

Jung, Gustavo

Resultados de una intervención educativa en razonamiento clínico para residentes de medicina interna evaluados con Script Concordance Test

Maestría en Educación para Profesionales de la Salud

Tesis 2022

Cita sugerida: Jung G. Resultados de una intervención educativa en razonamiento clínico para residentes de medicina interna evaluados con Script Concordance Test [tesis de maestría] [Internet]. [Buenos Aires]. Instituto Universitario Hospital Italiano de Buenos Aires; 2022 [citado AAAA MM DD]. 87 p. Disponible en: <https://trovare.hospitalitaliano.org.ar/descargas/tesisytr/20220711103808/tesis-jung-gustavo.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/>





Tesis de Maestría en Educación para Profesionales de la Salud

***Resultados de una intervención educativa en
razonamiento clínico para residentes de medicina
interna evaluados con script concordance test***

Candidato: Gustavo Jung

Director de Tesis: Dr. Carlos Brailovsky

Instituto Universitario Escuela de Medicina Hospital Italiano de Buenos Aires

Cohorte 2018

Ciudad Autónoma de Buenos Aires

06 de mayo de 2022

Abstract

Introducción

La competencia en razonamiento clínico es crítica en la práctica médica profesional. La enseñanza y evaluación de esta competencia es particularmente desafiante, porque se trata de una habilidad compleja, que además es contenido y contexto específica.

Al momento de la presente investigación, son pocas las investigaciones que evaluaron en forma robusta los resultados de una intervención educativa en razonamiento clínico, y ninguna en nuestro contexto.

Objetivos:

La presente investigación busca promover el aprendizaje del razonamiento clínico basado en la teoría y la mejor evidencia disponible.

Métodos:

Se llevó adelante una investigación experimental, prospectiva, controlada y no randomizada para determinar los resultados de una intervención educativa en razonamiento clínico para un grupo de 21 residentes de medicina interna de un hospital universitario de Buenos Aires, Argentina.

Se buscó determinar el nivel de competencia en ciertos aspectos del proceso de razonamiento clínico, mediante el *script concordance test* antes y después de la intervención educativa, en relación a un grupo control, de otro hospital universitario de Buenos Aires.

Resultados:

Luego de 22 semanas no fue posible demostrar diferencias significativas entre los grupos. La diferencia absoluta de las medias entre los grupos no alcanzó el límite pre especificado de 8 puntos, de acuerdo al diseño y poder de la muestra.

Sin embargo, la presente investigación busca aportar una mirada amplia y análisis profundo sobre las bases del aprendizaje, enseñanza y evaluación del razonamiento clínico; como así también, profundizar la discusión acerca del desarrollo de la competencia en razonamiento clínico de los residentes de nuestro país.

Palabras clave

Razonamiento clínico; educación médica; competencias; evaluación; residencias médicas; medicina interna; script concordance test.

Introducción

El razonamiento clínico es una habilidad, proceso o resultado a través del cual los médicos clínicos observan, recolectan e interpretan información para diagnosticar y tratar pacientes (Daniel et al., 2019). La competencia en razonamiento clínico es central en la práctica médica.

Existe abundante literatura médica acerca de la importancia del desarrollo del razonamiento clínico durante la formación de grado y posgrado. Tanto los Colegios de Médicos de Familia y de Médicos y Cirujanos de Canadá a través de la propuesta de competencias CANMEDS (J. Frank, Snell, & Sherbino, 2015; Shaw E, Oandasan I, Fowler N, 2017), la Asociación Americana de Facultades de Medicina (AAMC) a través de las “EPAs para ingresar al programa de residencias” (AAMC, 2014), como el Proyecto Tuning en Europa (Cumming & Ross, 2007), incluyen y destacan el razonamiento clínico como una habilidad compleja, transversal y fundamental.

En una encuesta a encargados docentes universitarios en EEUU, la mayoría considero que debe enseñarse el razonamiento clínico desde el inicio de la formación, sin embargo el 57% de las instituciones no cuenta con sesiones para el aprendizaje de esta competencia (Joseph Rencic, Trowbridge, Fagan, Szauter, & Durning, 2017).

Autores en el ámbito de la Medicina Interna alertan sobre una desconexión entre el enfoque de la educación médica actual y la importancia de las habilidades diagnósticas del razonamiento clínico (Simpkin, Vyas, & Armstrong, 2017).

No es de sorprender entonces, que las iniciativas desarrolladas para la enseñanza del razonamiento clínico sean consideradas “innovadoras” (Gay, Bartlett, & Mckinley, 2013).

Asimismo, poder demostrar intervenciones educativas que favorezcan el desarrollo del razonamiento clínico en médicos sería deseable éticamente, porque tendría un gran valor científico, clínico, y más importante: social (Emanuel, 1999).

A partir de mi experiencia cotidiana de trabajo con médicos residentes, observé dificultades en múltiples instancias del razonamiento clínico: desde la obtención de información relevante, su jerarquización, síntesis, abstracción, comunicación, y en la capacidad de generar hipótesis diagnósticas y luego verificarlas, teniendo en cuenta los riesgos y sesgos más comunes en el proceso.

El desarrollo de la competencia en razonamiento clínico es desafiante. Para volver factible la enseñanza, evaluación y diseño curricular de las competencias en general, y una compleja en particular como el razonamiento clínico, ten Cate acuñó el concepto de actividades profesional confiables (del inglés: *entrustable professional activities* o EPAs) (ten Cate, 2005). Las EPAs representan lo que hacen los médicos en su práctica cotidiana, por lo que este nuevo lenguaje resulta más breve y menos complicado (ten Cate, 2013). Las EPAs definen las tareas o responsabilidades que pueden ser confiadas a un aprendiz, una vez que alcanzó suficiente competencia específica, que permite su ejecución con diferentes grados de supervisión (Ten Cate, 2018).

Como primera aproximación al análisis de la competencia en razonamiento clínico, se invitó a residentes a responder una encuesta anónima. Se les pidió que auto evalúen sus competencias al comenzar la residencia, utilizando una escala tipo Likert de cuatro opciones, expresando el grado de acuerdo con la sentencia inicial. Cada una de las preguntas fue diseñada a partir de las EPAs clave relacionadas con el razonamiento clínico, que la AAMC establece como necesarias para el ingreso a las residencias.

Vale aclarar que, al ser ya residentes al inicio del primer y segundo año, el ejercicio implicaba que se evalúen a sí mismos, pero retrospectivamente, con los sesgos que esto que puede implicar. A pesar de ello, como la hipótesis previa implícita de la evaluación era evidenciar baja auto-percepción de sus competencias en razonamiento clínico al comenzar la residencia, en el peor de los casos ellos se sobre-estimarían, producto del tiempo ya transcurrido.

Los detalles de la encuesta de autoevaluación se encuentran en el [apéndice I](#). Los resultados globales de la encuesta muestran que *la gran mayoría de los residentes no se consideraba a sí mismo competente en razonamiento clínico al comenzar la residencia*. Este trabajo en terreno me permitió enriquecer la perspectiva teórica antes mencionada (pobre educación formal en razonamiento clínico), con mis observaciones no sistemáticas (pobre competencia en razonamiento clínico en residentes), junto con la propia mirada de la población objetivo de investigación. De dicha amalgama teórico-práctica es que surge la necesidad de un curso de formación en razonamiento clínico, y la evaluación de sus resultados.

Durante los años 2017 y 2018 realicé dos pruebas piloto de un curso de formación en razonamiento clínico para médicos residentes, que me permitieron adquirir experiencia en el diseño curricular, elaboración y perfeccionamiento de las estrategias de enseñanza, adaptación de los contenidos a las características particulares de los residentes, y mejora en la elaboración e implementación del instrumento de evaluación.

Los resultados de las autoevaluaciones de competencias de los residentes durante dichas pruebas piloto fueron similares a las del presente proyecto. A pesar de las importantes limitaciones que tienen las autoevaluaciones (K. W. Eva & Regehr, 2008), hay consistencia de los resultados a lo largo de 3 años consecutivos con casi 60 residentes involucrados, en cuanto a una *baja autopercepción de sus competencias en razonamiento clínico*.

Marco Conceptual

La enseñanza del razonamiento clínico debe estar basada en las teorías que nos permiten entender cómo el conocimiento de los problemas médicos evoluciona en la mente de los estudiantes (Schmidt & Mamede, 2015).

Se detallan a continuación el marco conceptual que guía la presente investigación.

Teoría de la Carga Cognitiva

Una de las teorías que permite abordar la enseñanza de habilidades complejas es la teoría de la carga cognitiva (Van Merriënboer & Sweller, 2005). La misma distingue la memoria de trabajo, de muy corto plazo (segundos), de la memoria de largo plazo.

La memoria de trabajo, de acuerdo con esta teoría, no puede abordar más de siete elementos al mismo tiempo. Aunque habitualmente lo hace con tan sólo dos a cuatro elementos, durante unos pocos segundos, pasados los cuales toda la información se pierde a menos que ocurra un nuevo estímulo. La facilidad con la cual la información puede ser procesada en la memoria de trabajo es el foco de la teoría de la carga cognitiva (Van Merriënboer & Sweller, 2005).

La memoria de largo plazo, en cambio, consiste en esquemas que varían en grados de complejidad y automatización. La experticia humana resulta, de acuerdo con esta teoría, de la organización, riqueza y complejidad del conocimiento acumulado en los esquemas.

Estos esquemas organizan y almacenan el conocimiento, y a su vez reducen la sobrecarga de la memoria de trabajo, porque aún esquemas muy sofisticados pueden ser tratados como un único elemento por la memoria de trabajo. A su vez, los esquemas pueden automatizarse si se aplican repetidamente.

Una elevada carga cognitiva –la dificultad intrínseca o extrínseca de la tarea–, suele aumentar el número de errores y enlentece el desempeño (Van Merriënboer & Sweller, 2005). Entonces, la automatización de los esquemas evita la sobrecarga de la memoria de trabajo, y permite ejecutar tareas con facilidad y sin aparente esfuerzo, características del desempeño experto.

Ericsson analizó cómo la experiencia, ejercida como práctica deliberada, se relaciona con el desarrollo de experticia en Medicina y otros dominios relacionados, y describió cómo la automatización también conlleva el riesgo de la detención del desarrollo, al perderse el control sobre la ejecución de las habilidades, haciendo dificultosas las modificaciones intencionales (Ericsson, 2004).

De acuerdo a esta teoría los expertos se diferencian de los novatos porque:

- han automatizado muchas tareas rutinarias;
- su más profunda comprensión del dominio (a través de esquemas cognitivos más ricos y sofisticados), les permite a su vez:
- reconocer e interpretar los problemas nuevos en términos más generales;
- pueden monitorear y reflexionar sobre el propio desempeño para, en última instancia:
- detectar y corregir errores.

(Young, Van Merriënboer, Durning, & Ten Cate, 2014).

Por tanto, la automatización de las tareas, una red de conocimientos amplios, elaborados, integrados y profundos, el reconocimiento de patrones, la auto-regulación, reflexión crítica y meta-cognición son características del hacer experto.

La teoría de la carga cognitiva permite entonces enfocar los diseños de instrucción desde una perspectiva que no sólo favorezca el desarrollo de esquemas, sino también su automatización. La nueva información debe ser primero procesada por la memoria de trabajo para construir esquemas en la memoria de largo plazo.

Por tanto, para el desarrollo de experticia diagnóstica, en vez de mejorar las estrategias generales de resolución de problemas, las estrategias de enseñanza basadas en promover la calidad y la organización del conocimiento pueden ser valiosas (Monteiro & Norman, 2013). Como los novatos no pueden reconocer patrones en la mayoría de los problemas clínicos que enfrentan, deben usar predominantemente el razonamiento analítico. La práctica repetitiva es clave para el desarrollo de tareas recurrentes (Van Merriënboer & De Croock, 2002). Para las tareas no recurrentes, es clave que los aprendices practiquen sobre problemas con una variabilidad similar a la del mundo real (Young et al., 2014).

En resumen, de acuerdo con esta teoría la práctica recurrente, repetitiva y deliberada de tareas de aprendizaje ricas, verosímiles, y adaptadas a la etapa del desarrollo del estudiante, son la base del aprendizaje complejo. Su repetición y variabilidad a lo largo de los programas de entrenamiento permite el desarrollo de esquemas mentales y su automatización, liberando la memoria de trabajo y reduciendo, por tanto, la carga cognitiva del proceso.

La experticia se desarrolla a lo largo del tiempo, pero el tiempo dedicado al entrenamiento *por sí sólo* no puede ser un marcador subrogado de competencia (J. R. Frank, Snell, Englander, & Holmboe, 2017)

Teoría de los Scripts

Esta teoría deriva de las ciencias cognitivas, y busca explicar cómo las personas comprenden los eventos de la vida cotidiana, y por qué suele ocurrir en forma automática y sin esfuerzo (Charlin, Boshuizen, Custers, & Feltovich, 2007). De acuerdo a esta teoría, nuestra mente interpreta la información del mundo comparando sus atributos con los modelos mentales creados con anterioridad, verificando sus similitudes y diferencias, patrones o irregularidades (Lubarsky, Dory, Audétat, Custers, & Charlin, 2015a). Cuando el resultado de esa comparación “encaja” lo suficiente, la interpretación del objeto ocurre en forma automática, rápida y sin esfuerzo. A este proceso se lo denomina “reconocimiento de patrones”, y para ejemplificarlo se describe a continuación una situación de la vida cotidiana.

Cuando un adulto se encuentra con un animal de cuatro patas, que es paseado con correa por su dueño por el parque y ladra, esa persona fácilmente reconocerá al animal como un perro, aunque quizás nunca antes lo haya visto, o no conozca esa raza específica de perro. Es el conjunto de características (cuatro patas, ladra), atributos (uso de correa), elementos contextuales (paseo por el parque, probable mascota), relacionados e integrados en red lo que permite el reconocimiento automático.

Por el contrario un niño pequeño -digamos de 2 años de edad-, podría confundir a un mastín napolitano con un caballo, o a un chihuahua con una rata, porque no ha tenido la experiencia suficiente para reconocer sus similitudes y diferencias. Cuando pregunte: *¿es*

eso un perro?, la respuesta lo ayudará a complejizar su representación mental de lo que un perro puede ser. Y si además lo escuchase ladrar, o lo acariciase, la integración sensorial haría su parte en la construcción del conocimiento.

En el ámbito de las profesiones de la salud los *illness scripts* (guiones de enfermedad) son redes de conocimiento vinculadas a los aspectos clínicos, características patológicas de una entidad y sus diferentes opciones de manejo, que los profesionales de la salud adquieren a lo largo de su formación, y se van perfeccionando con cada encuentro clínico (Fournier, Demeester, & Charlin, 2008). Es conocimiento clínico organizado y orientado al paciente (Custers, 2014).

La acumulación de conocimientos es insuficiente para la construcción de los *scripts*, porque los mismos derivan su valor de su aplicabilidad en la práctica (Keemink, Custers, van Dijk, & Ten Cate, 2018). La repetición de la experiencia clínica ofrece múltiples oportunidades para su validación, ajuste y reestructuración.

La mayoría de las veces los *scripts* se activan cuando los individuos se encuentran en un contexto apropiado (Custers, 2014). En medicina, la activación de los *illness scripts* ocurre en respuesta a estímulos tempranos que se van revelando a medida que se desenvuelve el encuentro clínico (Charlin et al., 2007). Esos estímulos pueden ser de muy variado origen: síntomas, signos del examen físico, o actitudinales, por mencionar algunos. Aunque los *illness scripts* reúnen una red de conocimientos variados, su activación ocurre en forma integrada (Custers, 2014), “en bloque”.

En ocasiones el reconocimiento de patrones ocurre tan rápido, en forma inconsciente y automática, que el médico hace diagnóstico de un problema con sólo ver al paciente entrar al consultorio. Por ejemplo, un adulto mayor que ingresa caminando con pasos cortos, reducido balanceo de brazos, facies poco expresiva y temblor distal de reposo, muy probablemente tenga alguna forma de parkinsonismo.

De acuerdo con esta teoría, la experticia en medicina depende entonces no sólo del número de *illness scripts* disponibles, sino también del contenido y organización del repertorio individualizado de *scripts* en la memoria de largo plazo, adquirido a partir del aprendizaje y la experiencia (Custers, 2014). Acorde a un mayor número, pero también riqueza, complejidad y profundidad de *illness scripts*, los médicos expertos pueden entonces desempeñarse con alto grado de automatismo, eficiencia y precisión. El “ojo clínico” que esas representaciones mentales les permiten reconocer patrones para la mayoría de los problemas clínicos con los que se enfrentan.

Los *scripts* maduros son redes de conocimiento ricas, flexibles y organizadas, que permiten una interpretación rápida y una acción eficiente (Lubarsky et al., 2015a). Por ejemplo, frente a un adulto con dolor urente severo en el tórax, con lesiones vesiculosas agrupadas de instalación aguda, el diagnóstico de herpes zóster es inmediato y permite su tratamiento precoz con antivirales y analgésicos.

Sin embargo, lograr la madurez de los *illness scripts* es un proceso largo y esforzado, que se inicia en los primeros años de la facultad de medicina, se afina a medida que aumenta la exposición a la práctica clínica, y continua a lo largo de la práctica profesional.

Lubarsky y col. identificaron algunas actividades de enseñanza alineadas con la teoría de los scripts (Lubarsky et al., 2015a), entre ellas: el aprendizaje basado en problemas, las auto-explicaciones, los mapas conceptuales, el pensamiento y reflexión en voz alta, y la lectura intencionada y dirigida para el desarrollo de los scripts. En cambio, la presentación de casos complejos o raros –por muy interesantes que puedan ser–, no contribuyen al desarrollo de los *illness scripts* en estudiantes (Custers, 2014).

A partir de la enseñanza basada en casos, se han realizado esfuerzos por evaluar la riqueza y madurez de los *scripts* de estudiantes de medicina en el ciclo pre clínico (Keemink et al., 2018), como objetivo subrogado de la evaluación de la competencia en razonamiento clínico.

También, la teoría de los *illness scripts* derivó en el desarrollo de un tipo específico de evaluación del razonamiento clínico (Custers, 2014): el *script concordance test*, que se analiza en extenso más adelante en este trabajo (ver [Evaluación del Razonamiento Clínico](#) y [Evaluación](#)).

Teoría del Procesamiento Dual

La teoría del procesamiento dual describe la arquitectura del pensamiento y la memoria. De acuerdo a la misma, la mente contiene una memoria de corto plazo, de capacidad limitada, que permite acceder a la memoria de largo plazo, de naturaleza asociativa y capacidad, en principio, ilimitada. Los recuerdos son recuperados de acuerdo a la fortaleza de su asociación con la información nueva (G. R. Norman et al., 2017).

A partir de ello, se describen dos mecanismos de procesamiento mental de la información. Uno rápido, intuitivo y cognitivamente no demandante, llamado Tipo 1. Y por otro lado, un mecanismo analítico, deductivo, y por tanto lento y cognitivamente esforzado, llamado tipo 2.

El procesamiento de tipo 1 busca rápidas asociaciones entre la nueva información y ejemplos similares en la memoria. La capacidad de establecer asociaciones depende de múltiples factores, tanto del objeto de investigación (forma de presentación, complejidad), como del sujeto que procesa la información (Richmond, Cooper, Gay, Atiomo, & Patel, 2020), no sólo de tipo cognitivo (amplitud y profundidad del conocimiento), como también emocionales (estado de ánimo, tolerancia a la incertidumbre, nivel de aversión a la toma de riesgo, entre otros) y contextuales (tiempo disponible, condiciones de trabajo, estímulos positivos y negativos), por mencionar algunos. Como el sistema 1 es intuitivo, operando por asociación con la memoria de largo plazo, permite resolver problemas de manera rápida, eficiente y no esforzada, sin sobrecargar la memoria (limitada) de trabajo.

El sistema del pensamiento analítico, o tipo 2, opera en la memoria de trabajo utilizando reglas lógicas para el procesamiento analítico de la información. Es, por lo tanto, un

proceso mucho más lento, esforzado y de capacidad limitada (ver [Teoría de la Carga Cognitiva](#)).

La teoría del procesamiento dual ha sido tenida en cuenta para el análisis de los errores en medicina (Croskerry, 2003; Graber, 2009; Graber et al., 2012; G. R. Norman et al., 2017). Aunque la mayoría de los errores diagnósticos son de origen multifactorial, los errores cognitivos son una causa común (Croskerry, 2003; Graber, Franklin, & Gordon, 2005). Diferentes sesgos cognitivos, como por el ejemplo el sesgo de confirmación y la satisfacción de la búsqueda (“*search satisficing*”), contribuyen al cierre precoz (Croskerry, 2003), y son mecanismos relacionados con el procesamiento intuitivo, no analítico y automático de la información clínica.

La mayoría de los médicos clínicos operan con niveles intermedios de esfuerzo y propensión al error, utilizando una combinación simultánea e interactiva del Sistema 1 y Sistema 2 (Graber, 2009). A partir de ello, muchos autores han propuesto estrategias para la enseñanza del razonamiento clínico y la reducción de los errores cognitivos, basados en estos dos mecanismos de procesamiento de la información descritos en esta teoría (Graber, 2009; J Rencic & Trowbridge, 2011; Trowbridge, 2008; Trowbridge, Dhaliwal, & Cosby, 2013; Trowbridge & Olson, 2018).

Lambe y col. (Lambe, O’Reilly, Kelly, & Curristan, 2016) realizaron una revisión sistemática de las intervenciones cognitivas basadas en la teoría del procesamiento dual para mejorar el razonamiento clínico. Entre las intervenciones educativas dirigidas a la adquisición de conocimientos o habilidades –con evidentes diferencias en el diseño de la investigación (población, intervención, resultados)-, ninguno de estos estudios demostró impacto significativo de las intervenciones en los objetivos evaluados. Dentro de las intervenciones educativas en el ámbito de trabajo, la reflexión guiada emergió como la más promisoría de las intervenciones.

En una investigación sobre los mecanismos cognitivos que influyen las respuestas de residentes y cirujanos mediante *script concordance test*, Gawad y col. describieron cómo los respondedores confían en los *scripts* formados a partir de la experiencia pasada, cuando están disponibles (Gawad, Wood, Cowley, & Raiche, 2021), y toman decisiones usando la deliberación consciente, en ausencia de la misma. Estos resultados son consistentes con el sistema 1 y 2 del procesamiento de la información, respectivamente, y suman al conjunto de evidencia para la validez de este instrumento de evaluación del razonamiento clínico.

La teoría del procesamiento dual ha sido criticada, porque representa sólo los extremos de un proceso que –en realidad- ocurriría a lo largo de un continuo (Custers, 2013). Desde el punto de vista educativo, el objetivo es preparar a los aprendices para adaptarse apropiada y efectivamente entre estos dos modos (Young et al., 2014).

La Competencia en Razonamiento Clínico

A lo largo de la historia ha habido notorios cambios en los discursos que definen el constructo de competencia en Medicina (Hodges, 2006). Por lo tanto, las definiciones de competencia han sido, son y seguirán siendo variables.

El marco de referencia CanMEDS para la educación basada en resultados, define al constructo de competencia como “*una habilidad observable de un profesional de la salud, que se desarrolla por etapas desde novato hasta clínico experto*” (J. Frank et al., 2015).

La dificultad con el lenguaje utilizado hasta la fecha para definir las competencias, es que no ha sido adecuado para facilitar la traducción de los dominios de la competencia en prácticas de entrenamiento, el monitoreo y evaluación de los aprendices (ten Cate, 2013). La cuestión lingüística no es en absoluto una cuestión menor, porque implica formular, explícitamente, en los currículos de las carreras de medicina y en los programas de formación de posgrado qué implica ser médico, o médico especialista. La identidad médica está en cuestionamiento (Ten Cate, 2014). Para la formulación de los objetivos de formación, existe una fuerte corriente que propone operacionalizar los conocimientos, habilidades, actitudes y competencias de una manera distinta, a saber: cuáles son las tareas que los graduados deben ser capaces de hacer en su trabajo diario (Ten Cate, 2018). Su acrónimo es, como ya se dijo, EPAs (*entrustable professional activities*), y su traducción al castellano es dificultosa. Se ha utilizado para ello la sigla APROC, en relación a “*actividades profesionales confiables*”.

Independientemente del marco de referencia utilizado para el diseño curricular, existe amplio consenso en que el razonamiento clínico es una competencia fundamental en la práctica (Amey, Donald, & Teodorczuk, 2017; Croskerry, 2009; Delany & Golding, 2014; Dhaliwal, 2013; Dumas, Torre, & Durning, 2018; Gay et al., 2013; Kassirer, 2010; Levin, Cennimo, Chen, & Lamba, 2016; J Rencic & Trowbridge, 2011; Schmidt & Mamede, 2015; Simpkin et al., 2017). Para ser un médico competente, éste debe ser competente en razonamiento clínico.

El razonamiento clínico es una habilidad compleja, lo cual representa un desafío para su enseñanza, porque tiene sus desafíos particulares y requiere de competencias docentes específicas (Janssen-Noordman, Merriënboer, van der Vleuten, & Scherpbier, 2006).

La enseñanza del razonamiento clínico requiere, por tanto, el abordaje de la complejidad. Los paradigmas cognitivista y constructivista son los más susceptibles de favorecer el desarrollo de conocimientos, habilidades y actitudes que sean significativos, reutilizables y transferibles (Durante, 2012).

Desde la perspectiva cognitivista, las experiencias concretas y completas (*whole-task*), verosímiles, auténticas y significativas son las máximas promotoras del aprendizaje (Van Merriënboer & De Croock, 2002).

Consistentemente a lo largo de los últimos 40 años de investigación en medicina en general, y en razonamiento clínico en particular, no fue posible demostrar transferencia de un contenido a otro (Keemink et al., 2018; Lineberry, Kreiter, & Bordage, 2013; G. Norman, Bordage, Page, & Keane, 2006), al punto que muchos autores cuestionan la propia existencia y la posible enseñanza de una habilidad general del razonamiento clínico (Hodges, 2006; Monteiro, Sherbino, Sibbald, & Norman, 2020; G. Norman, 2006). En ese sentido, Norman afirma que “la experticia no es una cuestión de adquirir

cierta clase de estrategia general de razonamiento” en la que esté todo incluido, sino que “no hay sustituto para la experiencia” (G. Norman, 2006).

La observación de que la solución de un problema predice pobremente la solución de otro problema similar fue descrito por primera vez por Elstein y col. (Elstein, Shulman, & Sprafka, 1990), quienes lo denominaron “especificidad de caso”.

La visión actual del problema es que **la experticia en razonamiento clínico es contenido-específica** (Keemink et al., 2018; Moulton, Regehr, Mylopoulos, & MacRae, 2007; C. P. Van Der Vleuten, 1996) **y contexto-específica** (Bowen & Ilgen, 2014; Brailovsky, 2001; Croskerry, 2009; S. Durning, Artino, Pangaro, van der Vleuten, & Schuwirth, 2011; K. Eva, 2005; Moulton et al., 2007).

Respecto al contenido-especificidad de la competencia, esto significa que alguien puede alcanzar altos niveles de desempeño, autonomía y metacognición en un determinado campo, pero ello puede no ser trasladable a otro, incluso dentro de la misma disciplina. Por ejemplo, un médico internista puede ser experto en el tratamiento de la enfermedad tromboembólica venosa, pero enfrentar dificultades en el manejo de un paciente con sepsis.

No menos importante, el desempeño individual en un problema particular predice pobremente el desempeño del mismo individuo con el mismo problema, pero en una situación diferente. A este fenómeno se lo denomina “especificidad de contexto”.

La epistemología del razonamiento clínico determina cómo se conceptualiza el constructo “contexto”. El significado de “contexto” se produce, negocia y legitima dentro de los diferentes tradiciones discursivas (Koufidis, Manninen, Nieminen, Wohlin, & Silén, 2021). Las diferentes conceptualizaciones tienen importantes implicancias tanto para el fenómeno observado de lo *contexto-específico* de la competencia, como para el aprendizaje del razonamiento clínico, y la investigación en educación médica.

Además, desde el punto de la evaluación de la competencia, es crítico comprender si el razonamiento clínico es una disposición relativamente estable o, en cambio, un estado dependiente del contexto (S. J. Durning, 2010; K. W. Eva, 2003).

Se ofrecen a continuación algunos ejemplos de cómo, en la práctica profesional, la experticia en razonamiento clínico es contexto-específica.

Primero, porque la *probabilidad pretest* de una entidad depende de la prevalencia (Silva Fuente-Alba & Molina Villagra, 2017). La epidemiología, entonces, es un elemento clave para la enseñanza del razonamiento clínico (Joseph Rencic, 2011), pero la misma es variable de acuerdo al espacio y tiempo. Frente a un paciente joven con hemoptisis crónica, por ejemplo, podría ser válido considerar en primer término el diagnóstico de tuberculosis en un país endémico como la Argentina, pero sería inadecuado en otros países de baja prevalencia, como Canadá o Suecia.

Segundo, porque aún dentro de una misma región geográfica, puede haber enormes diferencias de acuerdo al estrato socioeconómico de la población atendida en un

determinado centro. En tanto los factores socio-económico-ambientales tienen un enorme impacto en la salud de la población, la atención de un determinado grupo socio-económico (por ejemplo, en un hospital privado vs público) modificará la prevalencia de patologías en dicho centro.

Tercero, aún dentro una misma región, y un determinado estrato socio-económico atendido en cierto centro de salud, habrá diferencias según el rango etario de la población. La asistencia de pacientes añosos es fundamentalmente diferente de los pacientes adultos jóvenes.

Cuarto, la práctica médica asistencial requiere de habilidades comunicacionales y vinculares entre el médico, los pacientes y el resto del equipo médico, que están atravesadas por las diferencias culturales del lenguaje, el idioma y las normas sociales, entre otras.

Por ejemplo, el estado emocional del médico puede modificar su atención, reflexión y meta cognición. La capacidad de reconocer, evaluar y manejar las propias emociones se denomina inteligencia emocional (Murinson, Agarwal, & Haythornthwaite, 2008), y es otra fuente común de errores (Croskerry & Norman, 2008; M. Nendaz & Perrier, 2012).

La experticia, por lo tanto, no es un rasgo de la personalidad ni un estado, sino un *atributo*, dinámico y flexible, que puede aplicarse o no, de acuerdo al contexto (S. J. Durning, 2010; K. W. Eva, 2003).

Intervenciones Educativas en Razonamiento Clínico

La educación médica tradicional organiza la enseñanza en forma desintegrada y descontextualizada del paciente, generando grandes dificultades para organizar el conocimiento de forma relevante para la práctica profesional. La memoria requiere de la organización del conocimiento en forma contextualizada (C. P. Van Der Vleuten, 1996). Del mismo modo, ese tipo de enseñanza no facilita el desarrollo de competencias, porque el conocimiento desintegrado no puede ser transferido a la práctica médica (Van Merriënboer & De Croock, 2002).

Teniendo en cuenta el nivel de desarrollo de los estudiantes, sus conocimientos previos y las necesidades intelectuales acordes, se podría favorecer el desarrollo de la competencia en razonamiento clínico.

La presente investigación no pretende realizar una revisión sistemática de la bibliografía sobre enseñanza del razonamiento clínico, puesto que está por fuera de los objetivos de la misma. Se analizarán algunos elementos clave en la enseñanza del razonamiento clínico de acuerdo con la teoría, y la mejor evidencia disponible.

Como el razonamiento clínico es una habilidad compleja, la enseñanza del razonamiento clínico requiere, por tanto, el abordaje de la complejidad. Esto tiene desafíos particulares y requiere de competencias docentes específicas (Janssen-Noordman et al., 2006; Van Merriënboer & De Croock, 2002; Van Merriënboer & Kirschner, 2007; Young et al., 2014).

La epistemología del razonamiento clínico se relaciona en esencia con diferentes teorías del conocimiento. Para el diseño de las actividades de aprendizaje, la presente investigación se apoya principalmente en los paradigmas cognitivista y constructivista como teorías del aprendizaje (Illing, 2010). Se reconoce, sin embargo, la importancia de otras teorías (como la Cognición Situada, la Cognición Distribuida, el Aprendizaje Situado y la Psicología Ecológica), su potencial en la conceptualización del marco de referencia del razonamiento clínico, y sus implicancias para el aprendizaje.

Los paradigmas cognitivista y constructivista son susceptibles de favorecer el desarrollo de conocimientos, habilidades y actitudes que sean significativos, reutilizables y transferibles (Durante, 2012).

Desde la perspectiva constructivista, el conocimiento se construye a partir de la interacción y la comprensión del que aprende con el objeto de conocimiento (Brailovsky, 2001). Por tanto, se buscó generar experiencias auténticas en las que los estudiantes asuman un rol central y protagónico en su desarrollo, para lograr aprendizajes significativos (Ausubel, 1976).

Desde la educación basada en resultados, el rol docente consiste en maximizar las oportunidades de aprendizaje para los estudiantes. Por lo tanto, la educación no está centrada en el docente, como en la enseñanza tradicional, sino en el logro de resultados (aprendizajes) observables por parte de los estudiantes. El desarrollo de currículos y sistemas de evaluación basados en resultados son desafíos vigentes de la educación médica.

Como se mencionó antes, el razonamiento clínico es una habilidad compleja, y el desarrollo de experticia es contenido y contexto-específica. Por lo tanto, la enseñanza y evaluación del razonamiento clínico son procesos igualmente desafiantes, al estar influenciada por aquellos mismos determinantes (Kassirer, 2010).

El error médico es la tercera causa de muerte en EEUU (Makary & Daniel, 2016). En virtud de que los mecanismos del error médico involucran aspectos cognitivos en un 75% de los casos (M. Nendaz & Perrier, 2012), y que “la mayoría de los errores cognitivos no están relacionados con un déficit de conocimiento, sino con fallas en la recolección, integración y verificación de la información...” (M. Nendaz & Perrier, 2012, p. 1), las estrategias de enseñanza dirigidas a minimizar este tipo de errores serían de gran valor. En 2015, la Academia Nacional de Medicina (EE UU) instó a generar un enfoque amplio para evitar los errores y propone estrategias para la mejora en los diagnósticos (Balogh, Miller, & Ball, 2015), entre ellas la “mejora en la educación para profesionales de la salud, y entrenamiento en el proceso diagnóstico”.

Existe controversia, sin embargo, acerca de cuáles son los mejores mecanismos para enseñar razonamiento clínico (Amey et al., 2017; Bowen, 2006; Gay et al., 2013; Kassirer, 2010; Levin et al., 2016; Monteiro & Norman, 2013; Joseph Rencic, 2011; Schmidt & Mamede, 2015), y cómo reducir los errores en dicho proceso (Bowen, 2006; Monteiro & Norman, 2013; G. Norman, 2005; G. R. Norman et al., 2017).

Una revisión sistemática y síntesis narrativa (Thistlethwaite et al., 2012) sobre la efectividad del aprendizaje basado en casos en la educación de los profesionales de la salud demostró abrumadoramente que tanto los estudiantes, como los docentes, disfrutaban esta estrategia. El disfrute podría tener un refuerzo positivo sobre la motivación de los estudiantes y, al mismo tiempo, mejorar el uso del tiempo y la experiencia docente.

Asimismo, el aprendizaje basado en casos parece favorecer el aprendizaje efectivo en grupos pequeños (Thistlethwaite et al., 2012).

Una revisión sistemática de diferentes intervenciones educativas en razonamiento clínico (Lambe et al., 2016) identificó muchas recomendaciones, pero relativamente pocos ensayos empíricos que examinen su efectividad.

En 2020 Richmond y col. publicaron una revisión realista de las intervenciones educativas para desarrollar las habilidades del razonamiento analítico y no analítico (Richmond et al., 2020). Los autores reconocen y enfatizan en este trabajo la importancia de los elementos contextuales (conocimientos previos, autoestima, estrategias de resolución de problemas) de los estudiantes y los docentes, cuando se implementan y evalúan intervenciones educativas en razonamiento clínico.

Schaye et al. analizaron los resultados de implementar un currículo basado en el ámbito de trabajo para residentes de Medicina Interna, con intervenciones basadas en teorías del razonamiento clínico, como la teoría del procesamiento dual y la teoría de los scripts (Schaye, Elias, Janjigian, & Stern, 2019). Aunque hubo diferencias estadísticamente significativas en la aplicación de conocimientos relacionados al razonamiento clínico, no pudieron demostrar diferencias a través del instrumento de evaluación utilizado (DTI, o *Diagnostic Thinking Inventory*), que auto evalúa la aproximación al razonamiento clínico sobre un caso específico.

Se destacan a continuación otras dos investigaciones experimentales sobre las intervenciones educativas en razonamiento clínico, por sus relaciones con la presente investigación.

Compère y col. evaluaron los resultados de una intervención educativa controlada para el desarrollo del razonamiento clínico en residentes de anestesiología, a través de un programa de entrenamiento guiado por un experto, y basado en la discusión entre pares de casos clínicos (Compère et al., 2016). En esa investigación los resultados se obtuvieron a partir de una evaluación posterior a la intervención, sobre una población seleccionada, lo cual introduce un importante sesgo en la interpretación de sus resultados.

Se realizaron otras intervenciones educativas, pero en forma no controlada (Boulouffe, Charlin, & Vanpee, 2010; Kazour, Richa, Zoghbi, El-Hage, & Haddad, 2017), ofreciendo una menor calidad de evidencia. También existe evidencia de que la participación de ciertos programas de formación profesional en ciencias de la salud, se asocia con mejores resultados en razonamiento clínico, aunque la calidad de la evidencia es débil (Otterman, Maas, Schiemanck, van der Wees, & Kwakkel, 2019).

Evaluación del Razonamiento Clínico

Las mismas dificultades que impiden generalizar el diseño de un programa de entrenamiento en razonamiento clínico, también son aplicables a su evaluación. El razonamiento clínico no es una habilidad discreta y binaria. Tampoco un rasgo de la personalidad, persistente e inmutable.

En términos prácticos y generales, la evaluación del razonamiento clínico debe incorporar diferentes tipos de problemas diagnósticos y contextos variables, con el objetivo de capturar al amplio rango de conocimientos y habilidades necesarios para manejar escenarios comunes y relevantes de una determinada disciplina clínica (Monteiro et al., 2020).

Difícilmente evaluable en forma global, obliga a realizar inferencias a partir de evaluaciones parciales, que –analizadas individualmente– pueden resultar poco confiables, o sesgadas (Bowen & Ilgen, 2014).

Por este motivo, se han propuesto reglas para desarrollar *programas* de evaluación de la competencia (C. P. M. van der Vleuten & Schuwirth, 2019), entendiendo las limitaciones y fortalezas individuales de cada uno de los diferentes instrumentos de evaluación del razonamiento clínico.

Existen instrumentos que permiten evaluar el razonamiento clínico tanto en el entorno clínico (miniCex, notas escritas, pensar en voz alta), como fuera del mismo (*multiple choice questions, key-feature examinations, script concordance test*). También en ambientes simulados a través de OSCE y del uso de la tecnología (maniqués de alta fidelidad, realidad virtual, entre otros) (Daniel et al., 2019).

Es ilusorio pretender la perfección en un único instrumento de evaluación (C. P. Van Der Vleuten, 1996). Cada equipo docente deberá tomar decisiones respecto a la validez, confiabilidad e impacto educacional del instrumento a utilizar, considerando también su aceptabilidad y costo, de acuerdo con sus propios saberes, preferencias y posibilidades (Durante, 2012). Como se mencionó, la evaluación de las habilidades complejas en el contexto de la educación basada en competencias, requiere una aproximación más holística e integral (C. P. M. van der Vleuten & Schuwirth, 2019).

En el caso de la presente investigación, se eligió utilizar *script concordance test* (SCT) por su validez, confiabilidad, factibilidad, aceptabilidad y costo. Se reconoce que este instrumento, como todos, tiene limitaciones, sobre todo en los procesos de calificación (Lineberry et al., 2013) y en las respuestas del panel de expertos como standard de referencia (Kreiter, 2012; Lineberry et al., 2019).

Script Concordance Test

El *script concordance test* (SCT) es un instrumento de evaluación del razonamiento clínico basado en la simulación escrita de casos clínicos (Charlin, Roy, Brailovsky, Goulet, & Van Der Vleuten, 2000), y se utiliza para evaluar un aspecto específico de la competencia en razonamiento clínico: la habilidad para interpretar la información médica, bajo condiciones de incertidumbre (Lubarsky, Dory, Duggan, Gagnon, & Charlin, 2013).

El objetivo es explorar la estructura y organización del conocimiento, estableciendo relaciones aplicadas a problemas médicos.

El SCT es un instrumento de evaluación del razonamiento clínico fuera del ámbito de trabajo (Daniel et al., 2019), que ha demostrado adecuada validez de constructo en un amplio espectro de diferentes profesiones de la salud y en diversas especialidades clínicas (Dory, Gagnon, Vanpee, & Charlin, 2012), incluyendo la Medicina Interna (Marie et al., 2005).

Estudios de diferentes especialidades médicas han demostrado que el SCT permite discriminar entre diferentes *niveles de experticia* en razonamiento clínico, entre ellas Pediatría (Carrière, Gagnon, Charlin, Downing, & Bordage, 2009), Emergentología (Humbert, Besinger, & Miech, 2011), Psiquiatría (Kazour et al., 2017), Diagnóstico por Imágenes (Brazeau-Lamontagne, Charlin, Gagnon, Samson, & Van Der Vleuten, 2004), Anestesiología (Compère et al., 2016), entre otras, aportando a la validez de constructo del instrumento.

También demostró adecuada confiabilidad y validez en especialidades quirúrgicas, en un estudio de alto impacto educativo a nivel nacional canadiense (Nouh et al., 2012).

Aunque el SCT fue desarrollado inicialmente para médicos residentes, su uso se ha extendido también a alumnos de la facultad de Medicina (Humbert, Johnson, et al., 2011).

Diseño del Instrumento de Evaluación

Se les presenta a los evaluados una serie de escenarios clínicos breves, seguidos de preguntas que requieren del juicio sobre posibilidades diagnósticas, terapéuticas o de manejo. Los casos deben ser problemáticos, ofreciendo cierto grado de imprecisión, incertidumbre o indefinición intencionada, que caracterizan los encuentros clínicos de la vida real (Lubarsky, Dory, et al., 2013)

A cada escenario le sigue una hipótesis, y una porción de información adicional que puede modificar la probabilidad de la hipótesis. Se ofrecen ejemplos en el [apéndice III](#). Este formato refleja la manera en que la información es procesada en situaciones clínicas auténticas (Lubarsky, Gagnon, & Charlin, 2013). Las hipótesis pueden ser de diagnóstico, investigación o tratamiento.

Cuando el evaluado responde este tipo de evaluación, se le está pidiendo que realice varias operaciones del razonamiento clínico en forma sucesiva (Robert Gagnon, Lubarsky, Lambert, & Charlin, 2011). Primero, debe identificar el/los problemas y establecer probabilidades previas (probabilidad pre test). A continuación, se debe juzgar e interpretar la información adicional, para determinar cómo modifica la probabilidad de la hipótesis dada, para ese caso en particular (probabilidad pos test).

Se le pide al evaluado que responda en una escala tipo Likert (-2: fuertemente en contra de la hipótesis, -1: parcialmente en contra de la hipótesis, 0: ni a favor ni en contra, +1: parcialmente a favor de la hipótesis, +2: fuertemente a favor de la hipótesis), cómo

modifica la información adicional la probabilidad de la hipótesis presentada, para el caso en cuestión.

Por este mecanismo de respuesta es que se considera que “no hay una respuesta correcta”, sino que se asigna diferente puntaje de acuerdo a la distribución de respuestas del panel de expertos. En resumen, a mayor concordancia con las respuestas de los expertos, mayor puntaje. Se desarrollará con más detalle el sistema de puntuación más adelante (ver [Proceso de Calificación](#)).

La principal ventaja de este formato de respuesta radica en su capacidad de realizar preguntas clínicas útiles para evaluar la estructura del conocimiento clínico del respondedor (Kreiter, 2012).

Respecto al contenido y el número de casos, el instrumento de evaluación se diseñó de acuerdo a las recomendaciones vigentes (Charlin et al., 2010; Dory et al., 2012; Fournier et al., 2008; Robert Gagnon et al., 2011; Wilson, Pike, & Humbert, 2014). Es recomendable confeccionar un esquema o tabla resumen (*test blueprint*, en inglés) antes de confeccionar los scripts, para asegurar que están cubiertos los dominios del conocimiento esperados (Lubarsky, Dory, et al., 2013). En el esquema se detallan la complejidad de los síntomas, las enfermedades específicas consideradas, la toma de decisiones clínica (diagnóstico, investigación o manejo), el ámbito de práctica simulado (ambulatorio, emergencias, internación, etc) y otros elementos contextuales.

Teniendo en cuenta que el razonamiento clínico es contenido-específico, el número de casos debe ser suficiente para garantizar la validez del instrumento (Dory et al., 2012; Fournier et al., 2008). R. Gagnon y col. analizaron mediante el estudio psicométrico de tres evaluaciones, que 25 casos con tres ítems cada uno resultan en pruebas con confiabilidades adecuadas, con coeficientes G de entre 0,75 y 0,86 (Robert Gagnon, Charlin, Lambert, Carrière, & Van Der Vleuten, 2009). Por tanto, la confiabilidad de la calificación es óptima con 3 ítems por cada caso (Dory et al., 2012; Fournier et al., 2008; G. Norman et al., 2006), con una duración estimada de hasta 90 minutos.

Aspectos clave del diseño de los scripts

Especial atención requiere no sólo el número de casos e ítems, sino también el diseño del *contenido* de los scripts. Un aspecto crítico del SCT es el juicioso diseño, selección y organización de las viñetas, hipótesis e información adicional de cada script. Existen una serie de recomendaciones para los desarrolladores (Charlin et al., 2000; Dory et al., 2012; Fournier et al., 2008; Lubarsky, Dory, et al., 2013):

- La viñeta debe ser breve y contener suficiente información para representar el problema, pero al mismo tiempo tener cierto grado de imprecisión o incertidumbre propias de la práctica profesional.
- El caso/viñeta debe ser auténtico y verosímil, representativo de la práctica habitual;
- La construcción de los ítems requiere considerar los “elementos clave” (*key features*) del caso (G. Bordage & Page, 2018). Es decir, la elección de la pregunta

se debe enfocar en aquellos elementos que son críticos y desafiantes para la resolución de ese caso en particular.

- La hipótesis debe ser válida para la viñeta en cuestión;
- Cada pregunta debe generar una reflexión, y la respuesta de un ítem debe ser independiente de los otros ítems del caso. Las respuestas no deben estar “encadenadas” dentro de un mismo caso.
- Las preguntas no deben evaluar sólo conocimiento;
- Tanto la viñeta como de la información adicional deben ser necesarias para ser necesarias para responder la pregunta. Una forma de reconocer este aspecto, es ocultar la viñeta, y considerar si es posible responder la pregunta sin ella. En caso afirmativo, el script debe modificarse.

Cualquier fallo en atender a estos puntos puede comprometer la validez o confiabilidad del instrumento.

Es posible incluir material audiovisual (por ejemplo: signos clínicos, electrocardiogramas, radiografías de tórax) para aumentar la *autenticidad* de la experiencia del evaluado (Lubarsky, Chalk, Kazitani, Gagnon, & Charlin, 2009).

Es de particular relevancia evitar diseñar scripts que evalúen conocimiento factual, porque los mismos tienen alta concordancia (baja o nula varianza) y pobre poder de discriminación entre niveles de experticia. En el análisis estadístico posterior, los ítems con varianza nula no son tenidos en cuenta. También es posible que a posteriori se eliminen algunos de ellos, en favor de una mayor confiabilidad del instrumento (ver [Optimización del Test](#) para más detalle).

Todo instrumento de evaluación debe tener en cuenta las características de los examinados, sus conocimientos previos y el impacto educativo del test (Durante, 2005). La dificultad del examen se ajusta a estos parámetros. El SCT no es la excepción.

Es recomendable que al menos dos autores sean responsables del desarrollo de los scripts (Lubarsky, Dory, et al., 2013). Los desarrolladores deben estar familiarizados con el propósito, la población objetivo y tener experiencia/supervisión en el diseño del instrumento. Se recomienda enviar un set preliminar a dos o tres revisores independientes, para recibir feedback acerca de la claridad y relevancia de los scripts (Fournier et al., 2008).

El SCT es susceptible de sesgos en el mecanismo de respuesta que son irrelevantes para el constructo (Lineberry et al., 2013). Los autores analizan cómo, por ejemplo, adivinar respondiendo al centro (0) de la escala es superior a utilizar la escala como se espera. También demuestran cómo se pueden perjudicar los respondedores de los extremos (-2 / +2).

Aunque no hay evidencia de que responder al azar sea una estrategia de respuesta ventajosa, si es posible que responder “siempre al 0” genere resultados artificialmente elevados (Lubarsky, Dory, Meterissian, Lambert, & Gagnon, 2018), incluso similares a los de estudiantes avanzados de Medicina que responden fielmente el examen.

Aunque no es posible eliminar por completo estos sesgos utilizando el formato de calificación agregada, se los tuvo especialmente en cuenta para el diseño de los casos. Primero, evitando diseñar scripts de conocimiento factual, o que pueden ser respondidos sin tener en cuenta la viñeta. Segundo, anticipando y distribuyendo las respuestas esperadas a lo largo de todo el espectro de la escala de Likert (Wan, Tor, & Hudson, 2018).

Por último, limitando la proporción de ítems con respuesta estimada modal '0' del panel de expertos a menos del 10%, se podría eliminar el beneficio de responder siempre "al centro (0)" (Lubarsky et al., 2018).

El Panel de Expertos

El SCT no dispone de una "respuesta correcta" para cada pregunta, sino que las respuestas de los evaluados se compara contra el conjunto de respuestas de un panel de expertos.

Primero, es fundamental considerar cuál es el objetivo e impacto educativo de la evaluación (R. Gagnon, Charlin, Coletti, Sauvé, & Van Der Vleuten, 2005). De acuerdo a ello, la selección de las características y número del panel de expertos son elementos clave del diseño de este instrumento.

Respecto al mecanismo de selección de los integrantes del panel de expertos existe controversia, precisamente por las dificultades en definir y medir la experticia en razonamiento clínico a lo largo del espectro de la disciplina evaluada (G. Norman, 2005). En vista de la falta de consenso respecto a aquello que constituye la experticia en las profesiones de la salud, los "expertos" son simplemente aquellos individuos que tienen reputación de un conocimiento superior, agudeza clínica, y experiencia en su dominio, y cuyas opiniones son requeridas en casos desafiantes (Lubarsky, Gagnon, et al., 2013).

En cuanto a su composición, es fundamental que el panel esté integrado por representantes de la profesión o especialidad correspondiente a ser evaluada (R. Gagnon et al., 2005; Petrucci, Nouh, Boutros, Gagnon, & Meterissian, 2013). Es recomendable que los médicos del panel deben tener buena experiencia *general* en el campo a ser evaluado, más que expertos en partes estrechas del conocimiento médico (Fournier et al., 2008).

Diferentes composiciones de paneles de expertos generan diferentes claves de calificación. Cuando el panel de expertos es seleccionado dentro de ámbitos clínicos relevantes a los objetivos de la evaluación, las notas resultantes son más altas (Dory et al., 2012). Por lo tanto, es importante destacar que inevitablemente cada panel de expertos incorpora cierta cantidad de error de medida al instrumento. Con paneles mayores o iguales a 15 expertos, las respuestas modales reciben proporcionalmente más peso comparado con las respuestas infrecuentes (ver [Proceso de Calificación](#)).

En lo concerniente a la cantidad de expertos que deben integrar el panel, el número recomendado depende del impacto educativo de la evaluación (Dory et al., 2012). Gagnon y col. demostraron que cualquier número por encima de 10 integrantes está asociado con

una confiabilidad adecuada, y con correlaciones mayores al 95% respecto a paneles más numerosos (R. Gagnon et al., 2005).

Proceso de Calificación

La calificación del SCT busca reflejar el grado de concordancia entre las habilidades interpretativas de los examinados y los expertos, bajo condiciones de cierta incertidumbre clínica (Robert Gagnon et al., 2011). El proceso de calificación se basa en el principio de que cualquier respuesta de un experto tiene valor intrínseco, aún si otros expertos no están de acuerdo (Charlin et al., 2000).

Una vez que los expertos respondieron individualmente los exámenes, se obtiene una clave de calificación. Las claves de calificación son específicas para cada examen, y para cada panel de expertos dado. Esto es relevante porque diferentes composiciones del panel de expertos generarán diferentes claves de calificación, y por lo tanto diferentes calificaciones. De allí, nuevamente, la importancia de la selección, número y composición del panel de expertos.

Calificación agregada

Para obtener el puntaje de los examinados, se le asigna un puntaje a cada ítem de cada caso clínico, de acuerdo a la concordancia con las respuestas del panel. La respuesta más elegida entre los expertos (moda) recibe la nota máxima; las respuestas menos elegidas reciben crédito parcial.

Para cada respuesta, el puntaje es el número de expertos que eligieron esa opción, dividido por el valor modal de esa pregunta. Se muestra un ejemplo del sistema de calificación en la siguiente tabla:

Sistema de calificación agregada

	-2	-1	0	+1	+2
Número de expertos que eligieron la respuesta	0	0	1	8	3
Número de expertos que eligieron la respuesta, dividido la respuesta modal	0/8	0/8	1/8	8/8	2/8
Calificación para este ítem	0	0	0.125	1	0.25

De este modo, se obtienen las calificaciones individuales de cada ítem, para cada evaluación. La suma de las calificaciones individuales de todo el examen se denomina 'calificación agregada' (*aggregate scoring*) (Charlin, Desaulniers, Gagnon, Blouin, & van der Vleuten, 2002), porque asume que las variaciones de respuesta de los expertos son todas válidas, y por lo tanto asigna diferentes puntajes a los estudiantes que respondieron en concordancia con dicha variabilidad (Lubarsky, Dory, et al., 2013).

El sistema de calificación agregada demostró mayor validez de constructo que otros sistemas de calificación, como el alcanzado por consenso de expertos (Charlin et al., 2002). A su vez, Wilson et al. analizaron seis métodos de calificación diferentes para

SCT, concluyendo que este método posee la mejor combinación de confiabilidad y validez (Wilson et al., 2014)

En el SCT, a diferencia de otros instrumentos de evaluación, el resultado final busca reflejar la diferencia en el desempeño entre los examinados y el del panel de referencia como conjunto. Para ello, es necesario conocer el desempeño de cada experto y su distribución estándar (*DE*) alrededor de la media (\bar{x}) del panel.

Para calificar individualmente a los expertos, el puntaje se calcula tratando a cada uno como si fuese un examinado, calificándolo con una clave de corrección (que lo excluye) del resto de los expertos (Charlin et al., 2010). A partir de los puntajes individuales, se calcula con facilidad la \bar{x} del panel de expertos, y su *DE*.

Una vez obtenidos los puntajes de los expertos, su media global y su distribución estándar, es posible también comparar el desempeño de los diferentes grupos participantes de la investigación.

El resultado de la evaluación de los estudiantes está compuesto por la suma de los puntajes obtenidos en cada ítem. El puntaje máximo para el test resulta de la suma de los puntajes máximos disponibles en cada ítem, que corresponde a la moda. Los resultados se pueden expresar en números absolutos (suma de puntajes individuales) o en forma relativa (porcentaje del total disponible).

Optimización del Test

Las preguntas que generan menos discordancia entre los expertos (menor dispersión a lo largo de la escala), no son muy diferentes de una pregunta de opción múltiple (de respuesta única). Las preguntas con una distribución de respuestas muy amplia pueden ser demasiado ambiguas. Desde el punto de vista psicométrico, las preguntas tipo SCT ideales son aquellas que producen un rango acotado de respuestas de expertos alrededor de una respuesta modal (Lubarsky, Dory, et al., 2013).

El sistema de calificación agregada puede ser problemático (Lineberry, Kreiter, & Bordage, 2014). Como se mencionó anteriormente, este sistema de calificación implica que es posible otorgar créditos parciales a respuestas desviadas de la moda o inapropiadas según el juicio de expertos (Robert Gagnon et al., 2011), o simplemente equivocadas de acuerdo a la medicina basada en la evidencia (Ahmadi et al., 2014).

Al respecto, las guías para la confección de SCT vigentes al inicio de la presente investigación (Dory et al., 2012; Fournier et al., 2008; Lubarsky, Dory, et al., 2013) describen la posibilidad de “optimizar” o “refinar” el test. Dicho procedimiento busca aumentar la confiabilidad del instrumento, a partir del análisis psicométrico de las respuestas del panel de expertos, y de los evaluados.

Quienes argumentan a favor de eliminar ítems con correlación ítem-total negativa o positiva muy baja, buscan aumentar la confiabilidad del instrumento. De allí, por tanto, su “optimización”.

Sin embargo, la remoción de ítems o scripts completos puede implicar una amenaza para la validez de contenido (Lineberry et al., 2013; Lubarsky, Dory, et al., 2013), por la especificidad de caso discutida previamente. En síntesis: un número insuficiente de casos podría invalidar la prueba.

En ocasiones puede ocurrir, además, que los ítems con respuestas extremas (-2, +2) sean preguntas más básicas, o que evalúen conocimiento factual y, en consecuencia, respuestas más definidas. Por tanto, dichas preguntas tendrán un menor poder discriminativo, y menor consistencia interna con el resto del examen, haciéndolas más susceptibles a ser eliminadas (Lineberry et al., 2013), porque los ítems con baja varianza en el panel de expertos tienden a ser fáciles y a discriminar pobremente. En consecuencia, la remoción de ítems podría estar sesgada en contra de los extremos de la escala. Pero al mismo tiempo, no remover dichos ítems compromete la consistencia interna y el poder discriminativo del test (Lineberry et al., 2013).

De acuerdo a la teoría de los scripts, los mismos se desarrollan a partir de la experiencia *individual* con casos *particulares*. Esto lleva al desarrollo de un cierto grado de idiosincrasia, particularmente con casos no prototípicos. Precisamente, el SCT utiliza casos no tan definidos, que incluyen dosis auténticas y verosímiles de incertidumbre o imprecisión, para evaluar la habilidad para interpretar la información médica, bajo dichas condiciones.

La variabilidad, por tanto, está en el centro de la argumentación a favor de la validez de contenido del script concordance test (Charlin et al., 2002, 2006), en términos de su habilidad para evaluar el razonamiento clínico en casos poco definidos (Dory et al., 2012).

Variabilidad en las respuestas del panel de expertos y el sistema de calificación agregada: efectos en la confiabilidad y validez del test

Existe acalorada controversia en cuanto a la interpretación y manejo estadístico de la variabilidad de respuestas del panel de expertos, a partir de las cuales se conforma la clave de corrección.

Hay, en síntesis, dos posiciones contrapuestas: por un lado aquellos que argumentan en favor de analizar la confiabilidad con el alfa de Cronbach de los ítems en relación al total del examen, y optimizar la confiabilidad del test eliminando aquellos ítems con correlación negativa o positiva baja.

Por otro lado, otros autores alertan sobre el riesgo de eliminar ítems en favor del (deseable) objetivo de aumentar la confiabilidad, en la medida que puede comprometer la validez del instrumento. En particular, al limitar la validez de contenido. En síntesis: menos ítems, menos contenido-específico a evaluar, y por tanto menor validez.

Se exponen a continuación algunos argumentos en favor y en contra de cada uno de ellas.

El uso tradicional del sistema de calificación agregada se basa en la asunción de que las diferencias en las respuestas no se deben *necesariamente* a errores o ideas falsas, sino potencialmente a la riqueza del pensamiento sobre casos clínicos auténticos y, por tanto, poco definidos (Lubarsky, Gagnon, et al., 2013).

Otros autores, en cambio, consideran que el proceso de respuesta del panel de expertos es inconsistente con esa clásica interpretación y uso del sistema de calificación agregada del SCT (Lineberry et al., 2019). Por ejemplo, porque fuerza a los respondedores a interpretar la información dada, salteándose la posibilidad de una mayor búsqueda de información adicional. También, porque la noción de la teoría de los scripts requiere de considerar la interacción de los pedazos de información clínica con su *contexto* (Lineberry et al., 2019), algo que los breves textos presentados en los casos no suelen tener (Lineberry et al., 2014).

C. Kreiter considera que suele haber poca discrepancia en cuanto a la *dirección* (a favor o en contra) del impacto de la información adicional en la probabilidad de la hipótesis, pero si en la *magnitud* de ese cambio de probabilidades (algo o muy en contra, y algo o muy a favor) (Kreiter, 2012). Por el formato de respuesta del SCT, sólo conocemos cómo el respondedor asigna una probabilidad *posterior* a la información adicional provista. Sin embargo, el autor conjetura que es necesario conocer cuál es la probabilidad pre test asignada por el experto, para interpretar el significado de la respuesta indicada (más o menos probable) (Kreiter, 2012).

Cuando ocurre que en el panel de expertos la información adicional de un ítem genera respuestas opuestas en la escala de Likert (por ejemplo -1 y +1), hay autores que cuestionan la validez del instrumento, porque “una porción de información no puede simultáneamente hacer la hipótesis más y menos probable” (Lineberry et al., 2013), y porque “hay una respuesta objetivamente correcta para cada ítem del SCT” (Lineberry et al., 2013). Desde el punto de vista estadístico, las respuestas desviantes de la moda pueden potencialmente disminuir la confiabilidad de la calificación, al introducir un error de medida (Robert Gagnon et al., 2011).

Como se mencionó anteriormente, el proceso de optimización del test puede realizarse sobre las respuestas de los panelistas *previo* a la administración de la evaluación a la población objetivo, y también puede realizarse *a posteriori*, sobre los ítems de los evaluados. Considero importante discriminar ambos procedimientos.

El análisis y optimización de la clave de corrección permitiría asegurar el control de la calidad del *desarrollo* de los SCT, asegurando tanto la validez de constructo como de contenido de la evaluación (Wan et al., 2018). Para ello, Wan y col. proponen descartar los ítems con: 1) respuestas bimodales, 2) divergencia uniforme, 3) respuestas con *outliers* discretos. De acuerdo a este proceso, se conservan sólo aquellos ítems en los que los expertos están mayoritariamente de acuerdo, y responden de manera “psicométricamente ideal” (próximos alrededor de una moda) (Wan et al., 2018).

Al mismo tiempo, Gagnon y col. determinaron que para paneles de al menos 15 integrantes no está justificado desde el punto de vista psicométrico la remoción de panelistas desviantes (Robert Gagnon et al., 2011), porque el error de medida que agregan es despreciable.

Dory y col. concluyen que “los diseñadores de test no deben necesariamente buscar su eliminación, en tanto no hay beneficios psicométricos para hacerlo” (Dory et al., 2012).

Medidas de confiabilidad y fuentes de error

Así como el coeficiente de alfa de Cronbach es una medida de confiabilidad, no es *equivalente* a confiabilidad. Tampoco es una medida exhaustiva de la misma para los SCT, por lo que varios autores –¡incluyendo el propio Cronbach!- (Lineberry et al., 2014), favorecen el uso de la *teoría de la generalizabilidad* para medir adecuadamente *todas* las fuentes de error.

Teoría de la Generalizabilidad

De acuerdo a la teoría clásica de los test, cualquier medida tiene dos partes: la verdadera varianza, y el error de la varianza. Existen, sin embargo, múltiples fuentes de error: factores del candidato, factores de la tarea, factores del sistema de calificación, factores situacionales, y factores por verdadero azar (Bloch & Norman, 2012). El problema con la teoría clásica de los test es que permite tratar con las múltiples fuentes de error de a uno a la vez. Lo que hace la teoría de la generalizabilidad es incluir todas las fuentes de varianza en un solo análisis (Bloch & Norman, 2012). Una de las preguntas que permite responder este análisis es “*¿cuáles son las fuentes de error más probables en esta situación particular evaluada?*” (Bloch & Norman, 2012, p. 967).

El análisis por generalizabilidad permite medir, entonces, las fuentes de varianza verdadera entre los respondedores y, por lo tanto, discriminar su desempeño.

Pregunta de Investigación

¿Cuáles son los resultados de un curso de entrenamiento en razonamiento clínico para residentes de medicina interna, evaluados mediante script concordance test?

Objeto de Investigación

La competencia en razonamiento clínico

Propósito

- Promover el aprendizaje del razonamiento clínico.

Objetivo General

- Evaluar los resultados de un curso de entrenamiento en razonamiento clínico para residentes de Clínica Médica de un hospital universitario de Buenos Aires.

Objetivos Específicos

- Determinar el nivel de competencia de un grupo de residentes, en ciertos aspectos del proceso de razonamiento clínico, mediante el *script concordance test* antes y después de la intervención educativa.
- Comparar los resultados del grupo intervención con los de un grupo control, en relación a un único panel de expertos.
- Conocer la factibilidad, aceptabilidad y valoración del uso del *script concordance test* como parte de una intervención educativa en razonamiento clínico.

Diseño de Investigación

Para responder a la pregunta de investigación se desarrolló un **estudio cuantitativo, experimental, prospectivo, de intervención educativa controlada**. Por ello la presente investigación se posiciona desde el punto de vista epistemológico desde una perspectiva positivista.

Asimismo, como la investigación en educación médica explora aspectos sociales, diversos y contexto-dependientes, se anticipan limitaciones -propias de toda intervención educativa- para la determinación de causalidad de los resultados a partir de la intervención, precisamente porque el aprendizaje no es un cambio lineal, sino complejo e inestable (Bunniss & Kelly, 2010). Las mismas serán analizadas más adelante.

Se determinó que una población de 18 residentes tendría un poder del 85% para detectar una diferencia absoluta de la media de 8 puntos entre los grupos, con un alfa global de 0.05. Se estimó un valor de sigma de 8 a partir de estudios previos (Charlin et al., 2010, 2000; Nouh et al., 2012).

Las residencias médicas son sistemas formativos insertos en un espacio asistencial. A partir del desarrollo y auge de la educación médica basada en competencias (EMBC) se ha demostrado que el mero paso del tiempo es un criterio inválido para certificar competencia (Ten Cate et al., 2015).

Al mismo tiempo, es de esperar cierto grado de desarrollo de las competencias –incluido el razonamiento clínico-, a medida que los residentes avanzan en el programa de formación. Por lo tanto, para poder discriminar si las eventuales diferencias entre las evaluaciones (antes/después) se deben a la intervención educativa, o simplemente a la evolución de sus competencias a partir de su experiencia formativa como residentes, se consideró necesario contar con un grupo control.

La hipótesis de investigación es que la intervención educativa puede mejorar en ciertos aspectos la competencia de residentes de medicina interna en razonamiento clínico, evaluados mediante el *script concordance test*.

Aspectos éticos

La competencia en razonamiento clínico es central en el proceso diagnóstico. Como los errores diagnósticos pueden causar daño a los pacientes, la Academia Nacional de Ciencias (Balogh et al., 2015) propuso diferentes estrategias para reducir el error diagnóstico. Debido a que la seguridad del paciente es una prioridad en los sistemas de atención médica, el desarrollo de esta habilidad compleja es un tema prioritario en educación médica (Richmond et al., 2020). Por lo tanto, poder demostrar intervenciones educativas que favorezcan el desarrollo del razonamiento clínico en médicos residentes tiene un gran valor social (Emanuel, 1999).

Desde el punto de vista científico, se han publicado a la fecha pocas investigaciones experimentales educativas de intervención controlada para el desarrollo del razonamiento clínico, y ninguna en la Argentina. En este contexto, la presente investigación busca aportar información original, contribuyendo a la comunidad científica y educativa.

El proyecto de investigación fue notificado por anticipado y aceptado para su realización por el Comité de Revisión Institucional.

Los datos a analizar en la presente investigación la generaron los propios participantes con su juicio clínico, a partir de evaluaciones con casos clínicos simulados (ver [diseño del instrumento de evaluación](#)). No participaron pacientes reales ni durante la intervención educativa, como tampoco de las evaluaciones.

Cuando se utilizaron archivos multimedia (fotos de signos clínicos, estudios radiológicos, electrocardiogramas, etc.) para el diseño de los casos clínicos simulados de las evaluaciones, se tuvo en cuenta que no haya información personal que permita identificar a ninguna persona.

Aunque se trata de una población seleccionada, la participación en el estudio de investigación por parte de los residentes y del panel de expertos fue voluntaria y sin ningún tipo de contraprestación económica ni de otro tipo. Todos los participantes expresaron su consentimiento informado escrito digitalizado para participar del mismo.

A los efectos de poder controlar y verificar la correcta participación y obtención de los datos, se les solicitó a todos los participantes la dirección de correo electrónico durante las encuestas y evaluaciones. Solamente el investigador principal tiene acceso a los resultados de las evaluaciones junto con el correo electrónico de los participantes. A los efectos del análisis y presentación de los datos, los resultados fueron anonimizados para preservar la confidencialidad.

Los resultados de las evaluaciones no tuvieron ninguna consecuencia en la aprobación o desaprobación del programa de residencia de los participantes, sino que se trató de un programa netamente formativo. Al mismo tiempo el investigador principal, quien llevó adelante la intervención educativa, no tenía un rol en la evaluación del desempeño de los residentes en el marco de sus respectivos programas de residencia.

Declaración de Interés

El autor de la investigación no tiene conflictos de interés para declarar. Sólo el autor es responsable por el contenido y la escritura del manuscrito.

Materiales y Métodos

Población

Grupo Intervención

Como grupo intervención, se invitó a participar en la investigación a 21 residentes de primer y segundo año de un hospital universitario de la Ciudad Autónoma de Buenos Aires, Argentina, asumiendo una posible pérdida en el seguimiento del 10%.

Se consideró seleccionar en forma no probabilística a los residentes de primer y segundo año de la residencia de Medicina Interna por varios motivos. En primer lugar, porque desde el punto de vista de la competencia en razonamiento clínico los residentes

“inferiores” (R1-R2) no difieren significativamente de los estudiantes de grado (M. R. Nendaz et al., 2006), pero sí se diferencian significativamente de los residentes “superiores” (R3-R4) (Carrière et al., 2009; Nouh et al., 2012).

Por último pero no menos importante, por las dificultades y limitaciones propias de realizar un estudio de intervención controlada en educación médica, considerando la factibilidad del mismo y el poder de la muestra.

Como ya se mencionó antes, la participación de los residentes en el programa de entrenamiento fue voluntaria. Se les solicitó información demográfica (edad, género, estudios universitarios, etc.) que permitan describir, caracterizar y comparar al grupo control con el grupo intervención. Dicha información fue anonimizada para preservar la confidencialidad de los participantes.

Todos los participantes del grupo intervención firmaron consentimiento informado escrito para participar del estudio de investigación.

Grupo Control

Se seleccionaron 21 residentes de primer y segundo año de *otro* hospital universitario de la Ciudad Autónoma de Buenos Aires, cuyas características demográficas fueran comparables. Por los mismos motivos que fueron expuestos anteriormente, se desprende que el grupo control tampoco podrá estar muestreado al azar, sino por conveniencia.

Todos los participantes del grupo control firmaron consentimiento informado escrito para participar del estudio de investigación.

Actividades Docentes en Servicio

En paralelo –y en forma independiente– a la intervención, cada grupo de residentes participó de las habituales actividades docentes del programa de formación de residencia, a cargo de los servicios en los que practicaban.

Ambos grupos realizaron: recorridas (diarias) de sala con médicos de staff, ateneos de servicio de casos clínicos (semanales), ateneos institucionales multidisciplinarios (mensuales), pases de guardia (dos veces al día), clases magistrales (2 o 3 veces por semana).

El tipo y frecuencia de actividades docentes como parte de los programas de residencia en Medicina Interna fueron similares entre los grupos. El contenido, estrategias de enseñanza y evaluación de los mismos fueron gestionados en forma independiente por cada uno de los servicios. El autor no tuvo influencia en la planificación, programación o puesta en marcha de ninguno de los mismos.

Intervención Educativa

Para el diseño de actividades de enseñanza, se tuvieron en cuenta algunas de las ya mencionadas teorías cognitivas aplicadas al razonamiento clínico, como la teoría de los scripts (Keemink et al., 2018; Levin et al., 2016; Lubarsky et al., 2015a), la teoría de la carga cognitiva (Monteiro & Norman, 2013; Van Merriënboer & Sweller, 2005), y la

teoría del procesamiento dual (Graber, 2009; J Rencic & Trowbridge, 2011; Richmond et al., 2020; Trowbridge, 2008; Trowbridge et al., 2013; Trowbridge & Olson, 2018).

Se realizó particular énfasis en la búsqueda intencionada y dirigida de información (Georges Bordage, 1995), el uso de calificadores semánticos (Georges Bordage, 1994) para la representación del problema (Chang, Bordage, & Connel, 1998; M. R. Nendaz & Bordage, 2002) y su comunicación (Georges Bordage, 2007); la identificación de banderas rojas y elementos clave (G. Bordage & Page, 2018; Bowen, 2006), y su contrastación (Mandin, Jones, Woloschuk, & Harasym, 1997) para la generación de hipótesis diagnósticas y el desarrollo de esquemas mentales (Lubarsky, Dory, Audétat, Custers, & Charlin, 2015b), y en los procesos metacognitivos (Croskerry, 2003; Croskerry & Norman, 2008; Graber, 2009; Graber et al., 2012; G. R. Norman et al., 2017).

En la siguiente figura se describe el marco de referencia utilizado para englobar las estrategias de enseñanza utilizadas en el presente estudio, de acuerdo al marco teórico descripto previamente:

Búsqueda de Información

Identificación y jerarquización de las diferentes fuentes de información disponibles.

Investigación basada en hipótesis.

Técnicas de entrevista y fuentes de errores en la adquisición de información clínica.

Representación del Problema

Resumen del problema clínico en una frase breve, completa y contextualizada.

Utilización de calificadores semánticos y otras abstracciones.

Identificación de banderas rojas.

Generación de Hipótesis Diagnósticas

Aproximación al diagnóstico en forma probabilística, pronóstica y pragmática.

Construcción y refinamiento de esquemas mentales (*illness scripts*).

Integración del conocimiento biomédico, la fisiopatología y los datos clínicos.

Verificación Diagnóstica

Contrastación de hipótesis a partir de elementos clave.

Considerar las banderas rojas y “el peor escenario posible”.

Uso de los umbrales diagnóstico y terapéutico.

Metacognición

Reconocimiento de fuentes de errores comunes en el razonamiento clínico.

Estrategias para minimizar los sesgos cognitivos.

Figura 1: Marco de Referencia para la Enseñanza del Razonamiento Clínico.

Las estrategias de enseñanza buscaron favorecer el desarrollo de las diferentes actitudes, habilidades y conocimientos que integran la competencia en razonamiento clínico. Para ello, se diseñaron actividades de colaboración, tanto con tareas en grupos pequeños como

en plenario, la reflexión crítica, la revisión por pares y el feedback multidireccional. No se realizaron clases magistrales, porque las mismas no están alineadas con marco teórico que soporta la intervención educativa.

Se desarrolló un programa del curso para los residentes. En el mismo se especificaban los propósitos, objetivos de aprendizaje, modalidad de trabajo, cronograma y unidades, contenidos, bibliografía recomendada, tipos de evaluaciones, entre otros datos relevantes.

Los residentes participaron en las siguientes actividades de aprendizaje: el uso de casos simulados (escritos con multimedia), la discusión de ateneos de casos clínicos, el aprendizaje basado en problemas (ABP), los juegos de roles, y el análisis y reflexión sobre resúmenes de casos clínicos (propios de los residentes) extraídos de su práctica cotidiana. También la evaluación entre pares y feedback formativo. Se muestran ejemplos de algunas de las actividades realizadas en el [apéndice II](#).

Se planificó además trabajar en forma semi-presencial, con el soporte de un aula virtual (Google Classroom®). El objetivo era utilizar el aula invertida, para maximizar las oportunidades de aprendizaje en los encuentros presenciales. Al mismo tiempo, ofrecía la potencialidad de favorecer la individualización de los procesos de aprendizaje, evaluación y feedback. En el aula virtual se ofreció disponibilidad permanente y acceso a un repositorio digital con la bibliografía recomendada para cada etapa del programa. Los residentes contaron con acceso a internet por wifi institucional.

Se muestra a continuación un ejemplo de una actividad en el aula virtual, en este caso, para favorecer el análisis crítico de la bibliografía.

Razonamiento Clínico 2019

Instrucciones

Trabajo de los alumnos



Debate: Errores del Razonamiento Clínico



Gustavo Jung · 10 sept 2019

Fecha de entrega: 23 sept 2019

Consigna: al menos dos comentarios en el foro aportando análisis y reflexiones sobre la bibliografía obligatoria del módulo. Se valora la integración de los comentarios con los de los compañeros.

Preguntas disparadoras:

- 1) ¿Cuáles son las instancias clave del proceso diagnóstico? ¿Por qué?
- 2) ¿Cuáles son las fuentes más comunes e importantes de errores en el proceso de razonamiento clínico? ¿Por qué?
- 3) A partir de lo anterior, ¿qué estrategias permitirían reducir los errores más comunes e importantes del razonamiento clínico?

Plazo: Dos semanas

 Comentarios de la clase



Añade un comentario de clase...



Figura 2: ejemplo de actividad de aprendizaje en el aula virtual.

Durante la intervención educativa, el rol docente fue llevado adelante principalmente por el autor de la investigación. El mismo contó la colaboración de tres auxiliares docentes, quienes –alternativamente y durante algunos de los encuentros presenciales-, participaron en la discusión de casos clínicos y en la moderación de la reflexión en grupos pequeños.



Los tres auxiliares eran especialistas en Medicina Interna, con experiencia en la discusión de casos clínicos.

Las tareas fundamentales del docente a cargo fueron: presentación de los objetivos y consignas de aprendizaje; organización de los estudiantes para las tareas; presentación de los materiales de estudio y trabajo; moderación de la reflexión y debate, tanto en el aula virtual como en las actividades presenciales; presentación de ateneos de casos; uso de preguntas para favorecer la reflexión crítica; evaluación formativa y feedback sobre el desempeño individual (juegos de roles, aula virtual); evaluación formativa y feedback sobre el desempeño grupal (ABP, tareas en grupos pequeños).

El programa de entrenamiento inició en septiembre de 2019, con encuentros semanales de 60 minutos a lo largo de 30 semanas, hasta abril de 2020. La carga horaria semanal estimada en el aula virtual era de 60 minutos semanales, proyectando un total de 60 horas de intervención educativa.

Luego de 10 semanas (noviembre de 2019) se discontinuó el uso del aula virtual por baja participación de los residentes. A partir de ello, se realizaron adaptaciones en las actividades presenciales para favorecer la participación y el aprendizaje. Una de dichas adaptaciones fue favorecer el uso de resúmenes de casos clínicos reales -aportados por los mismos los residentes-, como material de insumo para el análisis, reflexión y trabajo en los encuentros presenciales. Asimismo, se limitó la carga de materiales bibliográficos complementarios requeridos previo a los encuentros.

En marzo de 2020 se impusieron en la Argentina severas restricciones sociales, profesionales e institucionales relacionadas con la pandemia Covid19, que modificaron en gran medida las condiciones laborales y educativas del ámbito de práctica de los médicos residentes involucrados en el proyecto. Entre otros cambios, todas las reuniones presenciales docentes fueron suspendidas y los residentes fueron divididos en equipos de trabajo alternantes cada 14 días. Debido a estas circunstancias, y en virtud de que no fue posible en ese momento generar la adaptación necesaria para virtualizar las actividades de enseñanza, la duración de la intervención educativa fue menor a lo anticipado.

De este modo, la intervención educativa consistió en reuniones presenciales semanales, con una duración aproximada de una (1) hora, a lo largo de 22 semanas entre septiembre de 2019 y marzo de 2020. El presentismo promedio fue del 69%. La carga horaria estimada en el aula virtual fue de una (1) hora, a lo largo de 10 semanas entre septiembre y noviembre de 2019.

En resumen, la carga horaria total estimada de la intervención educativa fue de 32 horas a lo largo de 6 meses. En la siguiente tabla se detallan las actividades y número de actividades de aprendizaje en las que participaron los residentes del grupo intervención durante dicho período:

Tabla resumen de las actividades de aprendizaje de la intervención

Actividad de Aprendizaje	Estímulos
Aprendizaje Basado en Problemas (ABP)	Dos (2)
Ateneos de Casos Clínicos	Uno (1)
Juegos de Roles	Tres (3)

Uso de Casos Simulados	Dos (2)
Uso de Resúmenes de Casos Reales	Siete (7)
Reflexión Crítica de la Bibliografía	Siete (7)

Evaluaciones

Se administraron dos evaluaciones diferentes a ambos grupos, una antes y otra después de la intervención educativa.

El autor principal de la presente investigación fue el responsable principal de la planificación, diseño y desarrollo de los *blueprints* y de los scripts de ambas evaluaciones, bajo la supervisión del director de la tesis. El Dr. Brailovsky posee una extensa experiencia y trayectoria en la concepción, el desarrollo y análisis de los SCT (Charlin et al., 2000; Hornos et al., 2013; Lineberry et al., 2019; Pleguezuelos et al., 2013).

Para el mapeo de los contenidos a evaluar, su distribución y desarrollo, se diseñaron *blueprints* de ambos exámenes. En ellos el autor clasificó subjetivamente los problemas clínicos de acuerdo a su relevancia clínica, teniendo en cuenta la epidemiología, severidad, tratabilidad y pronóstico, así como sus elementos claves.

Algunos ejemplos de problemas clínicos de alta relevancia evaluados fueron: abdomen agudo, neumonía, infección urinaria, dolor precordial, insuficiencia cardíaca, enfermedad trombo embólica, síncope, hipertensión arterial, insuficiencia respiratoria, asma, EPOC, cefalea, enfermedad vascular cerebral aguda, hemorragia digestiva alta, insuficiencia renal aguda, trastornos del medio interno.

También se consideraron muchos otros problemas clínicos, con una relativamente menor representación en las evaluaciones, por menor prevalencia, adecuación a la instancia formativa, severidad, por ejemplo: síndrome confusional agudo, pérdida de peso involuntaria, poliadenopatías, síndromes edematosos, ictericia, síndrome anémico, traumatismo encéfalo craneano, lumbalgia, síndrome febril agudo, diarrea aguda, HIV, hepatitis aguda, meningitis aguda, arritmias, shock, paro cardíaco, derrame pleural, vértigo, paraparesia, deterioro cognitivo, amaurosis fugax, vómitos, hemorragia digestiva baja, ascitis, insuficiencia renal crónica, enfermedad tiroidea, cetoacidosis diabética, alcoholismo, intoxicación por monóxido de carbono, pancitopenia, entre otros.

Para mayor verosimilitud y autenticidad con la complejidad de la práctica profesional médica, varios problemas clínicos pudieron estar integrados dentro de un mismo *script*, por lo que el número total de problemas considerados es mayor al número de casos diseñados.

El diseño del *contenido* de los casos (viñetas, hipótesis e información adicional) fue realizado teniendo especialmente en cuenta el nivel de formación de la población evaluada y sus conocimientos previos, así como el resto de las consideraciones detalladas previamente.

En el [apéndice III](#) se ofrecen ejemplos de casos/scripts utilizados.

Panel de Expertos

Para conformar el panel de expertos se invitó a 15 especialistas en Medicina Interna, de los cuales doce (12) participaron de la presente investigación. El autor/investigador no fue parte del panel. En la [tabla 1](#) se resumen las características del grupo de referencia.

Tabla 1: Características demográficas de los expertos

	Expertos n = 12
Participantes de la Encuesta	12 (100%)
Edad*	36,3 (3,2)
Género Femenino	10 (83%)
Años de Egresado de la Facultad de Medicina*	10,8 (3,5)
Años de Práctica Profesional en Medicina Interna*	6,5 (3,4)
Ámbito de Práctica:	
● asistencial ambulatoria y/o en emergencias	8 (66,7%)
● asistencial de pacientes internados	6 (50%)
● docencia de pregrado	4 (33,3%)
● docencia de posgrado	5 (41,7%)

* Expresada como: media (desvío estándar)

Las mujeres fueron representadas mayoritariamente (83,3%), con un promedio de edad de 36,3 años. El grupo contaba con una media de 10,8 años de egresados de la Facultad de Medicina, y una media de 6,5 años de práctica profesional en la especialidad, ya finalizada la residencia.

Las principales actividades profesionales que desempeñaban los expertos al momento de la evaluación eran: práctica asistencial ambulatoria y/o en emergencias (66,7%), práctica asistencial de pacientes internados (50%), docencia de pregrado (33,3%) y docencia de posgrado (41,7%), en forma no excluyente.

Optimización pretest

A partir del feedback recibido del panel de expertos, *previo* a la administración de la evaluación a los residentes, se eliminaron 4 ítems del examen pre intervención por consideraciones de validez de contenido. En el examen posterior a la intervención se eliminaron 2 ítems, por el mismo motivo. Por lo tanto, el número total de ítems de las evaluaciones pre y post intervención fueron 71 y 73, respectivamente.

Distribución de respuestas modales de expertos

A continuación, se reportan las distribuciones absolutas de las respuestas modales de los expertos a lo largo de la escala Likert en las evaluaciones pre y post intervención, y se calcula su participación relativa en cada una de ellas:

Tabla 2: Distribución de respuestas modales de expertos. SCT pre intervención

TEST PRE INTERVENCIÓN	-2	-1	0	+1	+2
Totales de respuestas modales de expertos	18	14	20	10	13
Extremas (-2 /+2):	31 (41.3%)				
Centrales (-1 / 0 / +1):	44 (58.7%)				

*NOTA: el número de respuestas modales (75) es mayor al número de ítems (71), porque algunos ítems tuvieron más de una respuesta modal.

Tabla 3: Distribución de respuestas modales de expertos. SCT post intervención

TEST POST INTERVENCIÓN	-2	-1	0	+1	+2
Totales de respuestas modales de expertos	7	19	18	12	21
Extremas (-2 /+2):	28 (36.4%)				
Centrales (-1 / 0 / +1):	49 (63.6%)				

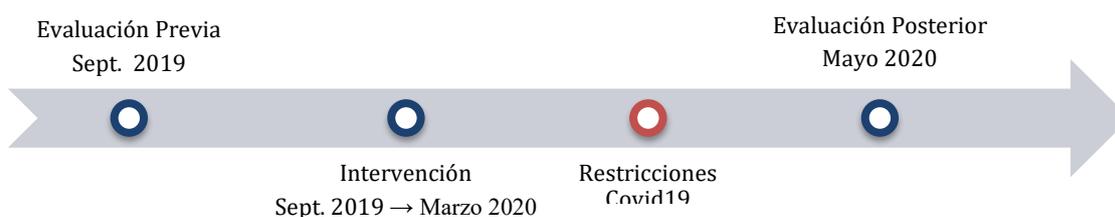
*NOTA: el número de respuestas modales (77) es mayor al número de ítems (73), porque algunos ítems tuvieron más de una respuesta modal.

Administración de los Exámenes

Las evaluaciones a residentes antes de la intervención se llevaron adelante durante un plazo de 2 semanas, entre fines de septiembre y comienzos de octubre de 2019. La intervención educativa se realizó entre septiembre y marzo de 2020.

Aunque se estimaba completar la segunda evaluación hasta un mes posterior a finalizada la intervención, debido a las restricciones antes mencionadas, las evaluaciones finales a los residentes se completaron a fines de mayo 2020.

Figura 3: Esquema del desarrollo de la investigación



La evaluación previa a la intervención se realizó de manera presencial con los residentes, en un aula compartida. La evaluación posterior a la intervención tuvo que realizarse en un aula virtual por las restricciones impuestas por la pandemia de Covid19. Independientemente del lugar donde se llevaron adelante, tanto los expertos como los residentes respondieron individualmente el examen bajo condiciones similares. Todos realizaron las evaluaciones en un formulario digital y dispusieron de un máximo de 90 minutos. Ambas evaluaciones se realizaron en idioma castellano.

Todos los participantes recibieron instrucciones sobre las características de la evaluación, el mecanismo propuesto de respuesta y los errores más frecuentes en la interpretación general de la consigna. Asimismo, tanto los expertos como los residentes tuvieron la oportunidad de practicar con casos clínicos de prueba el formato del SCT, antes de completar las evaluaciones formales.

Como en la mayoría de los casos el SCT resultaba novedoso para los participantes, y con el objeto de identificar dificultades y prevenir errores por inadecuada comprensión de las consignas, el investigador principal estuvo presente y disponible durante ambas instancias de evaluación, para resolver potenciales diferencias en la interpretación del mecanismo de respuesta del instrumento.

Formato y Soporte Digital

La evaluación se administró en formato digital utilizando la aplicación Formularios de Google®, que es de acceso gratuito irrestricto (<https://docs.google.com/forms>). Se eligió de esta manera porque facilita el control del flujo de datos, minimiza las omisiones de datos con respecto al papel, evita errores de transcripción del papel a digital y facilita el procesamiento de los datos (Fournier et al., 2008). Asimismo, las evaluaciones digitales permiten la incorporación de fotos, imágenes, videos y otros estudios complementarios que enriquecen y dan mayor verosimilitud a los scripts (Boulouffe et al., 2010; Lubarsky et al., 2009).

Punto de corte

En el caso de la presente investigación, como la evaluación no tenía carácter sumativo, ni tampoco se utilizó para la selección de candidatos, por lo que no fue necesario establecer un punto de corte o nota de aprobación.

Encuestas de valoración del instrumento de evaluación

Al finalizar las evaluaciones se les pedirá a todos los participantes (expertos y residentes) que respondan una encuesta. El objetivo de la encuesta es obtener información sobre la factibilidad, aceptabilidad y experiencia, a partir de las siguientes dimensiones: tiempo disponible, grado de dificultad y verosimilitud.

Se les pedirá a los participantes que respondan tres preguntas a lo largo de una escala de Likert, para identificar posibles fortalezas y/o debilidades del instrumento de evaluación:

1. *¿Considera que el tiempo disponible para el examen fue suficiente?*

En una escala de 1 a 10, donde 1 = muy insuficiente y 10 = muy suficiente

2. *¿Cuál considera que fue el grado de dificultad de los casos clínicos presentados?*

En una escala de 1 a 10, donde 1 = muy fácil y 10 = muy difícil

3. *¿Considera que los casos clínicos fueron representativos de situaciones clínicas reales?*

En una escala de 1 a 10, donde 1 = poco realista y 10 = muy realista

Análisis de los datos

Los datos de las evaluaciones se exportaron desde los Formularios® originales, y se resumieron y analizaron en hojas de cálculo Excel®. El análisis estadístico se realizó sobre dichas bases de datos con los programas estadísticos IBM SPSS Statistics® Versión 20 y EduG® Versión 6.1 – f.

De acuerdo al proceso de calificación agregada, se analizaron los resultados de los residentes antes y después de la intervención educativa. Para ello se calcularon las medias de cada uno de los grupos y sus respectivos desvíos estándar, para determinar si existen diferencias estadísticamente significativas entre el grupo intervención y el control, como consecuencia de la intervención educativa.

Para obtener lo mejor de cada uno de los análisis posibles, se realizaron análisis de confiabilidad con alfa de Cronbach basados en ítems y también basados en casos completos. Se calcularon los coeficientes individuales de cada ítem/caso y su contribución global a la confiabilidad de las dos evaluaciones, tanto para el panel de expertos, como para los médicos residentes. Asimismo, se realizaron los cálculos de la *G* (por generalizabilidad) y las diferentes fuentes de error.

Resultados

Datos Poblacionales

Participaron 21 residentes en cada grupo. En la [tabla 4](#) se resumen las características demográficas de los dos grupos de residentes. La edad media fue de 27,1 y 26,5 en los grupos intervención y control, respectivamente. En ambos estaban mayoritariamente representadas las mujeres (74% y 81%). Del grupo intervención participaron 15 residentes de primer año y 6 residentes de segundo año. En el grupo control participaron 13 residentes de primer año y 8 residentes de segundo año. La Universidad de Buenos Aires fue el ámbito de formación médica mayoritaria en ambos grupos. Hubo sólo dos residentes del grupo intervención que no participaron de la encuesta demográfica.

Tabla 4: Características demográficas de los residentes

	Grupo Intervención n = 21	Grupo Control n = 21	Valor de p
Participantes de la encuesta	19 (90%)	21 (100%)	-
Edad*	27,1 (1,9)	26,5 (1,9)	
Género femenino	14 (74%)	17 (81%)	0.29
Año de Residencia	R1 = 15 (71%) R2 = 6 (29%)	R1 = 13 (62%) R2 = 8 (38%)	0.51
Universidad de Origen	U.B.A. = 8 (42%) U.SAL. = 1 (5%) F.B. = 1 (5%) Otras = 9 (47%)	U.B.A. = 12 (57%) U.SAL. = 2 (10%) F.B. = 2 (10%) Otras = 5 (23%)	

U.B.A. = Universidad de Buenos Aires;

U.SAL. = Universidad del Salvador;

F.B. = Fundación Barceló.

* Edad = media (desvío estándar), expresada en años.

El valor de p fue calculado por chi2.

De ambos grupos, participaron la totalidad de residentes correspondientes a los primeros dos años de cada residencia. El número de residentes de segundo año es menor al número de residentes de primer año, porque luego de completar el primer año de Medicina Interna, algunos de ellos continúan su formación en Cardiología o Terapia Intensiva.

Se calculó el valor de *p* por *chi2* para el género y año de residencia de los participantes. Hubo significativamente más mujeres en el grupo control. Las diferencias en cuanto al año de residencia de los participantes entre ambos grupos no fueron estadísticamente significativas.

Encuestas sobre los Instrumentos de Evaluación

Como se explicó [previamente](#), al finalizar ambas evaluaciones se les pidió a todos los participantes que respondieran una encuesta sobre las mismas. En la [tabla 5](#) se resumen los resultados:

Tabla 5: Resultados de Encuestas sobre las Evaluaciones

Instancia	Expertos		Grupo Intervención		Grupo Control		Global
	Pre	Post	Pre	Post	Pre	Post	
Medida	Media	Media	Media	Media	Media	Media	Media (SD)
¿Considera que el tiempo disponible para el examen fue suficiente?	8,6	8,2	8,3	9,8	9,8	10	9,1 (0,8)
¿Cuál considera que fue el grado de dificultad de los casos clínicos presentados?	6,7	6,9	6,6	6,3	6,3	6,5	6,6 (0,2)
¿Considera que los casos clínicos fueron representativos de situaciones clínicas reales?	9,7	9,5	8,4	9,0	9,1	8,6	9,1 (0,5)

Globalmente, tanto los residentes como los expertos consideraron el tiempo de examen como muy suficiente, con una media de 9.1 (rango 8.2 - 10). Subjetivamente evaluaron la dificultad de los casos clínicos presentados en los scripts con una media de 6.6 (rango 6.3 - 6.9). Y respecto a la verosimilitud, se les preguntó si los casos eran representativos de situaciones clínicas reales, y el promedio de las medias de cada grupo fue 9.1 (rango 8.4 - 9.7).

Fuentes de datos

Por un fallo electrónico, se perdió un único registro de una evaluación del grupo control pre intervención. Por lo tanto, el número registrado de evaluaciones pre intervención fueron 21 (100%) y 20 (95%) en los grupos de estudio y control, respectivamente. El número de evaluaciones registradas posterior a la intervención fue del 100% (21), para ambos grupos.

En lo que respecta a cada uno de los exámenes, no hubo datos faltantes en ningún ítem, tanto de los residentes como los expertos. Por lo tanto, no fue necesario excluir ningún participante del análisis.

Análisis de la Confiabilidad, Varianza y Generalizabilidad

Como se mencionó en los [métodos de análisis de los datos](#), se llevó a cabo un análisis de Generalizabilidad para determinar las diferentes fuentes de varianza y error. Se realizaron análisis de varianza y generalizabilidad discriminados por ítems y por casos.

El n de residentes analizado fue 21 para todas las instancias, excepto el grupo control pre intervención, con un $n = 20$, por los motivos ya mencionados.

Análisis de varianza por ítems

El análisis de varianza por *ítems* permite discriminar las fuentes de error correspondientes a cada una de las evaluaciones, tomando como unidad de análisis cada uno de los ítems del SCT. El número de ítems analizado fue de 71 para el SCT pre intervención, y de 73 para el SCT post intervención. Se analizan las fuentes de error correspondientes a los Residentes (R), los Ítems (I), y la interacción de los residentes con los ítems (R/I).

Es posible tener una aproximación a la confiabilidad global del test, a partir de la teoría de la generalizabilidad. Para ello se calcula el coeficiente G relativo. Se presentan también en la [tabla 6](#) los resultados del coeficiente G relativo para cada uno de los grupos, en ambas evaluaciones.

Tabla 6: Análisis de varianza y generalizabilidad por ítems

ANÁLISIS DE VARIANZA POR ÍTEMS				
	Grupo Intervención		Grupo Control	
	PRE	POST	PRE	POST
Residentes (R)	4.5%	3%	0.9%	1.4%
Ítems (I)	13.2%	17.8%	14.8%	17%
Residentes/Ítems (R/I)	82.2%	79.2%	84.3%	81.6%
Coeficiente G relativo	0.80	0.74	0.42	0.55

La principal fuente de varianza resulta de la R/I, para ambas evaluaciones, en ambos grupos. La varianza correspondiente a los ítems (I) es similar entre los grupos intervención y control (PRE: 13.2% vs 14.8% y POST: 17.8% vs 17%, respectivamente). La varianza correspondiente a los residentes (R) es diferentes entre los dos grupos, con una menor varianza en el grupo control (PRE: 4.5% vs 0.9% y POST: 3% vs 1.4%, respectivamente).

Se destaca la diferencia de confiabilidad del instrumento medida por generalizabilidad entre los dos grupos. Mientras que en el grupo intervención, el G relativo fue de 0.80 y 0.74 (pre y post intervención, respectivamente), para el grupo control fue de sólo 0.42 y 0.55, respectivamente.

Análisis de varianza por scripts

El análisis de varianza por *scripts* ([Tabla 7](#)) permite también discriminar las fuentes de error correspondientes a cada una de las evaluaciones, pero en este caso de acuerdo a cada uno de los casos (o *scripts*) evaluados.

Los casos fueron diseñados para representar un determinado “guión de enfermedad” (*illness script*). A pesar de que las respuestas de cada ítem dentro de un caso no están encadenadas entre sí, y aunque cada ítem debería evaluar diferentes elementos clave de dicho *script*, se asume cierta unidad lógica, que permite su análisis conjunto.

Se analizan la fuente de error correspondiente a los Residentes (R), los casos = *Scripts* (S), y la interacción de los residentes con los *scripts* (R/S). Los S analizados fueron 25 para ambos SCT.

Se presentan también en la [tabla 7](#) los resultados del coeficiente G relativo para cada uno de los grupos, en ambas evaluaciones, sobre el análisis basado en *scripts*.

Tabla 7: Análisis de Generalizabilidad por scripts

ANÁLISIS DE VARIANZA POR SCRIPTS				
	Grupo Intervención		Grupo Control	
	PRE	POST	PRE	POST
Residentes (R)	11.5%	6.6%	2.4%	3.4%
Scripts (S)	10.8%	19.7%	17.1%	12.9%
Residentes/Scripts (R/S)	77.6%	73.7%	80.5%	83.7%
Coefficiente G relativo	0.79	0.69	0.43	0.50

En el análisis de varianza por scripts, la principal fuente de varianza resulta también de R/I, para ambas evaluaciones, en ambos grupos. La varianza correspondiente a S no es homogénea entre los grupos intervención y control (PRE: 10.8% vs 17.1% y POST: 19.7% vs 12.9%, respectivamente). Una vez más, el componente R es diferente entre los dos grupos, con una menor varianza en el grupo control (PRE: 11.5% vs 2.4% y POST: 6.6% vs 3.4%, respectivamente).

Al igual que en el análisis por ítems, en el análisis de la confiabilidad medida por generalizabilidad basado en scripts, se identifican también diferencias significativas entre los grupos. En el grupo intervención el G relativo fue de 0.79 y 0.69 para los SCT pre y post intervención, respectivamente. Mientras que para el grupo control fue de sólo 0.43 y 0.50. La diferencia sugiere diferencias poblaciones que serán abordadas en la [discusión](#).

Si se compara la varianza aportada por los residentes, cuando se analiza por ítems versus por *scripts*, queda claro que **el análisis de varianza por *scripts* discrimina mejor entre los residentes de ambos grupos**. Por lo tanto, se utilizará el análisis basado en *scripts* para presentar los [resultados de las evaluaciones](#).

Análisis de la confiabilidad

Es posible tener otra aproximación a la confiabilidad de los test a partir de la Teoría Clásica de los Test, mediante el Alfa de Cronbach. Como se mencionó, este coeficiente es una medida de confiabilidad, por lo que se presenta en la [Tabla 8](#) junto con el coeficiente G de generalizabilidad:

Tabla 8: Análisis de confiabilidad por scripts

ANÁLISIS DE CONFIABILIDAD POR SCRIPTS				
	Grupo Intervención		Grupo Control	
	PRE	POST	PRE	POST
Alfa de Cronbach normalizado	0.79	0.72	0.40	0.51
Coefficiente G relativo	0.79	0.69	0.43	0.50

Se observa que el alfa de Cronbach y el coeficiente G relativo son muy similares, lo cual es coherente con el constructo analizado (confiabilidad).

En conclusión, el análisis de confiabilidad demostró mayor discriminación cuando se realiza basado en *scripts* que por ítems, alcanzando niveles muy buenos de confiabilidad en el grupo intervención, y significativamente menores en el grupo control.

Resultados de las evaluaciones

Para el panel de expertos como para ambos grupos de residentes, se calcularon tanto las calificaciones individuales, como la media global y su correspondiente desvío estándar, de acuerdo al procedimiento estadístico descrito anteriormente en [Proceso de Calificación](#).

A los efectos de una mayor claridad e interpretación de los resultados, teniendo en cuenta que los puntajes máximos son variables en cada evaluación, se prefirió expresar en términos porcentuales. Entonces, el resultado de cada participante expresa el porcentaje obtenido sobre un máximo de 100%. Por ejemplo, un resultado de 70 expresa que el residente obtuvo un 70% del puntaje ideal, determinado por el panel de expertos como conjunto.

Por lo anterior, no se consideró necesario realizar la estandarización de los resultados.

En el [apéndice IV](#) se detallan todos los resultados individuales de los participantes.

A partir de la tabla y la figura que siguen, se desprenden los principales resultados de la investigación. En la [tabla 9](#) se resumen los resultados basados en el análisis de scripts de las evaluaciones, antes y después de la intervención educativa:

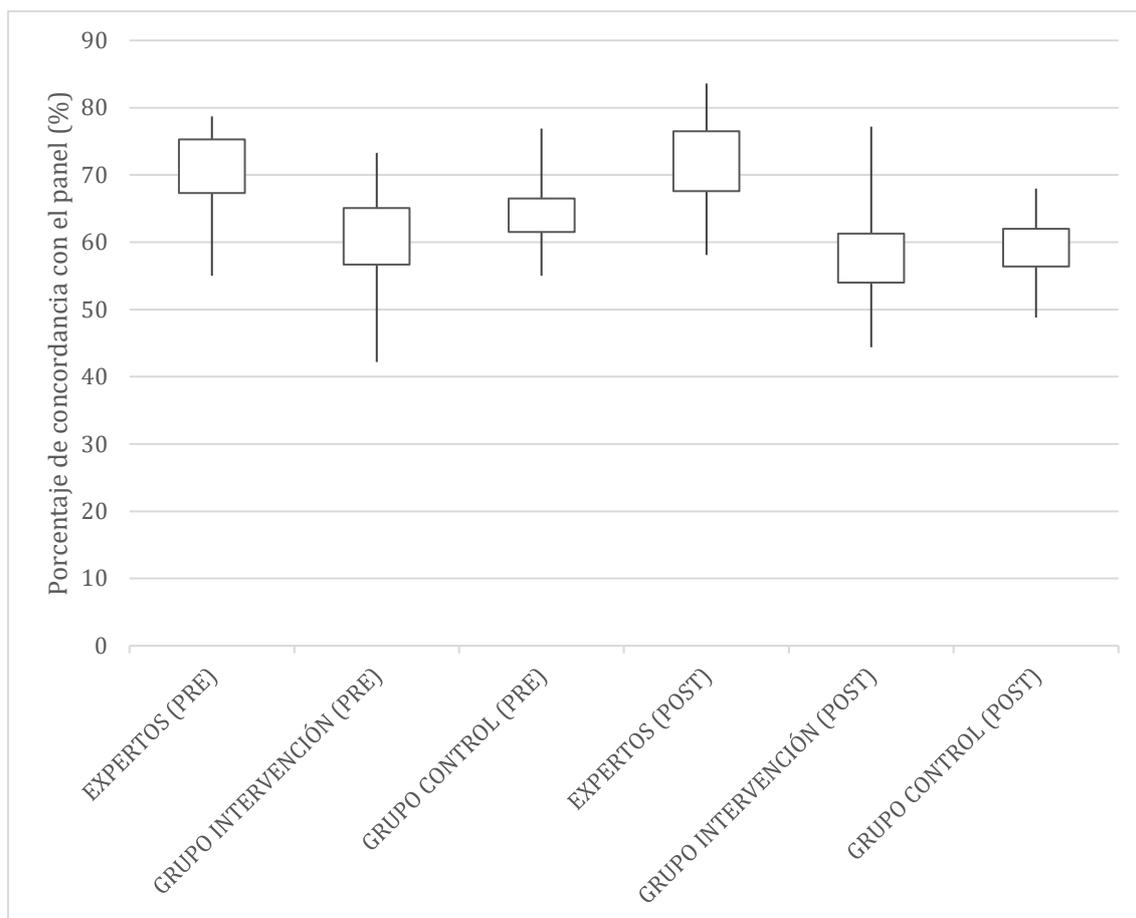
Tabla 9: Calificaciones de las evaluaciones

	PRE INTERVENCIÓN				POST INTERVENCIÓN			
	<i>n</i>	media (%)	DE	IC (95%)	<i>n</i>	media (%)	DE	IC (95%)
Panel de Expertos	12	71,3	6,3	67,3 – 75,3	12	72,0	7,0	67,6 – 76,5
Grupo Intervención	21	60,9	9,3	56,7 – 65,1	21	57,7	8,0	54,0 – 61,3
Grupo Control	20	64,0	5,4	61,5 – 66,5	21	59,2	6,1	56,4 – 62,0

n: Tamaño de la muestra; DE: Desvío estándar; IC: Intervalo de confianza

En la [figura 4](#) se muestra una comparación de los resultados entre los grupos, antes (primeras 3 columnas) y después de la intervención educativa (segundas 3 columnas). Cada caja representa el intervalo de confianza del 95% de la media, y las líneas los rangos mínimos y máximos de distribución de las calificaciones.

Figura 4: Comparación de resultados entre los grupos



Se observa que el desempeño del panel de expertos fue consistente a lo largo de ambas evaluaciones, con una media de 71,3% (IC 95%: 67,3% – 75,3%) para el primer SCT y 72% (IC 95%: 67,6% – 76,5%) para el segundo SCT.

Previo a la intervención educativa, la diferencia del IC 95% de la media entre los expertos y la del grupo intervención fue de 10,4% (10,6% a 10,2%), con un rango de diferencias entre mínimos de 12,8% y máximos de 5,4%.

Previo a la intervención educativa, la diferencia del IC 95% de la media entre los expertos y la del grupo *control* fue de 7,3% (5,8% a 8,8%), sin diferencia entre los mínimos (0%) y con una diferencia entre máximos de 1,8%.

El grupo intervención logró una media de 60,9% (IC 95%: 56,7% – 65,1%) *previo* a la intervención educativa. El grupo control, a su vez, alcanzó una media de 64% (IC 95%: 61,5% – 66,5%) en el primer SCT. Como los intervalos de confianza se superponen, no hay diferencias basales entre el grupo intervención y el grupo control.

Posterior a la intervención educativa, la diferencia del IC 95% de la media entre los expertos y la del grupo intervención fue de 14,3% (13,6% a 15,2%), con un rango de diferencias entre mínimos de 13,7% y máximos de 6,4%.

Posterior a la intervención educativa, la diferencia del IC 95% de la media entre los expertos y la del grupo *control* fue de 12,8% (11,2% a 14,5%), con un rango de diferencias entre mínimos de 9,3% y máximos de 15,6%.

Al analizar los grupos intervención y control posterior a la intervención educativa, se observa que ambos grupos *disminuyeron* su rendimiento. Mientras que los expertos mantuvieron el desempeño, la media (IC95%) de los grupos intervención y control disminuyeron a 57,7% (54,0% – 61,3%) y 59,2% (56,4% – 62,0%), respectivamente. Las diferencias de la media antes y después de la intervención educativa fueron de –3,2% para el grupo intervención, y –4,8% para el grupo control. La caída de la media fue 1,6% menor en el grupo intervención que en el grupo control.

Discusión

La presente investigación se propuso evaluar los resultados de una intervención educativa en razonamiento clínico, evaluados mediante script concordance test en forma prospectiva y controlada. No fue posible demostrar diferencias entre ambos grupos, de acuerdo al poder de la muestra y la magnitud del efecto estimado.

Este es –hasta donde se sabe- el primer estudio experimental cuantitativo en la Argentina de intervención educativa prospectiva y *controlada* que utiliza el script concordance test para la evaluación de residentes de Clínica Médica. Se trata, por lo tanto, de una experiencia innovadora y enriquecedora para la comunidad de educación médica.

La intervención educativa fue realizada de acuerdo a un marco teórico relevante, moderno y basado en la mejor evidencia disponible. El marco de referencia para la enseñanza del razonamiento clínico incluye, en la visión del autor, los conocimientos, habilidades y actitudes clave para el desarrollo de la competencia en razonamiento clínico.

Se tuvo en cuenta la importancia crítica de la especificidad de contenido y contexto para el diseño de las actividades de aprendizaje. Hubo una alineación constructiva entre los objetivos de aprendizaje, las estrategias de enseñanza y la evaluación para promover el aprendizaje.

La intervención educativa se sustentó en un programa particularmente diseñado y específico a la población objetivo, con objetivos explícitos de aprendizajes y con resultados esperados observables. Al mismo tiempo, se tuvieron en cuenta factores contextuales para la adaptación de las estrategias a la población objetivo, y sus características particulares al momento de la implementación.

Respecto al instrumento de evaluación, el SCT cuenta con adecuada validez de constructo, confiabilidad, factibilidad, aceptabilidad y costo relativamente bajo. Se puede administrar en forma digital, lo cual tiene ventajas operativas. Y el mecanismo de respuesta es relativamente intuitivo, porque se asemeja a la toma de decisiones de la práctica profesional, por lo que no requiere gran entrenamiento de los respondedores.

Por otra parte, se reconocen los grandes desafíos que implica su concepción y desarrollo, a mencionar: la tabla de especificaciones; el diseño de un número suficiente de casos auténticos pero poco definidos, basados en elementos clave, que incluyan incertidumbre, con información adicional relevante e hipótesis válidas; el requerimiento de anticipar y “distribuir” las respuestas esperadas a lo largo de toda la escala de Likert; la necesidad de contar un número relativamente grande de expertos; la optimización de casos previo a la administración a los evaluados, contemplando no sólo la confiabilidad sino también la

validez de contenido; el análisis de los resultados requiere conocimientos avanzados en estadística y psicometría, por lo que suele ser necesario contar con un asesor estadístico.

En síntesis, se trata de una herramienta sumamente interesante y valiosa, pero compleja y laboriosa. A su vez, si se desea capturar la competencia completa en razonamiento clínico, debe ser parte de un repertorio o programa de evaluaciones más amplias y sofisticadas.

Se desarrollará a continuación la discusión de los resultados expuestos.

Los resultados de los grupos de residentes fueron más bajos que los del panel de expertos, aportando a la validez de constructo del instrumento, porque discrimina entre niveles de experticia.

El tiempo disponible para completar las evaluaciones fue muy satisfactorio tanto para el panel de expertos como para ambos grupos, apoyando las recomendaciones vigentes al respecto.

Asimismo, otra fortaleza del diseño de las evaluaciones de la presente investigación parece haber sido su verosimilitud con casos reales, evaluados mediante la encuesta de calidad ([tabla 5](#)). Vale recordar que la autenticidad, así como el ajuste de la dificultad a la población objetivo son elementos críticos del diseño, que afectan fuertemente la validez y confiabilidad del instrumento.

El análisis discriminado de la confiabilidad demostró una mayor discriminación cuando se realiza basado en *scripts* que por ítems, alcanzando niveles muy buenos de confiabilidad en el grupo intervención, y significativamente menores en el grupo control. Por lo tanto, se discuten los resultados de las evaluaciones basados en el análisis de *scripts*.

Respecto a la optimización de los SCT, parecen ser claros los beneficios estadísticos de eliminar *scripts durante el desarrollo de la clave de corrección* para aumentar la confiabilidad y reducir el error de medida aportado por el panel de expertos y la variabilidad de sus respuestas. Asimismo, se podría favorecer la validez de contenido del instrumento, en la medida que –si y sólo si– el número final de *scripts* sea suficiente para la instancia de evaluación e implicancias educativas correspondientes.

Es importante discriminar este proceso de optimización *a priori* de la administración del test a los evaluados, de aquel realizado “canónicamente” (Wan et al., 2018) sobre los resultados de los evaluados (post hoc).

Vale recordar en este punto que hubiese sido posible aumentar la confiabilidad *a posteriori* descartando los *scripts* con baja correlación *script/total*. En efecto, la remoción selectiva de ítems o *scripts* con correlación negativa o pobre positiva, aumenta el alfa de Cronbach del test, como medida de confiabilidad de la misma. Así, se podrían identificar y eliminar ítems o *scripts* para aumentar una medida estadística (importante, pero no la única) del valor de la prueba. Dicho proceso hubiese representado, en la presente investigación, una cierta amenaza a la validez (de contenido) del instrumento.

Cabe preguntarse si, una vez completados los procesos de aseguramiento de la calidad del diseño de los *scripts* a evaluar, en cuanto a su validez de constructo y contenido, y

conociendo los efectos que genera la selección negativa de *scripts* con baja correlación, dicho proceso de “optimización” no podría incurrir en una simplificación estadística, que desconoce la complejidad e incertidumbre subyacente de los procesos de respuesta y amenaza la validez del instrumento.

Por todo lo analizado anteriormente, se decidió no eliminar *scripts a posteriori*, a partir del análisis individualizado del alfa de Cronbach de cada uno los *scripts*.

Se interpreta que hubo una adecuada distribución de las respuestas a lo largo de la escala de Likert ([tablas 2 y 3](#)), limitando el riesgo de ventaja de “responder al centro”.

La mayoría de las investigaciones que utilizaron SCT analizan la confiabilidad utilizando solamente el coeficiente alfa de Cronbach. La presente investigación incorpora además el análisis de la varianza y la confiabilidad de acuerdo a la teoría de la generalizabilidad. Una ventaja de esta aproximación estadística es el reconocimiento y discriminación de las diferentes fuentes de varianza: la varianza de las puntuaciones verdaderas, y la varianza por error aleatorio de medida.

Al respecto, las diferencias de confiabilidad entre los grupos medidas tanto de acuerdo a la teoría clásica de los test como por generalizabilidad, sugieren más diferencias poblacionales que del diseño del instrumento, porque cuanto más variable sea un grupo, mayor será la confiabilidad. Se discuten esas diferencias a continuación.

Primero, es interesante analizar las diferencias de los resultados entre los grupos, en relación con sus características contextuales ([tabla 4](#)). Para género y año de residencia, se calculó el valor de p por χ^2 , sin demostrar diferencias estadísticamente significativas. Eso significa que estas dos características poblacionales no permiten diferenciarlos. Sin embargo, pese a que desde el punto de sus características demográficas ambos grupos son similares, su comportamiento en las evaluaciones fue diferente.

En el análisis de la varianza por generalizabilidad basado en *scripts* ([tabla 7](#)), se observa que la varianza real aportada por los residentes del grupo control es muy baja, y mucho menor al grupo intervención. A su vez, la interacción de los residentes con los casos muestra que la especificidad de caso tuvo mayor peso en la varianza del grupo control que en el grupo intervención.

Podría interpretarse de lo anterior, que ambos grupos se comportan psicométricamente diferente, probablemente por una mayor homogeneidad de los integrantes del grupo control, en quienes tuvo más peso en la varianza la especificidad de casos. En consecuencia, la generalizabilidad de la prueba para ese grupo es significativamente menor que en el grupo intervención ([tabla 8](#)).

Los IC (95%) de la media son notablemente menores en el grupo control comparado con el grupo intervención, lo cual expresaría, del mismo modo que el análisis de la varianza, una mayor homogeneidad poblacional.

Otro factor que modifica la confiabilidad del instrumento es la dificultad de los casos. Subjetivamente, ambos grupos evaluaron la dificultad de las evaluaciones en forma casi idéntica ([tabla 5](#)). Por lo tanto, es poco probable que la dificultad de los casos haya influido significativamente en las diferencias observadas entre los grupos. Sin embargo,

se reconocen las limitaciones de una encuesta para la evaluación del grado de dificultad de una prueba.

Por otra parte, teniendo en cuenta la especificidad de contenido de la experticia en razonamiento clínico, podría considerarse si, más allá del nivel de dificultad, los contenidos específicos evaluados afectaron el desempeño de los grupos. Porque “lo que se está midiendo depende más del contenido del método o de la tarea propuesta al examinado, que de cualquier característica del método en sí mismo” (C. P. Van Der Vleuten, 1996, p. 51).

Como se describió en el diseño del instrumento, se consideraron síndromes y problemas clínicos de alta relevancia, y otros de menor. Los primeros, tuvieron mayor representación en el diseño de casos, con una presencia mayor al 90% en cada una de las evaluaciones. Los de menor relevancia fueron incluidos en una proporción menor, entre el 60 y 70%. Por lo tanto, los contenidos evaluados fueron similares en ambas evaluaciones pero no idénticos, por lo que es posible que cierta diferencia en los conocimientos sobre alguno de ellos haya generado diferencias entre los examinados, y los grupos. Sin embargo, como ya se expuso previamente, el instrumento fue diseñado de acuerdo a las guías vigentes, que demostraron que 25 casos con tres ítems cada uno tienen alta confiabilidad (Dory et al., 2012; Fournier et al., 2008; Lubarsky, Dory, et al., 2013).

A pesar de que el desarrollo de la experticia en razonamiento clínico no es lineal, llama la atención que los resultados del examen post intervención –para *ambos* grupos- fueron menores que en el examen pre intervención. Uno podría intuir que –más allá de la intervención educativa de la investigación-, los residentes desarrollarían su capacidad de interpretación de la información y la toma de decisiones, entre otras habilidades claves. Eso resultaría necesariamente en puntuaciones más altas, a medida que los residentes avanzan en su formación.

Asimismo, si la evaluación previa a la intervención pudiese haber tenido algún efecto en el desarrollo de la competencia -como lo tienen las evaluaciones formativas-, el resultado hubiese sido en la misma dirección: puntuaciones más altas en la evaluación posterior.

A lo largo de la investigación en razonamiento clínico y el uso SCT, se ha demostrado una y otra vez que este instrumento de evaluación es válido, porque discrimina entre niveles de experticia. Por tanto, no podría atribuirse al instrumento de medida la baja en el desempeño. Tampoco es probable, aunque es posible, que la especificidad de contenidos de la segunda evaluación sea responsable de la caída en las notas, por los mismos motivos desarrollados en el párrafo anterior. En tercer lugar, el desempeño del panel de expertos fue similar en ambas evaluaciones, lo cual también argumenta en contra de una dificultad mayor de la segunda.

A partir de lo expuesto, se podría conjeturar que la baja en el desempeño de los aspectos de la competencia en razonamiento clínico evaluados por script concordance test en ambos grupos podría ser *real*. Si esto fuese así, se abriría la necesidad de una mayor reflexión acerca del desarrollo de las competencias médicas en general, y del razonamiento clínico en particular, en los ámbitos de residencia. Sin embargo, serían necesarios más estudios para determinar con certeza dicha conjetura.

La educación tradicional organiza el currículo basado en el tiempo y en los contenidos/disciplinas a aprender. Por ejemplo, en la Facultad de Medicina más grande de la Argentina, se debe completar un curso de 4 meses por Medicina Interna para poder avanzar en la carrera. Pasada esa instancia, y durante los dos años siguientes de formación médica, el estudiante no vuelve a enfrentarse a esos contenidos. El foco está puesto en el tiempo y en los contenidos disciplinares, en vez de los *resultados* esperados de la formación médica.

A lo largo de los años, se ha recolectado evidencia de que algunos graduados no están preparados para la práctica profesional en forma efectiva y segura, y que existe una inaceptable variabilidad en las habilidades de los mismos (J. R. Frank et al., 2017). Los resultados de la auto evaluación de competencias desarrollada al inicio de este documento (ver [Introducción](#) y [Apéndice I](#)) presentan un panorama similar.

En las últimas décadas, con el advenimiento de la educación basada en resultados en Medicina, se ha puesto mayor atención en que los programas de formación de los profesionales de la salud respondan con mayor responsabilidad frente a las demandas y necesidades de los pacientes, y la población general. Por lo tanto, los currículos se centran en el estudiante y sus aprendizajes individuales, con una consecuente mayor flexibilidad en sus diseños (J. R. Frank et al., 2010).

El modelo educacional según el cual el tiempo dedicado al entrenamiento de una tarea es un marcador subrogado de competencia es hoy insostenible (J. R. Frank et al., 2017).

Volviendo a los resultados de la presente investigación, uno podría conjeturar que los programas de educación médica de los residentes evaluados no han logrado alcanzar los resultados observables esperables: el desarrollo de ciertos aspectos de la competencia en razonamiento clínico, de acuerdo al instrumento utilizado. Porque -más allá de la intervención educativa llevada adelante en el grupo de estudio-, los residentes deberían haber *mejorado* su desempeño en las habilidades evaluadas a lo largo del tiempo.

La baja del rendimiento de ambos grupos en la evaluación post intervención no es fácilmente justificable, porque el diseño de la investigación tampoco se planteó como objetivo la investigación de las *causas* del desempeño en las evaluaciones.

Es posible que hayan participado factores contextuales, como el agotamiento físico o mental, en una población particularmente susceptible al *burnout*, por encontrarse durante los primeros dos años de residencias en nuestro país, durante los cuales la carga de trabajo es mucho mayor. Como se mencionó antes, dichos factores no fueron medidos ni evaluados en la presente investigación.

Finalmente, no es descabellado considerar que la baja en el rendimiento sea real, abriendo un gran interrogante y, al mismo tiempo, generando oportunidades para futuras investigaciones acerca de la evaluación del desarrollo de la competencia en razonamiento clínico en el ámbito de las residencias.

Podría ser interesante complementar el abordaje experimental cuantitativo de la presente investigación con otras perspectivas y metodologías (por ejemplo, cualitativas o mixtas), para la construcción de datos sobre el desarrollo y evaluación de la competencia y experticia en razonamiento clínico.

Limitaciones

La capacidad de un estudio de intervención educativa de generar diferencias estadísticamente significativas depende de múltiples factores contextuales. Existen factores dependientes del docente, de los estudiantes, del instrumento de evaluación y de la interacción entre dichos sistemas (Bloch & Norman, 2012). Además, el aprendizaje no es un cambio lineal, sino complejo e inestable (Bunniss & Kelly, 2010).

El presente estudio presenta varias limitaciones. Se analizan a continuación algunos de los factores contextuales que limitan al presente estudio.

El presentismo del grupo intervención en las actividades de aprendizaje fue sub óptima (69%). Un mayor presentismo hubiese maximizado las oportunidades de aprendizaje y, por lo tanto, de determinar diferencias entre los grupos.

Asimismo, la participación activa de los residentes en las actividades de aprendizaje es clave en el diseño constructivista de la intervención. La posibilidad de los residentes de primer y segundo año de participar, su motivación y la predisposición para el aprendizaje activo son variables complejas, difíciles de medir, y fuertemente condicionadas por características personales y del ámbito de trabajo.

La factibilidad de realizar una intervención educativa presencial con residentes limitó el poder de la muestra. Aunque una diferencia de 3% entre los grupos ya ha sido considerada como significativa (Nouh et al., 2012), el número de participantes de esta investigación sólo permitiría detectar diferencias estadísticamente grandes ($\geq 8\%$) entre las medias (Rosner, 2013). Por lo tanto, es posible que haya habido diferencias pedagógicamente significativas, que quedaron fuera de la capacidad del estudio de detectarlas, por falta de poder.

Debido a las severas restricciones relacionadas con la pandemia Covid19 que se impusieron en Argentina en marzo de 2020, que obligaron a suspender todas las reuniones presenciales, y en virtud de que no fue posible realizar una rápida transición para virtualizar los contenidos y actividades de enseñanza en ese momento, la duración de la intervención educativa fue menor a lo anticipado. Es posible que si la intervención hubiese sido como fue programada originalmente, se podrían haber alcanzado mayores diferencias entre los grupos.

Al mismo tiempo, y teniendo en cuenta la especificidad de contenido, es posible que un mayor número de casos analizados durante la intervención educativa podría haber logrado mejores resultados.

La intervención educativa de este estudio se realizó fuera del ámbito de trabajo. El impacto educativo de la misma podría haber sido mayor, de haberse realizado en el mismo ámbito de práctica e involucrando a los médicos con funciones docentes a cargo de las áreas asistenciales.

Aunque la evaluación implementada en la presente investigación no tenía consecuencias en la trayectoria académica de los residentes, y por lo tanto la composición del panel de expertos fue suficiente, hubiese sido óptimo contar con al menos 15 (en vez de 12) expertos en el panel. Por tanto, es posible que el sistema de calificación agregada le haya

otorgado un mayor valor relativo a las respuestas menos seleccionadas y, por tanto, desviantes. Cuando el número de expertos aumenta, el valor relativo de las respuestas desviantes disminuye, aumentando la validez de contenido.

El rendimiento de un integrante del panel de expertos estuvo muy por debajo de lo esperado ([apéndice IV](#), E12). No hay sustento en la literatura que justifique su exclusión (Robert Gagnon et al., 2011), sin embargo, esta situación evidencia la dificultad de identificar y capturar la experticia en habilidades complejas -contenido y contexto específicas- como el razonamiento clínico.

Por último, el *script concordance test* permite evaluar sólo ciertos aspectos de la competencia en razonamiento clínico, que es una competencia compleja. Sería deseable poder combinar múltiples métodos de evaluación del razonamiento clínico, y realizarlos en forma combinada, continua, y longitudinal (un *sistema* de evaluación del razonamiento clínico), para tener una perspectiva más amplia y completa del resultado de las intervenciones educativas en razonamiento clínico.

Conclusiones

Este estudio busca apoyar y colaborar con los esfuerzos globales de implementar intervenciones educativas longitudinales para el desarrollo de la competencia en razonamiento clínico, de acuerdo con las principales teorías que la sustentan, y en forma adecuada al contexto de nuestros estudiantes.

Aunque no fue posible demostrar diferencias entre los grupos de acuerdo a los parámetros establecidos, el proceso de investigación y los resultados obtenidos resultaron sumamente enriquecedores y buscan aportar al colectivo de investigación en educación médica en general, y a la evaluación del razonamiento clínico, en particular.

Futuras Direcciones

Teniendo en cuenta que la organización del conocimiento médico es un proceso continuo que se inicia en la facultad de medicina, continua en los sistemas de residencias y se prolonga en la práctica profesional continua, ninguna estrategia de enseñanza de la competencia en razonamiento clínico debería estar limitada a cierta etapa del desarrollo profesional.

Los conocimientos, habilidades y actitudes que conforman la competencia en razonamiento clínico deben estar integrados en todas las etapas de la educación médica. Por tanto, es un desafío la conformación de un cuerpo de docentes clínicos que integren el razonamiento clínico a lo largo de todas las etapas de formación profesional. Son necesarios más estudios para conocer el impacto educativo de este tipo de intervenciones educativas.

La falta de poder para detectar diferencias estadísticamente menores podría sobreponerse, en caso de que se incluyese el presente estudio de investigación con otros similares, para su revisión sistemática y meta análisis.



Agradecimientos

Quiero agradecer la enorme paciencia, soporte y apoyo del Dr. Carlos Brailovsky a lo largo de más de dos años del desarrollo del presente trabajo. Así también a la Maestría en Educación para Profesionales de la Salud del Instituto Universitario del Hospital Italiano, a los dos hospitales escuela, los residentes y sus responsables donde se llevaron adelante las evaluaciones, sin quienes hubiese sido imposible este trabajo. Y por último, pero no en menor medida, a mi familia y amigos por el acompañamiento y cuidado durante todo el proceso.

Bibliografía

- AAMC. (2014). Core Entrustable Professional Activities for Entering Residency: Faculty and Learner's Guide. Retrieved from www.aamc.org/91514/reproductions.html
- Ahmadi, S. F., Khoshkish, S., Soltani-Arabshahi, K., Hafezi-Moghadam, P., Zahmatkesh, G., Heidari, P., ... Lotfipour, S. (2014). Challenging script

- concordance test reference standard by evidence: Do judgments by emergency medicine consultants agree with likelihood ratios? *International Journal of Emergency Medicine*, 7(1), 1–7. <https://doi.org/10.1186/s12245-014-0034-3>
- Amey, L., Donald, K. J., & Teodorczuk, A. (2017). Teaching clinical reasoning to medical students. *British Journal of Hospital Medicine*, 78(7), 399–401. <https://doi.org/10.12968/hmed.2017.78.7.399>
- Ausubel, D. (1976). *Significado y aprendizaje significativo*.
- Balogh, E., Miller, B., & Ball, J. (2015). *Improving Diagnosis in Health Care* (E. P. Balogh, B. T. Miller, & J. R. Ball, Eds.). <https://doi.org/10.17226/21794>
- Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Medical Teacher*, 34(11), 960–992. <https://doi.org/10.3109/0142159X.2012.703791>
- Bordage, G., & Page, G. (2018, December 17). The key-features approach to assess clinical decisions: validity evidence to date. *Advances in Health Sciences Education*, Vol. 23, pp. 1005–1036. <https://doi.org/10.1007/s10459-018-9830-5>
- Bordage, Georges. (1994). Elaborated Knowledge: A Key to Successful Diagnostic Thinking. *Academic Medicine*, 69(11), 883–885.
- Bordage, Georges. (1995). Where are the History and the Physical? In *CAN MED ASSOC J * MAY* (Vol. 15).
- Bordage, Georges. (2007). Prototypes and semantic qualifiers: From past to present. *Medical Education*, Vol. 41, pp. 1117–1121. <https://doi.org/10.1111/j.1365-2923.2007.02919.x>
- Boulouffe, C., Charlin, B., & Vanpee, D. (2010). Evaluation of clinical reasoning in basic emergencies using a script concordance test. *American Journal of Pharmaceutical Education*, 74(10), 194. <https://doi.org/10.5688/aj7410194>
- Bowen, J. L. (2006). Educational Strategies to Promote Clinical Diagnostic Reasoning. *New England Journal of Medicine*, 355(21), 2217–2225. <https://doi.org/10.1056/NEJMra054782>
- Bowen, J. L., & Ilgen, J. S. (2014). Now You See It, Now You Don't: What Thinking Aloud Tells Us About Clinical Reasoning. *Journal of Graduate Medical Education*, 6(4), 783–785. <https://doi.org/10.4300/jgme-d-14-00492.1>
- Brailovsky, C. (2001). Educación médica, evaluación de las competencias. In *Aportes para un cambio curricular en Argentina 2001* (pp. 103–122). Retrieved from 950-710-071-7
- Brazeau-Lamontagne, L., Charlin, B., Gagnon, R., Samson, L., & Van Der Vleuten, C. (2004). Measurement of perception and interpretation skills during radiology training: Utility of the script concordance approach. *Medical Teacher*, 26(4), 326–332. <https://doi.org/10.1080/01421590410001679000>
- Bunniss, S., & Kelly, D. R. (2010). Research paradigms in medical education research. *Medical Education*, 44(4), 358–366. <https://doi.org/10.1111/j.1365-2923.2009.03611.x>
- Carrière, B., Gagnon, R., Charlin, B., Downing, S., & Bordage, G. (2009). Assessing

- Clinical Reasoning in Pediatric Emergency Medicine: Validity Evidence for a Script Concordance Test. *Annals of Emergency Medicine*, 53(5), 647–652.
<https://doi.org/10.1016/j.annemergmed.2008.07.024>
- Chang, R., Bordage, G., & Connel, K. (1998). The importance of early problem representation during case presentations. *Academic Emergency Medicine : Official Journal of the Society for Academic Emergency Medicine*, 73(10), S109–S111.
- Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning. *Medical Education*, 41(12), 1178–1184.
<https://doi.org/10.1111/j.1365-2923.2007.02924.x>
- Charlin, B., Desaulniers, M., Gagnon, R., Blouin, D., & van der Vleuten, C. (2002). Comparison of an Aggregate Scoring Method With a Consensus Scoring Method in a Measure of Clinical Reasoning Capacity. *Teaching and Learning in Medicine*, 14(3), 150–156. https://doi.org/10.1207/S15328015TLM1403_3
- Charlin, B., Gagnon, R., Lubarsky, S., Lambert, C., Meterissian, S., Chalk, C., ... van der Vleuten, C. (2010). Assessment in the context of uncertainty using the script concordance test: More meaning for scores. *Teaching and Learning in Medicine*, 22(3), 180–186. <https://doi.org/10.1080/10401334.2010.488197>
- Charlin, B., Gagnon, R., Pelletier, J., Coletti, M., Abi-Rizk, G., Nasr, C., ... van der Vleuten, C. (2006). Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. *Medical Education*, 40(9), 848–854. <https://doi.org/10.1111/j.1365-2929.2006.02541.x>
- Charlin, B., Roy, L., Brailovsky, C., Goulet, F., & Van Der Vleuten, C. (2000). The Script Concordance Test: A Tool to Assess the Reflective Clinician. *Teaching and Learning in Medicine*, 12(4), 189–195.
https://doi.org/10.1207/S15328015TLM1204_5
- Compère, V., Abily, J., Moriceau, J., Gouin, A., Veber, B., Dupont, H., ... Dureuil, B. (2016). Residents in tutored practice exchange groups have better medical reasoning as measured by script concordance test: A controlled, nonrandomized study. *Journal of Clinical Anesthesia*, 32, 236–241.
<https://doi.org/10.1016/j.jclinane.2016.03.012>
- Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, Vol. 78, pp. 775–780.
<https://doi.org/10.1097/00001888-200308000-00003>
- Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine : Journal of the Association of American Medical Colleges*, 84(8), 1022–1028.
<https://doi.org/10.1097/acm.0b013e3181ace703>
- Croskerry, P., & Norman, G. (2008, May). Overconfidence in Clinical Decision Making. *American Journal of Medicine*, 121(5 SUPPL.).
<https://doi.org/10.1016/j.amjmed.2008.02.001>
- Cumming, A., & Ross, M. (2007). *Learning Outcomes/ Competences for Undergraduate Medical Education in Europe*. Retrieved from www.tuning-medicine.com
- Custers, E. J. F. M. (2013). Medical Education and Cognitive Continuum Theory.

- Academic Medicine*, 88(8), 1074–1080.
<https://doi.org/10.1097/ACM.0b013e31829a3b10>
- Custers, E. J. F. M. (2014). Thirty years of illness scripts: Theoretical origins and practical applications. *Medical Teacher*, 37(5), 457–462.
<https://doi.org/10.3109/0142159X.2014.956052>
- Daniel, M., Rencic, J., Durning, S. J., Holmboe, E., Santen, S. A., Lang, V., ... Gruppen, L. D. (2019). Clinical Reasoning Assessment Methods. *Academic Medicine*, 94(6), 902–912. <https://doi.org/10.1097/ACM.0000000000002618>
- Delany, C., & Golding, C. (2014). Teaching clinical reasoning by making thinking visible: An action research project with allied health clinical educators. *BMC Medical Education*, 14(1). <https://doi.org/10.1186/1472-6920-14-20>
- Dhaliwal, G. (2013). Developing teachers of clinical reasoning. *Clinical Teacher*, 10(5), 313–317. <https://doi.org/10.1111/tct.12082>
- Dory, V., Gagnon, R., Vanpee, D., & Charlin, B. (2012). How to construct and implement script concordance tests: Insights from a systematic review. *Medical Education*, 46(6), 552–563. <https://doi.org/10.1111/j.1365-2923.2011.04211.x>
- Dumas, D., Torre, D. M., & Durning, S. J. (2018, May 1). Using Relational Reasoning Strategies to Help Improve Clinical Reasoning Practice. *Academic Medicine*, Vol. 93, pp. 709–714. <https://doi.org/10.1097/ACM.0000000000002114>
- Durante, E. (2005). *La Evaluación de los conocimientos: lo que parece ser, ¿es realmente lo que es?* Retrieved from https://www1.hospitalitaliano.org.ar/multimedia/archivos/noticias_attachs/47/documentos/10469_18-23.Durante.pdf
- Durante, E. (2012). La enseñanza en el ambiente clínico: principios y métodos. *Revista de Docencia Universitaria*, 10, 149–175.
- Durning, S., Artino, A. R., Pangaro, L., van der Vleuten, C. P., & Schuwirth, L. (2011). Context and clinical reasoning: Understanding the perspective of the expert's voice. *Medical Education*, 45(9), 927–938. <https://doi.org/10.1111/j.1365-2923.2011.04053.x>
- Durning, S. J. (2010). Exploring the Influence of Contextual Factors of the Clinical Encounter on Clinical Reasoning Success (Unraveling context specificity). *Academic Medicine*, 85(5), 894.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical Problem Solving. *Evaluation & the Health Professions*, 13(1), 5–36.
<https://doi.org/10.1177/016327879001300102>
- Emanuel, E. (1999). ¿Que hace que la investigacion clinica sea etica? siete requisitos. *Investigación En Sujetos Humanos: Experiencia Internacional. Santiago de Chile: Programa Regional de Bioética OPS/OMS*, 33–46.
- Ericsson, K. A. (2004). Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Academic Medicine*, 17(2), 139–148. <https://doi.org/10.1097/00001888-200410001-00022>
- Eva, K. (2005). What every teacher needs to know about clinical reasoning. *Medical*



- Education*, 8(2), 59–68. <https://doi.org/10.4324/9781315113272-2>
- Eva, K. W. (2003). On the generality of specificity. *Medical Education*, 37(7), 587–588. <https://doi.org/10.1046/j.1365-2923.2003.01563.x>
- Eva, K. W., & Regehr, G. (2008). I'll never play professional football and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions*, 28(1), 14–19. <https://doi.org/10.1002/chp.150>
- Fournier, J. P., Demeester, A., & Charlin, B. (2008). Script concordance tests: Guidelines for construction. *BMC Medical Informatics and Decision Making*, 8. <https://doi.org/10.1186/1472-6947-8-18>
- Frank, J. R., Snell, L., Englander, R., & Holmboe, E. S. (2017). Implementing competency-based medical education: Moving forward. *Medical Teacher*, 39(6), 568–573. <https://doi.org/10.1080/0142159X.2017.1315069>
- Frank, J. R., Snell, L. S., Cate, O. Ten, Holmboe, E. S., Carraccio, C., Swing, S. R., ... Harris, K. A. (2010). Competency-based medical education: theory to practice. *Medical Teacher*, 32(8), 638–645. <https://doi.org/10.3109/0142159X.2010.501190>
- Frank, J., Snell, L., & Sherbino, J. (2015). CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada. *CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada*, 1–30. Retrieved from www.royalcollege.ca/rcsite/documents/canmeds/canmeds-full-framework-e.pdf
- Gagnon, R., Charlin, B., Coletti, M., Sauv e, E., & Van Der Vleuten, C. (2005). Assessment in the context of uncertainty: How many members are needed on the panel of reference of a script concordance test? *Medical Education*, 39(3), 284–291. <https://doi.org/10.1111/j.1365-2929.2005.02092.x>
- Gagnon, Robert, Charlin, B., Lambert, C., Carri ere, B., & Van Der Vleuten, C. (2009). Script concordance testing: More cases or more questions? *Advances in Health Sciences Education*, 14(3), 367–375. <https://doi.org/10.1007/s10459-008-9120-8>
- Gagnon, Robert, Lubarsky, S., Lambert, C., & Charlin, B. (2011). Optimization of answer keys for script concordance testing: Should we exclude deviant panelists, deviant responses, or neither? *Advances in Health Sciences Education*, 16(5), 601–608. <https://doi.org/10.1007/s10459-011-9279-2>
- Gawad, N., Wood, T. J., Cowley, L., & Raiche, I. (2021). How do cognitive processes influence script concordance test responses? *Medical Education*, 55(3), 354–364. <https://doi.org/10.1111/medu.14416>
- Gay, S., Bartlett, M., & Mckinley, R. (2013). Teaching clinical reasoning to medical students. *Clinical Teacher*, 10(5), 308–312. <https://doi.org/10.1111/tct.12043>
- Graber, M. L. (2009). Educational strategies to reduce diagnostic error: Can you teach this stuff? *Advances in Health Sciences Education*, 14(1 SUPPL), 63–69. <https://doi.org/10.1007/s10459-009-9178-y>
- Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine*, 165(13), 1493–1499. <https://doi.org/10.1001/archinte.165.13.1493>

- Graber, M. L., Kissam, S., Payne, V. L., Meyer, A. N. D., Sorensen, A., Lenfestey, N., ... Singh, H. (2012). Cognitive interventions to reduce diagnostic error: A narrative review. *BMJ Quality and Safety*, 21(7), 535–557. <https://doi.org/10.1136/bmjqs-2011-000149>
- Hodges, B. (2006). Medical education and the maintenance of incompetence. *Medical Teacher*, 28(8), 690–696. <https://doi.org/10.1080/01421590601102964>
- Hornos, E. H., Pleguezuelos, E. M., Brailovsky, C. A., Harillo, L. D., Dory, V., & Charlin, B. (2013). The Practicum Script Concordance Test: An Online Continuing Professional Development Format to Foster Reflection on Clinical Practice. *Journal of Continuing Education in the Health Professions*, 33(1), 59–66. <https://doi.org/10.1002/chp.21166>
- Humbert, A. J., Besinger, B., & Miech, E. J. (2011). Assessing Clinical Reasoning Skills in Scenarios of Uncertainty: Convergent Validity for a Script Concordance Test in an Emergency Medicine Clerkship and Residency. *Academic Emergency Medicine*, 18(6), 627–634. <https://doi.org/10.1111/j.1553-2712.2011.01084.x>
- Humbert, A. J., Johnson, M. T., Miech, E., Friedberg, F., Grackin, J. A., & Seidman, P. A. (2011). Assessment of clinical reasoning: A Script Concordance test designed for pre-clinical medical students. *Medical Teacher*, 33(6), 472–477. <https://doi.org/10.3109/0142159X.2010.531157>
- Illing, J. (2010). Thinking about Research and Scholarship. In T. Swanwick (Ed.), *Research Methodologies in EU and International Law*. <https://doi.org/10.5040/9781472560919.ch-002>
- Janssen-Noordman, A. M. B., Merriënboer, J. J. G., van der Vleuten, C. P. M., & Scherpbier, A. J. J. A. (2006). Design of integrated practice for learning professional competences. *Medical Teacher*, 28(5), 447–452. <https://doi.org/10.1080/01421590600825276>
- Kassirer, J. P. (2010). Teaching clinical reasoning: Case-based and coached. *Academic Medicine*, 85(7), 1118–1124. <https://doi.org/10.1097/ACM.0b013e3181d5dd0d>
- Kazour, F., Richa, S., Zoghbi, M., El-Hage, W., & Haddad, F. G. (2017). Using the Script Concordance Test to Evaluate Clinical Reasoning Skills in Psychiatry. *Academic Psychiatry*, 41(1), 86–90. <https://doi.org/10.1007/s40596-016-0539-6>
- Keemink, Y., Custers, E. J. F. M., van Dijk, S., & Ten Cate, O. (2018). Illness script development in pre-clinical education through case-based clinical reasoning training. *International Journal of Medical Education*, 9, 35–41. <https://doi.org/10.5116/ijme.5a5b.24a9>
- Koufidis, C., Manninen, K., Nieminen, J., Wohlin, M., & Silén, C. (2021). Representation, Interaction, Interpretation. Making sense of the context in clinical reasoning. *Medical Education*, (7), medu.14545. <https://doi.org/10.1111/medu.14545>
- Kreiter, C. D. (2012, March). Commentary: The response process validity of a script concordance test item. *Advances in Health Sciences Education*, Vol. 17, pp. 7–9. <https://doi.org/10.1007/s10459-011-9325-0>
- Lambe, K. A., O'Reilly, G., Kelly, B. D., & Curristan, S. (2016). Dual-process



- cognitive interventions to enhance diagnostic reasoning: A systematic review. *BMJ Quality and Safety*, 25(10), 808–820. <https://doi.org/10.1136/bmjqs-2015-004417>
- Levin, M., Cennimo, D., Chen, S., & Lamba, S. (2016). Teaching Clinical Reasoning to Medical Students: A Case-Based Illness Script Worksheet Approach. *MedEdPORTAL Publications*, 12. https://doi.org/10.15766/mep_2374-8265.10445
- Lineberry, M., Hornos, E., Pleguezuelos, E., Mella, J., Brailovsky, C., & Bordage, G. (2019). Experts' responses in script concordance tests: a response process validity investigation. *Medical Education*. <https://doi.org/10.1111/medu.13814>
- Lineberry, M., Kreiter, C. D., & Bordage, G. (2013). Threats to validity in the use and interpretation of script concordance test scores. *Medical Education*, 47(12), 1175–1183. <https://doi.org/10.1111/medu.12283>
- Lineberry, M., Kreiter, C. D., & Bordage, G. (2014). Script concordance tests: strong inferences about examinees require stronger evidence. *Medical Education*, 48(4), 452–453. <https://doi.org/10.1111/medu.12417>
- Lubarsky, S., Chalk, C., Kazitani, D., Gagnon, R., & Charlin, B. (2009). The script concordance test: A new tool assessing clinical judgement in neurology. *Canadian Journal of Neurological Sciences*, 36(3), 326–331. <https://doi.org/10.1017/S031716710000706X>
- Lubarsky, S., Dory, V., Audétat, M.-C., Custers, E., & Charlin, B. (2015a). Using script theory to cultivate illness script formation and clinical reasoning in health professions education. *Canadian Medical Education Journal*, 6(62), 61–70. Retrieved from <http://www.cmej.ca>
- Lubarsky, S., Dory, V., Audétat, M.-C., Custers, E., & Charlin, B. (2015b). Using script theory to cultivate illness script formation and clinical reasoning in health professions education. *Canadian Medical Education Journal*, 6(2), e61–e70. <https://doi.org/10.36834/cmej.36631>
- Lubarsky, S., Dory, V., Duggan, P., Gagnon, R., & Charlin, B. (2013). Script concordance testing: From theory to practice: AMEE Guide No. 75. *Medical Teacher*, 35(3), 184–193. <https://doi.org/10.3109/0142159X.2013.760036>
- Lubarsky, S., Dory, V., Meterissian, S., Lambert, C., & Gagnon, R. (2018). Examining the effects of gaming and guessing on script concordance test scores. *Perspectives on Medical Education*, 7(3), 174–181. <https://doi.org/10.1007/s40037-018-0435-8>
- Lubarsky, S., Gagnon, R., & Charlin, B. (2013). Scoring the Script Concordance Test: not a black and white issue. *Medical Education*, 47(12), 1159–1161. <https://doi.org/10.1111/medu.12362>
- Makary, M. A., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. *BMJ (Online)*, 353. <https://doi.org/10.1136/bmj.i2139>
- Mandin, H., Jones, A., Woloschuk, W., & Harasym, P. (1997). Helping students learn to think like experts when solving clinical problems. *Academic Medicine*, 72(3), 173–179. <https://doi.org/10.1097/00001888-199703000-00009>
- Marie, I., Sibert, L., Roussel, F., Hellot, M. F., Lechevallier, J., & Weber, J. (2005). The script concordance test: A new evaluation method of both clinical reasoning and skills in internal medicine. *Revue de Medecine Interne*, 26(6), 501–507.

<https://doi.org/10.1016/j.revmed.2004.10.028>

Monteiro, S., & Norman, G. (2013). Diagnostic Reasoning: Where We've Been, Where We're Going. *Teaching and Learning in Medicine*, 25(SUPPL.1), 37–41.

<https://doi.org/10.1080/10401334.2013.842911>

Monteiro, S., Sherbino, J., Schmidt, H., Mamede, S., Ilgen, J., & Norman, G. (2020). It's the destination: diagnostic accuracy and reasoning. *Advances in Health Sciences Education*, 25(1), 19–29. <https://doi.org/10.1007/s10459-019-09903-7>

Monteiro, S., Sherbino, J., Sibbald, M., & Norman, G. (2020). Critical thinking, biases and dual processing: The enduring myth of generalisable skills. *Medical Education*, 54(1), 66–73. <https://doi.org/10.1111/medu.13872>

Moulton, C. A. E., Regehr, G., Mylopoulos, M., & MacRae, H. M. (2007, October). Slowing down when you should: A new model of expert judgment. *Academic Medicine*, Vol. 82, pp. S109–S116.

<https://doi.org/10.1097/ACM.0b013e3181405a76>

Murinson, B. B., Agarwal, A. K., & Haythornthwaite, J. A. (2008). Cognitive Expertise, Emotional Development, and Reflective Capacity: Clinical Skills for Improved Pain Care. *Journal of Pain*, 9(11), 975–983.

<https://doi.org/10.1016/j.jpain.2008.07.010>

Nendaz, M., & Perrier, A. (2012). Diagnostic errors and flaws in clinical reasoning: Mechanisms and prevention in practice. *Swiss Medical Weekly*, Vol. 142.

<https://doi.org/10.4414/smw.2012.13706>

Nendaz, M. R., & Bordage, G. (2002). Promoting diagnostic problem representation. *Medical Education*, 36(8), 760–766. <https://doi.org/10.1046/j.1365-2923.2002.01279.x>

Nendaz, M. R., Gut, A. M., Perrier, A., Louis-Simonet, M., Blondon-Choa, K., Herrmann, F. R., ... Vu, N. V. (2006). Brief report: Beyond clinical experience: Features of data collection and interpretation that contribute to diagnostic accuracy. *Journal of General Internal Medicine*, 21(12), 1302–1305.

<https://doi.org/10.1111/j.1525-1497.2006.00587.x>

Norman, G. (2005, April). Research in clinical reasoning: Past history and current trends. *Medical Education*, Vol. 39, pp. 418–427. <https://doi.org/10.1111/j.1365-2929.2005.02127.x>

Norman, G. (2006). Building on Experience — The Development of Clinical Reasoning. *New England Journal of Medicine*, 355(21), 2251–2252.

<https://doi.org/10.1056/NEJMe068134>

Norman, G., Bordage, G., Page, G., & Keane, D. (2006). How specific is case specificity? *Medical Education*, 40(7), 618–623. <https://doi.org/10.1111/j.1365-2929.2006.02511.x>

Norman, G. R., Monteiro, S., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2017). The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking. *Academic Medicine*, 92(1), 23–30.

<https://doi.org/10.1097/ACM.0000000000001421>

Nouh, T., Boutros, M., Gagnon, R., Reid, S., Leslie, K., Pace, D., ... Meterissian, S. H.

- (2012). The script concordance test as a measure of clinical reasoning: A national validation study. *American Journal of Surgery*, 203(4), 530–534. <https://doi.org/10.1016/j.amjsurg.2011.11.006>
- Otterman, N., Maas, M., Schiemanck, S., van der Wees, P., & Kwakkel, G. (2019). Development and validity of an innovative test to assess guideline-consistent clinical reasoning by physical therapists in stroke rehabilitation. *Journal of Rehabilitation Medicine*, 0. <https://doi.org/10.2340/16501977-2562>
- Petrucci, A. M., Nouh, T., Boutros, M., Gagnon, R., & Meterissian, S. H. (2013). Assessing clinical judgment using the Script Concordance test: The importance of using specialty-specific experts to develop the scoring key. *American Journal of Surgery*, 205(2), 137–140. <https://doi.org/10.1016/j.amjsurg.2012.09.002>
- Pleguezuelos, E. M., Hornos, E., Dory, V., Gagnon, R., Malagrino, P., Brailovsky, C. A., & Charlin, B. (2013). Script concordance testing in continuing professional development: local or international reference panels? *Advances in Health Sciences Education*, 18(4), 779–785. <https://doi.org/10.1007/s10459-012-9427-3>
- Rencic, J., & Trowbridge, R. (2011). A novel curriculum in clinical reasoning. *Journal of General Internal Medicine*, 26, S612–S613. Retrieved from <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L70654434>
- Rencic, Joseph. (2011). Twelve tips for teaching expertise in clinical reasoning. *Medical Teacher*, Vol. 33, pp. 887–892. <https://doi.org/10.3109/0142159X.2011.558142>
- Rencic, Joseph, Trowbridge, R. L., Fagan, M., Szauter, K., & Durning, S. (2017). Clinical Reasoning Education at US Medical Schools: Results from a National Survey of Internal Medicine Clerkship Directors. *Journal of General Internal Medicine*, 32(11), 1242–1246. <https://doi.org/10.1007/s11606-017-4159-y>
- Richmond, A., Cooper, N., Gay, S., Atiomo, W., & Patel, R. (2020, August 1). The student is key: A realist review of educational interventions to develop analytical and non-analytical clinical reasoning ability. *Medical Education*, Vol. 54, pp. 709–719. <https://doi.org/10.1111/medu.14137>
- Rosner, B. (2013). Fundamentals of Biostatistics, Eighth Edition. In *Cengage Learning* (Vol. 261).
- Schaye, V., Elias, K. L., Janjigian, M., & Stern, D. T. (2019). Theory-guided teaching: Implementation of a clinical reasoning curriculum in residents. *Medical Teacher*, 41(10), 1192–1199. <https://doi.org/10.1080/0142159X.2019.1626977>
- Schmidt, H. G., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: A narrative review and a proposal. *Medical Education*, Vol. 49, pp. 961–973. <https://doi.org/10.1111/medu.12775>
- Shaw E, Oandasan I, Fowler N, E. (2017). CanMEDS-FM 2017: A competency framework for family physicians across the continuum. *The College of Family Physicians of Canada*. Retrieved from https://www.cfpc.ca/uploadedFiles/Resources/Resource_Items/Health_Professionals/CanMEDS-Family-Medicine-2017-ENG.pdf
- Silva Fuente-Alba, C., & Molina Villagra, M. (2017). Likelihood ratio (razón de

- verosimilitud): definición y aplicación en Radiología. *Revista Argentina de Radiología*, 81(3), 204–208. <https://doi.org/10.1016/j.rard.2016.11.002>
- Simpkin, A. L., Vyas, J. M., & Armstrong, K. A. (2017). Diagnostic Reasoning: An Endangered Competency in Internal Medicine Training. *Annals of Internal Medicine*, 167(7), 507. <https://doi.org/10.7326/M17-0163>
- ten Cate, O. (2005). Entrustability of professional activities and competency-based training. *Medical Education*, 39(12), 1176–1177. <https://doi.org/10.1111/j.1365-2929.2005.02341.x>
- ten Cate, O. (2013). Competency-Based Education, Entrustable Professional Activities, and the Power of Language. *Journal of Graduate Medical Education*, 5(1), 6–7. <https://doi.org/10.4300/jgme-d-12-00381.1>
- Ten Cate, O. (2014). What is a 21st-century doctor? rethinking the significance of the medical degree. *Academic Medicine*, Vol. 89, pp. 966–969. <https://doi.org/10.1097/ACM.0000000000000280>
- Ten Cate, O. (2018). A primer on entrustable professional activities. *Korean J Med Educ*, 30(1), 1–10. <https://doi.org/10.3946/kjme.2018.76>
- Ten Cate, O., Chen, H. C., Hoff, R. G., Peters, H., Bok, H., & Van Der Schaaf, M. (2015). Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Medical Teacher*, 37(11), 983–1002. <https://doi.org/10.3109/0142159X.2015.1060308>
- Thistlethwaite, J. E., Davies, D., Ekeocha, S., Kidd, J. M., MacDougall, C., Matthews, P., ... Clay, D. (2012). The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Medical Teacher*, 34(6), e421–e444. <https://doi.org/10.3109/0142159X.2012.680939>
- Trowbridge, R. L. (2008). Twelve tips for teaching avoidance of diagnostic errors. *Medical Teacher*, 30(5), 496–500. <https://doi.org/10.1080/01421590801965137>
- Trowbridge, R. L., Dhaliwal, G., & Cosby, K. S. (2013). Educational agenda for diagnostic error reduction. *BMJ Quality and Safety*, 22(SUPPL.2), 28–32. <https://doi.org/10.1136/bmjqs-2012-001622>
- Trowbridge, R. L., & Olson, A. P. J. (2018). Becoming a teacher of clinical reasoning. *Diagnosis (Berlin, Germany)*, 5(1), 11–14. <https://doi.org/10.1515/dx-2018-0004>
- Van Der Vleuten, C. P. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education : Theory and Practice*, 1(1), 41–67. <https://doi.org/10.1007/BF00596229>
- van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2019). Assessment in the context of problem-based learning. *Advances in Health Sciences Education*, 24(5), 903–914. <https://doi.org/10.1007/s10459-019-09909-1>
- Van Merriënboer, J. J. G., & De Croock, M. (2002). Performance-based ISD 10 steps to complex learning. *Performance Improvement*, 41(7), 35–40. <https://doi.org/10.1002/pfi.4140410708>



- Van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten Steps to Complex Learning: A systematic Approach to Four-Component Instructional Design*. 1–29.
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. <https://doi.org/10.1007/s10648-005-3951-0>
- Wan, M. S. H., Tor, E., & Hudson, J. N. (2018). Improving the validity of script concordance testing by optimising and balancing items. *Medical Education*, *52*(3), 336–346. <https://doi.org/10.1111/medu.13495>
- Wilson, A. B., Pike, G. R., & Humbert, A. J. (2014). Analyzing Script Concordance Test Scoring Methods and Items by Difficulty and Type. *Teaching and Learning in Medicine*, *26*(2), 135–145. <https://doi.org/10.1080/10401334.2014.884464>
- Young, J. Q., Van Merriënboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive Load Theory: Implications for medical education: AMEE Guide No. 86. *Medical Teacher*, *36*(5), 371–384. <https://doi.org/10.3109/0142159X.2014.889290>

Apéndice I: Autoevaluación de competencias

Se presentan las encuestas a 20 residentes al comienzo del primer y segundo año de sus residencias básicas. El objetivo era la autoevaluación de competencias profesionales críticas.

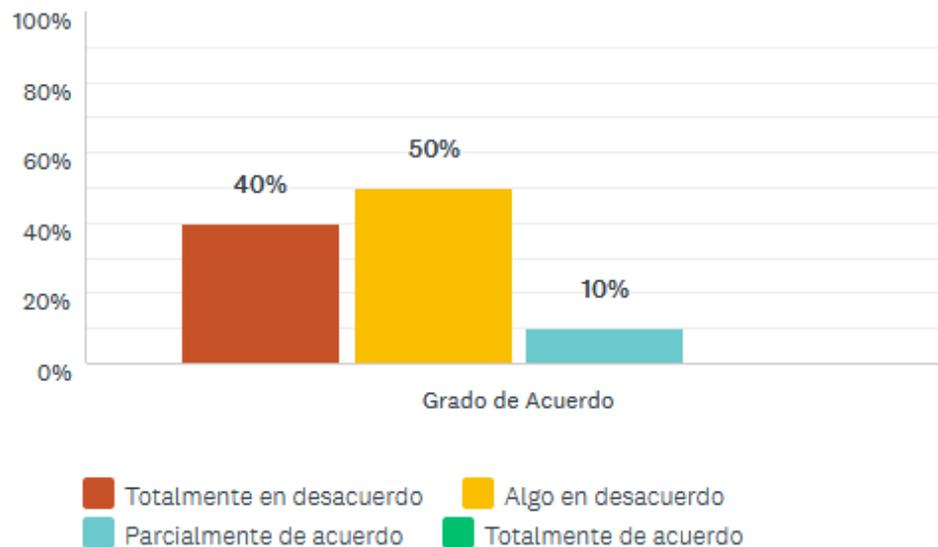
Pregunta No 1:

Al comenzar el primer día de la residencia, yo creo que era capaz de realizar el interrogatorio y el examen físico de mis pacientes *sin supervisión directa*¹:

¹ “Supervisión directa” significa tener a un residente superior o staff físicamente a tu lado.

Al comenzar el primer día de la residencia, yo creo que er...

Respondidas: 20 Omitidas: 0



Autoevaluación de Competencias - Residentes HB 2019-2020

🔍 (0)

Análisis

El **90%** (18 sobre 20) estuvo “totalmente” o “algo en desacuerdo” con “ser capaces de **realizar el interrogatorio y el examen físico** sin supervisión directa”, y solo el 10% (2 sobre 20) estuvo “parcialmente de acuerdo”.

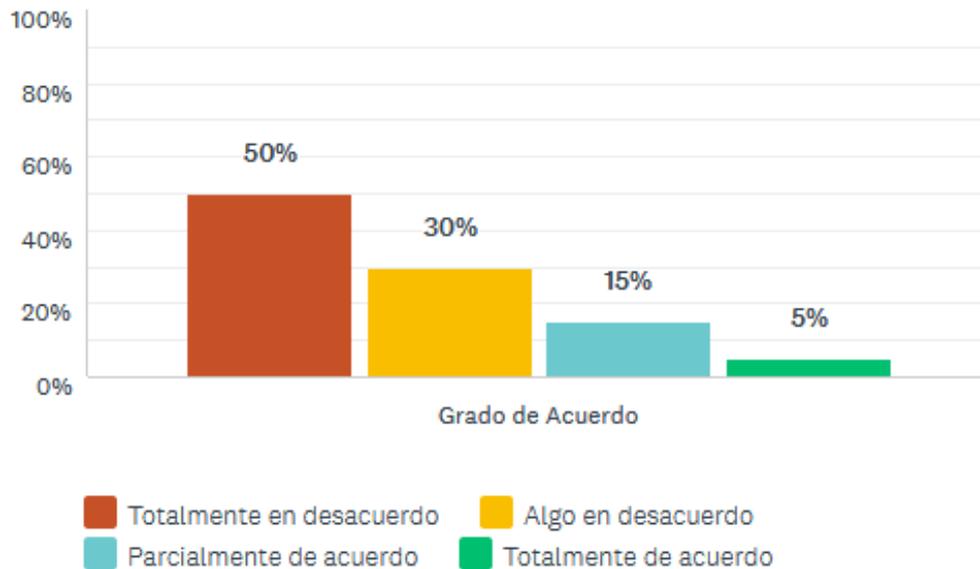
Pregunta No 2:

Al comenzar el primer día de la residencia, yo era capaz de hacer una adecuada presentación resumida del encuentro con mi paciente sin supervisión directa¹:

¹ “Supervisión directa” significa tener a un residente superior o staff físicamente a tu lado.

Al comenzar el primer día de la residencia, yo era capaz d...

Respondidas: 20 Omitidas: 0



Autoevaluación de Competencias - Residentes HB 2019-2020

🔍 (0)

Análisis

El 80% (16 sobre 20) estuvo “totalmente” o “algo en desacuerdo” con “ser capaces de **hacer una presentación resumida del encuentro con el paciente**, sin supervisión directa*”, el 15% (3 sobre 20) estuvo “parcialmente de acuerdo” y sólo 1 de 20 (5%) estuvo “totalmente de acuerdo”.

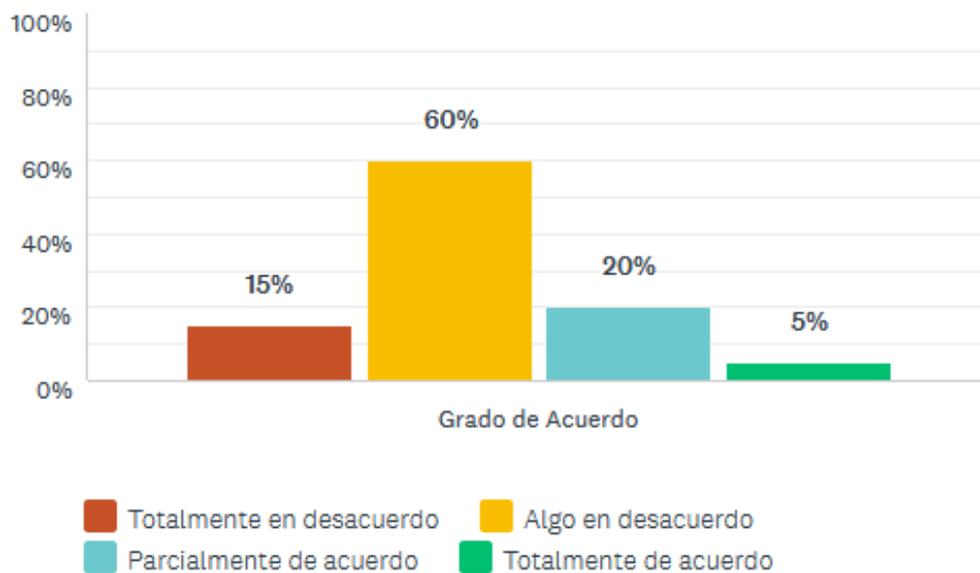
Pregunta No 3:

Al comenzar el primer día de la residencia, yo era capaz de **establecer un panel de diagnósticos diferenciales** de mis pacientes a cargo, *sin supervisión directa*¹:

¹ “Supervisión directa” significa tener a un residente superior o staff físicamente a tu lado.

Al comenzar el primer día de la residencia, yo era capaz d...

Respondidas: 20 Omitidas: 0



Autoevaluación de Competencias - Residentes HB 2019-2020

🔍 (0)

Análisis

El 75% (15 sobre 20) estuvo “totalmente” o “algo en desacuerdo” con “ser capaces de **establecer un panel de diagnósticos diferenciales** de los pacientes a cargo, sin supervisión directa”, 4 de 20 (20%) estuvo “parcialmente de acuerdo” y sólo 1 de 20 (5%) estuvo “totalmente de acuerdo”.

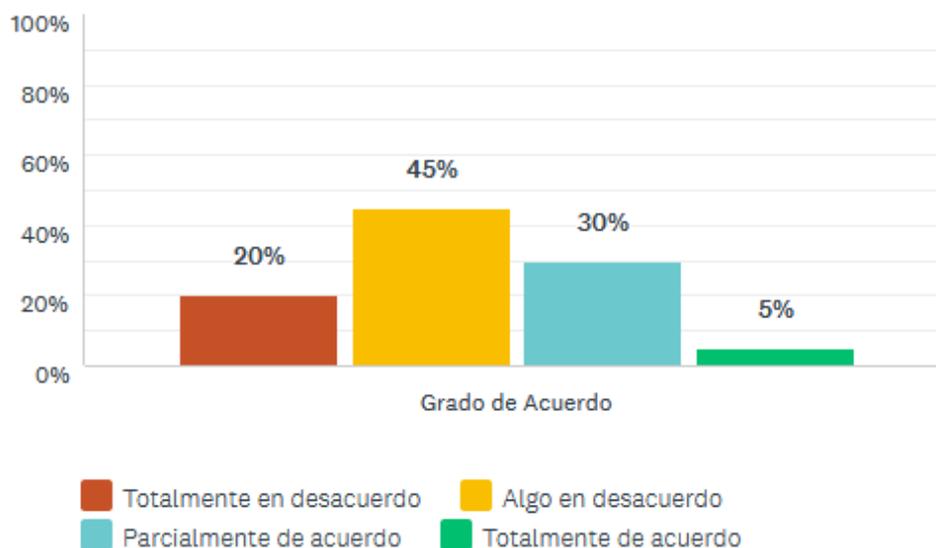
Pregunta No 4:

Al comenzar el primer día de la residencia, yo era capaz de **seleccionar, justificar e interpretar estudios complementarios comunes** para mis pacientes *sin supervisión directa*¹:

¹ “Supervisión directa” significa tener a un residente superior o staff físicamente a tu lado.

Al comenzar el primer día de la residencia, yo era capaz d...

Respondidas: 20 Omitidas: 0



Autoevaluación de Competencias - Residentes HB 2019-2020

🔍 (0)

Análisis

El **65%** (13 sobre 20) estuvo “totalmente” o “algo en desacuerdo” con “ser capaces de **seleccionar, justificar e interpretar estudios complementarios comunes** para sus pacientes, sin supervisión directa”, 6 de 20 (30%) estuvo “parcialmente de acuerdo”, y sólo 1 de 20 (5%) estuvo “totalmente de acuerdo”.

Pregunta No 5:

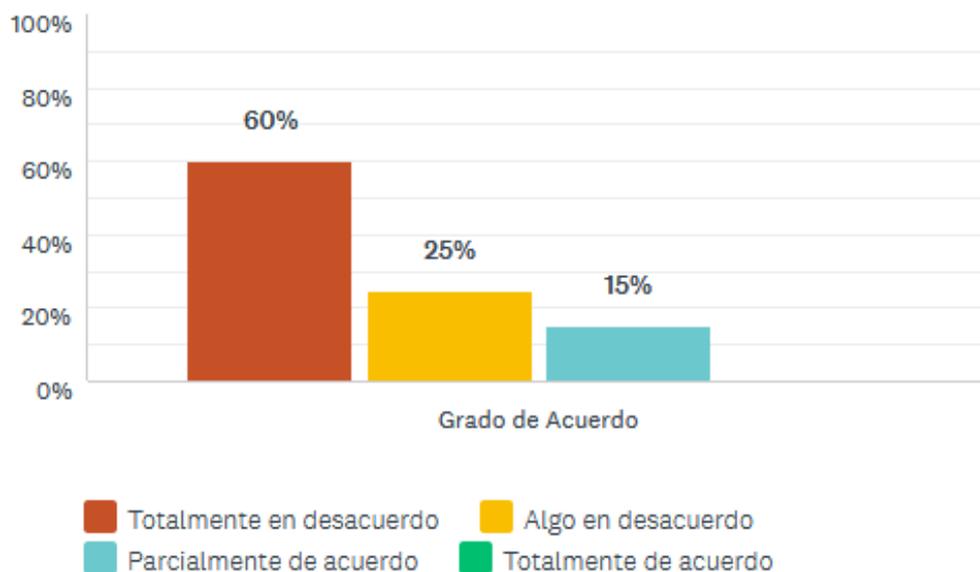
Al comenzar el primer día de la residencia, yo era capaz de **identificar fuentes de errores comunes¹ y contaba con estrategias para minimizarlos, sin supervisión directa²:**

¹ Se refiere a errores comunes en el interrogatorio, examen físico, selección e interpretación de estudios complementarios.

² “Supervisión directa” significa tener a un residente superior o staff físicamente a tu lado.

Al comenzar el primer día de la residencia, yo era capaz d...

Respondidas: 20 Omitidas: 0



Autoevaluación de Competencias - Residentes HB 2019-2020

🔍 (0)

Análisis

El **85%** (17 sobre 20) estuvo “totalmente” o “algo en **desacuerdo**” con “ser capaces de **identificar fuentes de errores comunes y contaba con estrategias para minimizarlos, sin supervisión directa**”, 3 de 20 (15%) estuvo “parcialmente de acuerdo” y ninguno estuvo “totalmente de acuerdo”.

Apéndice II: Ejemplos de intervenciones educativas y estrategias de enseñanza.

Aprendizaje Basado en Problemas

Ejemplo 1:

ABP – Discusión clínica y toma de decisiones.

Un varón de 18 años es atendido en su servicio, donde ocurre la siguiente discusión:

Médico A: *El paciente comienza su enfermedad actual posterior a un viaje en avión de tres tramos -17 hs en total- con dolor en el hombro derecho, de inicio rápido, sin mediar trauma, y que empeora parcialmente con el movimiento, asociado a un dolor punzante torácico lateral derecho, de intensidad moderada-severa, que limita y dificulta la respiración, persistente durante 48 hs, motivo por el cual consulta. No fuma, no usa drogas, ni tuvo episodios previos similares. Niega otros síntomas, como tos, expectoración, fiebre. No usa medicación habitual. En el examen físico no presentaba taquicardia ni taquipnea, estaba normotenso y sin desaturación (98%). El dolor no reproduce a la palpación del hombro ni del tórax, limita parcialmente la inspiración profunda, y no tiene ruidos pulmonares agregados. El murmullo vesicular está conservado. No presenta edemas, ingurgitación yugular ni soplo cardíaco. No tiene asimetría de miembros inferiores. La inspección del hombro es normal, y no se palpan adenomegalias.*

Médico B: *Creo que en este punto sería importante establecer diagnósticos de trabajo y la probabilidad clínica de cada uno de ellos, para guiar la aproximación diagnóstica.*

Médico C: *Yo preferiría conocer los resultados de los estudios que se solicitaron, para discutir luego, con más información pertinente.*

Médico A: *Se realizaron laboratorio general, ECG y Rx Tx, sin alteraciones. Se realizó una tomografía de tórax sin contraste, que no demostró alteraciones. Se solicitó un dímero D que resultó positivo (1100 ug/dl). El proBNP y las troponinas fueron negativas. Se realizó una angiografía de tórax que resultó negativa para TEP, aunque la calidad del estudio es sub-óptima.*

Médico B: *Me resulta difícil interpretar estos resultados... creo que no alcanzamos el umbral diagnóstico, y debemos tomar decisiones de manejo y tratamiento.*

Médico C: *Yo creo que con esos resultados negativos podemos darle el alta. Hemos descartado las causas potencialmente graves.*

Médico A: *No sé si nos vamos a poner de acuerdo, ¿y se le preguntamos al jefe para que defina?*

Ejemplo 2:

Razonamiento Clínico

ABP – 23/12/2019

Son las 7.30 am, Ud. está post-guardia por comenzar el pase de guardia, cuando recibe un llamado de Roberto, su paciente de 45 años:

Roberto: *Hola Doc., ¿cómo estás? Necesito verte porque estoy con un cólico renal terrible.*

Ud.: *Hola Roberto, contame qué te pasó, por favor.*

Roberto: *Ayer estaba mirando el fulbo... estaba perfecto (hasta que nos hizo el gol Boca), cuando a las 9 y 25 de la noche me agarró un dolor tremendo en las lumbares. Pensé que era una contractura, pero no me mejoró al levantarme. No paraba.... Entonces me tomé dos Dolten de 20, y no me mejoró nada... entonces me tomé dos más. Creo que me cayeron un poco mal, porque me agarraron náuseas y Lili -mi esposa- me dijo que me puse pálido. Por suerte no vomité. Entonces me acosté en la cama, ¡pero no se me iba! ¡Estaba doblado en la cama! Llamamos a la ambulancia a casa, vino el médico, pero ni me revisó. Me tomó la presión, la temperatura, y aunque no tenía fiebre, me dijo que tenía la presión alta (19 – 10), por un cólico renal. Luego me aplicó un intramuscular (no sé de qué) y se fue. Pero a mí no me mejoró nada, nunca se me fue el dolor y no pude pegar un ojo en toda la noche.... Un dolor espantoso...*

Ud.: *¿Vos tuviste cálculos en los riñones, Roberto?*

Roberto: *Sí, Doqui, hace muchos años, es el único problema de salud que tuve. Creo que fue del lado derecho... o del izquierdo, no me acuerdo... el urólogo me dijo que tome mucha agua así no me ponía un catéter... y zafé. Ahora no sé cuál de los riñones será porque me duelen los dos, pero por suerte orino bien. ¿Cuándo te puedo ver en el consultorio? ¿O voy a verlo directo al urólogo?*

Juegos de Roles

- Ejemplo 1: libreto para paciente simulado:

Role Playing – Paciente con cefalea

CONTEXTO

Ud. es una paciente que concurre al consultorio de un médico clínico en el hospital.

Puede usar el nombre que Ud. desee y tiene 35 años. Es vendedor/a de comercio y trabaja 12 horas por día. Tiene 3 hijos de 16, 11 y 8 años. El mayor, adolescente, le trae muchos problemas.

Fuma ocasionalmente uno o dos cigarrillos y toma alcohol (cerveza) los fines de semana, no más de 2 litros. Hizo su control ginecológico hace 1 año. No hace ejercicio.

SITUACIÓN CLÍNICA

Ud. consulta porque en las últimas semanas duerme muy mal, está muy cansada todo el día y siente que “no puede más”. Lo que más le molesta es que le volvió a agarrar un dolor de cabeza que había tenido años atrás. Pero hace foco en los problemas que tiene con su hijo adolescente: por ejemplo, que “sale a bailar 3 veces por semana y vuelve borracho”, “que tuvo que ir a buscarlo en la madrugada a la comisaría”, que “trae a la casa a la novia sin su consentimiento”, que “no estudia”, y que “mi marido no me ayuda”.

No tuvo fiebre, el apetito está conservado y no perdió peso. En ocasiones tuvo náuseas y un vómito con el dolor de cabeza. Cualquier otro síntoma que se le consulte, no lo tiene.

ACTITUD

Ud. debe mostrarse agotada, sin energía, triste y al borde de las lágrimas. Puede llorar, si le parece. No se enfoque en la descripción del dolor de cabeza, sino en todos sus problemas laborales (“trabajo demasiado”), de pareja (“mi marido no me ayuda”) y familiares (su hijo adolescente).

Sólo si el médico le pregunta en forma directa las características del dolor, Ud. contesta:

- ✓ Localización: toda la cabeza, pero **más del lado izquierdo**.
- ✓ Frecuencia: dos o tres veces por semana.
- ✓ Intensidad: bastante intensa (8/10), a veces impide seguir trabajando. No la despierta en la noche.
- ✓ Carácter: “como que late”
- ✓ Mejora solo parcialmente con ibuprofeno 600 mg. El paracetamol no le hace nada. En ocasiones tomó “migral”. Le ayuda acostarse con la luz apagada en el cuarto y “estar tranquila”.
- ✓ Síntomas asociados: náuseas. Alguna vez vomitó “porque le cayó algo mal”.
- ✓ No tuvo pérdida de fuerza en los miembros, cambios en la visión, movimientos anormales u otros síntomas neurológicos.
- ✓ **Ha tenido este dolor hace muchos años atrás**, “cuando era joven” (20 años aprox).

- Ejemplo 1: libreto para médico:

Role Playing - Médico

CONTEXTO

Ud. es un médico/a residente en su rotación por consultorios externos del hospital.

Puede usar su propio nombre y edad.

Se dispone a atender a un/a paciente por primera vez, por un turno programado.

Dispone de 15 minutos. No se puede exceder de ese tiempo de consulta.

SITUACIÓN CLÍNICA

Ud. debe asistir al paciente en cuestión, realizando las acciones que considere necesarias para establecer un diagnóstico y, de ser necesario, puede planificar exámenes complementarios y/o tratamientos, según corresponda.

ACTITUD

Actúe de acuerdo a su práctica y costumbre habitual.

- Ejemplo de consigna para trabajo en grupos pequeños sobre el juego de roles:

Discusión Grupal

1. Definan el problema de la paciente en una frase breve.
2. Generen hipótesis diagnósticas: al menos dos.
3. Comparen las hipótesis con los elementos clave de cada una presentes.
4. Establezcan una probabilidad clínica para cada una de las hipótesis.
5. Indiquen qué haría para verificar o descartar las hipótesis mencionadas. Justifiquen.

- Ejemplo 2: libreto para paciente simulado:

Role Playing - Paciente - Caso Dolor Lumbar

CONTEXTO

Ud. es un/a paciente que concurre al consultorio de su médico clínico de cabecera. Puede usar el nombre que Ud desee y tiene 55 años. Es abogado/a y hace ejercicio regularmente (bicicleta). No tiene antecedentes relevantes, ni toma medicación. Conoce a su médico de cabecera desde hace muchos años, y tiene buena relación. Hace los controles clínicos habituales.

SITUACIÓN CLÍNICA

Ud consulta por un control de salud habitual. Concorre para “hacerse un chequeo”. El único síntoma que tiene es un dolor lumbo-sacro, que comenzó hace 2 meses atrás. Al comienzo era intermitente y de intensidad leve, actualmente es continuo y de intensidad moderada-severa, y mejora solo parcialmente con ibuprofeno (toma 600 mg cada 8 horas). El dolor es en la línea media, región lumbo-sacra. No limita sus actividades laborales, pero tuvo que limitar la bicicleta. No mejora al acostarse, e incluso algunas noches lo ha despertado el dolor. No tiene problemas urinarios ni defecatorios. Tampoco debilidad en miembros inferiores, ni trastornos sensitivos. No tuvo fiebre, el apetito está conservado y no perdió peso. Cualquier otro síntoma que se le consulte, no lo tiene.

ACTITUD

Ud no le da importancia al dolor lumbar. De hecho no debe mencionarlo al inicio de la consulta: concurre para un “chequeo”. Solo debe mencionarlo si se le pregunta: ¿tiene alguna molestia? (O similar), o en caso de que esté terminando la consulta, el médico no se lo haya preguntado, y Ud debe decirlo desde la puerta, justo antes de retirarse, como al pasar.

CONDUCTA

Si el médico lo revisa, Ud no evoca ningún síntoma, EXCEPTO un dolor exquisito a la palpación de la apofisis espinosa de L4 (aproximadamente). Si el médico no hace esa maniobra, Ud no manifiesta dolor. Si el médico le pide que ubique el dolor, Ud puede señalar la zona. Si el médico le plantea la posibilidad de hacer estudios por el dolor lumbar, Ud DEBE preguntar qué es lo que piensa su médico que tiene. Por ejemplo: “¿qué piensa que tengo, Doc? ¿es serio esto? ¿Puedo tener cáncer?”

- Ejemplo 2: libreto para médico:

Role Playing - Médico

CONTEXTO

Ud. es un médico/a en su consultorio particular, fuera del ámbito hospitalario. Puede usar su propio nombre y edad. Se dispone a atender a un/a paciente que conoce hace muchos años, por un turno programado. Dispone de 15 minutos. No se puede exceder de ese tiempo de consulta.

SITUACIÓN CLÍNICA

Ud debe asistir al paciente en cuestión, realizando las acciones que considere necesarias para establecer un diagnóstico, y comunicárselo al paciente.

ACTITUD

Actúe de acuerdo a su práctica y costumbre habitual.

Uso de Casos Virtuales

Ejemplo de uso de casos virtuales para el uso de calificadores semánticos, representación del problema, generación de hipótesis y verificación diagnóstica:

Parte 1

Razonamiento Clínico

G. Jung

2019

TALLER MÓDULO 2: CALIFICACIÓN SEMÁNTICA

La Dra. P. es médica de familia y está atendiendo pacientes en el consultorio de demanda espontánea como parte de un equipo de Clínica Médica en un hospital de tercer nivel. Dispone de 15 minutos por paciente.

Al primer turno concurre el Sr. M. Chan, cuyo documento de triaje de enfermería informa como queja principal tos. Es la primera consulta del Sr. Chan, quien está bastante abrigado -es invierno- y está sentado en la camilla cuando la Dra. P. entra a la habitación. La Dra. P. cierra la puerta y se para, apoyada en la pared, con un cuaderno de notas en la mano. Comienza un interrogatorio detallado sobre sus antecedentes, incluidos medicamentos, uso de tóxicos y sustancias, y detalla un árbol genealógico. Incluye el estado civil, la situación de vida actual y el historial de internaciones. No interroga antecedentes ocupacional o de viajes. Continúa: "Las enfermeras dijeron que tiene tos. ¿Hace cuánto tiempo le ocurre?". Sigue con una serie de preguntas con respecto a la descripción y progresión de la tos. Descubre que el paciente tiene tos crónica que parece haber empeorado mucho. Pregunta sobre síntomas asociados y factores de empeoramiento o alivio.

La Dra P. se incomoda en varias ocasiones durante el interrogatorio porque el Sr Chan "no sabe explicar lo que le pasa". Resulta que el Sr Chan es chino, llegó al país hace 2 años, y aún no domina bien el castellano. Parece estar algo preocupado o asustado, la Dra P. no sabría decir, y desconoce las creencias sobre la propia salud del paciente.

Al terminar con el interrogatorio, la Dra P. dice: "Está bien, señor Chan, voy a echarle un vistazo". Comienza por auscultar los pulmones en seis áreas, primero debajo de la camisa y luego moviéndose sobre la camisa para las zonas pulmonares posteriores superiores. Ausculta murmullo vesicular, sin ruidos pulmonares agregados. Escucha el corazón en cuatro áreas, sin oír soplos ni ruidos cardíacos agregados. Toma el otoscopio en la pared y lo usa para verificar la reacción pupilar a la luz y los movimientos oculares, mira dentro de la orofaringe y examina el oído. Hace un examen breve pero apropiado del abdomen y revisa la piel en busca de erupciones y pies en busca de pulsos. Ella no encuentra otros hallazgos relevantes.

Al finalizar, la Dra P. le dice al paciente que lo discutirá con su equipo y que regresará. Cuando sale de la habitación, el Sr. Chan pregunta tímidamente: "¿Qué cree que está causando mi tos?" La Dra P. se da vuelta y responde: "Estoy seguro de que no es nada grave, probablemente una infección respiratoria superior o bronquitis. Hay algunos medicamentos que causan tos, pero no los usa. Probablemente hagamos una radiografía de tórax". Luego sale de la habitación.

Actividad Individual. Tiempo: hasta 10 minutos.

1. **Escriba el problema del paciente en una frase breve.**

Actividad Grupal (3/4 personas). Tiempo: hasta 15 minutos.

1. Comparen y **seleccionen la versión individual que mejor represente el problema de la paciente.**
En base a esa representación, discutan y enumeren:
 - A. **¿Qué hipótesis** generarían que les permitan guiar la búsqueda de información adicional?
 - B. **¿Qué datos clave adicionales** del interrogatorio y el examen clínico consideran necesarios para reforzar o debilitar esas hipótesis?

Parte 2: luego del trabajo individual y en grupos pequeños, se realizó una puesta en común en plenario, y continua con la siguiente información adicional y consigna:

Razonamiento Clínico

G. Jung

2019

TALLER MÓDULO 2: CALIFICACIÓN SEMÁNTICA

Información Adicional Sr Chan.

La tos tiene al menos 1 año de evolución, y empeoró en el último mes. El señor Chan niega haber fumado. Tiene 45 años, trabajó como administrativo en una pequeña empresa toda su vida, y vivía en una pequeña ciudad en China, antes de venir al conurbano bonaerense (Lomas de Zamora) hace 2 años. Niega contacto con otras personas con tos crónica, ni ha tenido viajes recientes. Aunque no se ha pesado, varias le han mencionado que lo ven más delgado, pero él responde no cree, porque "tiene hambre como siempre o más que antes". Niega tener acidez, reflujo, historia de broncoespasmo, ni enfermedades importantes previas dignas de mencionar. No usa esteroides, ni otras medicaciones inmunosupresoras. En alguna ocasión, recuerda haber presentado expectoración con sangre, aunque no le dió importancia.

Al examen físico lo encuentra delgado, pero luce en buen estado general. Ausculta crepitantes en la región anterior, superior derecha. La percusión pulmonar es sonora, al igual que la columna. Se palpan adenopatías cervicales inferiores duro-elásticas, ligeramente dolorosas, móviles.

Parte 2. Actividad Grupal (3/4 personas). Tiempo: hasta 20 minutos.

1. Re-escriban la síntesis del caso en una nueva frase breve, con la información adicional obtenida.
 2. ¿Cómo modificaron estos datos sus hipótesis preliminares? Ordenen las hipótesis en **orden de probabilidad**.
 3. ¿Cómo intentaría **verificar su hipótesis**? Comparen su estrategia con la de la Dra P. ¿En qué se parecen? ¿En qué se diferencian?
 4. ¿Cuáles creen que son los potenciales **errores** que creen que evitaron Uds en este caso, al realizar el interrogatorio dirigido y el examen físico basado en hipótesis?
-

Ateneos de Discusión de Casos Clínicos (reales)

Ejemplo 1: En cada una de las preguntas, se genera espacio y oportunidades para la reflexión.

ATENEO CASOS
RAZONAMIENTO CLÍNICO

Gustavo Jung Marzo 2020

Mujer de 58 años

Enfermedad Actual

- Fiebre y somnolencia de 12hs de evolución.

¿Cuáles son sus hipótesis diagnósticas? ¿Por qué?

Ordenen las hipótesis en probabilidades

- Antecedentes Personales:
 - Hipotiroidismo.
 - Pericarditis Constrictiva (2013), Cx.
 - EPOC
- Medicación Habitual:
 - Levotiroxina.
 - Amoxicilina Clavulanico 1 gramo cada 12 hs (+6), por "bronquitis aguda".

¿Cómo modifican estos datos sus hipótesis previas?

¿Por qué?

¿Qué espera encontrar en el examen físico?

¿Banderas rojas?

Examen Físico

- TA 100/60 FC 90 FR 18 T 38,0 SO2 95% 0,21
- Somnolienta, hiporeactiva.
- GCS 11/15 (M5V4O2).
- Protege la vía aérea. Buena mecánica ventilatoria.
- Rigidez de nuca en los dos planos del movimiento.
- Kernig y Brudzinski negativos.
- Resto sin hallazgos a destacar.

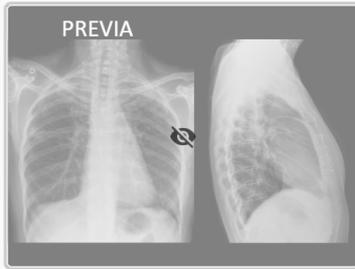
¿Cómo modificó el EF su presunción diagnóstica?

¿Estudios?

¿Tratamiento?

LABORATORIO	
Hto	33 %
GB	15.700 (66/20/14)
PlaQ	264.000/mm ³
Na	134 mEq/l
K	3,6 mEq/l
Gluc	88 mg/dl
Urea	36 mg/dl
Creat	0,85 mg/dl
Hepatograma	Normal
Calcio	9,1 mg/dl
EAB: 7.44/37/65/24.5/0.6/94% (0,21)	
Lactato	15,3 mg/dl

Orina	
Sedimento	No Inflamatorio
UC	Directo Negativo
Tóxicos	Negativo



13

TC cerebro

NORMAL

14

LCR	
Aspecto	Limpido, incoloro
Células	1
Gluc	51
Prot	0.30
Cl	118
Directo	Negativo

15

¿Cómo continuaría el proceso diagnóstico?
¿Por qué?

16

Aspecto	LCR		
	Día 0	Día 1	Día 3
Aspecto	Limpido, incoloro	Rosado, Turbio, Botón Hemático	Limpido, incoloro
Células	1	18 (>PMN)	4
Gluc	51 mg/dl	43 mg/dl	97 mg/dl
Prot	0.30 g/L	0.40 g/L	0,59
Cl	118	120	127
Directo	Negativo	Negativo	Negativo
Cultivo y PCRs	Negativo	Negativo	En curso

17

RMN cerebro

NORMAL

18



19

EEG

Actividad Epileptiforme Bilateral

20



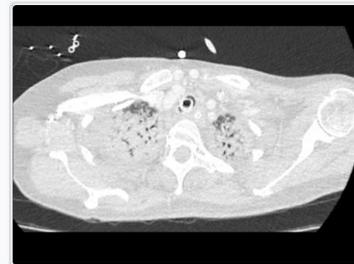
21

➡ ¿POR QUÉ?

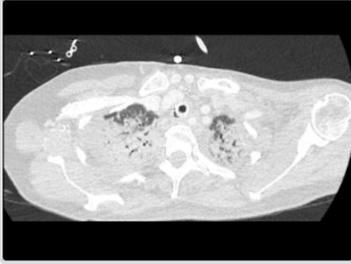
22



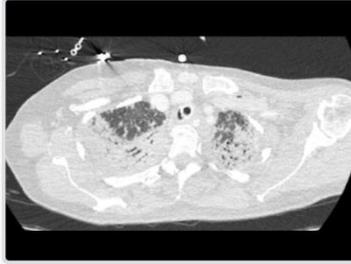
23



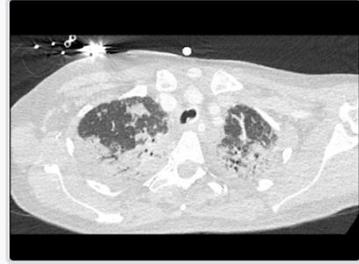
24



25



26



27



28



29



30

Apéndice III: Ejemplos de Scripts/Casos

- Se detalla un ejemplo de script utilizado en los casos de prueba.

Mujer de 47 años que consulta por astenia, adinamia, dolor abdominal leve en hipocondrio derecho e ictericia progresiva de 1 semana de evolución.

Descripción (opcional)

Ud. considera la hipótesis: hepatitis aguda *

Descripción (opcional)

Y recibe los siguientes resultados de laboratorio: Bilirrubina Total 3.4 mg/dl, Bilirrubina directa 2.1 mg/dl, GOT 456 mg/dl, GPT 678 mg/dl, FAL 167 mg/dl

Valores normales: GOT hasta 34 mg/dl; GPT hasta 55 mg/dl; FAL hasta 150 mg/dl

S2i1 - Entonces la hipótesis* se vuelve: *

1: Casi descartada o descartada; 2: Poco probable; 3: Ni más ni menos probable; 4: Más probable; 5: Casi confirmada o confirmada

	1	2	3	4	5	
Casi descartada o descartada	<input type="radio"/>	Casi confirmada o confirmada				

- Ejemplos de Scripts de las Evaluaciones:
 - Casos 1 y 13: de la evaluación pre intervención.
 - Casos 19 y 24: de la evaluación pos intervención.



Caso 1



Ítem 1/3

Se presenta en Emergencias una mujer de 36 años, con dolor abdominal agudo, continuo, a predominio en fosa iliaca derecha, no irradiado. Tiene nauseas, y no tuvo fiebre ni escalofrios. Niega flujo vaginal o dispareunia. Su fecha de última menstruación fue hace 4 semanas atrás.

Descripción (opcional)

Ud. considera solicitar una tomografía de abdomen con contraste*

Descripción (opcional)

Y obtiene una B-HCG positiva.

Descripción (opcional)

S1i1 - Ese estudio se vuelve: *

1: Contraindicado o casi contraindicado; 2: Menos Indicado; 3: Ni más ni menos indicado; 4: Más indicado; 5: Completamente indicado o casi completamente indicado

	1	2	3	4	5	
Contraindicado o casi contraindicado	<input type="radio"/>	Completamente Indicado o casi completamente indicado				

Caso 1



Item 2/3

Se presenta en Emergencias una mujer de 36 años, con dolor abdominal agudo, continuo, a predominio en fosa iliaca derecha, no irradiado. Tiene nauseas, y no tuvo fiebre ni escalofrios. Niega flujo vaginal o dispareunia. Su fecha de última menstruación fue hace 4 semanas atrás.

Descripción (opcional)

Ud. considera la hipótesis de un embarazo ectópico*

Descripción (opcional)

Y la paciente le comenta que se realizó la ligadura de trompas hace 3 años.

Descripción (opcional)

S1i2 - Entonces la hipótesis se vuelve: *

1: Casi descartada o descartada; 2: Poco probable; 3: Ni más ni menos probable; 4: Más probable; 5: Casi confirmada o confirmada

	1	2	3	4	5	
Casi descartada o descartada	<input type="radio"/>	Casi confirmada o confirmada				

Caso 1



Item 3/3



Se presenta en Emergencias una mujer de 36 años, con dolor abdominal agudo, continuo, a predominio en fosa iliaca derecha, no irradiado. Tiene nauseas, y no tuvo fiebre ni escalofrios. Niega flujo vaginal o dispareunia. Su fecha de última menstruación fue hace 4 semanas atrás.

Descripción (opcional)

Ud. considera la hipótesis de apendicitis aguda *

Descripción (opcional)

La paciente no tiene nauseas, vómitos ni hiporexia.

Descripción (opcional)

S1i3 - Entonces la hipótesis se vuelve: *

1: Casi descartada o descartada; 2: Poco probable; 3: Ni más ni menos probable; 4: Más probable; 5: Casi confirmada o confirmada

	1	2	3	4	5	
Casi descartada o descartada	<input type="radio"/>	Casi confirmada o confirmada				

Caso 13



Ítem 1/3

Mujer de 37 años que cursa el segundo día de internación por cesárea electiva con anestesia peridural, de un embarazo controlado. Intercurre con cefalea holocraneana, continua y progresiva hasta máxima intensidad, que no mejora con analgésicos ni con reposo. Los signos vitales y el examen físico no demuestran alteraciones. Se le realizó una TC de cerebro sin contraste, que resultó normal. Evoluciona con alteración progresiva del estado mental, foco motor en hemisferio izquierdo y una convulsión tónico clónica, que requiere su intubación orotraqueal

Descripción (opcional)

Ud considera la posibilidad de administrar sulfato de magnesio *

Descripción (opcional)

Y la tensión arterial es 140/90

Descripción (opcional)

S13i1 - Entonces el tratamiento se encuentra: *

1: Contraindicado o casi contraindicado; 2: Menos indicado; 3: Ni más ni menos indicado; 4: Más indicado; 5: Completamente indicado o casi completamente indicado

Contraindicada o casi contraindicada 1 2 3 4 5 Completamente indicada o casi completamente indicada

Caso 13



Ítem 2/3

Mujer de 37 años que cursa el segundo día de internación por cesárea electiva con anestesia peridural, de un embarazo controlado. Intercurre con cefalea holocraneana, continua y progresiva hasta máxima intensidad, que no mejora con analgésicos ni con reposo. Los signos vitales y el examen físico no demuestran alteraciones. Se le realizó una TC de cerebro sin contraste, que resultó normal. Evoluciona con alteración progresiva del estado mental, foco motor en hemisferio izquierdo y una convulsión tónico clónica, que requiere su intubación orotraqueal

Descripción (opcional)

Ud. considera la posibilidad de iniciar anticoagulación por sospecha de trombosis venosa central *

Descripción (opcional)

Y recibe el siguiente informe de resonancia de cerebro: edema cerebral difuso y hemorragia intracerebral focal

Descripción (opcional)

S13i2 - Entonces el tratamiento se encuentra: *

1: Contraindicado o casi contraindicado; 2: Menos indicado; 3: Ni más ni menos indicado; 4: Más indicado; 5: Completamente indicado o casi completamente indicado

	1	2	3	4	5	
Contraindicada o casi contraindicada	<input type="radio"/>	Completamente indicada o casi completamente indicada				

Caso 13



Ítem 3/3



Mujer de 37 años que cursa el segundo día de internación por cesárea electiva con anestesia peridural, de un embarazo controlado. Intercurre con cefalea holocraneana, continua y progresiva hasta máxima intensidad, que no mejora con analgésicos ni con reposo. Los signos vitales y el examen físico no demuestran alteraciones. Se le realizó una TC de cerebro sin contraste, que resultó normal. Evoluciona con alteración progresiva del estado mental, foco motor en hemisferio izquierdo y una convulsión tónico clónica, que requiere su intubación orotraqueal

Descripción (opcional)

Ud. considera la necesidad de realizar una punción lumbar *

Descripción (opcional)

Y la paciente resulta ser HIV positiva, con CD4 = 250, carga viral indetectable

Descripción (opcional)

S13i3 - Entonces la conducta se encuentra: *

1: Contraindicada o casi contraindicada; 2: Menos indicada; 3: Ni más ni menos indicada; 4: Más indicada; 5: Completamente indicada o casi completamente indicada

Contraindicada o casi contraindicada 1 2 3 4 5 Completamente indicada o casi completamente indicada

Caso 19

Ítem 1/3

Una mujer de 55 años se encontraba sentada trabajando sentada, cuando comenzó bruscamente con palpitaciones, debilidad, sudoración, náuseas, visión borrosa, y pérdida de la conciencia. Se recuperó completamente en pocos segundos, pero persiste con palpitaciones, motivo por el cual consulta. Es tabaquista actual, hipotiroidea y tiene ataques de pánico ocasionales. Su medicación incluye levotiroxina 88 µg/día, clonazepam 0,5 mg por día. A su ingreso se encuentra normotensa, afebril, y presenta el siguiente ECG:



Ud. considera realizar: cardioversión farmacológica*

Descripción (opcional)

Y recibe laboratorio: K⁺ 3,0 mEq/dL

K⁺ significa: kalemia sérica

S19it - Entonces la conducta* se encuentra: *

1: Contraindicada o casi descartada; 2: Menos indicada; 3: Ni más ni menos indicada; 4: Más indicada; 5: Completamente indicada o casi completamente indicada

	1	2	3	4	5	
Contraindicada o casi descartada	<input type="radio"/>	Completamente indicada o casi completamente indicada				

Caso 19

Ítem 2/3

Una mujer de 55 años se encontraba sentada trabajando sentada, cuando comenzó bruscamente con palpitaciones, debilidad, sudoración, náuseas, visión borrosa, y pérdida de la conciencia. Se recuperó completamente en pocos segundos, pero persiste con palpitaciones, motivo por el cual consulta. Es tabaquista actual, hipotiroidea y tiene ataques de pánico ocasionales. Su medicación incluye levotiroxina 88 µg/día, clonazepam 0,5 mg por día. A su ingreso se encuentra normotensa, afebril, y presenta el siguiente ECG:



Ud. considera realizar: cardioversión eléctrica*

Descripción (opcional)

Y la paciente comienza con dolor precordial opresivo irradiado a cuello.

Descripción (opcional)

S192 - Entonces la conducta* se encuentra: *

1: Contraindicada o casi descartada; 2: Menos indicada; 3: Ni más ni menos indicada; 4: Más indicada; 5: Completamente indicada o casi completamente indicada

Contraindicada o casi descartada 1 2 3 4 5 Completamente indicada o casi completamente indicada

Caso 19



Ítem 3/3

Una mujer de 55 años se encontraba sentada trabajando sentada, cuando comenzó bruscamente con palpitaciones, debilidad, sudoración, náuseas, visión borrosa, y pérdida de la conciencia. Se recuperó completamente en pocos segundos, pero persiste con palpitaciones, motivo por el cual consulta. Es tabaquista actual, hipotiroidea y tiene ataques de pánico ocasionales. Su medicación incluye levotiroxina 88 µg/día, clonazepam 0,5 mg por día. A su ingreso se encuentra normotensa, afebril, y presenta el siguiente ECG:



Ud. considera administrar: bisoprolol 5 mg via oral*

Descripción (opcional)

Y la paciente recuerda episodios previos de palpitaciones, pero sin pérdida de la conciencia.

Descripción (opcional)

S19i3 - Entonces la conducta* se encuentra: *

1: Contraindicada o casi descartada; 2: Menos indicada; 3: Ni más ni menos indicada; 4: Más indicada; 5: Completamente indicada o casi completamente indicada

	1	2	3	4	5	
Contraindicada o casi descartada	<input type="radio"/>	Completamente indicada o casi completamente indicada				

Caso 24

Ítem 1/3

Julia tiene 75 años y está jubilada. Consulta por "dolor en la parte baja de la espalda" desde hace un mes. No recuerda haber tenido ningún golpe. Nota que "el dolor está presente todo el tiempo, pero es peor a la mañana, y cuando estoy quieta por mucho tiempo. No aguanto estar acostada, y no me deja dormir." Toma tamoxifeno desde hace 4 años por un adenocarcinoma de mama que requirió cuadrantectomía y radioterapia localizada.

Descripción (opcional)

Ud. considera solicitar una tomografía de columna lumbar *

Descripción (opcional)

Y el dolor reproduce a la palpación de las apófisis espinosas L3 y L4.

Descripción (opcional)

S24i1 - Entonces el estudio* se encuentran: *

1: Contraindicado o casi descartado; 2: Menos Indicado; 3: Ni más ni menos indicado; 4: Más indicado; 5: Completamente indicado o casi completamente indicado

1 2 3 4 5

Contraindicado o casi descartado Completamente Indicados o casi completamente indicados

Caso 24

Ítem 2/3

Julia tiene 75 años y está jubilada. Consulta por "dolor en la parte baja de la espalda" desde hace un mes. No recuerda haber tenido ningún golpe. Nota que "el dolor está presente todo el tiempo, pero es peor a la mañana, y cuando estoy quieta por mucho tiempo. No aguanto estar acostada, y no me deja dormir." Toma tamoxifeno desde hace 4 años por un adenocarcinoma de mama que requirió cuadrantectomía y radioterapia localizada.

Descripción (opcional)

Ud. considera la hipótesis: metástasis ósea *

Descripción (opcional)

Y la fosfatasa alcalina sérica (FAL) es normal.

Descripción (opcional)

S24i2- Entonces la hipótesis* se vuelve: *

1: Casi descartada o descartada; 2: Poco probable; 3: Ni más ni menos probable; 4: Más probable; 5: Casi confirmada o confirmada

1 2 3 4 5

Casi descartada o descartada Casi confirmada o confirmada

Caso 24

Ítem 3/3

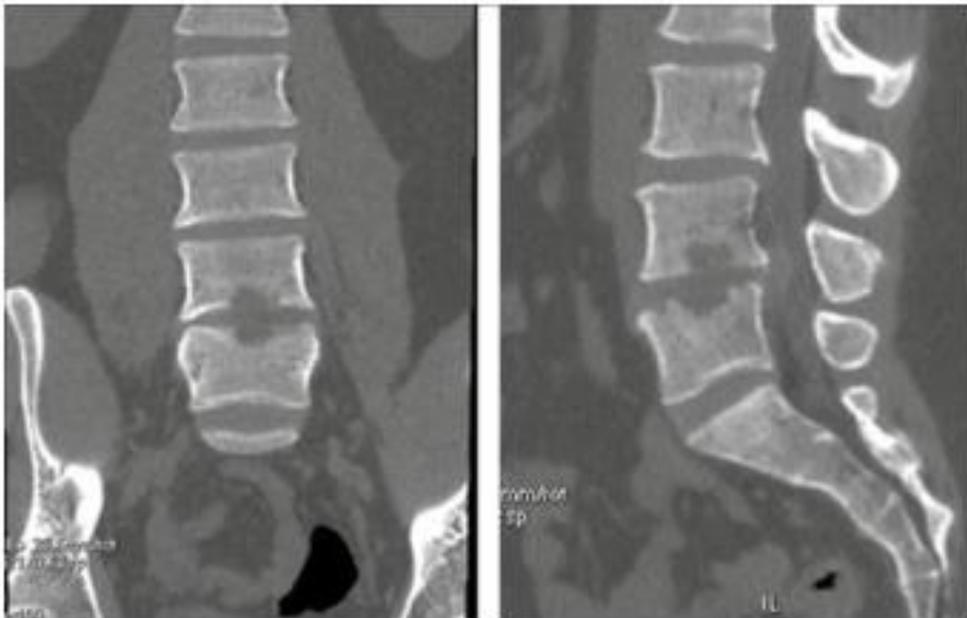
Julia tiene 75 años y está jubilada. Consulta por "dolor en la parte baja de la espalda" desde hace un mes. No recuerda haber tenido ningún golpe. Nota que "el dolor está presente todo el tiempo, pero es peor a la mañana, y cuando estoy quieta por mucho tiempo. No aguanto estar acostada, y no me deja dormir." Toma tamoxifeno desde hace 4 años por un adenocarcinoma de mama que requirió cuadrantectomía y radioterapia localizada.

Descripción (opcional)

Ud. considera la hipótesis: fractura patológica *

Descripción (opcional)

Y recibe el siguiente estudio de la paciente:



S24i3- Entonces la hipótesis* se vuelve: *

1: Casi descartada o descartada; 2: Poco probable; 3: Ni más ni menos probable; 4: Más probable; 5: Casi confirmada o confirmada

	1	2	3	4	5	
Casi descartada o descartada	<input type="radio"/>	Casi confirmada o confirmada				

Apéndice IV: Resultados individuales de SCT

Se informan a continuación los resultados de todos los participantes, de acuerdo al análisis basado en scripts.

Panel de Expertos: Resultados individuales de los SCT.

PANEL DE EXPERTOS	PRE	POST	MEDIA
E01	76,6	69,6	73,1
E02	74,3	78,1	76,2
E03	70,6	65,3	68,0
E04	71,7	72,4	72,1
E05	74,1	73,6	73,9
E06	71,8	75,3	73,6
E07	78,7	64,9	71,8
E08	75,3	79,2	77,3
E09	67,7	83,6	75,7
E10	65,5	69,9	67,7
E11	73,8	74,1	74,0
E12	55	58,1	56,6
MEDIA	71,3	72,0	71,6
DE	6,3	7,0	

DE = desvío estándar

Grupo Control: Resultados individuales de los SCT.

GRUPO CONTROL	PRE	POST	MEDIA
C01	60,0	52,7	56,3
C02	64,2	59,1	61,6
C03	58,2	51,7	55,0
C04	66,4	65,0	65,7
C05	61,8	68,0	64,9
C06	58,1	48,8	53,5
C07	55,0	50,5	52,8
C08	67,1	60,1	63,6
C09	69,5	66,1	67,8
C10	61,6	57,0	59,3
C11	76,9	57,1	67,0
C12	61,4	60,1	60,8
C13	68,2	51,8	60,0
C14	61,4	65,7	63,6
C15	65,1	53,1	59,1
C16	61,6	59,3	60,5

C17	60,0	63,1	61,5
C18	65,2	57,8	61,5
C19	64,2	67,1	65,6
C20	74,7	67,2	70,9
C21	-	61,9	61,9
MEDIA	64,0	59,2	62,6
DE	5,4	6,1	

Nota: El C21 pre intervención es el único dato faltante, por error de registro electrónico.

DE = desvío estándar

Grupo Intervención: Resultados individuales de los SCT.

GRUPO INTERVENCIÓN	PRE	POST	MEDIA
I01	72,8	67,2	70,0
I02	73,3	77,2	75,3
I03	69,0	61,9	65,5
I04	58,7	53,1	55,9
I05	72,8	62,3	67,6
I06	69,5	63,6	66,5
I07	61,8	50,7	56,3
I08	62,9	49,7	56,3
I09	69,6	61,7	65,6
I10	65,6	66,9	66,3
I11	65,6	56,4	61,0
I12	52,3	44,4	48,4
I13	60,0	62,2	61,1
I14	46,5	47,2	46,8
I15	60,8	58,9	59,8
I16	60,6	58,2	59,4
I17	61,2	58,1	59,7
I18	48,7	49,0	48,9
I19	44,7	60,1	52,4
I20	60,5	52,8	56,7
I21	42,2	49,1	45,7
MEDIA	60,9	57,7	55,5
DE	9,3	8,0	

DE = desvío estándar