

Rowensztein, Hernán

Métodos y puntos de corte de un examen clínico estructurado y objetivo en una residencia de pediatría

Maestría en Educación para Profesionales de la Salud

Tesis 2018

Cita sugerida: Rowensztein H. *Métodos y puntos de corte de un examen clínico estructurado y objetivo en una residencia de pediatría [tesis de maestría] [Internet]. [Buenos Aires]: Instituto Universitario Hospital Italiano de Buenos Aires; 2018 [citado AAAA MM DD]. 46 p. Disponible en: <http://trovare.hospitalitaliano.org.ar/descargas/tesisytr/20211228121411/tesis-rowensztein-hernan.pdf>*

Este documento integra la colección Tesis y trabajos finales de Trovare Repositorio del Institucional del Instituto Universitario Hospital Italiano de Buenos Aires y del Hospital Italiano de Buenos Aires. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

Para más información visite el sitio <http://trovare.hospitalitaliano.org.ar/>





Métodos y puntos de corte de un Examen Clínico Estructurado y Objetivo en una residencia de pediatría

por

Dr. Hernán Rowensztein

Dirigida por

Dr. Carlos Alberto Brailovsky

Presentado en cumplimiento de los requisitos para la obtención del Título de Magister en Educación para los Profesionales de la Salud

ante el

Instituto Universitario del Hospital Italiano de Buenos Aires

Buenos Aires – Noviembre de 2018

© Copyright: Hernán Rowensztein, 2018

Agradecimientos

- **Al Prof Dr. Eduardo Durante:** Director de la Maestría en Educación para Profesionales de la Salud. Quién me ha orientado en el camino de la elección del tema de tesis.
- **Al Prof. Dr. Carlos Alberto Brailovsky:** Co-Director de la Maestría y Director de mi tesis. Quien ha dedicado tanto tiempo y paciencia para asesorarme a la distancia. Por sus sustanciales aportes, no solo a la investigación objeto de mi tesis, sino también a la implementación del método de evaluación en el Hospital de Pediatría “Prof. Dr. Juan P. Garrahan”.
- **A la Dra. Susana P. Rodríguez,** Directora Asociada de Docencia e Investigación y al **Dr. Juan Carlos Vassallo,** Coordinador Docente del Hospital de Pediatría “Prof. Dr. Juan P. Garrahan”, quienes no solo me han permitido desarrollar el proyecto de tesis en el Hospital, sino que también lo han tomado como una prioridad en la evaluación de los residentes.
- **A la Dra. Carolina Cernadas,** Coordinadora de Investigación del Hospital de Pediatría “Prof. Dr. Juan P. Garrahan”, quien me ha asesorado en el análisis estadístico de la tesis.
- **A mi familia:** que han sabido tenerme la paciencia necesaria para que pueda dedicar el tiempo que una maestría requiere.
- **A los residentes del Hospital de Pediatría “Prof. Dr. Juan P. Garrahan”:** fuente de inspiración para la mejora constante de mis actividades diarias.
- **A los niños:** por quienes siento la responsabilidad de formar el mejor recurso humano posible, capacitado para asumir con solidez y una mirada integral, el rol de médico de cabecera desde el nacimiento hasta la adolescencia.

Resumen

Introducción: En el Hospital Garrahan se ha incorporado el *Examen Clínico Estructurado y Objetivo (en inglés OSCE)* como evaluación formativa, y más recientemente sumativa al final de cada uno de los 4 años de la residencia de clínica pediátrica. Se trata de una evaluación por estaciones que representan situaciones clínicas o profesionales que el examinado debe resolver. Este tipo de evaluación bien realizada tiene alta confiabilidad y validez.

Objetivo: Comparar 4 métodos y definir puntos de corte que permitan aprobar o reprobar el OSCE.

Resultados: Se analizó un OSCE con 9 estaciones tomado a 62 residentes de pediatría de 2º año. Se compararon 4 métodos: 1) alcanzar 60% de una lista de cotejo SI/NO, 2) alcanzar 60% de una lista de cotejo ponderada, 3) aprobación de al menos 6 de las 9 estaciones basándose en la teoría de los grupos límite, 4) obtener un puntaje total en la prueba por encima de la suma de los puntos de corte de todas las estaciones basándose en la teoría de los grupos límite. Se obtuvo un coeficiente Alfa de Cronbach de 0.621 para el examen. Se halló una relación positiva entre el rendimiento en el OSCE y el desempeño académico y asistencial de los examinados.

Conclusión: Hay 2 métodos que solos o combinados permiten alcanzar una proporción aceptable de aprobados/no aprobados (método de grupos límites sobre la mediana de la prueba, tomado como criterio absoluto, o combinado con que los examinados deban aprobar un mínimo de 6 estaciones), y con buena correlación con el desempeño demostrado en las otras evaluaciones realizadas durante el período lectivo.

Palabras clave

Examen Clínico Estructurado y Objetivo. Competencias clínicas. Evaluación sumativa. Confiabilidad y validez. Puntuación de aprobación.

Tabla de contenidos

Introducción	1
Marco teórico	2
Evaluaciones formativas	2
Evaluaciones sumativas	2
Historia del examen clínico estructurado y objetivo (OSCE)	4
EI OSCE	5
Construcción de un OSCE	11
1. Planificación	11
2. Ejecución	13
3. Análisis de resultados	14
4. Retroalimentación o “Feedback”	14
Objetivo	15
Objetivo general	15
Objetivos específicos	15
Materiales y métodos	15
Métodos de punto de corte	17
1. Porcentaje de la lista de cotejo	17
2. Lista de cotejo ponderada	17
3. Grupo límite sobre estaciones	17
4. Grupo límite sobre mediana del test	17
Análisis estadístico	17
Resultados	18
Análisis general del OSCE	18
Conclusión	23
Discusión	23
Anexo I: Análisis de las 9 estaciones	i
Estación Nro 1: Vacunas.....	i
Estación Nro 2: Emergencia metabólica	ii
Estación Nro 3: Diagnóstico por imágenes	iii
Estación Nro 4: Interpretación de una espirometría	iv
Estación Nro 5: Entrevista alimentaria	v
Estación Nro 6: Maduración y desarrollo	vi

Estación Nro 7: Antropometría	vii
Estación Nro 8: Comunicación	viii
Estación Nro 9: Reanimación cardiopulmonar	ix
Bibliografía	x
Curriculum vitae (abreviado)	xv

Introducción

En el Hospital Garrahan de Buenos Aires existe la Residencia de Pediatría más grande del país, ingresan 70 residentes por año dependientes del Ministerio de Salud de la Nación (60 de pediatría y 10 de residencias articuladas de Neonatología y Terapia Intensiva Infantil que realizan los 3 primeros años de pediatría), y entre 5 y 10 becarios extranjeros adscriptos al programa. La residencia de pediatría dura 4 años, y durante este tiempo los residentes son sometidos a evaluaciones, algunas son exámenes de selección múltiple (al finalizar los módulos teóricos del curso superior universitario), evaluaciones del desempeño al finalizar las distintas rotaciones interna y externas del hospital (ej.: internación, neonatología, atención primaria de la salud, etc.), Mini-Cex en algunas rotaciones (consultorio de seguimiento y algunos Centros de Salud asociados), y OSCE una vez por cada año de residencia. Desde hace dos años se ha integrado un grupo multidisciplinario (pediatras clínicos, especialistas en salud mental, paliativos, cuidados críticos y neonatales, pedagogos) a los fines de desarrollar recursos instruccionales y proponer estrategias de enseñanza-evaluación en la residencia, además del personal dedicado full time a la docencia.

El OSCE es organizado por la Dirección Asociada de Docencia e Investigación (DADI) y participan médicos del staff de hospital, Jefes de residencia de clínica pediátrica, miembros del comité de residencia del hospital y particularmente en la estación sobre habilidades para la comunicación, algunos médicos integrantes del grupo de trabajo en Comunicación.

Hace poco más de tres años se ha comenzado a utilizar el OSCE como evaluación formativa. El ejercicio y puesta a punto del equipo docente en la utilización del OSCE por un lado, y la buena aceptación de los residentes por el otro, ha llevado a considerar su implementación como evaluación sumativa al final de cada uno de los 4 años de la residencia de clínica pediátrica, incluso como examen final integrador anual para lograr la promoción de año.

De la complejidad del hecho de evaluar, teniendo en cuenta cuales son los conocimientos mínimos que se le exigen a los médicos residentes y, por otro lado, cuáles son las competencias que han llegado a adquirir gracias a las estrategias de enseñanza, surge este proyecto de investigación.

Marco teórico

La toma de una evaluación puede ser decidida en función de un objetivo de diagnóstico académico o para la toma de decisión acerca de la aprobación de contenidos, selección de aspirantes e incluso la promoción curricular. Otras veces estos objetivos se suman y combinan entre sí. Podemos definir entonces las evaluaciones según el objetivo con el que se las emplee como formativas o sumativas.

Evaluaciones formativas. Estas evaluaciones procuran contribuir a la mejora de los aprendizajes de los alumnos realizándose de manera coordinada con la actividad de este durante su desarrollo. El residente es el centro de la evaluación formativa, receptor y partícipe activo de los procesos de retroalimentación, monitoreo y autorregulación de los aprendizajes. Aquellos estudiantes que reciben retroalimentación de manera sistemática desarrollan una conciencia metacognitiva más profunda y se van haciendo cargo de la autorregulación de su aprendizaje (Anijovich et al., 2010). La evaluación formativa es un concepto más complejo de lo que parece, y puede ser emprendida tanto por docentes formales del programa (evaluación formativa formal) como por cualquier otro como pacientes, familiares de éstos, u otros estudiantes (evaluación formativa informal) (Yorke, 2003). Las evaluaciones formativas informales, son evaluaciones que tienen lugar durante el curso de las actividades diarias, y que pueden no estar específicamente estipuladas en el diseño de un programa. En cambio, las evaluaciones formativas formales, se realizan -aunque no exclusivamente- por el personal académico o por los supervisores de las actividades dentro de una institución. En cuanto al OSCE, su capacidad formativa ha sido estudiada desde la percepción de los examinados con buenos resultados, siempre y cuando esté claro para ellos el propósito con el que se realiza el mismo (Chisnall, Vince, Hall, & Tribe, 2015).

Evaluaciones sumativas. Siguiendo a la distinción entre la evaluación formativa y sumativa de Bloom y col. (Bloom, Hastings, & Madaus, 1971) podemos decir que esta última se refiere típicamente a determinar en qué medida un alumno ha alcanzado los objetivos curriculares. Entonces su finalidad es averiguar si se han cumplido los objetivos finales, a largo plazo planteados, y saber si el programa de métodos y contenidos ha resultado satisfactorio para

las necesidades del grupo al que se destinó. Si bien será utilizada para certificar los estudios, al igual que cualquier evaluación, no puede ser un fin en sí misma, sino que debe promover nuevos aprendizajes. Hay indicios de que la provisión de feedback junto con una decisión sumativa limita su valor, pero se sabe poco acerca de qué factores contribuyen a esto (Schuwirth & Van Der Vleuten, 2011).

Cuando un examen es utilizado como herramienta para tomar decisiones, se debe pasar por un proceso previo de elección del instrumento más adecuado, y también haber decidido el puntaje que sirve como punto de corte para decidir entre aquellos que tienen un buen desempeño y aquellos que no lo tienen.

Norcini (Norcini, 2003) propone 6 pasos para establecer lo que él llama “estándares” de los exámenes escritos -para nosotros “punto de corte o puntaje de aprobación”:

Paso 1: decidir sobre **el tipo de estándar** de los cuales hay dos tipos: relativos y absolutos, y la decisión entre ellos está relacionada con el propósito del examen. Los *estándares relativos* se expresan como número o porcentaje de examinados, por lo que el punto de corte habrá que establecer, por ejemplo, en el puntaje que pasan los 50 mejores o el 20% superior. Los estándares relativos son los más apropiados para exámenes donde el propósito es identificar un cierto número de examinados. Esto incluye pruebas que se usan para seleccionar los puntadores más altos o más bajos para las admisiones, donde un número limitado de estudiantes puede ser seleccionado. Los estándares absolutos se expresan como un número o porcentaje de las preguntas de la prueba, de modo que el punto de corte sea establecer, por ejemplo, 70 respuestas correctas de las 100 preguntas (70%) en la prueba. Los estándares absolutos son más apropiados para las pruebas de competencia, cuando el propósito es establecer que los examinados saben lo suficiente.

Paso 2: decidir **el método para el estándar**. habría cuatro categorías para clasificar los métodos: métodos relativos (ej. en función de la media del grupo), métodos absolutos basados en el test (ej. Angoff), métodos absolutos

basados en el desempeño de los examinados (ej. contraste de grupos, borderline regression) y métodos de compromiso (ej. Hofstee).

Paso 3: seleccionar los **evaluadores**. Son los que determinarán los puntos de corte.

Paso 4: celebrar una reunión para consensuar los propósitos de la prueba, las características de los examinados y los criterios de aprobación.

Paso 5: establecer los puntos de corte de aprobación.

Paso 6: el después de la prueba. Asegurarse que los resultados se hayan ubicado dentro de los estándares habituales. (Norcini, 2003)

Historia del Examen Clínico Estructurado y Objetivo

En 1972, Harden y sus colegas realizaron el primer examen clínico estructurado y objetivo (OSCE), afirmando que cumplía con todos los criterios de un método ideal de evaluación de las competencias clínicas (Harden, Stevenson, Wilson Downie, & Wilson, 1975). En esta versión original descrita por Harden, los estudiantes se movieron alrededor de 18 estaciones de prueba y 2 de descanso en una sala de hospital. Cada estación tenía una duración de 4 minutos y medio, con una pausa de 30 segundos entre las estaciones. El tiempo total de examen fue de 100 min. Cada estación probó una sola competencia; por ejemplo, la realización de un procedimiento, la historia clínica o el examen clínico de un paciente. Desde entonces ha crecido sin pausa su utilización en diversos programas de formación tanto clínicos como quirúrgicos (Harden, 2016).

Harden reflexiona años después acerca del origen del OSCE (Harden, 2016), y afirma que a finales de los años sesenta y principios de los setenta, como profesor titular de medicina en Glasgow y más tarde en Dundee, era responsable de la evaluación de los estudiantes. Como tal, tres cosas le impactaban: La primera fue que la evaluación de las habilidades clínicas de un estudiante se consideraba importante y que un estudiante no podía graduarse sin aprobar el examen clínico. En segundo lugar, había importantes deficiencias en el examen clínico representado por el impacto de la suerte tanto en el tipo de paciente visto por el estudiante en el caso clínico, como en los dos examinado-

res asignados para evaluar la competencia del estudiante. En el examen clínico tradicional, las calificaciones otorgadas por un examinador a menudo varían considerablemente de las otorgadas por otro examinador que observa el mismo desempeño. En tercer lugar, a Harden le pareció que debería ser posible construir un examen que evaluara confiablemente el rango de competencias esperadas del estudiante, en el cual lo que se evaluaría en cada estación se definiría claramente de antemano y se reflejaría en una lista de verificación y calificación (Harden, 2015). A partir del trabajo de Barrows y Abrahamson sobre el uso de pacientes simulados, pensó que éstos podrían reemplazar a un paciente real en el examen (Barrows & Abrahamson, 1964). El OSCE entonces se diseñó para introducir la estandarización y reducir el número de variables que podrían afectar la evaluación del desempeño. Por lo tanto, en un OSCE bien diseñado, las calificaciones de los examinados deberían reflejar predominantemente el desempeño de los candidatos únicamente, con un efecto mínimo de otras fuentes de variación (Khan, Ramachandran, Gaunt, & Pushkar, 2013).

El Examen Clínico Estructurado y Objetivo

En adelante explicaré más en detalle en que consiste un OSCE: éste se trata de una evaluación por estaciones que representan situaciones clínicas o profesionales que el examinado debe resolver (Khan, Ramachandran, et al., 2013). El conjunto de los casos/situaciones clínicas debe permitir la evaluación de las competencias seleccionadas que se espera que el profesional posea. En todas las situaciones clínicas, se debe actuar de la manera más correcta posible. No se trata solamente de "acertar" el diagnóstico o de prescribir el tratamiento apropiado sino de ejercer como un buen profesional en todos los casos incluyendo no solamente los aspectos clínicos, en sentido estricto, sino también aquellos necesarios para el desempeño de la profesión, como la relación médico paciente, la consideración de la dinámica familiar, la gestión clínica, el trabajo en equipo, etc. EL OSCE permite evaluar las competencias de un profesional de la salud, mientras que el examinado "muestra como" se desempeña frente a diversos desafíos que se le plantean en "estaciones" diseñadas específicamente a tal fin.

Este tipo de evaluación, bien realizada, tiene alta confiabilidad y validez (Durante, 2006) (Brailovsky & Grand'Maison, 2000) (Tudiver, Rose, Banks, & Pfortmiller, 2009) y, con todas sus fortalezas y debilidades, su aplicación ha crecido en todo el mundo tanto como evaluación de rutina en programas de formación con criterio de promoción (Grand'Maison, Lescop, & Brailovsky, 1993), como también para otorgar la licencia médica en diversos países (Turner & Dankoski, 2008) (McCoy & Merrick, 2010) (Violato, Marini, & Lee, 2003) . En la provincia de Quebec (Canadá) se utilizó por primera vez en el mundo el OSCE para dar la certificación de especialistas en el año 1991 (Brailovsky, Grand'Maison, & Lescop, 1992).

Dentro de las debilidades señaladas al OSCE, se pueden mencionar que algunos autores refieren una baja validez de constructo, debido a factores como la limitación en el tiempo para el desarrollo de un caso clínico, a que se examina un pequeño espectro de conocimientos y habilidades, y que en el caso de estudiantes de posgrado no los evalúa con la profundidad suficiente. Con respecto a la confiabilidad, las críticas se refieren a que la necesidad de respuestas rápidas por parte de los examinados puede afectarla. Lo mismo que la implementación de un bajo número de estaciones, que los "pacientes estandarizados" sean poco creíbles..., un staff desorganizado y/o habitaciones ruidosas donde se desarrolla el OSCE (Barman, 2005). Dicho de otra forma, los OSCE producen en algunos examinados un nivel extra de estrés comparado a otro tipo de exámenes, y un agotamiento importante cuando son prolongados por un alto número de estaciones, y pierden confiabilidad cuando el número de estaciones es bajo. Además, comparado con otro tipo de exámenes estructurados orales, el OSCE consume más tiempo y es mucho más costoso (Smith, Price, & Houston, 1984). Sin embargo, conociendo estas debilidades, podemos anticiparnos y evitar que influyan negativamente tanto en la validez y confiabilidad de los resultados, como en el ánimo de los examinados (en un OSCE tradicional ven menos pacientes que en la mitad de una guardia que realizan). Harden por su parte, critica en primer lugar los costos, que en Gran Bretaña equivalen a £355 por candidato (dólares estadounidenses 458 aproximadamente) y en Canadá llega a \$CAN 1200 (dólares estadounidenses 900 aproximadamente) (Brailovsky et al., 1992), un segundo aspecto es la compartimentali-

zación de la medicina, hecho que se trata de salvar con casos integrados. En tercer lugar, otro argumento relacionado contra el uso del OSCE es que el examen no es una evaluación auténtica de la competencia de los estudiantes y que las estaciones en el examen no evalúan las habilidades esperadas de un médico en la práctica clínica, pero esto en realidad depende de cómo se desarrollan e implementan las estaciones en la práctica. Si el uso de pacientes se restringe a pacientes simulados, no se evaluará el acercamiento del estudiante a pacientes con hallazgos físicos reales, aunque podemos agregar que los equipos de simulación modernos han mejorado mucho en este aspecto.

Un cuarto concepto erróneo sobre el OSCE se relaciona con el rol del examinador. La tarea del examinador puede verse como un mecanismo con la función restringida a la observación de un estudiante y la calificación de su desempeño utilizando una lista de verificación y casillas de verificación, cada una representando un elemento de la performance. Sin embargo, además de completar una lista de verificación, se le puede pedir al examinador que llegue a un juicio general sobre la competencia del estudiante y que califique una escala de calificación global. También pueden tener la libertad de otorgar puntos de bonificación, donde se demuestra un excelente desempeño. Esto, veremos más adelante, toma enorme importancia en nuestro OSCE al optar por el método de los “grupos límites” para establecer el estándar de aprobación (Harden, 2015).

EL OSCE lleva la palabra objetividad en el propio nombre, lo que se basa en la estandarización de los casos y en el uso de las listas de cotejo, pero los críticos dicen que esto no siempre es así. Hay estudios que muestran que las escalas de calificación global calificadas por los expertos mostraron una mayor confiabilidad entre estaciones, una mejor validez de constructo y una mejor validez concurrente que las listas de cotejo. Además, la presencia de las listas de cotejo, en estos estudios no mejoró la confiabilidad o validez de la escala de calificación global sobre la de la escala de calificación global solamente. Estos resultados sugieren que las escalas de calificación globales administradas por expertos son una medida sumativa más apropiada al evaluar a los candidatos en los exámenes basados en el rendimiento (Regehr, MacRae, Reznick, & Szalay, 1998).

Como se nombró previamente un paso importante en la planificación de un examen -y el objeto de esta tesis- es establecer el **estándar o punto de corte** para decidir entre aquellos que tienen un buen desempeño y aquellos que no lo tienen. Podríamos tomar la definición sobre puntos de corte de Cizek (Cizek & Bunch, 2007): *"El seguimiento adecuado de un sistema prescrito y racional de reglas o procedimientos que resulta en la asignación de un número para diferenciar entre dos o más estados o grados de desempeño"*. La decisión sobre el estándar o punto de corte cambia de acuerdo con la finalidad de test, en cuanto a si se utilizará para promoción, para admisión o como examen formativo. En el caso de nuestro programa de formación, *el OSCE consiste en un examen integrador anual* que se utiliza para decidir la promoción de año de los examinados. Para evaluar el mejor método hemos comparado durante varios exámenes métodos diferentes: que podríamos clasificar como métodos absolutos, métodos relativos y basados en criterios (Cusimano, 1996) (McKinley & Norcini, 2014). A continuación, detallaré cada uno de ellos:

- 1) Métodos absolutos: en estos métodos el punto de corte se establece previamente, teniendo en cuenta un criterio absoluto, de conocimientos mínimos aceptados. Por ejemplo, aprobar el 60% de las preguntas o ítems de la lista de cotejo en el caso del OSCE, u obtener un mínimo de 60 puntos sobre 100 en una prueba. Este criterio no tiene en cuenta el nivel alcanzado por el grupo ni la dificultad del examen. Con los métodos absolutos se sabe de antemano que nota deben obtener para aprobar o reprobar el examen.
- 2) Métodos relativos: en este método el estándar se establece en base a una comparación de aquellos que toman la evaluación entre sí. Es decir que el valor del punto de corte no se sabe hasta la finalización del examen, y surge del rendimiento de los examinados. Por ejemplo, cuando el puntaje de aprobación se establece en función del promedio de los examinados, el percentil 25, etc., el estándar es relativo.
- 3) Métodos basados en criterios: en este caso el estándar de aprobación se establece en el examen mismo y en base a la observación de los candidatos. Se establecen pautas de conocimientos mínimos

aceptados (criterios), se observa a los candidatos y se los encuadra en alguno de ellos.

Para el caso del OSCE se han utilizado distintos estándares de evaluación, para el proyecto aquí presentado he comparado 4 métodos diferentes, 2 métodos absolutos y 2 basados en criterios. Ambos basados en criterios utilizan el método de los "Grupos límite" (en inglés Borderline groups) con una variante entre ambos que detallaremos más adelante. Primero detallaré el marco teórico de la utilización de los grupos límites.

El método del grupo límite requiere la identificación de las características del examinado "límite" (por ejemplo, conocimiento y habilidades). El examinado "límite" es aquel cuyo conocimiento y habilidades no son del todo adecuados, pero tampoco son lo suficientemente inadecuados como para no pasar la prueba (Livingston & Zieky, 1982) (Hejri, Jalili, Muijtjens, & Van Der Vleuten, 2013). Los observadores clasifican las actuaciones reales de los examinados en claramente desaprobado, límite, aprobado o sobresaliente. El puntaje de aprobación se establece con la puntuación mediana (es decir, percentil 50) del grupo límite. Este concepto es importante, ya que algunos autores no recomiendan tomar la media debido a la influencia que pueden tener los examinados con valores extremos (Rothman & Cohen, 1996). Una modificación de este método que se ha utilizado en los OSCE es utilizar las evaluaciones recogidas durante la administración del mismo y no predeterminadas (Reznick, Blackmore, Dauphinée, Rothman, & Smee, 1996). En esta modificación, no se utiliza un panel de jueces; en cambio, los mismos observadores proporcionan información utilizada para derivar el puntaje de aprobación para cada estación. Si se utilizan expertos (por ejemplo, médicos, miembros de la facultad) para calificar las estaciones del OSCE, se les puede preguntar si el rendimiento que han visto se consideraría "límite". También se han desarrollado otras variantes, como por ejemplo el Medical Council of Canada donde tienen una escala de 6 niveles en lugar de 4, y dos de ellas son borderline ("borderline unsatisfactory", "borderline satisfactory") y se promedia para obtener el puntaje de aprobación de la estación (Smee, 2001). Estos métodos funcionan bien cuando hay un número suficiente de examinados que se asignan al grupo límite, cuando este número es bajo, el valor no es estable. En estos casos -OSCE a pequeña escala- el

método de grupo borderline tiene muchas ventajas porque es fácil de usar, no requiere mucho apoyo estadístico y es un método eficiente y defendible de configuración estándar. A pesar de estas ventajas, pueden surgir problemas potenciales cuando se aplica a OSCE en pequeña escala. Un enfoque con regresión lineal, similar en naturaleza al método de grupo borderline, demostró todos los beneficios del último método de configuración estándar y también demostró tener menos errores estadísticos asociados (Wood, Humphrey-Murto, & Norman, 2006). En el Hospital Garrahan hemos seleccionado una escala de Likert de 4 valores: Insuficiente, límite, suficiente y sobresaliente.

Para la definición de las competencias que serán evaluadas en el OSCE del Hospital Garrahan, se tiene en cuenta el curriculum del Programa de la Residencia de Pediatría y el “Global Pediatric Education Consortium”, iniciativa que ha desarrollado un conjunto de estándares centrales globalmente pertinentes para el adiestramiento, la evaluación, la acreditación médica individual y el desarrollo profesional continuo que servirán como estímulo para mejorar la asistencia pediátrica en todo el mundo. Estos estándares definidos por expertos reconocidos en educación médica pediátrica describen las competencias, los conocimientos y las habilidades que deben dominarse para asegurar que el aprendizaje y el ejercicio sean parejos a través de las fronteras. En este sentido han definido las 12 áreas de competencia para el pediatra en ejercicio. Una competencia central en los OSCE es una de ellas (la N° 8) que contempla “La comunicación y habilidades interpersonales: la capacidad del residente para comunicarse de forma efectiva con pacientes, familiares y otros profesionales sanitarios; y de demostrar escucha activa”. (Global Pediatric Education Consortium. (Adaptación al español), 2012). Lo mismo el Consejo de Acreditación de Educación Médica de Posgrado (en inglés ACGME) de los Estados Unidos, establece dentro de sus seis competencias requeridas que los residentes de pediatría deben demostrar “habilidades interpersonales y de comunicación que resulten en el intercambio efectivo de información y la colaboración con los pacientes, sus familias y los profesionales de la salud” (Accreditation Council for Graduate Medical Education, 2013).

En la actualidad se evalúan aspectos de la comunicación en varias de las estaciones del OSCE, pero se dedican en nuestro caso 1 o 2 estaciones para su exclusiva evaluación.

Construcción de un OSCE

El OSCE consta de 4 etapas en su realización:

1. Planificación
2. Ejecución
3. Análisis de resultados
4. Retroalimentación o “Feedback”

1. Planificación

El OSCE lleva un prolongado tiempo de planificación, que se extiende habitualmente desde 1 a 2 meses previos al día del examen. Durante el mismo se debe:

- a) Conformar el equipo de trabajo (recursos humanos) y los colaboradores que participarán el día del examen (especialistas que aportan el estándar diagnóstico de referencia, actores, secretarías/personal administrativo, técnico de simulación).
- b) Elaboración de la tabla de especificaciones: es el documento básico de un OSCE. Representa el diseño general de la prueba, y está constituida por una parrilla de filas y columnas, de manera que se relacionan los casos con los instrumentos evaluativos que utilizan y con los componentes competenciales que evalúan. Cada una de las filas corresponde a una situación clínica, de manera que habrá tantas como estaciones compongan la OSCE. Las columnas se dividen en dos grupos. Las primeras de la izquierda contienen algunos datos de los casos: nombre, instrumento evaluativo, etc. El resto se corresponden con cada uno de los componentes y subcomponentes competenciales que queremos evaluar, reflejando en ellas el porcentaje que cada caso va a aportar para la evaluación de cada componente competencial. Generalmente cada caso sirve

para evaluar no más de dos o tres de estos componentes, y de igual manera cada componente puede ser, aunque no necesariamente, evaluado por dos o tres casos (DeSerdio, 2002). Ver más adelante la tabla de especificaciones utilizada (Tabla 1).

- c) Elaboración de las estaciones: para confeccionar las estaciones se utilizan uno o varios instrumentos de evaluación combinados, por ejemplo, una madre o padre simulado con quien debe entrevistarse, la interpretación de imágenes de métodos diagnósticos, realizar maniobras de reanimación cardiopulmonar a un maniquí en el contexto de un paro cardiorrespiratorio, una comunicación telefónica con un paciente, realizar un procedimiento manual como suturar o realizar una punción lumbar, etc. EL OSCE puede estar diseñado por diferente número de estaciones según los contenidos a evaluar, siendo lo más habitual contar desde 8 hasta 20 estaciones. Cada estación se diseña con un instructivo para el residente, un instructivo para el actor o actriz, un instructivo para el observador, y finalmente una lista de cotejo de las acciones, conceptos y/o ítems que se espera que el examinado cumpla. En la figura 1 se puede ver la planilla de planificación del OSCE.
- d) Establecer el criterio de evaluación: como se citó en el “Paso 5” de Norcini comentado previamente, el criterio de aprobación debe ser establecido previamente por un grupo docente seleccionado para este fin.

Figura 1. Planilla general para planificar un OSCE

Evaluación de Competencias
Planificación del ECEO

Aula Estación Resumen contenido Docentes a cargo

	GH	I	J	KK	CeSim	EF
A	① Hidratación DSH: Revisar la <u>indicación</u> de OMS e <u>indicación</u> soluc <u>Poliject</u> .	③ Antropometría Medir talla y calcular velocidad (caso normal)	④ Imágenes Rx de torax con cardiomegalia y congestión pulm	⑥ Alimentación 6m <u>Indicación de</u> form continuac madre que empieza a trabajar.	⑧ Comunicación Pase a UCI de un lactante con bronquiolitis	⑨ RCP Reanimación de niño en insuficiencia respiratoria> PCR
	Docentes: XX y XX	Docentes: XX y XX	Docentes: XX y XX	Docentes: XX y XX	Docentes: XX y XX	Docentes: XX y XX
B	Docentes: XX y XX	Docentes: XX y XX	Docentes: XX y XX	Docentes: XX y XX	Docentes: XX y XX	Docentes: XX y XX
A	② Medio Interno Corrección de K		⑤ Vacunas Carnet de vacunas incompletas. 18m	⑦ Maduración Describir hitos madurat 5ª, video y dibujo		
	Docentes: XX y XX		Docentes: XX y XX	Docentes: XX y XX		
B	Docentes: XX y XX		Docentes: XX y XX	Docentes: XX y XX		

2. Ejecución

Se cita a los examinados al menos 30 minutos antes del comienzo del examen, y se les brinda una charla introductoria donde se explica la dinámica del OSCE, confidencialidad de las observaciones, prohibición del uso de teléfonos celulares, y toda información relevante previa al examen. Luego el examinado comienza en una estación y va recorriendo todas. Una condición del OSCE es que no hay correlatividad entre las estaciones, ya que un examinado comienza por la estación 1, luego pasa a la 2, luego a la 3 y así hasta terminar de recorrer todas. Pero otro examinado comienza por la 2, y luego pasa a la 3 y finaliza en la 1, así cada estación está permanentemente ocupada por un examinado.

Cada estación es evaluada mediante un observador que cuenta con una lista de cotejo que incluye las competencias o componentes de ellas que son evaluadas en esa estación y las acciones, respuestas, actitudes que van a ser valoradas. Esta lista de cotejo se encuentra “ciega” en cuanto al puntaje de la ponderación para el observador.

3. Análisis de resultados

Inmediatamente finalizado el OSCE procedemos al análisis de los resultados. Para ello contamos con planillas Excel diseñadas para tal fin, montadas en computadoras portátiles conectadas a una red wifi que nos brindan la información de las observaciones en tiempo real. En cada estación el docente carga la lista de cotejo observada sin poder visualizar las otras estaciones ni los resultados. En la computadora “maestra” se puede observar todo el OSCE y detectar errores en la carga y los resultados se van calculando a medida que se cargan las observaciones.

Finalmente obtenemos quienes pasan y quienes no pasan la prueba, también el score obtenido en promedio para cada una de las rondas y así poder detectar si existió “contaminación” del examen por filtración de las estaciones.

4. Retroalimentación o “Feedback”

Esta instancia final del OSCE es de sumo valor, ya que brinda tanto a examinados como a docentes una retroalimentación muy valorada (Junod Perron et al., 2016). La devolución constructiva o feedback es definida como el acto de dar información a un estudiante o a un residente describiendo su desempeño en una situación clínica observada (Alves de Lima, 2008). La retroalimentación formativa se define como "información comunicada al alumno que pretende modificar su pensamiento o comportamiento con el fin de mejorar el aprendizaje". La retroalimentación formativa refuerza el aprendizaje apropiado y contribuye a la corrección de deficiencias de los alumnos (Kogan et al., 2012). En nuestra experiencia, el feedback no solo tiene buena recepción en los residentes, sino que también funciona como un cierre del examen que fortalece la aceptación del mismo por los examinados. Lo realizamos 2 horas posteriores a la finalización del recorrido de las estaciones, en un aula con la participación de todos los residentes examinados y del equipo docente y actores. La retroalimentación es de 360 grados, examinados hacia docentes y actores, docentes hacia los examinados brindando los resultados del desempeño del grupo en su promedio, actores hacia los residentes y también hacia el equipo docente en cuanto a la construcción de los escenarios.

Objetivo

1) Objetivo general: Comparar distintos métodos y definición de los puntos de corte que permitan aprobar o reprobado a los residentes en el examen final integrador anual.

2) Objetivos específicos:

a.- Comparar métodos estadísticos vs los basados en el rendimiento de los examinados.

b.- Establecer metodología y puntos de corte para aprobar o reprobado cada estación de OSCE.

c.- Establecer metodología y puntos de corte para aprobar o reprobado el Test en su totalidad

Materiales y métodos:

Se han llevado a cabo 4 exámenes OSCE por año en el Hospital Garrahan, uno por cada año de la residencia de pediatría. Con el objetivo del análisis que ocupa este proyecto, se ha seleccionado el examen tomado a los residentes de 2do año el día 22 de marzo de 2018 sobre 62 residentes, luego de evaluar la métrica de los nombrados exámenes teniendo en cuenta el de mejor performance (Pell, Fuller, Homer, & Roberts, 2010).

Cada OSCE consta de un circuito de 9 estaciones con una duración de 7 minutos cada una, donde no tienen correlatividad una con otra. Debido al elevado número de residentes implementamos la utilización de “estaciones duplicadas”, es decir que hay 2 estaciones de cada una exactamente iguales, de esta forma de cada ronda participan 18 residentes, con algunas rondas donde se implementa una estación de descanso y se incorporan 20 residentes. Es decir que hay 2 circuitos, uno A y otro B por donde circula cada residente de principio a fin.

Cada estación es desarrollada por el equipo docente del Hospital Garrahan, se comienza a trabajar unos 2 meses antes de la fecha del OSCE y las estaciones son luego ensayadas con los actores participantes y “refinadas” por el equipo coordinador del examen compuesto por el Coordinador de recursos humanos y una pedagoga.

La estación de comunicación transcurre aprovechando las instalaciones del Centro de Simulación (CeSim Garrahan) donde se consta de cámara Gesell, micrófonos, cámaras de video y posibilidad de grabar las entrevistas y es desarrollada con la presencia de un paciente estandarizado, que en nuestro caso es una madre o padre personalizada por un actor/actriz contratado.

Participaron del OSCE analizado unos 30 colaboradores entre actores, docentes observadores con sus listas de cotejo, auxiliar para el control del tiempo de las estaciones, etc.

Al finalizar las rondas (en este caso 3) se realiza la carga del desempeño de los residentes en planillas preconfiguradas, se obtienen los rendimientos globales del grupo y por la tarde en una reunión plenaria con todos los residentes se realiza una devolución bidireccional, los residentes en cuanto a su percepción del OSCE y desde el equipo docente a ellos en cuanto al rendimiento. Luego, las semanas siguientes, en forma individual se acercan a la oficina de Docencia para conocer su rendimiento individual.

La tabla de especificaciones con la cual se preparó el examen puede verse a continuación (Tabla 1). La tabla es genérica para todos los OSCE tomados en el Hospital Garrahan, con la ponderación seleccionada en este caso, para residentes de 2° año de pediatría.

Tabla 1. Tabla de especificaciones.

R2 22/03/2018			Componentes							
Estación	Area de conocimiento	Instrumento	1°	2°	3°	4°	5°	6°	7°	Total
			Anamnesis	Semiología	Diagnóstico	Manejo	Comunicación	Prevención	Hab técnicas	
1	Vacunas	Madre simulada				70	30			100
2	Emergencia metabólica	Paciente simulado			20	70	10			100
3	Radiología	Monitor imágenes			50	40	10			100
4	Neumonología - Espirometría	Pictorial			60	30	10			100
5	Nutrición - Alimentación	Entrevista madre simulada	20			60	10	10		100
6	Maduración - Desarrollo	Video + madre simulada		100						100
7	Comunicación	Cámara Gesell					100			100
8	Antropometría	Maniquí simulación			30	50	10		10	100
9	RCP	Maniquí simulación				55	5		40	100
		Puntuación	20	100	160	375	185	10	50	900
		Ponderación	2%	11%	18%	42%	21%	1%	6%	100%

- **Métodos de punto de corte:**

Para el análisis comparativo de métodos de punto de corte se utilizan planillas Excel desarrolladas especialmente para este proyecto. Se definen a continuación los 4 puntos de corte comparados:

- 1) Porcentaje de la lista de cotejo: en este método no ponderado, se toma como punto de corte obtener el 60% de respuestas correctas de la lista de cotejo. Tomado cada ítem de la lista como SI (lo cumple) o NO (no lo cumple).
- 2) Lista de cotejo ponderada: en este método se toma como punto de corte obtener el 60% del puntaje de la lista de cotejo. Es la misma lista que la anterior, pero cada ítem de la lista se encuentra ponderado luego de un acuerdo del grupo de expertos. Por lo tanto un ítem marcado como NO suma 0 (cero) y marcado como SI toma el puntaje de ese ítem (por ejemplo 25 puntos).
- 3) Grupo límite sobre estaciones: en este método se toma como punto de corte obtener la aprobación de al menos 6 de las 9 estaciones. Para ser aprobada cada estación se utiliza una lista de cotejo ponderada, y como punto de corte deben obtener 60% de los puntos de la estación. Luego si aprueban al menos 6 de las 9 estaciones aprueban el test.
- 4) Grupo límite sobre mediana del test: en este método se toma como punto de corte obtener un puntaje total en el test por encima de la suma de los puntos de corte de todas las estaciones. El punto de corte se obtiene por la suma de los puntos de corte de cada una de las 9 estaciones calculado por los grupos límite. Se utilizarán las medias y medianas para ser comparadas entre sí cuando se utilice la teoría de los grupos límite como punto de corte de cada estación y del OSCE en su totalidad, definiendo su utilización según la dispersión de los examinados "límites".

- **Análisis estadístico:**

Para el análisis estadístico de la consistencia interna del test (considerada como "confiabilidad") se utilizará el coeficiente Alfa de Cronbach. Esta es una medida por la cual, en una evaluación si la concordancia interna del test es

lo suficientemente sólida y se vuelve a tomar el test a otro grupo, se obtendrán resultados similares. Para verificar si cada caso simulado tenía su propia consistencia interna, también se calculó el coeficiente Alfa de Cronbach utilizando los ítems como si fuesen los casos independientes.

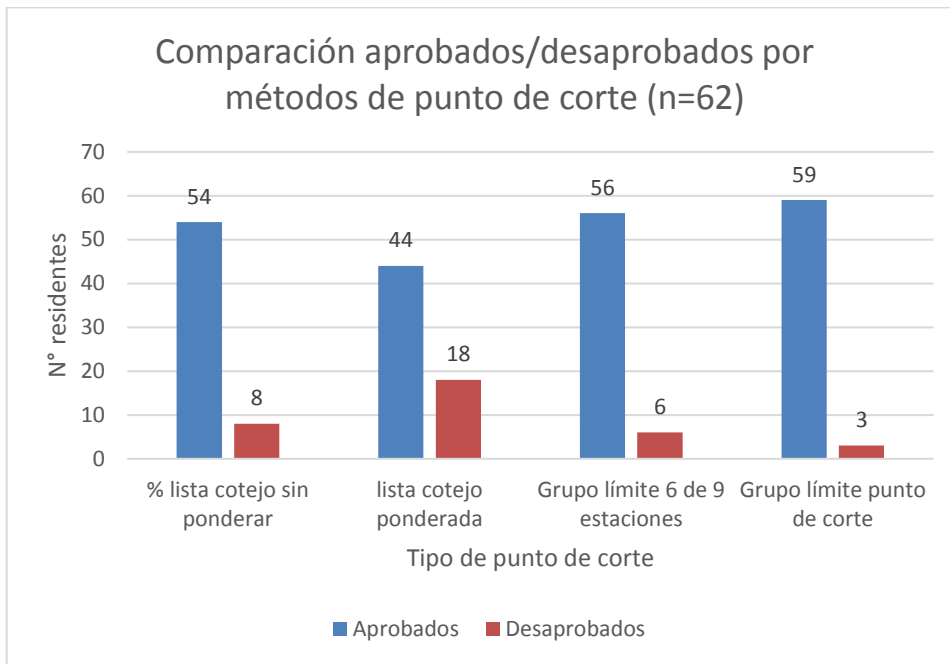
Para estimar el efecto de las variables predictoras sobre las de respuesta, se utilizó un modelo de regresión lineal simple. Se calcularon los intervalos de confianza a partir del error estándar del coeficiente.

- **Resultados: Análisis General del OSCE**

A continuación, analizaremos los resultados del mencionado examen tomado a los residentes de 2° año de pediatría. Han participado 62 residentes del OSCE completo, y se ha obtenido un coeficiente Alfa de **Cronbach de 0.621** para el examen.

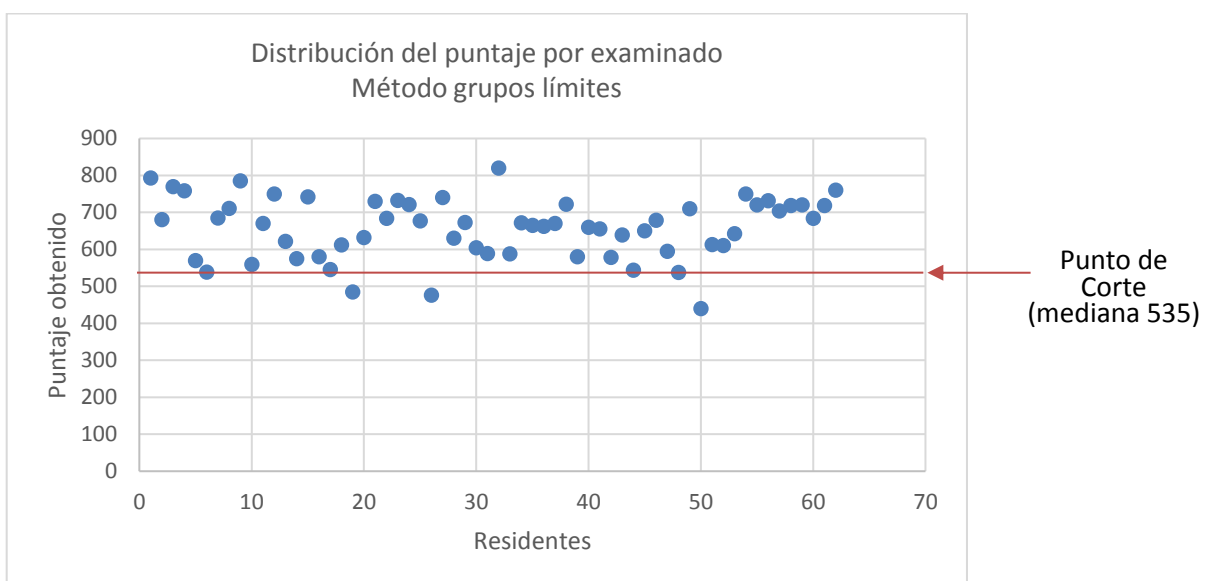
En primer lugar, se describirán los resultados globales del OSCE, y luego el análisis de cada estación en forma individual. En la figura 2 se puede ver la proporción de aprobados comparando los 4 métodos para puntos de corte ensayados. En el mismo se evidencia el alto número de candidatos que no pasan la prueba (n 18) si se tiene en cuenta el método absoluto de requerir el 60% del puntaje ponderado de cada estación, y aprobar al menos 6 de 9 estaciones. Y el método con menor proporción de desaprobados corresponde al de grupo límite sobre la mediana del test (n 3).

Figura 2. Relación entre aprobados y no aprobados según método de punto de corte



En la Figura 3. se puede ver la distribución por puntajes totales obtenidos en el test, por todos los examinados. El máximo puntaje al que se podía aspirar eran 900 puntos, que no fue logrado por ningún examinado. Por debajo de la línea del punto de corte se puede ver a los 3 desaprobados por el método de grupo límite sobre la mediana del test.

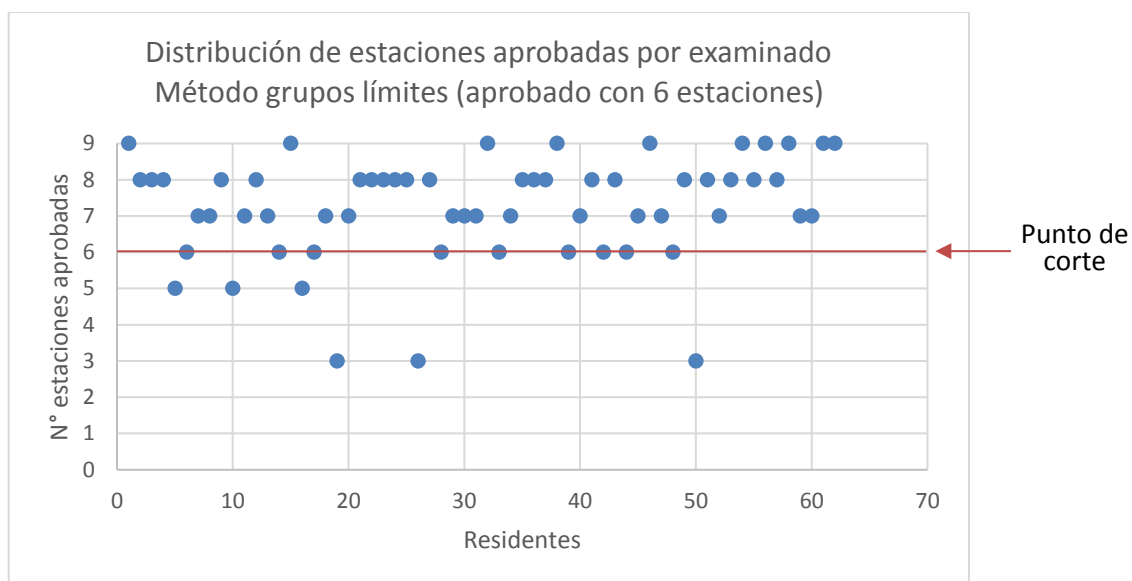
Figura 3. Puntaje por examinado, método grupos límites.



En la Figura 4. se puede ver la distribución de los examinados teniendo en cuenta el número de estaciones aprobadas, y donde se toma como punto de corte obtener como mínimo 6 de las 9 estaciones (66% del test).

En ella se puede ver que han quedado desaprobados 6 examinados, y 10 han logrado aprobar las 9 estaciones, aunque sin obtener el puntaje máximo.

Figura 4. Puntaje por examinado, método 6 de 9 estaciones.



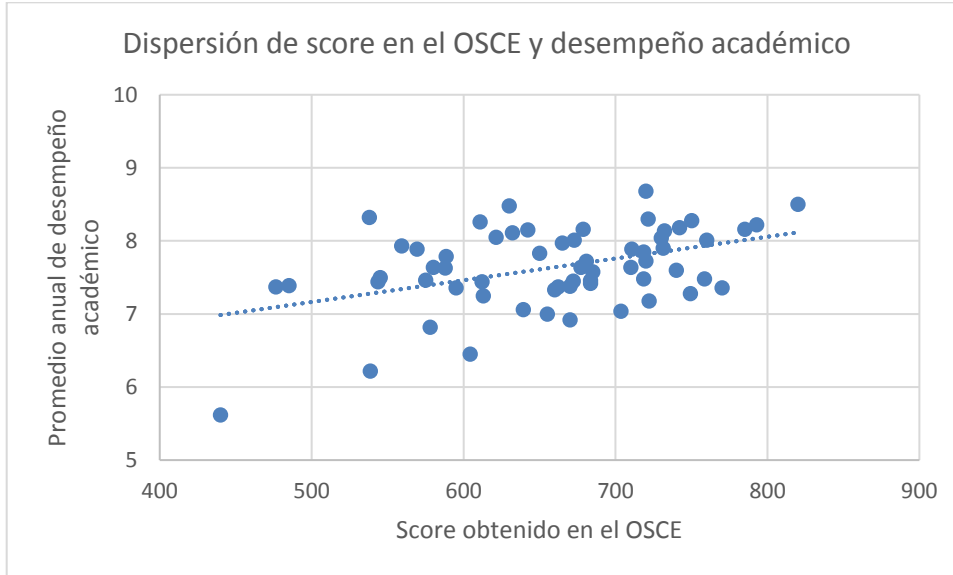
Para poder establecer la relación entre el rendimiento en el OSCE y el desempeño académico y asistencial de los examinados, calculamos el coeficiente de correlación.

En primer lugar, en la Figura 5. se puede ver la relación entre la variable “Desempeño académico” y la variable “Score del OSCE”. El **desempeño académico** de los examinados lo definimos de acuerdo con el resultado anual en los exámenes rendidos durante los módulos del curso universitario (1° año de la carrera de especialista universitario en pediatría) habiéndose calculado el promedio de las notas obtenidas durante el año académico comprendido entre junio de 2017 a mayo de 2018.

Como resultado se obtuvo que por cada unidad de aumento en el Score de OSCE se observó un efecto de 0,0032399 sobre el promedio de actividad académica, con una $P=0,0002$ IC 95% (0,001 - 0,0048).

Se observan algunos puntos alejados que corresponden a potenciales outliers o influyentes.

Figura 5. Relación entre score de OSCE y desempeño académico

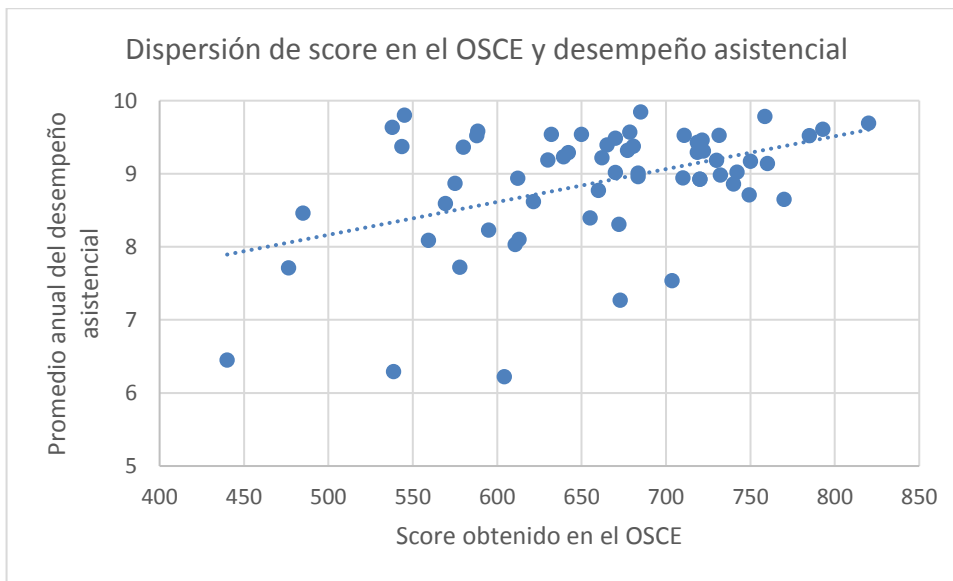


En la Figura 6. se puede ver la dispersión de los examinados de acuerdo con su “**desempeño asistencial**” en relación con el resultado obtenido en el OSCE. Se define desempeño asistencial al promedio de notas obtenidas durante el período junio de 2017 a mayo de 2018, en las evaluaciones de desempeño que se les realiza a los residentes al asistir pacientes en las diferentes áreas por donde se les asigna rotar.

En el gráfico se puede observar que la dispersión es algo mayor que la observada con los resultados obtenidos en los exámenes escritos.

En este caso por cada unidad de aumento en el Score de OSCE se observó un efecto de 0,0043137 sobre el promedio de actividad asistencial, con una $P=0,0005$ IC 95% (0,002 ; 0,0066).

Figura 6. Relación entre score de OSCE y desempeño asistencial



- **Conclusión**

El desarrollo del presente análisis permite concluir que hay 2 métodos que solos o combinados permiten alcanzar una proporción aceptable de aprobados/no aprobados, y con buena correlación con el desempeño demostrado por los examinados en las otras evaluaciones realizadas durante el período lectivo. El método de grupos límites sobre la mediana de la prueba, tomado como criterio absoluto, o combinado con que además los examinados deban aprobar un mínimo de 6 de 9 estaciones (o un porcentaje similar si varía el número de estaciones en otro OSCE), nos ha permitido detectar a quienes aún no han logrado el grado de autonomía deseado para la etapa en la que fueron evaluados. Por otro lado, podemos afirmar que existe adecuada concordancia entre las calificaciones de los observadores en las categorías de Insuficiente-límite- suficiente y sobresaliente con los puntajes obtenidos en las listas de cotejo y la aprobación o no de la estación.

En los próximos meses realizaremos nuevos OSCE con un importante número de residentes (aproximadamente 80 en cada uno) lo que nos posibilitará dar el siguiente paso, que será validar los métodos elegidos.

- **Discusión**

Como vimos, aunque existe una gran cantidad de métodos para establecer estándares en los exámenes de rendimiento (Cusimano, 1996), no existe un “estándar de oro”. Diversos estudios confirman que varios métodos producen diferentes puntajes de corte y los coordinadores de exámenes deben elegir un método defendible pero viable para su examen. Idealmente, el método elegido produciría el resultado más preciso.

El valor para alfa, que generalmente se considera aceptable en este tipo de evaluación donde se usan pacientes estandarizados y reales, es de 0.7 o superior. Pero un valor de alfa demasiado alto tampoco es bueno -por ejemplo >0.9- indicando posiblemente redundancia en la evaluación. Por otro lado, un bajo puntaje alfa de Cronbach que a veces se atribuye al resultado de estaciones mal diseñadas, en realidad puede ser por grandes diferencias en las pun-

tuaciones medias de las mismas. Si bien la confiabilidad obtenida en la prueba presentada se encuentra un poco por debajo de los estándares habituales deseados en educación médica, podemos hacer dos observaciones: en primer lugar se halla en ascenso en nuestros OSCE, ya que logramos en este examen analizado el valor más alto hasta ahora, y en segundo lugar en una interesante revisión sistemática sobre la confiabilidad de los resultados del OSCE medida en 98 publicaciones (Brannick, Erol-Korkmaz, & Prewett, 2011), se ha visto que entre las estaciones, el alfa de Cronbach (fiabilidad global) era igual a 0,66 (IC del 95%: 0,62-0,70). Este estudio también confirma un principio bien aceptado de que la confiabilidad aumenta con el número de estaciones. El alfa para los OSCE con menos de 10 estaciones fue de 0,56 y para los de más de 10 estaciones fue de 0,74. Aunque concluyen que, sin embargo, la variabilidad en esta medida de confiabilidad (incluso con un gran número de estaciones) aún es evidente. Otro aspecto importante por considerar cuando se selecciona el estándar de punto de corte para la aprobación, es el número de examinados. El método de grupos contrastados (en inglés *contrasting groups*), solo es confiable en cohortes importantes (Burrows, Bingham, & Brailovsky, 1999). Para el coordinador de OSCE a menor escala -como el que realizamos en el Garrahan- hay varios beneficios de usar el método de grupo límite (en inglés *borderline groups*) como método estándar, ya que no requiere ningún procedimiento estadístico complejo y el punto de corte es fácil de calcular. Este último también se basa en el rendimiento real del examinado en lugar de una revisión de los elementos de la lista de verificación como ocurre con los métodos centrados en las pruebas, que además requieren expertos en la materia (panelistas) para emitir juicios (Boulet, De Champlain, & Mckinley, 2003) (Wood et al., 2006).

Este tipo de evaluación, bien realizada, tiene alta confiabilidad y validez (Durante, 2006) (Brailovsky & Grand'Maison, 2000) (Tudiver et al., 2009). La validez de constructo del OSCE -entendida como el grado en que un instrumento cumple en medir precisamente aquello para lo que fue diseñado- ha sido demostrada en estudios como el de Brailovsky y col. sobre estudiantes del último año de medicina, residentes y especialistas en medicina familiar, donde los estudiantes de medicina obtuvieron puntajes significativamente menores que los otros grupos (Brailovsky, Grand'Maison, & Lescop, 1997).

Con respecto a la validez de contenido -entendido como que el test presenta una muestra adecuada de los contenidos a los que se refiere, sin omisiones y sin desequilibrios de contenido- también ha sido estudiada (Grand'Maison, Brailovsky, & Lescop, 1996). La validez de contenido descansa generalmente en el juicio de expertos (métodos de juicio). Se define como el grado en que los ítems que componen la prueba representan el contenido que el test trata de evaluar.

El OSCE, con todas sus fortalezas y debilidades, es en la actualidad utilizado como método de evaluación para otorgar la licencia médica en países como Estados Unidos de Norteamérica, Canadá y Gran Bretaña (Turner & Dankoski, 2008) (McCoy & Merrick, 2010) (Grand'Maison et al., 1996) (Grand'Maison, Brailovsky, Lescop, & Rainsberry, 1997) (J. Brown, Handsfield-Jones, Rainsberry, & Brailovsky, 1996) (Violato et al., 2003). En Argentina se utiliza en pocas instituciones, pero se encuentra recomendado por el Ministerio de Salud de la Nación para la evaluación del desempeño de los residentes (Davini, Rodríguez & Daverio, n.d.).

La incorporación al OSCE de la evaluación de las habilidades en comunicación ha sido explorada y aceptada por la comunidad médica dedicada a la educación (Turner & Dankoski, 2008). Incluso se han generado estrategias derivadas del OSCE enfocadas exclusivamente a la evaluación de habilidades comunicacionales (Caron et al., 2013), que consta de 4 estaciones con pacientes simulados. La utilización de pacientes estandarizados para la evaluación de habilidades comunicacionales ha sido explorada incluso en pediatría en la Argentina (Pol, Wainfeld, & Yulitta, 2015); hallándose justamente que las competencias más deficitarias fueron habilidades de la comunicación y actividades preventivas. En el OSCE presentado se pudo ver que la estación dedicada a comunicación debe ser mejorada en cuanto a su validez de constructo, nosotros hemos elegido como instrumento de observación el Kalamazoo (Duffy et al., 2004; Joyce, Steenbergh, & Scher, 2010; Peterson, Calhoun, & Rider, 2014; Schirmer et al., 2005), reconocido y utilizado internacionalmente, pero de todas formas esta estación no ha funcionado lo suficientemente bien en cuanto a la categorización de los examinados según la teoría de los grupos límite.

Otro aspecto no menos importante, es el costo de realizar un OSCE. Los reportes coinciden en el alto costo de este tipo de examen (Brown, Ross, Cleland, & Walsh, 2015) (Khan, Gaunt, Ramachandran, & Pushkar, 2013), que en Argentina -país en vías de desarrollo- no puede dejar de ser tenido en cuenta. Invertir en exámenes costosos debe estar respaldado por resultados favorables documentados.

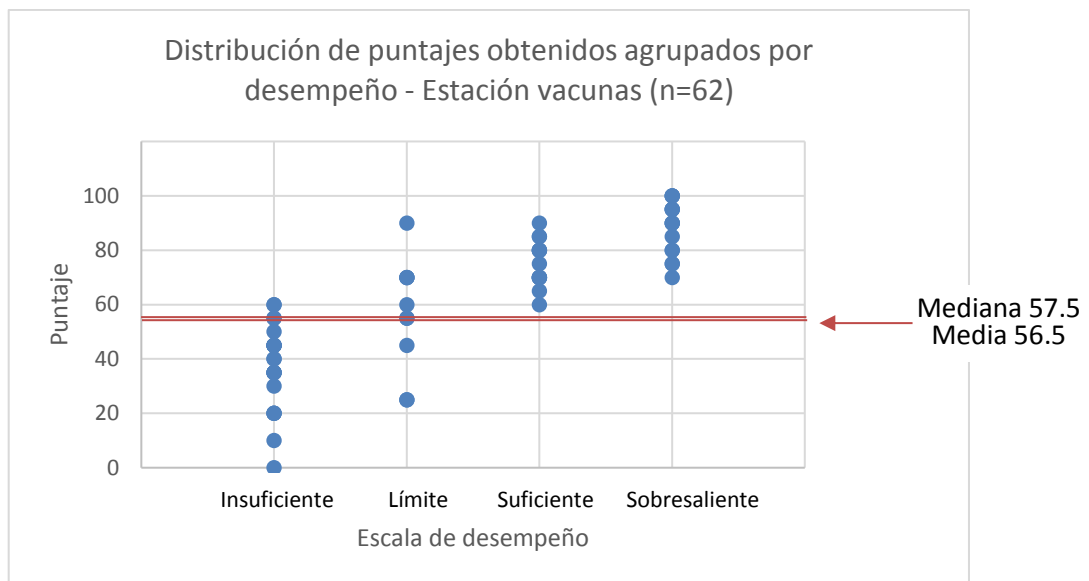
- **Anexo I: Análisis de las 9 estaciones**

Estación Nro 1: Vacunas

El objetivo de esta estación era que el residente demuestre ser capaz de indicar las vacunas correspondientes al ingreso escolar a un paciente de 6 años con diagnóstico de Síndrome Nefrótico corticosensible, que recibe cortico-terapia prolongada como parte de su tratamiento.

En esta estación se obtuvo un Coeficiente Alfa de Cronbach de 0.54, con una cantidad de 25 desaprobados utilizando la teoría de los grupos límite, la media y la mediana tuvieron puntajes muy cercanos como puntos de corte de aprobación (Figura 7).

Figura 7. Puntaje obtenido por los examinados en la Estación Vacunas

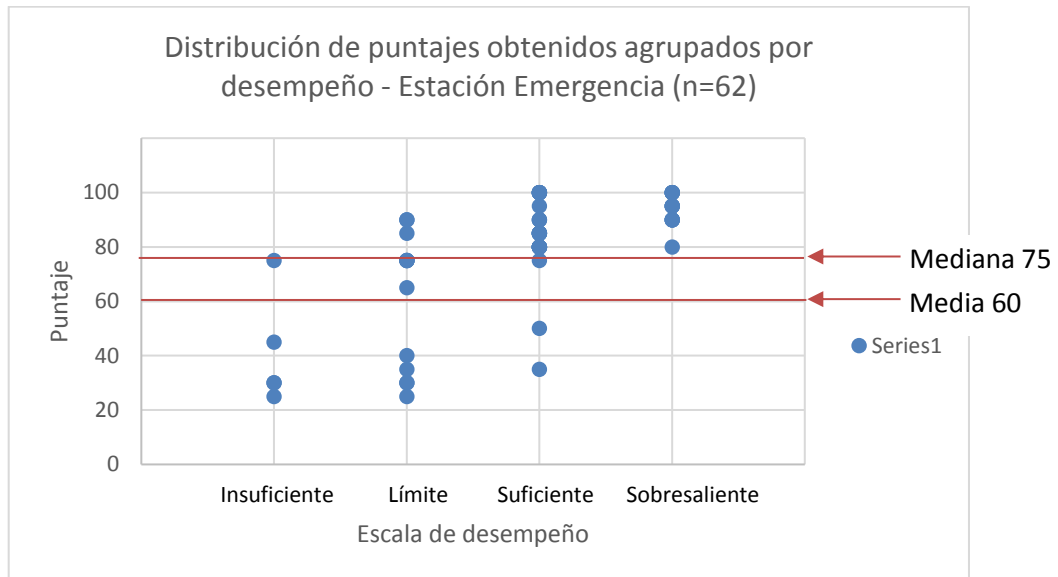


Estación Nro 2: Emergencia metabólica

El objetivo de esta estación era que el residente demuestre ser capaz de identificar y compensar una complicación grave de la diabetes (hipoglucemia).

En esta estación se obtuvo un Coeficiente Alfa de Cronbach 0.6, aquí la distribución tuvo un registro particular, como se puede ver el grupo desaprobado estuvo muy por debajo de la mediana, y a su vez esta estuvo 15 puntos por encima de la media (Figura 8). El grupo calificado como límite obtuvo puntajes muy dispersos, llegando en un caso a obtener 90 puntos sobre 100.

Figura 8. Puntaje obtenido por los examinados en la Estación Emergencia

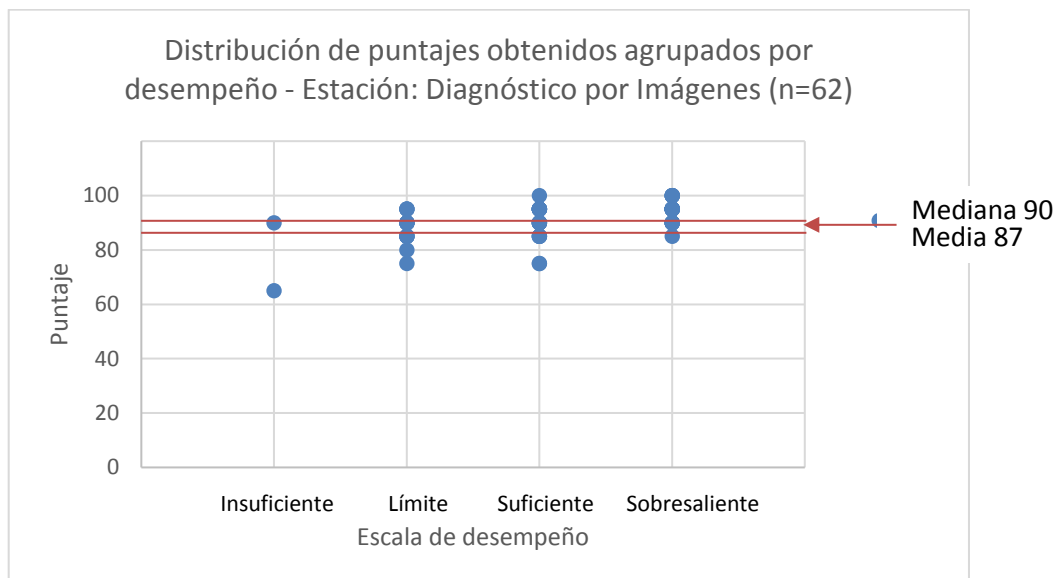


Estación Nro 3: Diagnóstico por imágenes

El objetivo de esta estación era que el residente demuestre ser capaz de interpretar una radiografía de abdomen, realizar el diagnóstico de neumoperitoneo e indicar las medidas terapéuticas iniciales.

En esta estación se obtuvo un Coeficiente Alfa de Cronbach 0.6 con una enorme concentración de examinados con puntajes entre 75 y 100, con un muy alto rendimiento del grupo, y con puntos de corte muy cercanos a los 90 puntos de mediana y media (Figura 9).

Figura 9. Puntaje obtenido por los examinados en la Estación Imágenes

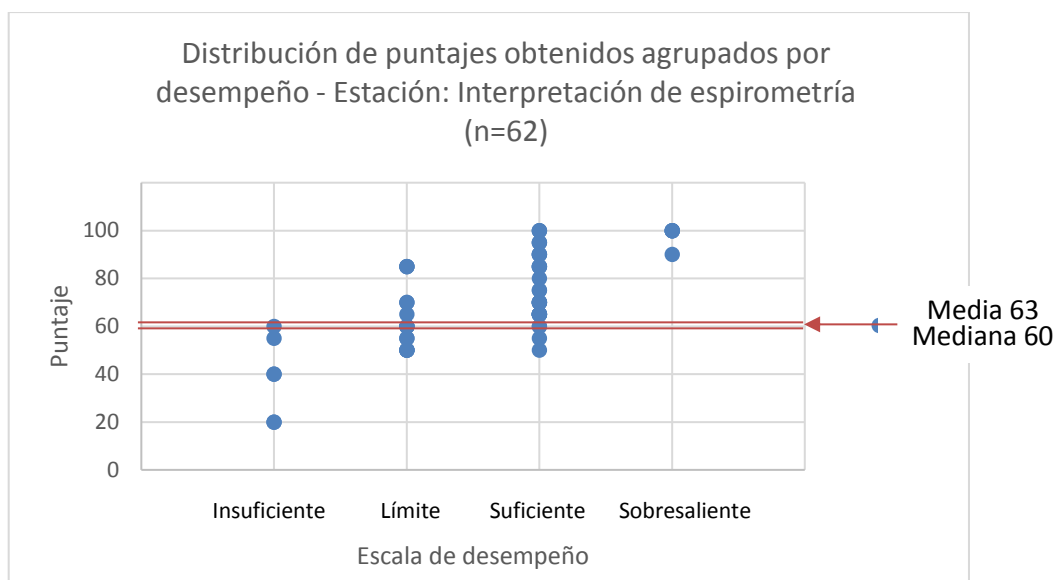


Estación Nro 4: Interpretación de una espirometría

El objetivo de esta estación era que el residente demuestre ser capaz de interpretar un examen funcional respiratorio (espirometría) de un asmático, y que decida la conducta a seguir.

En esta estación se obtuvo un Coeficiente Alfa de Cronbach 0.6, llama la atención que el grupo límite y el grupo que obtuvo suficiente, obtuvieron puntajes bastante similares (Figura 10). En este caso tomando la mediana como punto de corte, ningún examinado con “insuficiente” en la observación logró el puntaje necesario para aprobar.

Figura 10. Puntaje obtenido por los examinados en la Estación espirometría

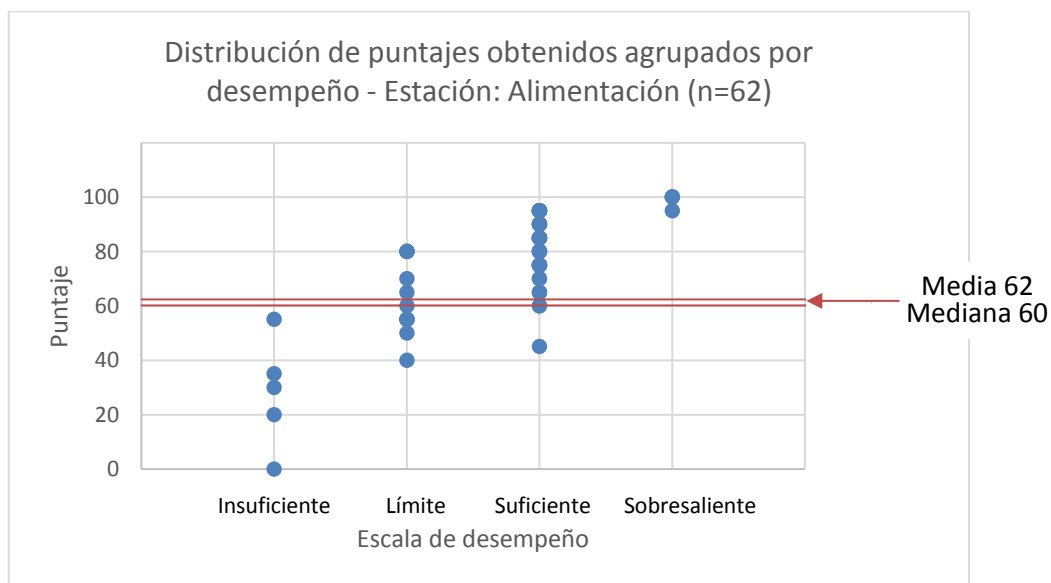


Estación Nro 5: Entrevista alimentaria

El objetivo de esta estación era que el residente demuestre ser capaz de entrevistar a una madre sobre la alimentación de un bebé de 14 días de vida, identificar mala técnica alimentaria y dar pautas de amamantamiento adecuadas.

En esta estación se obtuvo un Coeficiente Alfa de Cronbach 0.59; nuevamente en el grupo con desempeño insuficiente no hubo aprobados, y debido a la distribución de los examinados la media y la mediana coinciden en gran medida (Figura 11). En esta estación hubo alumnos con puntaje máximo (100 puntos) y mínimo (0 puntos).

Figura 11. Puntaje obtenido por los examinados en la Estación Alimentación

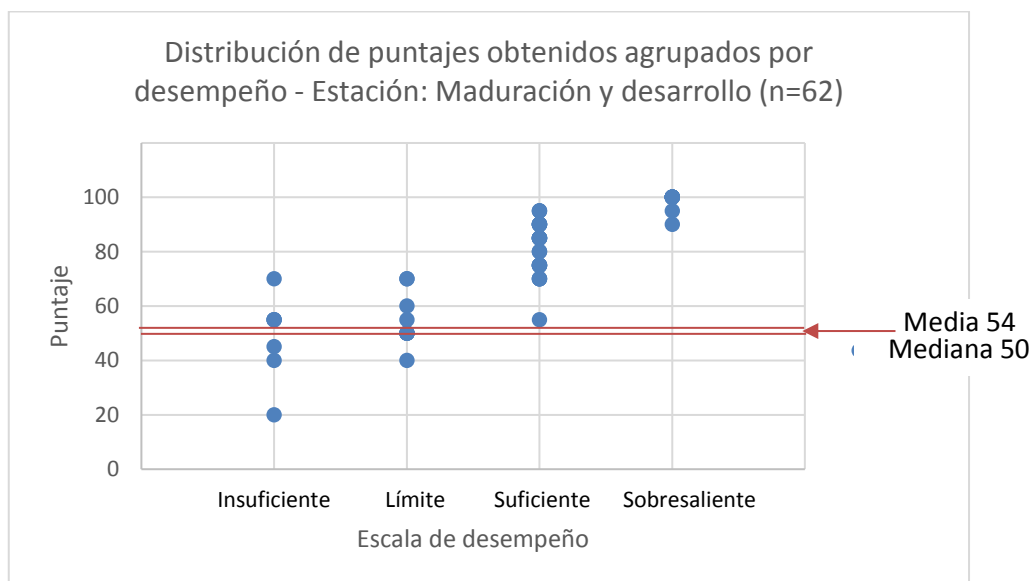


Estación Nro 6: Maduración y desarrollo

El objetivo de esta estación era que el residente demuestre ser capaz de reconocer las pautas madurativas en una niña de 2 años y 2 meses, y desarrollar este aspecto en la Historia Clínica.

En esta estación se obtuvo un Coeficiente Alfa de Cronbach 0.62. El desempeño de los residentes fue muy bueno, con solo 4 desaprobados (Fig 12)

Figura 12. Puntaje obtenido por los examinados en la Estación Maduración.

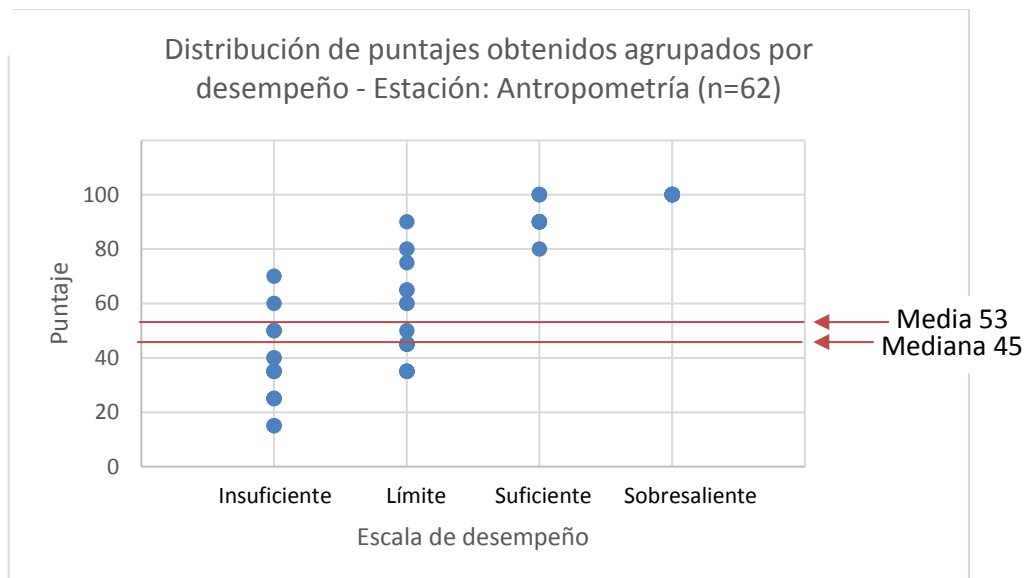


Estación Nro 7: Antropometría

El objetivo de esta estación era que el residente demuestre ser capaz de hacer el diagnóstico de madurador lento a un paciente que consulta por baja talla, utilizando los datos que se le brindan. A su vez, comunicarle el diagnóstico a la madre.

En esta estación se obtuvo un Coeficiente Alfa de Cronbach 0.52. En este caso el número de desaprobados alcanzó los 17, resultando en una estación de alta dificultad (Figura 13).

Figura 13. Puntaje obtenido por los examinados en la Estación Antropometría.

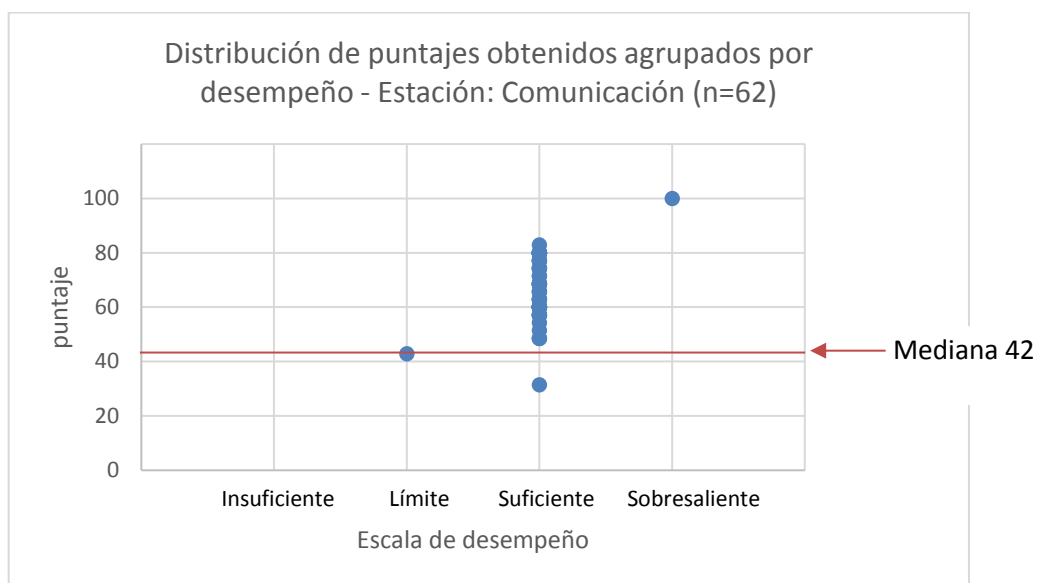


Estación Nro 8: Comunicación

El objetivo de esta estación era que el residente demuestre ser capaz de comunicar a los padres de un niño la necesidad de repetir una extracción de sangre para diagnóstico.

En esta estación se obtuvo un Coeficiente Alfa de Cronbach 0.61. Es una estación que con frecuencia hay pocos o ningún desaprobado, en este OSCE un solo residente obtuvo la calificación de "límite" por lo que su puntaje resulta en el punto de corte de aprobación de la estación (Figura 14). Y el único examinado reprobado había tenido una observación calificada como "Suficiente" pero no alcanzó el puntaje requerido para aprobar, y de hecho desaprobó el OSCE.

Figura 14. Puntaje obtenido por los examinados en la Estación Comunicación.

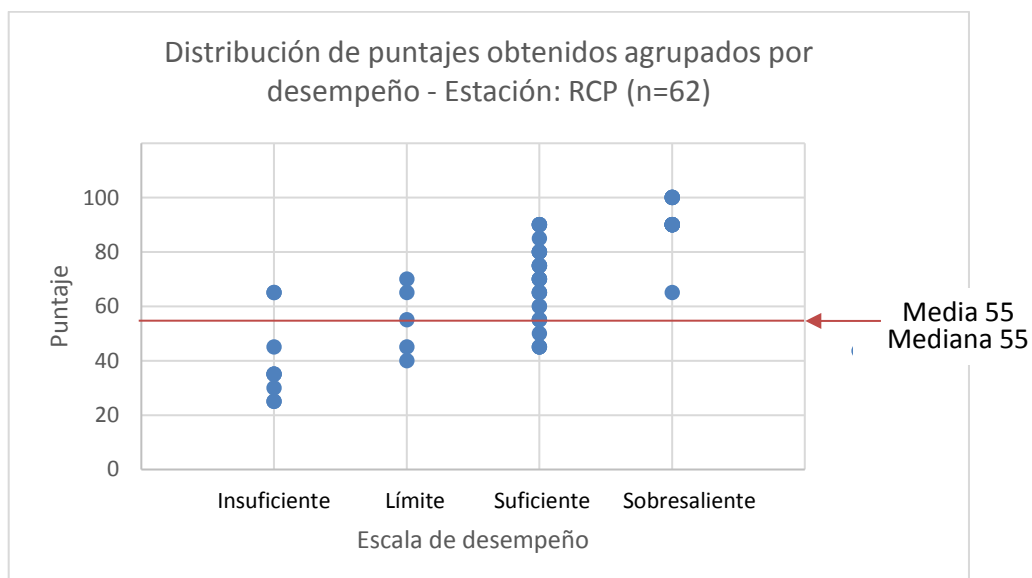


Estación Nro 9: Reanimación cardiopulmonar

El objetivo de esta estación era que el residente demuestre ser capaz de reconocer y asistir un paro respiratorio, y reconocer el paro cardíaco y comenzar con maniobras de resucitación avanzadas (colocar un acceso vascular intraóseo y administración de adrenalina EV).

En esta estación se obtuvo un Coeficiente Alfa de Cronbach 0.61. Se trata de una estación muy técnica, que es incluida en todos los OSCE con variantes en cuanto al caso clínico y lo que se le requiere al examinado (intubar, hacer masaje cardíaco, colocar una vía intraósea, etc.). De todas formas y a pesar de prepararse para el examen, el número de desaprobados fue de alto (n=13). Salvo uno de los residentes observado como “insuficiente” que obtuvo el puntaje suficiente para aprobar, los demás “insuficientes” no aprobaron la estación. Mientras que dos “suficientes” la desaprobaron (Figura 15).

Figura 15. Puntaje obtenido por los examinados en la Estación RCP.



Bibliografía

- Accreditation Council for Graduate Medical Education. (2013). ACGME Program Requirements for Graduate Medical Education in Pediatrics, 1–34.
- Anijovich, R., Camilloni, A., Cappelletti, G., Hoffmann, J., Katzkowicz, R., & Lopez Mottier, L. (2010). Evaluación formativa de los aprendizajes: síntesis crítica de los trabajos francófonos. In *La evaluación significativa* (1º, pp. 43–67). Buenos Aires.
- Barman, a. (2005). Critiques on the Objective Structured Clinical Examination. *Annals of the Academy of Medicine, Singapore*, 34(8), 478–482.
- Barrows, H., & Abrahamson, S. (1964). The programmed patient: a technique for appraising student performance in clinical neurology. *J Med Educ*, 39, 802–805.
- Bloom, B., Hastings, J., & Madaus, G. (1971). *Handbook on formative and summative evaluation of student learning*. (McGraw-Hill, Ed.). New York.
- Boulet, J. R., De Champlain, A. F., & Mckinley, D. W. (2003). Setting defensible performance standards on OSCEs and standardized patient examinations. *Medical Teacher*, 25(3), 245–249.
<https://doi.org/10.1080/0142159031000100274>
- Brailovsky, C. A., & Grand'Maison, P. (2000). Using Evidence to Improve Evaluation: A Comprehensive Psychometric Assessment of a SP-Based OSCE Licensing Examination. *Advances in Health Sciences Education*, 5(3), 207–219. <https://doi.org/10.1023/A:1009869328173>
- Brailovsky, C. A., Grand'Maison, P., & Lescop, J. (1992). A large-scale multicenter objective structured clinical examination for licensure. *Academic Medicine : Journal of the Association of American Medical Colleges*, 67(10 Suppl), S37-9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1388549>
- Brailovsky, C. A., Grand'Maison, P., & Lescop, J. (1997). Construct Validity of the Quebec Licensing Examination SP-Based OSCE. *Teaching and Learning in Medicine*, 9(1), 44–50.
- Brannick, M. T., Erol-Korkmaz, H. T., & Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Medical Education*, 45(12), 1181–1189. <https://doi.org/10.1111/j.1365-2923.2011.04075.x>
- Brown, C., Ross, S., Cleland, J., & Walsh, K. (2015). Money makes the (medical assessment) world go round: The cost of components of a summative final year Objective Structured Clinical Examination (OSCE). *Medical Teacher*, 37(7), 653–659.
<https://doi.org/10.3109/0142159X.2015.1033389>
- Brown, J., Handsfield-Jones, R., Rainsberry, P., & Brailovsky, C. A. (1996). Certification examination of the College of Family Physicians of Canada. Part 4-Simulated office orals. *Canadian Family Physician*, 42, 1539–1548.
- Burrows, P. J., Bingham, L., & Brailovsky, C. A. (1999). A Modified Contrasting Groups Method Used for Setting the Passmark in a Small Scale

- Standardised Patient Examination. *Advances in Health Sciences Education*, 4(2), 145–154. <https://doi.org/10.1023/A:1009826701445>
- Caron, A., Perzynski, A., Thomas, C., Saade, J. Y., McFarlane, M., & Becker, J. (2013). Development of the objective, structured communication assessment of residents (OSCAR) tool for measuring communication skills with patients. *Journal of Graduate Medical Education*, 5(4), 570–575. <https://doi.org/10.4300/JGME-D-12-00230.1>
- Chisnall, B., Vince, T., Hall, S. K., & Tribe, R. M. (2015). Evaluation of outcomes of a formative objective structured clinical examination for second-year UK medical students, 76–83. <https://doi.org/10.5116/ijme.5572.a534>
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards for tests*. London: Sage Publications.
- Cusimano, M. D. (1996). Standard setting in medical education : Academic Medicine. *Academic Medicine*, 71(10), S112-20. <https://doi.org/10.1097/00001888-199610000-00062>
- Davini, A. C., Rodríguez, G., María, C., & Daverio, D. (n.d.). *Guía para el desarrollo de la Evaluación del Desempeño en las Residencias del equipo de Salud*.
- DeSerdio, R. E. (2002). ECOE: Evaluación Clínica Objetiva Estructurada. *Medicina de Familia (And)*, 2, 127–132.
- Duffy, F. D., Gordon, G. H., Whelan, G., Cole-kelly, K., Frankel, R., & Skills, I. (2004). Assessing Competence in Communication and Interpersonal Skills: The Kalamazoo II Report, 495–507.
- Durante, E. (2006). Algunos métodos de evaluación de las competencias : Escalando la pirámide de Miller. *Rev. Hosp. Ital. B Aires*, 26(2), 55–61.
- Global Pediatric Education Consortium. (Adaptación al español) (2012). Retrieved from <http://www.globalpediatrics.org/globalcurriculum/curriculumdownloadpdf.html>
- Grand'Maison, P., Brailovsky, C. A., & Lescop, J. (1996). Content validity of the Quebec licensing examination (OSCE) - Assessed by practising physicians. *Canadian Family Physician*, 42(FEB.), 254–259.
- Grand'Maison, P., Brailovsky, C. A., Lescop, J., & Rainsberry, P. (1997). Using standardized patients in licensing/certification examinations: comparison of two test in Canada. *Family Medicine*, 29(1), 27–32. Retrieved from http://teacherweb.com/NY/Arlington/Losee/CE_Ch13_328-345.pdf
- Grand'Maison, P., Lescop, J., & Brailovsky, C. A. (1993). Canadian experience with structured clinical examinations. *Canadian Medical Association*, 148(9), 1573–1576.
- Harden, R. M. (2015). Misconceptions and the OSCE. *Medical Teacher*, 37(7), 608–610. <https://doi.org/10.3109/0142159X.2015.1042443>
- Harden, R. M. (2016). Revisiting “Assessment of clinical competence using an objective structured clinical examination (OSCE).” *Medical Education*, 50(4), 376–379. <https://doi.org/10.1111/medu.12801>

- Harden, R., Stevenson, M., Wilson Downie, W., & Wilson, G. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal*, 1(February), 447–451.
- Hejri, S. M., Jalili, M., Muijtjens, A. M. M., & Van Der Vleuten, C. P. M. (2013). Assessing the reliability of the borderline regression method as a standard setting procedure for objective structured clinical examination. *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences*, 18(10), 887–891. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24497861>
- Joyce, B. L., Steenbergh, T., & Scher, E. (2010). Use of the kalamazoo essential elements communication checklist (adapted) in an institutional interpersonal and communication skills curriculum. *Journal of Graduate Medical Education*, 2(2), 165–169. <https://doi.org/10.4300/JGME-D-10-00024.1>
- Junod Perron, N., Louis-Simonet, M., Cerutti, B., Pfarrwaller, E., Sommer, J., & Nendaz, M. (2016). The quality of feedback during formative OSCEs depends on the tutors' profile. *BMC Medical Education*, 16(1), 1–8. <https://doi.org/10.1186/s12909-016-0815-x>
- Khan, K. Z., Gaunt, K., Ramachandran, S., & Pushkar, P. (2013). The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration. *Medical Teacher*, 35(9), e1447–e1463. <https://doi.org/10.3109/0142159X.2013.818635>
- Khan, K. Z., 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(9). <https://doi.org/10.3109/0142159X.2013.818634>
- Kogan, J. R., Conforti, L. N., Bernabeo, E. C., Durning, S. J., Hauer, K. E., & Holmboe, E. S. (2012). Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*, 46(2), 201–215. <https://doi.org/10.1111/j.1365-2923.2011.04137.x>
- Lima, A. E. A. D. E., Cardiovascular, I., & Aires, D. B. (2008). La devolución constructiva o feedback. *Medicina*, (68), 88–92.
- Livingston, S., & Zieky, M. (1982). Passing scores: A manual for setting standards of performance on education and occupational tests. *Educ Testing Serv.*
- McCoy, J., & Merrick, H. (2010). The Objective Structured Clinical Examination. *Association for Surgical Education*, 20(2), 1–104. <https://doi.org/10.1192/pb.29.6.232-c>
- McKinley, D. W., & Norcini, J. J. (2014). How to set standards on performance-based examinations: AMEE Guide No. 85. *Medical Teacher*, 36(2), 97–110. <https://doi.org/10.3109/0142159X.2013.853119>
- Norcini, J. J. (2003). Standard setting on educational tests. *Medical Education*, 37, 464–469.
- Pell, G., Fuller, R., Homer, M., & Roberts, T. (2010). How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Medical Teacher*, 32(10), 802–811. <https://doi.org/10.3109/0142159X.2010.507716>

- Peterson, E. B., Calhoun, A. W., & Rider, E. a. (2014). The reliability of a modified Kalamazoo Consensus Statement Checklist for assessing the communication skills of multidisciplinary clinicians in the simulated environment. *Patient Education and Counseling*, 96(3), 411–418. <https://doi.org/10.1016/j.pec.2014.07.013>
- Pol, M. F., Wainfeld, M. G., & Yulitta, P. H. (2015). Diseño e implementación de una evaluación clínica objetiva y estructurada en pediatría de atención primaria. *Investigación En Educación Médica*, 4(14), e17–e18. [https://doi.org/10.1016/S2007-5057\(15\)30061-2](https://doi.org/10.1016/S2007-5057(15)30061-2)
- Regehr, G., MacRae, H., Reznick, R. K., & Szalay, D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine : Journal of the Association of American Medical Colleges*, 73(9), 993–997. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9759104>
- Reznick, R. K., Blackmore, D., Dauphinée, W. D., Rothman, A. I., & Smee, S. (1996). Large-scale high-stakes testing with an OSCE: report from the Medical Council of Canada. *Academic Medicine : Journal of the Association of American Medical Colleges*, 71(1 Suppl), S19-21. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8546771>
- Rothman, A., & Cohen, R. (1996). A comparison of empirically and rationally defined standars for clinical skills checklists. *Academic Medicine*, 71(10), S1–S3.
- Schirmer, J. M., Mauksch, L., Lang, F., Marvel, M. K., Zoppi, K., Epstein, R. M., ... Pryzbylski, M. (2005). Assessing communication competence: A review of current tools. *Family Medicine*, 37(3), 184–192. <https://doi.org/10.1016/j.jsurg.2010.12.005>
- Schuwirth, L. W. T., & Van Der Vleuten, C. P. M. (2011). General overview of the theories used in assessment: AMEE Guide No. 57. *Medical Teacher*, 33(10), 783–797. <https://doi.org/10.3109/0142159X.2011.611022>
- Smee, S. (2001). Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. *Medical Education*, 35(11), 1009–1010.
- Smith, L. J., Price, D. A., & Houston, I. B. (1984). Objective structured clinical examination compared with other forms of student assessment. *Archives of Disease in Childhood*, 59(12), 1173–1176. <https://doi.org/10.1136/adc.59.12.1173>
- Tudiver, F., Rose, D., Banks, B., & Pfortmiller, D. (2009). Reliability and validity testing of an evidence-based medicine OSCE station. *Family Medicine*, 41(2), 89–91.
- Turner, J. L., & Dankoski, M. E. (2008). Objective structured clinical exams: A critical review. *Family Medicine*, 40(8), 574–578.
- Violato, C., Marini, A., & Lee, C. (2003). PROCEDURES FOR SETTING CUTOFF SCORES EXAMINATIONS USING CLUSTER ANALYSIS, 26(1), 59–72. <https://doi.org/10.1177/0163278702250082>
- Wood, T., Humphrey-Murto, S., & Norman, G. (2006). Standard setting in a small scale OSCE: A comparison of the modified borderline-group method

and the borderline regression method. *Advances in Health Sciences Education, 11*(June), 115–122. <https://doi.org/10.1007/s10459-005-7853-1>

Yorke, M. (2003). Formative assessment in higher education : Moves towards theory and the enhancement of pedagogic practice, (1997), 477–501.

Curriculum Vitae



- Nombre y apellido: Hernán Rowensztein.
- Fecha de nacimiento: 25 de marzo de 1972.
- Nacionalidad: Argentino.
- DNI: 22.706.712
- Estado civil: Casado.
- Título: Médico (U.B.A. 1997)
- Matrícula Profesional: MN 97.343
- Especialidad: Pediatría.
- Residencia en Pediatría y Jefatura de Residencia en el Hospital de Pediatría "Prof. Dr. Juan P. Garrahan"
- Especialista Universitario en Pediatría (U.B.A.)
- Residencia posbásica en Medicina Interna Pediátrica en el Hospital de Pediatría "Prof. Dr. Juan P. Garrahan"
- Maestría en Educación para Profesionales de la Salud. Instituto Universitario del Hospital Italiano de Buenos Aires. Duración 2 años (240 horas presenciales, 340 horas a distancia) 2012-2013.
- Beca "Ramón Carrillo-Arturo Oñativia 2004", para investigación sobre "*Evaluación económica de la carga de enfermedad y costos asociados a las internaciones por Infección Respiratoria Aguda en Niños*". Duración: 1 año (Premio: X Congreso Argentino de Pediatría Social, 2005)
- Médico de Planta en Clínica Pediátrica en el Hospital de Pediatría "Prof. Dr. Juan P. Garrahan" desde 2005.
- Jefe de Clínica en el Hospital de Pediatría "Prof. Dr. Juan P. Garrahan". 2012-2015
- Coordinador de Unidad de Formación de Recursos Humanos del Hospital de Pediatría "Prof. Dr. Juan P. Garrahan", 2015 - hasta el presente.
- Docente en la Facultad de Medicina de la U.B.A. desde 1993. Cargo actual rentado por concurso de la cátedra de Pediatría.
- Director de 3 becas de investigación completas, y 1 en curso.
- Director Asociado de tesis de Maestría en Investigación en Ciencias Médicas de la Facultad de Medicina de la U.B.A.
- Autor de 5 capítulos de libros, y más de 20 publicaciones en revistas científicas.