aunque identificó una estación que debe ser enmendada, para un futuro "banco de estaciones". Es importante recalcar que los resultados del análisis de calidad de un OSCE son relativamente independientes del grado de dificultad del test y, por ende de la proporción de aprobados/reprobados (Triviño *et al.*, 2002), como puede apreciarse en nuestras tablas de resultados.

Es fundamental que un examen de alta responsabilidad sea prolijamente diseñado, por un equipo cohesionado, con visión amplia y miradas complementarias, así como competencias académicas y técnicas que se manifiesten en todas las facetas que componen el examen (Triviño *et al.*, 2002; Khan *et al.*, 2013a, Khan *et al.*, 2013b; Daniels & Pugh, 2018). En un proceso evaluativo de este tipo el enfoque académico organizativo no debe concentrarse sólo en seleccionar los contenidos específicos de la especialidad en cuestión. Si se relegan a un plano secundario -o se confía que se cumplan "por defecto"- aspectos como: la congruencia entre los temas a evaluar y las competencias esperadas, el adecuado diseño de ítems (instrucciones, preguntas, pautas y rúbricas), el entrenamiento de pacientes simulados y evaluadores, y el aseguramiento de la seguridad, objetividad y calidad psicométrica del examen, pueden verse seriamente comprometidas la validez y confiabilidad de éste (Turner & Dankoski, 2008, Khan *et al.*, 2013).

Para futuras instancias, es necesario un re-examen cuidadoso de nuestras estaciones, con un escrutinio crítico de los casos y de sus componentes esenciales. También es aconsejable aumentar el número de estaciones. Todo ello permitirá optimizar el grado de confiabilidad del examen. Por otra parte, el contar en el equipo organizador con representantes de la atención primaria de salud podría conferir más pragmatismo a la selección de temas representativos de la práctica de un médico general y, con ello, fortalecer la validez de resultados de los resultados del proceso evaluativo.

### Agradecimientos

Los autores agradecen las útiles sugerencias proporcionadas por Mr Godfrey Pell, Principal Research Fellow, Leeds Institute of Medical Education, U.K.

## Referencias

- Al-Naami MY. (2008). Reliability, validity, and feasibility of the Objective Structured Clinical Examination in assessing clinical skills of final year surgical clerkship. *Saudi Medical Journal* **29**, 1802-1807.
- Auewarakul C, Downing SM, Praditsuwan R & Jaturatamrong U. 2005. Item analysis to improve reliability for an internal medicine undergraduate OSCE. *Advances in Health Sciences Education. Theory and Practice* **10**, 105-113.
- Behrens C, Morales V, Parra P, Hurtado A, Fernández R, Giaconi E, Santelices L, Armijo S, Furman F. (2018). Diseño e implementación de OSCE para evaluar competencias de egreso en estudiantes de medicina en un consorcio de universidades chilenas. *Revista Médica de Chile* **146**, 1197-1204.
- Brannick MT1, Erol-Korkmaz HT & Prewett M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Medical Education* **45**, 1181-1189.
- Cronbach LJ & Shavelson RJ: (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement* **64**, 391-418.
- Daniels VJ & Pugh D (2018) Twelve tips for developing an OSCE that measures what you want. *Medical Teacher* **40**, 1208-1213.
- Epstein RM & Hundert EM. (2002). Defining and assessing professional competence. *JAMA* **287**, 226–235.
- Fuller R, Homer M, Pell G & Hallam J. (2016). Managing extremes of assessor judgment within the OSCE. *Medical Teacher* **27**, 1-9.
- Grand'Maison P, Lescop J, Rainsberry P, Brailovsky CA, Tudiver F, Rose D, Banks B & Pfortmiller D. (1992). Large-scale use of an objective, structured clinical examination for licensing family physicians. *CMAJ* **146**, 1735-1740.
- Harden R, Lilley P & Patricio P. (2016). The definitive guide to the OSCE: the objective structured clinical examination as a performance assessment. 1st ed. Churchill Livingstone: Edinburgh; New York.
- Harden RM, Stevenson M, Downie WW & Wilson GM. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal* **1**, 447–451.
- Hilliard RI & Tallett SE. 1998. The use of an objective structured clinical examination with postgraduate residents in pediatrics. *Archives of Pediatrics and Adolescent Medicine* **152**, 74-78.
- Homer M, Pell G, Fuller R & Patterson J. (2016). Quantifying error in OSCE standard setting for varying cohort sizes: A resampling approach to measuring assessment quality. *Medical Teacher* **38**, 181-188.
- Khan KZ, Ramachandran S, Gaunt K & Pushkar P. (2013). The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Medical Teacher* **35**, 437-446.
- Khan KZ, Gaunt K, Ramachandran S & Pushkar P. (2013). The objective structured clinical examination (OSCE): AMEE guide no. 81. Part II: organisation & administration. *Medical Teacher* **35**, 447-463.
- Kunakov N & Bozzo S. (2015). La revalidación práctica del título de médico cirujano a través de un método estandarizado: Experiencia de la Universidad de Chile. *Revista Médica de Chile* **143**, 1058-1064.
- Kunakov N, Moraga L & Ortiz L. (2018). Revalidación de títulos médicos extranjeros: eficacia y eficiencia de un examen colaborativo y estandarizado. *Revista Médica de Chile* **146**, 232-240.
- Medical Council of Canada – Le Conseil Médical du Canada (2019). Common mistakes made on Objective Structured Clinical Examination stations. Accedido en: <https://mcc.ca/examinations/common-mistakes-osce/> el 15 de febrero de 2018.
- Morales C, Álvarez N, Candia P, Celedón F, Meyer L, Daniel K, Lombardic N & Guiraldes E. (2018). Experiencia con el empleo del examen OSCE en el ciclo de evaluación de competencias clínicas de médicos que revalidan su título. *Revista Médica de Chile* **146**, 341-343.
- Patrício MF, Julião M, Fareleira F & Carneiro AV. (2013). Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Medical Teacher* **35**, 503-514.
- Pell G, Fuller R, Homer M & Roberts T. (2016). How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Medical Teacher*, **27**, 1-9.
- Roberts C, Newbie D, Jolly B, Reed M & Hampton K. (2006). Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Medical Teacher* **28**, 535–543.
- Sibert L1, Mairesse JP, Aulanier S, Olombel P, Becret F, Thiberville J, Peron JM, Doucet J & Weber J. (2001). Introducing the objective structured clinical examination to a general practice residency programme: results of a French pilot study. *Medical Teacher* **23**, 383-388.
- Tavakol M & Brennan RL. (2013). Medical education assessment: a brief overview of concepts in generalizability theory. *International Journal of Medical Education* **4**, 221-222.

Tavakol M & Dennick R. (2011). Post-examination analysis of objective tests. *Medical Teacher* **33**, 447-458.

Tavakol M & Dennick R. (2012). Post-examination interpretation of objective test data: monitoring and improving the quality of high-stakes examinations: AMEE Guide No. 66. *Medical Teacher* **34**, e161-e175.

Triviño X1, Vásquez A, Mena A, López A, Aldunate M, Varas M, Lillo R & Wright A. (2002). Aplicación del Examen Clínico Objetivo Estructurado (OSCE) en la evaluación final del internado de pediatría en dos escuelas de medicina. *Revista Médica de Chile* **130**, 817-824.

Turner JL1 & Dankoski ME. (2008). Objective structured clinical exams: a critical review **40**, 574-678.

Van der Vleuten C. (2000). Validity of final examinations in undergraduate medical training. *BMJ* **321**, 1217- 1219.

Walters K1, Osborn D & Raven P. (2005). The development, validity and reliability of a multimodality objective structured clinical examination in psychiatry. *Medical Education* **39**, 292-298.

Yeates P & Sebok-Syer S. (2016). Hawks, Doves and Rasch decisions: Understanding the influence of different cycles of an OSCE on students' scores using Many Facet Rasch Modeling. *Medical Teacher*; **39**, 92-99.