TROVARE
Repositorio Institucional
Instituto Universitario y Hospital Italiano de Buenos Aires

Giannasi, Sergio Eduardo

Adaptación transcultural y validación del Instrumento para la Evaluación del Desempeño del Docente Clínico de Maastricht

Maestría en Educación para Profesionales de la Salud

Tesis 2014

Cita sugerida (Vancouver): *Giannasi SE. Adaptación transcultural y validación del Instrumento para la Evaluación del Desempeño del Docente Clínico de Maastricht. [tesis de maestría] [Internet]. [Buenos Aires]: Instituto Universitario Hospital Italiano; 2014 [citado AAAA MM DD]. [viii]; 53; [xxi] p. Disponible en: http://trovare.hospitalitaliano.org.ar/descargas/tesisytr/20160203120232/tesis-giannasi-sergio.pdf*

HOSPITAL ITALIANO de Buenos Aires

Este documento integra la colección Tesis y trabajos finales de Trovare Repositorio Institucional del Instituto Universitario Hospital Italiano 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/

Esta obra está bajo licencia 2.5 de Creative Commons Argentina
Acceda a la licencia completa:
http://creativecommong.org/licenses/legal.htm

Adaptación Transcultural y Validación del Instrumento para la Evaluación del desempeño del Docente Clínico de Maastricht

por

Sergio Eduardo Giannasi

Tutor: Dr. Eduardo Durante

Dirigida por Dr. Carlos Brailovsky

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

ante el

Instituto Universitario Escuela de Medicina Hospital Italiano de Buenos Aires

Buenos Aires
Octubre 2014

© Copyright Sergio E. Giannasi, 2014

Agradecimientos:

- ✓ Al Prof. Dr. Eduardo Durante. Subjefe Servicio de Medicina Familiar HIBA. Director de la maestría en Educación para Profesionales de la Salud del Instituto Universitario Escuela de Medicina del HIBA. Por ser mí maestro en la docencia clínica y por haberme ayudado con sobrada paciencia en cada uno de los pasos de esta tesis.

- ✓ Al Prof. Dr. Carlos Brailovsky. Co-Director de la maestría en Educación para Profesionales de la Salud del Instituto Universitario Escuela de Medicina del HIBA. Por aceptar ser el director mi tesis, por ser un crítico justo y favorecedor de mi trabajo y por contribuir de modo fundamental en el análisis G. de los resultados.

- ✓ Al Dr. Fernando Ramón Vázquez. Médico del servicio de Medicina Familiar HIBA. Por su contribución cardinal en el análisis de los resultados, especialmente en su área de conocimiento ¨Análisis Factorial Confirmatorio¨.

- ✓ A mi familia, especialmente a Marcela, mi esposa. Por el apoyo que me brindaron para completar la maestría y esta tesis. No solo me regalaron el tiempo para trabajar, sino que me hicieron sentir feliz mientras lo hacía.

- ✓ A la Dra. Kecskes Claudia. Subjefa Servicio Terapia Intensiva Adultos HIBA. y a la Dra. Roberta Ladenheim. Médica del Servicio de Clínica Médica CEMIC, con quien hicimos en forma conjunta la adaptación transcultural del instrumento y el reclutamiento de los cuestionarios.

- ✓ Al Dr. Eduardo San Román. Jefe del servicio de Terapia Intensiva Adultos HIBA. Por apuntalar todos mis proyectos docentes dentro y fuera del servicio y por beneficiarme con tiempo protegido para la realización de esta investigación.

- ✓ A mis compañeros del CIEM HIBA: El Dr. Eduardo Durante, el Dr. Carlos Fustiñana, el Dr. Alfredo Eymann, el Dr. Marcelo Figari y el Dr. Gustavo Maid y, por los aportes invalorables que hicieron durante la validación de contenido y por las opiniones y críticas que me ayudaron a mejorar este trabajo.

- ✓ A las autoridades del Instituto Universitario Escuela de Medicina del HIBA, especialmente a los Dr. Marcelo Figari y Dr. Diego Fainglod, por favorecer el marco de excelencia que me permite y estimula para continuar mi formación y por facilitarme los medios para trabajar en docencia y difundir esta investigación.

- ✓ A los residentes del HIBA y del CEMIC que respondieron el cuestionario, dedicando el tiempo y la paciencia necesaria que esta tarea requirió. Sin su apoyo este trabajo hubiese sido imposible de concretar.

- ✓ A todos los docentes clínicos del HIBA y del CEMIC que en todo momento se prestaron en forma desinteresada y sin ningún tipo de condicionamientos a ser evaluados en su desempeño docente.

- ✓ A los docentes de la MEPS IUHIBA por sus excelentes cualidades profesionales, y por haber generado un clima de aprendizaje excelente para que podamos discutir, reflexionar y aprender los aspectos relevantes de la educación superior.

Introducción

La docencia clínica es reconocida a lo largo de la historia de la medicina como un componente fundamental de la formación médica. El esquema tradicional del maestro con su aprendiz, es una de las estrategias más usadas por los docentes para que los estudiantes aprendan en el ámbito clínico.

Sin embargo, para ser un buen docente clínico (DC), el médico, debe desarrollar otras competencias, que exceden el hecho de ser ¨un modelo de médico¨. Un buen DC debe ser además, un facilitador de la adquisición de las competencias del alumno, un buen supervisor de sus actividades y promover el desarrollo de un profesionalismo adecuado a cada especialidad.

Hoy en día son múltiples los factores que atentan para que la docencia clínica se desarrolle de manera adecuada. Ponen en riesgo la calidad de este proceso, la escasa formación formal en educación que tienen los supervisores de residentes, la falta de tiempo y el exceso de actividades extras que se suman a las ya difíciles tareas asistenciales y docentes.

Contar con una estrategia docente, que sea a la vez práctica y sustentada por una teoría educativa, es primordial para el desarrollo de esta actividad. Las investigaciones en educación clínica indican que el aprendizaje depende en gran medida de conductas de enseñanza que favorezcan la participación activa de los estudiantes y, de la atención puesta en los procesos cognitivos que subyacen a la tarea, tanto del profesor como del alumno.

Basados en un enfoque constructivista del aprendizaje del alumno y de los métodos de enseñanza del docente, Collins, Brown, y Newman (1989) desarrollaron la teoría del “aprendizaje cognitivo”. Este enfoque resume e integra varias de las fortalezas de las teorías en educación médica, en un solo modelo. Los métodos de enseñanza propuestos en esta teoría (modelado, coaching, andamiaje, articulación, reflexión y exploración), facilitan el aprendizaje de los alumnos desde las etapas iniciales, hasta avanzada la formación, cuando se espera que los residentes sean competentes para resolver problemas en forma independiente, autoevaluarse y formular sus propios objetivos de aprendizaje.

El DC además de estar capacitado en la enseñanza clínica siguiendo los principios de esta estrategia, requiere ser evaluado en forma continua. La evaluación y posterior feedback acerca de su trabajo docente, le permitiría mejorar en aquellas áreas específicas donde su desempeño es menor.
Para poder hacer una correcta evaluación se requiere de un instrumento que sea válido y confiable, y al mismo tiempo basado en una teoría educativa sólida. En la Universidad de Maastricht la Dra. Stalmeijer y colaboradores, desarrollaron y validaron un instrumento de evaluación de desempeño del DC (2008, 2010) basado en la teoría cognitiva del aprendizaje. En nuestro idioma no existe una herramienta de evaluación del DC que sea mismo tiempo válida, confiable y basada en la mencionada teoría, para usar con nuestros docentes.

Esta tesis, trata del trabajo llevado a cabo durante la validación transcultural del Instrumento de Evaluación del Desempeño Docente Clínico de Maastricht, desde la traducción del cuestionario original, hasta la valoración de sus propiedades psicométricas.

Resumen:

Introducción:
La evaluación del desempeño docente, permite a los mismos recibir feedback que los ayude a mejorar. Stalmeijer y col. (2010) desarrollaron un instrumento de evaluación de desempeño del Docente Clínico (IEDDC), construido sobre la base de la teoría del aprendizaje cognitivo (AC), que es válido y confiable. Para poder ser utilizado en otras poblaciones se requiere de una correcta adaptación y validación de sus propiedades psicométricas.

Objetivo:
Obtener una versión válida del IEDDC en español y Aportar evidencia de validez y determinar la confiabilidad del IEDDC, para que los residentes evalúen docentes durante sus rotaciones clínicas.

Material y Método:
Se realizó una adaptación transcultural siguiendo la Guía Internacional para la Adaptación de los Cuestionarios (I.T.Commission 2005).
Se investigaron 4 fuentes de validez. 1)Contenido: expertos evaluaron la pertinencia de los ítems y dominios con la teoría del AC. 2)Proceso de respuesta: se analizó la comprensión de cada uno de los ítems mediante entrevistas con residentes hasta saturar la muestra. 3)Constructo: mediante un Análisis Factorial Confirmatorio. 4)Confiabilidad: mediante un alfa de Cronbach y un estudio G. A través de una optimización de resultados del estudio G, se estimó la cantidad de observadores necesarios para una correcta evaluación del DC, para cada uno de los dominios y para el juicio global; y se determinaron las fuentes de error absoluto para una correcta generalización de los resultados.

Resultados:
Luego de realizar la adaptación transcultural, 1031 residentes de dos hospitales universitarios de la CABA, evaluaron a 187 DC.
La validación de contenido y del proceso de respuesta fue correcta y, realizada por expertos en docencia y residentes.
Constructo: el AFC realizado con una muestra de 151 DC con al menos cuatro evaluaciones cada uno, demostró que todos los índices y criterios de buen ajuste son correctos con un modelo de 5 factores latentes y 14 ítems.
Confiablidad: el alfa de Cronbach fue de 0.80. El estudio G determinó índices de confiablidad elevados y con bajos valores de varianza del error absoluto. Estudio D determinó que son suficientes entre 4 y 6 evaluaciones para un feedback al DC, respecto de cada dominio y del juicio global.

Conclusión:
El IEDDC parece ser un instrumento válido y confiable para que los residentes evalúen el desempeño de los DC. Los resultados pueden ser generalizables a DC de hospitales universitarios de Argentina en los cuales los residentes realizan rotaciones clínicas

Palabras Clave:
1. En español:
    a. Aprendizaje
    b. Argentina
    c. Cognición
    d. Cuestionarios / * normas
    e. Educación, Médica, Postgrado / *estándares
    f. Enseñanza / * normas
    g. Humanos
    h. Modelos, Educativo
    i. Profesores/Docentes, clínicos / * estándares
    j. Proyectos Piloto
    k. Reproducibilidad de Resultados
    l. Rotaciones clínicas
    m. Validación transcultural de cuestionarios.
    n. Validez, confiabilidad

2. En Inglés:
    a. Learning
    b. Argentine
    c. Cognition
    d. Questionnaires/*standards
    e. Education, Medical, Postgraduate/*standards
    f. Teaching/*standards
    g. Humans
    h. Models, Educational
    i. Faculty, Medical/*standards
    j. Pilot Projects
    k. Reproducibility of Results
    l. Clinical Clerkship
    m. Cross-cultural validation /questionnaires.
    n. Validity, reliability

Contents

FUNDAMENTACIÓN: 1

OBJETIVO DEL ESTUDIO 16

REALIZAR UNA VALIDACIÓN TRANSCULTURAL DEL INSTRUMENTO DE EVALUACIÓN DEL DESEMPEÑO DEL DOCENTE CLÍNICO DE MAASTRICHT AL IDIOMA ESPAÑOL. 16
OBJETIVOS ESPECÍFICOS:..... 16
✓ OBTENER UNA VERSIÓN VÁLIDA DEL IEDDC EN ESPAÑOL 16
✓ APORTAR EVIDENCIA PARA LA VALIDEZ Y DETERMINAR LA CONFIABILIDAD. 16

MATERIALES Y MÉTODO. 16

- DE LA ADAPTACIÓN TRANSCULTURAL DEL INSTRUMENTO. 16
- DEL APORTE DE EVIDENCIA SOBRE LA VALIDEZ Y DETERMINACIÓN DE LA CONFIABILIDAD DEL INSTRUMENTO. 17
  1. *Evidencia basada en el Contenido del test:* *17*
  2. *Evidencia basada en el Proceso de Respuesta:* *18*
  3. *Evidencia basada en la estructura interna del test y confiabilidad:* *18*
  4. *Evidencia basada en la Relaciones con otras variables:* *19*
- RESPECTO DE LAS CARACTERÍSTICAS DEL INSTRUMENTO DE EVALUACIÓN. 19
- CON RELACIÓN A LOS MÉDICOS RESIDENTES (EVALUADORES): 20
- CARACTERÍSTICAS DE LOS DOCENTES (EVALUADOS): 20
- RESPECTO DE LA PRESERVACIÓN DE LA INTEGRIDAD DEL ANONIMATO..... 21
- DETERMINACIÓN DE N DE EVALUACIONES. 21
- **C**OMITÉ DE INVESTIGACIÓN EN EDUCACIÓN MÉDICA (CIEM)**:** 22
- PRUEBAS ESTADÍSTICAS: 22
- COMITÉ DE ÉTICA DE PROTOCOLOS DE INVESTIGACIÓN: 23

RESULTADOS: 24

I. **OBJETIVO 1: OBTENER UNA VERSIÓN VÁLIDA DEL IEDDC EN ESPAÑOL:** 24
- *Traducción Directa*.....*24*
- *Traducción inversa o Retrotraducción (del español al inglés).*.....*25*

II. OBJETIVO 2: APORTE DE EVIDENCIA SOBRE LA VALIDEZ Y DETERMINACIÓN DE LA CONFIABILIDAD DEL IEDDC EN NUESTRO MEDIO. 26
1. *Aporte de evidencia para la validez de contenido:*.....*26*
2. *Aporte de evidencia del proceso de respuesta*.....*26*
- *Características de la administración de la prueba:* *27*
- *Recolección de la muestra:* *27*

3. *Aporte de validez basada en la estructura interna del test y de la confiabilidad.*.....*28*
- *Validez de constructo mediante un análisis factorial confirmatorio.*.....*28*
- *Análisis de las relaciones causales mediante el modelo de ecuaciones estructurales.* *31*

Determinación de la confiabilidad.....*33*
- *Consistencia Interna - Cálculo del coeficiente Alfa de Cronbach.* *33*
- *Análisis de Generalizabilidad.* *36*

4. Evidencia basada en la Relaciones con otras variables. Análisis de las correlaciones entre de cada uno de los factores y el juicio sobre el desempeño global 40

DISCUSIÓN 42

APÉNDICES

BIBLIOGRAFÍA

Tabla 1. Resultados del análisis factorial confirmatorio del Instrumento de evaluación Original de Maastricht con uno a cinco factores latentes. 13
Tabla 2. Componentes de la varianza del cuestionario IEDDC para cada uno de los factores y para el Juicio global * 14
Tabla 3. Guía orientadora para el proceso de adaptación transcultural de instrumentos 17
Tabla 4. Recomendaciones sobre el tamaño muestral mínimo cuando hay Normalidad Multivariante 22
Tabla 5. Test de valoración de la normalidad multivariada para datos continuos 28
Tabla 6. Índices de ajuste del Análisis Factorial Confirmatorio con 1, 3 y 5 factores latentes (método de estimación: DWLS). 29
Tabla 7. Índices de ajuste del Análisis Factorial Confirmatorio con modelos alternativos (método de estimación ULS) 31
Tabla 8. Resultados de ajuste del AFC y MEE 31
Tabla 9. Análisis de la confiablidad de la muestra mediante el Alfa de Cronbach para cada uno de los factores 33
Tabla 10. Cronbach: Estadísticos de todos los elementos. Muestra del HIBA..... 34
Tabla 11. Cronbach: Estadísticos de todos los elementos. Muestra del CEMIC 35
Tabla 12. Tabla del Estudio G correspondiente al Diseño de Medición: D/OI..... 37
Tabla 13. Diferencias en los resultados del estudio G al considerar el modelo de 14 o 15 ítems. 39
Tabla 14. Coeficientes G absolutos para cada uno de los Dominios y el Desempeño global ... 40
Tabla 15. Correlaciones entre cada uno de los factores y la valoración global..... 41

Fundamentación:

La enseñanza de la medicina fuera del aula y especialmente, junto a la cama del paciente ha sido reconocida a lo largo de la historia de la medicina como un componente fundamental de la formación médica (Whitman and Schwenk 1997). El esquema tradicional del maestro con su aprendiz fue por mucho tiempo la estrategia usada por los docentes para que los estudiantes aprendan en el ámbito clínico. Aquí, el médico en formación aprende siguiendo e imitando a un ¨modelo de médico¨ durante su práctica asistencial.

Si bien el modelado es una estrategia eficaz para la transmisión de valores, actitudes, patrones de pensamiento y comportamiento a los estudiantes y residentes (Elzubeir and Rizk 2001), por si solo es insuficiente. Hoy en día, además del modelado, se reconocen como importante: la facilitación de la adquisición de habilidades clínicas, de la comunicación y de la enseñanza de aspectos humanísticos, del autoaprendizaje y de la autoevaluación, entre otras. (LaCombe 1997; Thibault 1997; Ramani, Orlander et al. 2003)

Tradicionalmente, se espera que los profesionales de la salud que asisten pacientes sean capaces de supervisar la adquisición de habilidades clínicas de los estudiantes y residentes (Kilminster and Jolly 2000). Se entiende como supervisión del estudiante de medicina al *"seguimiento, orientación y feedback en cuestiones de desarrollo personal, profesional y educativo, en el contexto de la atención de los pacientes"* (Kilminster and Jolly 2000). Hay investigaciones que destacan los beneficios de la supervisión durante las rotaciones clínicas de los estudiantes de medicina (Dolmans, Wolfhagen et al. 2002; Alhaqwi, Kuntze et al. 2014).

En la formación de posgrado, las residencias médicas representan un modelo educativo muy depurado de la docencia clínica, en la cual la práctica guiada y la supervisión son las estrategias centrales para la formación (Morris 2010). Sin embargo, a pesar de su importancia, la supervisión clínica de buena calidad es poco frecuente (Grant, Kilminster et al. 2003). Hay dos posibles explicaciones para este fenómeno: por un lado, la escasez de tiempo disponible para enseñar y, por otro la pobre formación docente de los médicos. Ambas situaciones se

explican porque la actividad principal de los hospitales no es la enseñanza clínica sino la atención de los pacientes (Seabrook 2003; Hoffman and Donaldson 2004).

Las presiones provocadas por la falta de tiempo hacen que la asistencia, la investigación, la administración y la docencia compitan entre sí, conspirando para que un hospital pueda ser el ambiente ideal de aprendizaje altamente estructurado y complejo (Spencer 2003). Esto significa que las experiencias de aprendizaje de los estudiantes, sean alumnos o residentes, están determinados en gran medida por los acontecimientos del día a día y no por consideraciones pedagógicas (Deketelaere, Kelchtermans et al. 2006).

La problemática específica de los residentes y de los docentes merece una consideración especial (Shankel and Mazzaferri 1986). Como consecuencia de la aparición de leyes laborales que disminuyeron ostensiblemente la duración de la jornada laboral de los residentes[1], el tiempo para la formación es hoy en día más escaso y por lo tanto requiere de los supervisores una competencia docente aún mayor (Stanley 1998). Para facilitar la formación de los residentes en un tiempo limitado, David Hargreaves propone que las estrategias docentes se basen en el acompañamiento planificado, intencional, permanente y controlado (Hargreaves D.H. 1997).

Además, y posiblemente, la causa más importante para la falta de supervisión efectiva es la escasa o nula preparación formal que la mayoría de los médicos tienen en el ejercicio de la función de docentes clínicos(DC) (Cottrell, Kilminster et al. 2002). ¨*La tarea de enseñar a otros profesionales es, en sí misma, compleja y a menudo, vivida como una carga muy exigente y frustrante, sobre todo cuando se asume sin una adecuada preparación*¨ (McLeod 1986). Sara Frisch, una reconocida educadora afirma que "*sin una debida formación de los*

[1] En EEUU el Consejo de Acreditación para la Educación Médica de Posgrado (Accreditation Council for Graduate Medical Education) normatizó en Julio del 2003 el horario laboral del médico residente, limitándolo a 80 horas semanales. Además se estableció el horario de la jornada laboral diaria, de las guardias y los intervalos de descanso; todo con el objetivo de garantizar una atención más segura y eficiente del paciente y mejorar el aprendizaje del residente. Del mismo modo, la Regulación Europea del Tiempo del Trabajo médico (European Working Time Directive) y el Programa de Modernización de la Carrera Médica (Modernising Medical Careers) limitó el número de horas de trabajo a 48 horas por semana a partir del año 2009. En nuestro país, no existe una ley nacional regulatoria respecto a las horas de trabajo del residente, sin embargo algunas provincias como la de Buenos Aires, San Luis, la CABA y colegios médicos provinciales tienen regulaciones que van desde las 48 hs a las 64 hs semanales.

coordinadores, una recorrida de sala puede conllevar una carga aplastante y una pérdida de tiempo valioso para el docente, el paciente y los alumnos" (Frisch, Boucher et al. 1984).

El éxito de la educación clínica depende de un enfoque integral que incluye: un plan de estudios estructurado con actividades planificadas, la capacitación y capacidad de liderazgo del docente y la motivación de los estudiantes y docentes (Pathak, Pathak et al. 2000). En una investigación realizada en el Hospital General de Montreal, los alumnos consideraron que un buen DC debía planificar la tarea de los alumnos, identificar las metas individuales de cada uno de ellos, demostrar empatía con sus necesidades, promover el aprendizaje en todos los niveles y no ser meros expositores sino facilitadores del aprendizaje (McLeod 1986).

Otros autores afirman que el DC debiera ser, además de una fuente de información para el alumno, un facilitador de la resolución de los problemas del paciente con las fuentes de información a su alcance (Rees 1987).

Las investigaciones en educación clínica indican que el aprendizaje depende, en gran medida, de conductas de enseñanza que favorezcan la participación activa de los estudiantes y de la atención puesta en los procesos cognitivos que subyacen a la tarea, tanto del profesor como del alumno(Ericsson 2004; Moulaert, Verwijnen et al. 2004). Según Weinstein y Mayer, estas conductas de enseñanza representan *"secuencias integradas de procedimientos o actividades que son escogidas por el aprendiz para facilitar la adquisición, almacenamiento y recuerdo de la información"* (Weinstein and Mayer 1985).

Además, las estrategias de la enseñanza y del aprendizaje interactúan en un contexto determinado y si bien una determinada forma de enseñanza no necesariamente conduce a un tipo particular de aprendizaje, el modelo teórico docente subyacente tiene un impacto significativo en el aprendizaje. Knowles definió a una teoría educativa como *¨un sistema integral, coherente, e internamente consistente de ideas acerca de un conjunto de fenómenos¨* (Knowles M. 1973). Siguiendo los lineamientos de esta definición, Kaufman y Mann resumieron las evidencias de la educación médica en ocho enfoques teóricos(Kaufman D.M. and Mann K.V. 2010); estos son: ¨los principios del

aprendizaje en adultos¨(Knowles MS 1980), ¨la teoría cognitiva social¨ (Bandura A 1986), ¨la práctica reflexiva¨ (Schön DA 1983), ¨el aprendizaje transformativo¨ (Mezirow J 1994), ¨el aprendizaje auto dirigido¨ (Candy PC 1991), ¨el aprendizaje experiencial¨ (Kolb DA 1984), ¨el aprendizaje situado¨(Lave J. and Wenger E. 1991) y ¨el aprendizaje en comunidades de práctica¨ (Wenger E 1998).

El enfoque del aprendizaje cognitivo resume e integra varias de las fortalezas de esos enfoques en un solo modelo. En este mismo sentido y basado en un enfoque constructivista del aprendizaje y de los métodos de enseñanza Collins, Brown, y Newman (1989) desarrollaron la teoría del "aprendizaje cognitivo" (del inglés: *Cognitive Apprenticeship*).

Aquí es importante definir en forma clara el significado del término ¨aprendizaje¨ en nuestro idioma y ¨apprenticeship¨ del inglés.

En el idioma español, una de las acepciones de la expresión ¨aprendizaje¨ se utiliza para describir una forma de aprender en el campo laboral, en donde un maestro enseña en forma práctica y progresiva una determinada tarea[2]. Sin embargo, el uso más habitual es indistinto de ¨aprender¨ y no se refiere en forma específica a un determinado modelo educativo. En cambio, el término en ingles *¨apprenticeship¨,* define en forma concreta y específica un modo particular de aprender en el campo laboral[3].

---

[2] aprendizaje. (RAE)
(De aprendiz).

- ✓ m. Acción y efecto de aprender algún arte, oficio u otra cosa.
- ✓ m. Tiempo que en ello se emplea.
- ✓ m. Psicol. Adquisición por la práctica de una conducta duradera.

[3] Apprenticeship. (Logman)
The job of being an apprentice, or the period of time in which you are an apprentice.

Apprenticeship. (The Free Dictionary)
ap•pren•tice (ə-prĕn′tĭs)
n.1. One bound by legal agreement to work for another for a specific amount of time in return for instruction in a trade, art, or business.
2. One who is learning a trade or occupation, especially as a member of a labor union.
3. A beginner; a learner.
tr.v. ap•pren•ticed, ap•pren•tic•ing, ap•pren•tic•es
To place or take on as a beginner or learner

En su forma más simple, un aprendiz es alguien a quien se le paga por un determinado trabajo mientras lo aprende. A través de la historia la relación entre el empleador y el aprendiz se ha complejizado, proporcionado al aprendiz la oportunidad de aprender su oficio a través de la práctica supervisada (Souster V and Jackson N 2003). Antes de la aparición de la escolarización, el ¨apprenticeship¨ era el método más común de enseñanza; se utilizó para transmitir los saberes necesarios para la práctica de los expertos desde en los campos de la pintura y la escultura, hasta la medicina y el derecho. Incluso hoy en día muchas habilidades complejas e importantes se aprenden con métodos de enseñanza no didácticos del tipo ¨apprenticeship¨ como son la observación, el entrenamiento y el acercamiento progresivo a la actividad del experto, mediante la realización de una variedad de tareas de complejidad creciente.

Una de las diferencias más importantes entre la educación formal impartida en las escuelas y los métodos de enseñanza del ¨apprenticeship¨ es que las habilidades y el conocimiento impartido en las escuelas no tienen una relación directa con la práctica. En la enseñanza del tipo ¨apprenticeship¨ las habilidades no sólo son continuamente utilizadas por los profesionales calificados, sino que también son un instrumento fundamental para el logro tareas significativas. El aprendiz ve en forma tangible lo que debe aprender a ¨hacer¨, ve lo que sucede con su trabajo y sus progresos (Souster V and Jackson N 2003). En cambio, en las escuelas, la resolución de los ¨problemas prácticos¨ es a menudo invisible para los estudiantes y el profesor (Collins, Brown et al. 1989).

Además, el ¨apprenticeship¨ incorpora la enseñanza de las habilidades y conocimientos en el contexto social y funcional de su uso. Esta diferencia no es sólo académica, sino que tiene importantes consecuencias en el tipo de conocimientos que los estudiantes adquieren.

En esta tesis utilizaré la palabra *¨aprendizaje¨* con el significado de *¨la enseñanza en la práctica o en el campo laboral¨; y* el de *¨aprendizaje cognitivo¨* como el método descripto en el trabajo de *Collins, Brown, & Newman (1989)* que desarrollaré mas adelante.

El objetivo de la enseñanza a través del método de aprendizaje cognitivo es ayudar a los estudiantes a adquirir conocimientos a través de la observación y la práctica guiada, facilitar la resolución de los problemas y fomentar la autonomía del alumno (Collins, Brown et al. 1989). Para lograr este objetivo, los métodos de enseñanza deben ser diseñados para dar a los estudiantes la oportunidad de observar y descubrir las estrategias que utilizan los expertos en su actividad profesional. Tal enfoque permitirá a los estudiantes ver de qué modo estas tácticas se corresponden con sus propios conocimientos conceptuales y factuales, y de qué modo se hace uso de estos como recursos para resolver problemas en el entorno social y en el ámbito práctico.

La teoría cognitiva del aprendizaje propone 6 métodos de enseñanza. El modelado, coaching y andamiaje son el núcleo del aprendizaje y están diseñados para ayudar a los estudiantes a adquirir las habilidades cognitivas y metacognitivas a través de la observación y de la práctica guiada. La articulación y la reflexión, tienen como objetivo que los estudiantes observen cómo los expertos resuelven los problemas y a hacer consciente (y a controlar) sus propias estrategias de resolución de problemas. La exploración fomenta la autonomía del alumno no sólo en los procesos de resolución de problemas definidos por los docentes, sino también en la formulación y resolución de los problemas que son propios de los estudiantes.

El **modelado** (¨modelling¨ en ingles) implica la demostración que realiza un experto al llevar a cabo una tarea determinada, de manera que los alumnos puedan observarlo y puedan construir un modelo conceptual de los procesos que son necesarios para realizar dicha tarea (Carroll and Bandura 1987). Esto requiere la declaración de los procesos que están por detrás de una actividad específica, de la heurística y de los procesos de control, mediante el cual los expertos hacen uso del conocimiento conceptual y procedimental básico.

El ¨***coaching***¨[4] consiste en observar a los estudiantes mientras estos realizan una tarea y ofrecerles sugerencias, feedback[5](Fornellsa, Juliàb et al. 2008),

[4] La traducción de la palabra *coaching* al español puede tener las siguientes acepciones: entrenamiento, asesoría, consejería (RAE, 2014). Sin embargo, su significado en español no termina de reflejar todos los aspectos de la palabra *coaching* en inglés, motivo por el cual usaré la palabra en el idioma original.

modelado, recordatorios y nuevas tareas encaminadas a lograr un desempeño más cercano al desempeño del experto (Collins, Brown et al. 1989). Puede servir para recordarle al estudiante una parte de la tarea que ha olvidado o pasada por alto momentáneamente. Las acciones del coaching están directamente relacionadas con eventos específicos o problemas que surgen a medida que el alumno intenta llevar a cabo una tarea determinada.

El **andamiaje** (¨scaffolding¨ en ingles) se refiere al apoyo que el docente brinda al aprendiz mientras éste realiza la tarea. El docente ayuda al estudiante en el momento que éste no puede completar una o varias partes de la tarea general. Esta ayuda pueden ser sugerencias o del tipo de soporte físico para ayudar a completar una habilidad procedimental. Se debe estimular al estudiante para que asuma la responsabilidad de realizar en forma independiente la mayor parte de la tarea completa. Un requisito del andamiaje es que el docente haga un diagnóstico preciso de nivel de la habilidad actual del estudiante y del nivel de dificultad que la actividad requiere. A medida que la competencia del estudiante crece, el docente retirará en forma progresiva el soporte brindado (Collins, Brown et al. 1989).

La **articulación** (¨articulation¨ en inglés) incluye dos aspectos. En primer lugar, se refiere al concepto de separar los componentes de las diferentes habilidades. En segundo lugar, lograr que los individuos organicen de manera coherente y eficaz su conocimiento, el razonamiento o los procesos de resolución de problemas en un dominio específico (McLellan H 1996). Incluye las estrategias para que los estudiantes enuncien su conocimiento, razonamiento o los procesos que usan en la resolución de problemas (Collins A and Stevens A. L 1983). Este aspecto compromete desde procesos simples, por ejemplo que el alumno describa verbalmente cómo el resuelve un problema determinado, o que adopte el rol de supervisor en colaboración con el docente.

[5] Usarse el término ¨feedback¨ con el siguiente significado: ¨El concepto de feedback, información usada por un sistema para hacer ajustes en la consecución de un objetivo, fue ampliado por Weiner hacia un ámbito más humanístico: feedback es el control de un sistema mediante la reinserción en el propio sistema de los resultados de su actuación. Si esta información es capaz de producir cambios en el método general y en el de actuación, se trata ya de un proceso que se puede denominar aprendizaje¨. No usaré ¨Retroalimentación¨ por ser un término menos usado en al ámbito de la educación médica.

El objetivo final es lograr que el alumno formule, exprese y controle su conocimiento respecto de la resolución de problemas.

La **reflexión** involucra a los estudiantes para que comparen su propio proceso de resolución de problemas con los de un experto u otros estudiantes (Collins, Brown et al. 1989). Este es, en última instancia, un modelo cognitivo de cada experiencia. La reflexión ocurre cuando se reproducen las actuaciones tanto del experto como del novato y se las compara. Esto se puede hacer a través de una variedad de métodos[6].

La **exploración** implica comprometer a los estudiantes para que resuelvan los problemas con una metodología propia, y ya no con los métodos del DC (Collins, Brown et al. 1989). Este paso es fundamental para que los estudiantes aprendan a hacerse las preguntas necesarias para resolver los problemas que les son propios. La exploración incluye el proceso de retiro del apoyo brindado, en las etapas anteriores, por el docente para resolver los problemas.

Los docentes necesitan ser evaluados para determinar si son capaces de desempeñarse aplicando las conductas docentes que se derivan del modelo de aprendizaje utilizado y conocer cuáles son las áreas en las que deben mejorar (Litzelman, Stratos et al. 1998; Snell, Tallett et al. 2000; Ramani and Leinster 2008). Bowden y Marton (2003) observaron que, para que los resultados de la evaluación sean útiles a los profesores en la optimización de su práctica docente, el instrumento de evaluación debe tener fundamentos teóricos sólidos. ¨*Cuando las evaluaciones no se basan en las teorías de la enseñanza efectiva, es difícil establecer la dirección en la cual los esfuerzos deben ser puestos para mejorar la enseñanza y, en consecuencia, la mejora real es probablemente dudosa¨ (Bowden and Marton 2003)*. Del mismo modo, el instrumento utilizado para la evaluación del desempeño docente no solo debe parecer válido por la base teórica que la sostiene, sino también debe ser testeada su validez y confiabilidad en forma empírica (Downing 2003).

[6] Un ejemplo de exploración es realizar una especie de ¨autopsia¨ del proceso mental que hizo el experto y se lo compara con el del estudiante, permitiendo hacer un análisis paso a paso del proceso de resolución de problemas. Alternativamente, se pueden grabar las actuaciones del experto y del estudiante y compararlas. El nivel de detalle con el que se hace la repetición dependerá de la etapa del aprendizaje en la que se encuentra el estudiante.

Beckman y colaboradores listaron 21 artículos de validación que describen los instrumentos para proporcionar feedback a los DC, sobre las evaluaciones de los estudiantes de grado y posgrado, acerca su desempeño (Beckman, Ghosh et al. 2004). Los instrumentos de evaluación que más se adaptan a la teoría cognitiva del aprendizaje son tres. Uno de los más utilizados es el desarrollado en la Universidad de Stanford, EEUU, que contiene categorías basadas en teorías educativas y psicológicas, y también en observaciones empíricas de la enseñanza clínica (Litzelman, Stratos et al. 1998). Este instrumento fue validado para medir la eficacia de la enseñanza en diversos ámbitos de aprendizaje (por ejemplo al tutor de grupos de ABP) y no específicamente para evaluar el desempeño del DC. Otra herramienta de evaluación frecuentemente utilizada es el ¨Cleveland Clinical Teaching Effectiveness instrument (CCTEI)¨ (Copeland and Hewson 2000). Es un instrumento válido y confiable y su fortaleza radica en que fue diseñado con la participación de todos los interesados en el proceso educativo (Beckman, Ghosh et al. 2004). La desventaja fundamental es que las dimensiones teóricas que se evalúan no están claramente especificadas, por lo que el proceso de feedback se ve limitado (Beckman, Ghosh et al. 2004).

El tercer instrumento fundamentado en la teoría del aprendizaje cognitivo ¨The Maastricht Clinical Teaching Questionnaire¨ y traducido por nosotros como ¨Instrumento de Evaluación del Desempeño del Docente Clínico (IEDDC) de Maastricht¨ fue elaborado en la Universidad de Maastricht con el objetivo que los estudiantes avanzados de medicina evalúen a sus DC (Stalmeijer, Dolmans et al. 2010).

Dada las características recién enunciadas, tomé este último instrumento como la base de mi tesis. A continuación realizaré una descripción detallada de la creación del IEDDC y de su proceso de validación.

En la etapa inicial, los autores elaboraron un cuestionario con preguntas destinadas a investigar los 6 dominios del aprendizaje cognitivo y agregaron preguntas sobre la capacidad del docente para crear un clima seguro de aprendizaje y de opiniones generales sobre el desempeño docente (Stalmeijer,

Dolmans et al. 2008). De esta manera, el cuestionario inicial contaba con un total de 8 dominios y 30 ítems. El número de ítems se distribuyó del siguiente modo: 4 para cada uno de los 6 dominios originales (modelado, coaching, andamiaje, articulación, reflexión y exploración), 3 para la evaluación del clima de aprendizaje, 2 ítems abiertos de opinión, y una puntuación global sobre el desempeño docente. Luego, para investigar la validez del contenido, invitaron a representantes de las tres partes interesadas en el proceso educativo: los educadores, médicos DC y estudiantes, para que juzguen cada una de las preguntas en su relevancia y formulación. Si bien era esperable que solo los expertos estén en condiciones de evaluar la calidad de la enseñanza, los autores también pidieron la opinión de los DC sin formación docente y a estudiantes, no sólo para sumar validez de contenido, sino también para ampliar la aceptación y el apoyo de todas las partes involucradas en una posterior aplicación en terreno del mismo. De este modo, 12 educadores, 16 médicos DC con actividad asistencial y 12 estudiantes avanzados, evaluaron el instrumento. Los educadores, fueron elegidos de las ocho escuelas de medicina de Holanda con experiencia en la enseñanza clínica. Los DC participaban en la enseñanza durante las rotaciones de los estudiantes en diversas especialidades (medicina interna, pediatría, cirugía, obstetricia / ginecología, otorrinolaringología y dermatología), en diferentes hospitales de los Países Bajos. Los estudiantes fueron voluntarios de la Escuela de Medicina de Maastricht, habían completado al menos 6 de un total de 10 rotaciones clínicas y todos tenían experiencia en ser supervisados en sus prácticas clínicas. A todos los convocados, se les pidió que evalúen la relevancia de cada uno de los ítems del cuestionario mediante una puntuación del 1 (muy irrelevante) al 5 (muy relevante). Además, se les solicitó que, para aquellos ítems puntuados muy bajo (valores 1 y 2), explicaran el motivo de esa calificación y que en caso de considerarlo necesario agreguen otros nuevos.

En general, la pertinencia de los ítems fue calificada como muy buena por los tres grupos de interés, con puntuaciones promedio que oscilaron entre 3,4 y 4,6. Una vez analizadas todas las respuestas, se eliminaron del cuestionario los ítems calificados por debajo de 3.5 por los tres grupos de actores. También

fueron eliminadas las preguntas que dieron lugar a fuertes comentarios negativos sobre la calidad del enfoque o de la redacción. Todos los participantes evaluaron de forma similar los dominios ¨coaching¨, ¨articulación¨ y ¨misceláneas¨. En comparación con los otros grupos, los estudiantes calificaron como menos relevantes los ítems de ¨modelado¨ y ¨reflexión¨; los DC y educadores dieron calificaciones más altas para ¨andamiaje ¨ y ¨exploración¨.

Como consecuencia del análisis cualitativo y de los resultados cuantitativos, de los 30 ítems iniciales se eliminaron cuatro, trece fueron modificados en su redacción y se añadió una pregunta solicitada por los tres grupos respecto al clima del aprendizaje. En el apéndice 1, se presenta el cuestionario original con los 30 ítems con las evaluaciones promediadas en cada uno de ellos, los ítems eliminados y agregado.

Concluida la validación de contenido, los investigadores propusieron como objetivos futuros determinar la validez de constructo del MCTQ, generalizar los resultados del cuestionario para determinar el número de alumnos necesarios para proporcionar feedback a los médicos acerca de sus habilidades de supervisión y analizar el modo en el que los diferentes factores se relacionan entre sí y con la puntuación del desempeño o juicio global.

En año 2010 la Dra. Stalmeijer y colaboradores publicaron los resultados del estudio psicométrico del instrumento sobre el que se evaluó la validez de contenido en la etapa anterior (Stalmeijer, Dolmans et al. 2010). Esta nueva investigación aportó evidencia de validez de constructo y se probó la confiabilidad entre un grupo de estudiantes y docentes de dos hospitales afiliados a la Facultad de Medicina de la Universidad de Maastricht. En estos hospitales, las rotaciones clínicas de los estudiantes tienen una duración entre 5 y 10 semanas, se realizan durante el 4° y 5° año de la carrera (de un total de 6 años) y comprende los servicios de medicina interna, cirugía, pediatría, obstetricia y ginecología, neurología, dermatología, otorrinolaringología, oftalmología y psiquiatría. Los docentes fueron evaluados tanto en el ámbito ambulatorio como el de internación.

Se recibieron un total de 1315 evaluaciones de 291 DC. Se calculó para cada uno de los docentes evaluados el promedio en cada uno de los ítems y fueron incluidos en el análisis sólo aquellos que tenían 4 o más observaciones de los estudiantes. Luego de comprobar que la distribución de los datos tenía normalidad multivariada, se realizó una análisis factorial confirmatorio (AFC), siguiendo una estimación de máxima probabilidad, mediante el programa AMOS. Adicionalmente se usó el paquete estadístico SPSS15 para calcular las correlaciones entre las puntuaciones de cada uno de los factores y el juicio global.

Para estimar el mínimo número de evaluaciones necesarias para lograr una calificación confiable de un DC, se realizó un estudio de generalizabilidad (coeficiente G) de los puntajes mediante el SPSS versión 15.

Para el subgrupo de 126 DC que recibieron más de cuatro observaciones, se calcularon los coeficientes G -tanto para el juicio global de la enseñanza como para cada factor[7]. Se realizó un análisis de la consistencia interna del test mediante el cálculo del coeficiente alfa de Cronbach para cada una de las escalas, y se considero satisfactorio un valor de al menos 0,70.

Por último, probaron un modelo que comprendía todos los factores y el juicio general mediante el ajuste de un modelo estructural lineal a los datos. En este análisis como en el AFC se utilizaron los mismos índices y criterios[8].

El AFC demostró que el instrumento no presentaba un ajuste óptimo de los dominios propuestos en la teoría cognitiva. Por lo tanto, se eliminaron los ítems que mostraran posible superposición de texto y significado y / o los que fueron sugeridos para ser eliminados según los índices de modificación generados

[7] Este diseño permite estimar los componentes de la varianza de dos fuentes: (1) las diferencias existentes entre los maestros (T) (objeto de medición) y (2) las diferencias que existen entre las evaluaciones hechas por los alumnos anidados dentro de los docentes y del error general (S: T, e)

[8] χ^2 divided by the degrees of freedom (CMIN/*df*)is < 2;
the goodness-of-fit index (GFI) is >0.90;
the comparative fit index (CFI) is > 0.90;
the root mean square residual (RMSEA) is < 0.1; y
the PCLOSE value is > 0.5.

con el programa Amos 7.0[9]. Se probó el modelo con diferente número de factores latentes y los resultados demostraron que el modelo de cinco factores, en lugar de 6 como era originalmente, con un total de 14 ítems, proporcionaba un ajuste excelente con índices adecuados[10]. En la tabla1 se describe el resultado del AFC con los diferentes índices para los diferentes modelos. La medición de la confiabilidad mediante el análisis de la consistencia interna mostró coeficientes α de Cronbach muy altos para cada una de las cuatros variables latentes (modelado, coaching, articulación y exploración) surgidas del AFC, con valores que se encuentran entre 0.83 y 0.96.

Tabla 1. Resultados del análisis factorial confirmatorio del Instrumento de evaluación Original de Maastricht con uno a cinco factores latentes.

N° de Factores	χ^2	*df*	*P*	*CMIN/df*	*GFI*	*CFI*	*RMSEA*	*PCLOSE*
1	474.6	77	.00	6.16	0.58	0.76	0.20	0.00
2	264.3	76	.00	3.48	0.72	0.89	0.14	0.00
3	221.8	74	.00	3.00	0.78	0.91	0.13	0.00
4	157.6	71	.00	2.22	0.83	0.95	0.10	0.00
5	73.0	67	.30	1.09	0.92	1.00	0.03	0.85

El conjunto de datos constaba de cuatro o más evaluaciones anónimas de los alumnos para los 126 DC. Los resultados indican que el modelo de cinco factores tiene el mejor ajuste a los datos. Se utilizaron los siguientes índices y criterios de ajuste: (1) χ^2 dividida por los grados de libertad (CMIN / *df*) debe ser < 2; (2) el índice de bondad de ajuste (GFI) debe ser > 0,90; (3) el índice de ajuste comparativo (CFI) debe ser > 0,90; (4) la raíz cuadrada media residual (RMSEA) debe ser < 0.1; y (5) la pclose > 0.5.

Los resultados del estudio de generalizabilidad mostraron que la varianza asociada a los docentes en cada uno de los dominios fue baja (entre 0.11 y 0.14), y para el juicio global fue de 0,34. En cambio la varianza asociada a los estudiantes anidados dentro de los docentes varió según el factor evaluado, entre 0,44 y 0,74 (ver tabla 2).

[9] Se trata de los índices que indican los elementos que pueden ser destituidos para lograr un mejor ajuste del modelo

[10] CMIN/df 1.09, GFI 0.92, CFI 1.0, RMSEA 0.03, y PCLOSE 0.85.

Tabla 2. Componentes de la varianza del cuestionario IEDDC para cada uno de los factores y para el Juicio global *

Factor	Docentes Clínicos (T)	Estudiantes: DC (S:Te)
Modelado	0.12	0.44
Coaching	0.11	0.44
Articulación	0.12	0.46
Exploración	0.12	0.74
Clima de aprendizaje	0.14	0.48
Valoración Global	0.34	0.88

*El conjunto de datos constaba de cuatro o más evaluaciones anónimas de los alumnos para los 126 DC. El cuadro representa la interacción entre la varianza asociada al efecto de los DC (T) y a la interacción entre el DC y el estudiante (S: Te).

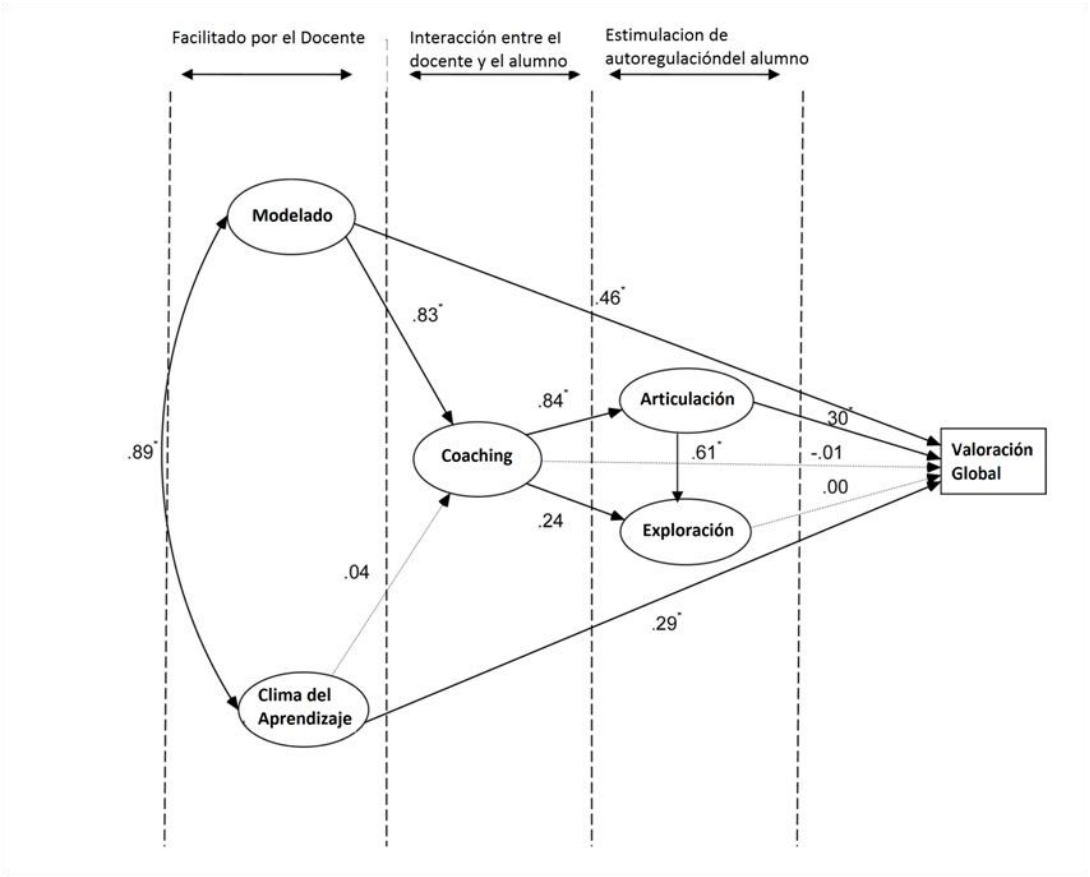
El número mínimo de observaciones necesarias para una evaluación confiable del docente, varió significativamente entre los diferentes dominios analizados. Considerando un Coeficiente-G de al menos 0,70, se requieren por lo menos siete respuestas de los estudiantes para el juicio global; entre 8 y 10 para modelado, coaching, articulación, y establecimiento de un ambiente de aprendizaje seguro, y 14 puntuaciones para exploración[11].

El análisis de las relaciones causales, entre los cinco factores (o métodos de enseñanza clínica) que surgieron del AFC, mediante el modelo de ecuaciones estructurales, dio lugar a un buen ajuste a los datos[12] (Fig.1). El análisis mostró que el modelado, el ambiente de aprendizaje, y la articulación afectan significativamente el juicio global de la enseñanza clínica. Por otra parte, el modelado desempeña un papel muy importante, ya que afecta sustancialmente el coaching, que a su vez afecta claramente a la articulación, y la articulación a su vez impacta sustancialmente sobre la exploración.

Aunque el ambiente de aprendizaje no afecta significativamente al coaching, tenía un efecto directo sobre el juicio global. Coaching y la exploración no parecen afectar directamente a juicio global, pero coaching tiene un efecto en el juicio global a través de la mediación de la articulación.

[11] Una explicación de los autores es que, posiblemente al haber solo dos ítems para este dominio se requieran muchas más evaluaciones que para el resto de los dominios evaluados.
[12] (CMIN/*df* 1.34, GFI 0.90, CFI 1.0, RMSEA 0.05, y PCLOSE 0.43)

Figura 1. Análisis de la relaciones entre factores latentes mediante modelo de ecuaciones estructurales (trabajo original)(Stalmeijer, Dolmans et al. 2010)



Modelo lineal de la enseñanza clínica explicando las relaciones entre los diferentes métodos de enseñanza pertenecientes al IEDDC. Los coeficientes Path (coeficientes de regresión estándar) muestran que el modelado, el ambiente de aprendizaje, y la articulación afectan significativamente el juicio global de la enseñanza clínica. Ver más explicación en el texto. El coeficiente de regresión estándar es significativa al nivel 0,01 (también indicada por líneas oscuras).

La fuerza de los resultados de las pruebas psicométricas y de la validación de contenido realizada por los educadores, DC y estudiantes conjuntamente con la solidez de la teoría subyacente hacen de este instrumento una herramienta muy valiosa. Además, en este cuestionario la presencia de ítems que exploran en forma separada las diferentes dimensiones permite una mejor individualización del desempeño del docente y facilita un feedback específico.

Las excelentes características del IEDDC nos motivaron a utilizarlo en nuestro medio luego de realizar la adaptación a nuestro idioma con su correspondiente validación.

Objetivo del Estudio

Realizar una validación transcultural del Instrumento de Evaluación del Desempeño del Docente Clínico de Maastricht al idioma español.

Objetivos específicos:

- ✓ Obtener una versión válida del IEDDC en español
- ✓ Aportar evidencia para la validez y determinar la confiabilidad.

Materiales y Método.

Para poder hacer uso del IEDDC en nuestro medio, es esencial realizar la adaptación transcultural del mismo y luego someterlo a una valoración de sus características psicométricas entre nuestros docentes y estudiantes.

La adaptación transcultural tiene como objetivo medir el mismo fenómeno en diferentes culturas. La utilización de instrumentos inadecuadamente adaptados puede conducir a conclusiones erróneas que son difíciles de detectar durante el análisis. Por este motivo la traducción simple no es aconsejable por las diferencias culturales y en el idioma. (Arribas 2006)

- **De la adaptación transcultural del instrumento.**

Luego de pedir autorización a los autores del trabajo original (Stalmeijer, Dolmans et al. 2010), realizamos la traducción del instrumento siguiendo las ¨Directrices para la traducción y adaptación de pruebas¨ de la ¨The International Test Commission¨(Commission 2005). En la tabla 3 se puede ver un resumen esquemático propuesto por Beaton, Bombardier et al. 2000. Nosotros nos basamos en este último para desarrollar y presentar el proceso de adaptación.

Tabla 3. Guía orientadora para el proceso de adaptación transcultural de instrumentos

Traducción	• **Producir varias traducciones.** • **Usar traductores calificados.**
Retrotraducción	• Producir tantas retrotraducciones como traducciones hechas. • Usar traductores calificados diferentes de los primeros.
Comité Revisor	• Constituir un comité de expertos para comparar ambas versiones del documento. • Los integrantes deberían ser multidisciplinarios. • Usar técnicas estructurales para resolver discrepancias • Modificar, corregir y/o hacer nuevas preguntas en caso de ser necesario • Asegurarse que la traslación sea completamente comprensible. • Verificar la equivalencia transcultural de las versiones original y final del instrumento.
Aportar evidencia para Validez aparente	• Evaluar las equivalencia de ambas versiones usando: • un pretest, o • alguna técnica de "sondeo", o • someter ambas versiones a personas bilingües no expertas.
Peso de la puntuación	• Considerar la adaptación de la puntuación original al contexto cultural.

Adaptado de (Arribas 2006)

- **Del aporte de evidencia sobre la validez y determinación de la confiabilidad del instrumento.**

  Se buscarán evidencias de 4 de las fuentes de validez definidas en ¨*las normas para pruebas en educación y en psicología*¨, desarrolladas en forma conjunta por la Asociación Americana de Investigación Educativa, la Asociación Americana de Psicología y el Consejo Nacional sobre Medición de la Educación(American-Education-Research-Association,American-Psychological-Association et al. 1999)

  1. **Evidencia basada en el Contenido del test:**

     Se debe demostrar que el contenido del test (ítems) elegido conforma una muestra representativa del universo de ítems del constructo a evaluar. Se debe buscar la pertinencia de los ítems en los dominios y de estos en el constructo. El aporte de evidencia en esta área se realizará mediante la valoración de expertos del ¨Comité de investigación en docencia del HIBA¨.

2. **Evidencia basada en el Proceso de Respuesta:**

   Se analiza el proceso de respuesta para proporcionar información sobre los factores que afectan a los datos obtenidos con el instrumento. Mediante la evaluación de todas estas fuentes se busca evitar sesgo durante la administración del instrumento. Se obtuvo preguntando a un grupo de residentes el modo en el cual articulan sus procesos de pensamiento durante la respuesta de cada uno de los ítems. Se evaluó una vez que el proceso de traducción y validación de contenido hubo finalizado. Buscamos que la redacción de los ítems sea la adecuada para que interpretación del residente sea relevante para el constructo que se medía. A esta fuente de validez también se la denomina ¨validez aparente¨. Otros aspectos del proceso de respuesta son el método de entrega de instrumentos, de puntuación y reporte.

3. **Evidencia basada en la estructura interna del test y confiabilidad:**

   Mediante un procesamiento estadístico de los datos se buscó determinar las relaciones existentes entre ítems y dominios del constructo a medir. Además se comprobó si los ítems se agrupaban del mismo modo a la predicha en el modelo teórico. Examinamos estas relaciones por medio de un análisis factorial confirmatorio. La confiabilidad se midió mediante el cálculo del coeficiente Alfa de Cronbach, y se consideró confiable un Alfa de Cronbach mayor de 0.80.

   Se realizó un estudio G de generalizabilidad para determinar la capacidad de generalización de los resultados (Carline 2004), identificación de las fuentes de error asociadas a las mediciones y la cantidad necesaria de estudiantes para evaluar en forma confiable un DC.

   Se realizaron dos planes de análisis buscando responder las siguientes preguntas:

   1. ¿En qué medida se pueden generalizar los resultados de la evaluación del desempeño de los DC obtenidos a través de una valoración hecha por los residentes usando el IEDDC?

2. ¿En qué medida las evaluaciones hechas por residentes son generalizables a través del uso de los ítems del IEDDC?
3. ¿Cuál es el número de residentes que se necesitan para una evaluación confiable de los DC en cada unos de los dominios del IEDDC y del desempeño global, para proveerles un feedback adecuado?

4. **Evidencia basada en la Relaciones con otras variables:**
   Se midieron las relaciones entre las puntuaciones recibidas por los DC en cada uno de los dominios y el desempeño global.

Existe una quinta fuente de evidencia de validez que se relaciona con las ¨Consecuencias de la aplicación del test¨, esta no fue objetivo de estudio de nuestra investigación.

- **Respecto de las Características del instrumento de evaluación.**

Se trata de un cuestionario estructurado que evalúa los diferentes métodos de enseñanza de la teoría cognitiva de la educación, a través de 14 ítems. La valoración de cada uno de las preguntas se realiza mediante una escala tipo Likert de 5 puntos con doble denominación en los extremos de la escala. La menor valoración es el puntaje ¨1¨ (o ¨totalmente en desacuerdo con la aseveración planteada¨), en el otro extremo, el máximo puntaje es un ¨5¨ (¨totalmente de acuerdo¨ con la aseveración planteada). En medio de esos valores se encuentran los puntajes ¨2¨, ¨3¨, y ¨4¨ con progresivo grado de acuerdo con la aseveración.

Los datos de la escala de Likert serán tratados como variable continua, lo que permite el análisis paramétrico, el AFC y el estudio de generalizabilidad (Norman and Streiner 2008).

El cuestionario finaliza con una valoración global del desempeño del docente con un puntaje que va de 1 (valor más bajo correspondiente al peor desempeño) a 10 (mejor desempeño), además presenta dos preguntas abiertas que investigan las áreas en las que el docente muestra su mejor

desempeño y aquellas en las que debe mejorar. Estas dos últimas preguntas no fueron tenidas en cuenta en el análisis de este trabajo.

- **Con relación a los médicos residentes (evaluadores):**

  Se invitó a participar del estudio a médicos residentes de las especialidades clínicas o quirúrgicas, durante sus rotaciones clínicas de ya sea de internación o de atención ambulatoria. Los residentes pertenecen al Hospital Italiano de Buenos Aires y al CEMIC en cualquier etapa de su formación.

  El residente tenía la opción de elegir el docente que deseaba evaluar, con la única condición de que hubiera cumplido una rotación de al menos 4 semanas con una interacción diaria frecuente con el mismo (4 hs). A cada uno de los residentes participantes se le entregó un sobre de papel madera que contenía una carta explicando los alcances de la investigación (ver apéndice 2) y tres cuestionarios del IEDDC para completar. Cada uno de los residentes podría evaluar hasta tres DC. Se habilitaron espacios protegidos en donde los residentes depositaran los sobres sellados con los cuestionarios.

  Durante una reunión con ellos o en cualquier otro momento, los investigadores respondieron todas las preguntas que los mismos realizaron en cuanto al instrumento de evaluación, a los objetivos de la investigación o a cualquier otro tipo de duda. La participación fue voluntaria, anónima y sin retribución de ningún tipo.

- **Características de los docentes (evaluados):**

  Se invitó a participar a los médicos especialistas con funciones docentes de residentes (DC) de varias especialidades del HIBA y del CEMIC. Para ser evaluados, los médicos debían haberse desempeñado como supervisores de residentes al menos cuatro horas al día durante cuatro semanas.

  A cada uno de los DC se le solicitó un consentimiento verbal para participar del estudio, se le explicó en forma individual los alcances de la investigación y se respondieron todas las preguntas que los mismos realizaron en cuanto al instrumento de evaluación o cualquier otro tipo de duda.

Debido a que este estudio solo tuvo como finalidad la de determinar la validez y confiabilidad del instrumento, no se realizó feedback de su desempeño docente.

- **Respecto de la preservación de la integridad del anonimato**

  Los datos de los docentes evaluados sólo fueron conocidos por el investigador principal, quien fue responsable directo de pasar los resultados desde el formulario de papel a la base de datos electrónica.

  Los resultados de las evaluaciones fueron almacenados de un modo seguro. La base de datos se encuentra en una computadora con una clave de acceso sólo conocida por los investigadores. Los archivos y carpetas tuvieron nombres que no se relacionaban con la investigación en cuestión.

- **Determinación de N de evaluaciones.**

  Tomamos como punto de partida para el cálculo del N la necesidad impuesta por el AFC.

  Dado que el instrumento de evaluación original fue diseñado para evaluar el desempeño de docentes clínicos individuales y ofrecerles feedback que ayuden a mejorar su desempeño, se acordó utilizar, como en el estudio original (Stalmeijer, Dolmans et al. 2010), sólo los datos de aquellos docentes clínicos que hubieran recibido cuatro o más calificaciones individuales de los estudiantes.

  El cálculo del n necesario de evaluaciones para determinar la validez del instrumento depende de la existencia o no de normalidad multivariada de los datos (Henson 2006). Se determinó según las siguientes reglas:

  a) en caso de existir normalidad, se utiliza la recomendación basada en el número de variables latentes, indicadores y comunalidades, ver tabla 4. (Hair, Tatham et al. 2006).
  b) en caso de no existir normalidad multivariada y utilizando el método de análisis conocido como Diagonally Weighted Least Squares, el n se estima mediante la siguiente fórmula: (k+1) (k+2) /2 (Jöreskog, Sörbom et al. 1996), donde k = nº de indicadores a analizar (ítems).

Un análisis de una muestra parcial reveló que no presentaba normalidad multivariada, y para nuestro caso en donde k = 14 (cantidad de ítems en el análisis), la fórmula se presentaba de la siguiente manera: (14 + 1) (14 + 2) / 2 = 240 / 2 = 120.

De este modo, el tamaño muestral mínimo de evaluaciones fue entonces de 120 docentes evaluados con al menos cuatro evaluaciones de residentes.

Tabla 4. Recomendaciones sobre el tamaño muestral mínimo cuando hay Normalidad Multivariante

Tamaño Mínimo	*Condiciones*
100	Modelos con cinco o menos variables latentes. Cada variable latente con más de tres indicadores, y con comunalidades entre los ítems elevadas (0.6 o mayor).
150	Modelos con siete o menos variables latentes. Las comunalidades modestas (0.5) y con tres o más indicadores por variable latente
300	Modelos con un máximo de siete variables latentes, comunalidades bajas (por debajo de 0.45), y/o con varias variables latentes medidas con menos de tres ítems.
500	Modelos con un elevado número de variables latentes, algunas con comunalidades bajas, y/o con variables medias con menos de tres indicadores.

Adaptado de (Hair, Tatham et al. 2006)

- **Comité de Investigación en Educación Médica (CIEM)**:

  Un comité de seis expertos en investigación de la educación médica, colaboró en la búsqueda de fuentes para aportar evidencia para la validez de contenido del instrumento, determinar la pertinencia y adecuación de las traducciones y retrotraducciones y de confeccionar el instrumento definitivo en español.

- **Pruebas Estadísticas:**

  La validez de constructo se determinó fundamentalmente mediante un análisis factorial confirmatorio. Como en el estudio original (Stalmeijer, Dolmans et al. 2010), se estudiaron las correlaciones existentes entre cada uno de los factores y la puntuación del desempeño global (correlación de Pearson). Se realizó también un Modelo de Ecuaciones Estructurales que determinó las relaciones

entre los factores latentes y el desempeño global, comparando con los resultados del estudio original (Stalmeijer, Dolmans et al. 2010).

Se promediaron las cuatro o más evaluaciones de cada uno de los DC, y se determinó la existencia o ausencia de normalidad multivariada de la distribución de resultados mediante el cálculo de asimetría y curtosis (Test de Normalidad Multivariada para variables continuas del programa Lisrel 9.10).

Se utilizó el programa Lisrel 9.10 (Jöreskog, Karl and Dag Sörbom. USA: SSI Scientific Software International), a través de un Análisis Factorial Confirmatorio y un Modelo de Ecuaciones Estructurales, para determinar si los datos confirman el modelo teórico. Se utilizaron los siguientes indicadores de buen ajuste de Análisis Factorial Confirmatorio: Chi2/grados de libertad, RMSEA, NFI, NNFI, CFI, RMR, GFI, AGFI.

No se utilizó la p de Chi 2, dado que no es confiable al no cumplirse el supuesto de normalidad multivariante (Romá and Segura 1994; Arias 2008; Norman and Streiner 2008).

La confiablidad (Consistencia interna) se determinó con el cálculo del Alpha de Cronbach para cada uno de los dominios del constructo.

Para el estudio G se utilizó el software EduG 6.1e (Generalizability Study. Working Group "Edumetrics - Quality of measurement in education" of the Swiss Society for Research in Education). Debido a que este programa estadístico trabaja con muestras balanceadas de datos, se decidió que se reducirían en forma aleatorizada a 4 las evaluaciones de aquellos DC con más de 4 de ellas. Aquellos cuestionarios que presentaban ausencia de puntuación en alguno de los ítems, el dato se completo con el valor promedio resultante del análisis al ítem correspondiente. En todos los casos se consideró un coeficiente G satisfactorio igual o mayor a 0.70.

- **Comité de Ética de Protocolos de Investigación:**

  El protocolo fue aprobado por el comité de ética de investigación del hospital Italiano de Buenos Aires. N° de aprobación 1776

Resultados:

I. Objetivo 1: Obtener una versión válida del IEDDC en español:

Mediante una comunicación vía mail con la Dra. René Stalmeijer, obtuvimos su autorización y colaboración para realizar la adaptación transcultural del instrumento. Además fue confirmada la ausencia hasta ese momento de otro proceso de validación en español del IEDDC.

- **Traducción Directa**

  Se realizó una traducción a un idioma español amplio, que fuese de fácil comprensión para los médicos o los estudiantes de medicina de habla hispana independientemente de las idiosincrasias.

  La traducción al español o ¨traducción directa¨ fue realizada por dos traductores expertos de lengua nativa española; a uno de ellos se le informó los objetivos del estudio y al otro no (traductor inocente). En el apéndice 3 constan ambas traducciones directas y el documento en Inglés enviado a los traductores.

  La revisión de los documentos enviados por los traductores fue realizada por un comité de expertos en educación médica, reunidos en el comité de investigación en educación médica del Hospital italiano de Buenos Aires (CIEM). Luego del análisis de cada una de las traducciones, se confeccionó un único instrumento en español que interpretó de la mejor manera posible el constructo de la teoría cognitiva de la educación y a nuestra lengua. Cada uno de los ítems volcados al instrumento se copió de forma textual a lo escrito por cada uno de los traductores, para luego poder comparar con las retrotraducciones y los comentarios de los autores del estudio original. En el apéndice 4 se puede ver el documento consensuado por los expertos del CIEM.

  Además, los miembros del CIEM, cada uno en forma independiente y luego en una reunión de consenso, evaluaron la pertinencia de los ítems a los modos de enseñanza y de estos con la teoría del aprendizaje cognoscitivo.

- **Traducción inversa o Retrotraducción** (del español al inglés).

  Otros dos traductores expertos e independientes a la investigación realizaron la traducción al idioma Inglés desde el documento unificado por el CIEM. Ambos traductores fueron elegidos por poseer a la lengua inglesa como nativa y por desconocer los objetivos de la investigación y el instrumento original. Ambas traducciones constan en el Apéndice 5. Con el objetivo de detectar errores de interpretación, estas dos versiones fueron enviadas a la Dra. René Stalmeijer. La autora elegió una de las traducciones, aquella que fue considerada como la mas adecuada. Sobre la base de la traducción elegida, la autora del estudio original sugirió algunas correcciones.

  Nuevamente el comité de expertos, teniendo en cuenta las modificaciones sugeridas por la autora del trabajo original, confeccionó el primer cuestionario definitivo en el idioma Español. Apéndice 6.

II. Objetivo 2: Aporte de evidencia sobre la validez y determinación de la confiabilidad del IEDDC en nuestro medio.

1. Aporte de evidencia para la validez de contenido:

Se solicitó al grupo de expertos reunidos en el CIEM, que evalúen los instrumentos tanto en inglés como en español. Se les pidió que evalúen la relación existente de cada uno de los ítems del cuestionario con el dominio correspondiente y con el constructo. También valoraron la calidad formal de los ítems en español respecto de la estructura gramatical, claridad y adecuación del contenido a la nueva población en la cual se usará el instrumento. Consideraron que los 14 ítems correspondientes a los 5 modos de enseñanza reflejaban adecuadamente la teoría del aprendizaje cognitivo y que cumplían con los requisitos necesarios para que IEDDC evalúe a los DC. También se estableció que no era necesario agregar ninguna de las preguntas eliminadas por los autores originales, ni agregar preguntas complementarias.

2. Aporte de evidencia del proceso de respuesta

El primer cuestionario definitivo en español fue sometido a una prueba de validez aparente. Esta forma de validez está muy emparentada con la validez de contenido, se refiere al grado en que un test da la impresión a los evaluadores que mide lo que se pretende medir. Se utilizó la técnica sondeo, en la cual se usa el cuestionario en una población de prueba, en este caso utilizamos un grupo de residentes de los hospitales intervinientes en la investigación, a quienes se los invitó a analizar el instrumento de evaluación docente. Luego de las entrevistas con los residentes se logró saturar la muestra respecto a la interpretación de cada uno de los ítems, comentarios y sugerencias de cambios o agregado de preguntas. En la mayoría de los casos la interpretación de cada una de las preguntas fue similar o igual a la interpretación hecha por los expertos del CIEM. Las modificaciones y /o agregados sugeridas por algunos de los entrevistados fueron menores y en ningún caso hubo coincidencia entre ellos. Dado que no fue necesario

modificar el instrumento ni la escala de puntuación, se confeccionó el cuestionario definitivo y se archivó en formato pdf seguro para evitar futuras modificaciones durante la prueba. Este cuestionario fue el utilizado en todas las evaluaciones del estudio. Ver apéndice 7.

- **Características de la administración de la prueba:**

Se obtuvo un listado con los nombres de todos los residentes de ambos hospitales, de allí se eligieron aquellas residencias que contaran con rotaciones clínicas. En cada uno de los servicios en donde se ejercía la docencia clínica, los investigadores nos reunimos con los residentes, se les explicaron los alcances de la investigación y se respondieron sus preguntas y sus dudas. Las preguntas más frecuentes estuvieron relacionadas con los objetivos del estudio y con la confidencialidad de las respuestas. A cada uno de los posibles evaluadores se le entregó un sobre con los cuestionarios y una nota explicativa con los alcances de la investigación. Del mismo modo a lo que ocurrió en el estudio original, los residentes eligieron en forma voluntaria hasta tres DC a evaluar. Una vez que los investigadores entregamos los sobres, los residentes evaluaban a los DC en el momento y lugar que consideraran adecuados y entregaban el sobre con los cuestionarios completos o vacios, en el sobre cerrado en un lugar designado por los propios residentes. De este modo nos aseguramos del anonimato y la libertad de los residentes de completar y entregar los cuestionarios sin que sintieran algún tipo de presión.

- **Recolección de la muestra:**

Se evaluó el desempeño de los DC de diferentes especialidades. Se evaluaron las actividades docentes llevadas a cabo en las salas de internación o en los consultorios ambulatorios de dos hospitales universitarios de la ciudad autónoma de Buenos Aires. Desde febrero del año 2012 a mayo del año 2013 se recibieron un total de 1078 evaluaciones correspondientes a 187 DC. Un total de 47 cuestionarios fueron eliminados del análisis por estar incompletos varios de los ítems o tener marcadas más de una opción en alguna de las preguntas (para el análisis fueron utilizados 1031 cuestionarios). El promedio

de evaluaciones por docente fue 5.51±3.4 con un rango de 1 - 21. Un 69,5% (130) de los DC evaluados correspondió a profesionales del HIBA y el resto (57) a DC del CEMIC. Del total de evaluaciones entregadas completas, un 66,14% (682) correspondieron a evaluaciones hechas por residentes del HIBA, el resto (349) a residentes del CEMIC. En el apéndice 8 se encuentran gráficos en donde se observan las evaluaciones según la especialidad del residente y las evaluaciones según la especialidad de la rotación o del DC.

La especialidad con más evaluaciones y evaluadores fue medicina interna y clínica médica, seguida por cardiología, terapia intensiva adultos y ortopedia/traumatología

3. Aporte de validez basada en la estructura interna del test y de la confiabilidad.

- **Validez de constructo mediante un análisis factorial confirmatorio.**

Del mismo modo que en el estudio original (Stalmeijer, Dolmans et al. 2010), para la determinación de la validez de constructo solo fueron considerados aquellos docentes que tuvieran 4 o más observaciones, quedando conformada la muestra por 151 docentes. La cantidad de evaluados fue considerada suficiente para este estudio (Jöreskog, Sörbom et al. 1996).

Se promediaron los valores de las respuestas de cada ítem para cada docente, Ej.: si el ítem numero 1, figura con un 3 para el docente X, este 3 es el promedio de los valores que colocaron los 4 o más alumnos que lo calificaron al respecto del ítem 1.

Mediante el test de normalidad multivariada para variables continuas se determinó que nuestros datos no tienen distribución normal (tabla 5).

Tabla 5. Test de valoración de la normalidad multivariada para datos continuos

Asimetría				*Curtosis*		*Asimetría y Curtosis*	
Valor	**Z-Score**	**P-Value**	**Value**	**Z-Score**	**P-Value**	**Chi-Square**	**P-Value**
77.134	**20.313**	**0.000**	**365.865**	**10.525**	**0.000**	**523.372**	**0.000 (*)**

(*) La p significativa indica que la distribución multivariada es distinta que la normal.

Dado que también la distribución univariada está lejos de ser normal para todos los ítems, no se justificó intentar "normalizar" la distribución de algunos de ellos utilizando un logaritmo o una potencia. De este modo utilizamos métodos de Análisis Factorial Confirmatorio (AFC) que no asumen la normalidad de los datos. Estos métodos son: Diagonally Weighted Least Squares (DWLS), y Unweighted Least Squares (ULS). Por lo cual se realizó el AFC mediante el método DWLS, por ser el más recomendado para estos casos.

Siguiendo el mismo procesamiento de los datos a lo realizado en el estudio original, buscamos determinar los índices de ajuste, ya sea considerando que todos los ítems correspondan a un solo factor latente o a varios (hasta 5). En la Tabla 6 de nuestro análisis se ven 3 modelos (Método de estimación utilizado: DWLS). El primer modelo corresponde a un solo factor latente (ítems del 1 al 14), el segundo a 3 factores latentes (ítems 1 a 3 en un factor, ítems 4 a 11 en un segundo factor, e ítems 12 a 14 en el último factor), y el tercer modelo son 5 factores (ítem1 a 3 en ¨modelado¨, 4 a 6 en ¨coaching¨, 7 a 9 en ¨articulación¨, 10 y 11 en ¨exploración¨ y 12 a 14 en ¨clima de aprendizaje seguro¨). Se puede ver como mejoran los índices de ajuste en la medida que se aumenta el número de factores latentes, hasta llegar al sugerido por el marco teórico del modelo original de 5 factores. En este último caso todos los valores de los índices de ajuste son correctos.

Tabla 6. Índices de ajuste del Análisis Factorial Confirmatorio con 1, 3 y 5 factores latentes (método de estimación: DWLS).

Modelos: n de Factores Latentes	*Chi2/g. de liber. *(≤ 3)*	*RMSEA *(< 0.05)*	*NFI *(>0.95)*	*NNFI *(>0.95)*	*CFI *(>0.95)*	*RMR *(< 0.08)*	*GFI *(>0.95)*	*AGFI *(>0.95)*
1 FL	367,85/7 : **4.77**	0.052	0.94	0.94	0.95	0.07	0.99	0.987
3 FL	255.97/74: **3.46**	0.045	0.959	0.96	0.97	0.0496	0.995	0.993
5 FL	179.62/67: **2.68**	0.04	0.97	0.975	0.98	0.03	0.998	0.997

*Debajo de cada uno de los índices de ajuste el valor mínimo para considera significancia. En todos los casos el modelo de cinco factores ajusta de la mejor manera.

(*) Encabezando la tabla 5 está abreviado cada índice de ajuste, y se especifica los valores que debería tener. Como en el estudio original, se puede ver como mejoran los índices de ajuste en la medida que aumentamos el número de factores latentes (FL), hasta llegar al sugerido por el marco teórico del modelo original de 5 factores latentes. En este último caso todos los valores de los índices de ajuste son correctos.

Para agregar mayor confianza en la correcta validez de constructo y en la estabilidad del modelo, realizamos AFC en cada mitad de la base, dividiéndola de dos maneras distintas. En el primer caso dividimos la base en los primeros 75 casos y los segundos 76 casos según el orden que estaban en la base de datos. En el segundo caso incluimos sólo los casos que se ubicaban en la base con número par y luego lo mismo con los que se ubicaban con número impar, obteniendo en las cuatro modalidades valores compatibles con un buen ajuste. Dado que al dividir la base de datos contábamos con menos N, utilizamos el método Robust Unweighted Least Squares (ULS). Este método requiere un N menor, y tampoco necesita distribución normal de los datos. Los resultados pueden verse en la tabla 7.

También puede percibirse la consistente estabilidad del modelo de 5 factores latentes, al observar los valores similares de los índices de ajuste, cuando los análisis se realiza con el método DWLS (Ver Tabla 6, cuarta fila (5 FL); y cuando se realiza sobre la base completa con el método ULS, ver Tabla 7, segunda fila (Toda la base), ya que ambos métodos tienen una dinámica diferente de análisis estadístico.

Tabla 7. Índices de ajuste del Análisis Factorial Confirmatorio con modelos alternativos (método de estimación ULS)

Con ULS	*N*	*Chi2/g. de liber. (< 3)*	*RMSEA (< 0.05)*	*NFI (>0.95)*	*NNFI (>0.95)*	*CFI (>0.95)*	*RMR (< 0.08)*	*GFI (>0.95)*	*AGFI (>0.95)*
Toda la Base	151	193.93/67: **2.89**	0.04	0.969	0.972	0.979	0.03	0.998	0.996
Los primeros 75m	75	139.73/67: **2.08**	0.000	0.946	0.96	0.97	0.048	0.99	0.99
Desde 76 a 151	76	162,43/67: **2.42**	0.000	0.953	0.96	0.97	0.034	0.998	0.996
Solo pares	75	130.33/67: **1.94**	0.000	0.95	0.966	0.97	0.04	0.995	0.993
Solo impares	76	158.43/67: **2.36**	0.000	0.955	0.964	0.97	0.035	0.998	0.996

AFC dividiendo la tabla de registro por mitades. Se usó el método ¨Unweighted Least Squares¨ en todos los casos

- **Análisis de las relaciones causales mediante el modelo de ecuaciones estructurales.**

Las pruebas del modelo con los cinco factores latentes basado en el AFC dio lugar a un buen ajuste a los datos, ver tabla 8.

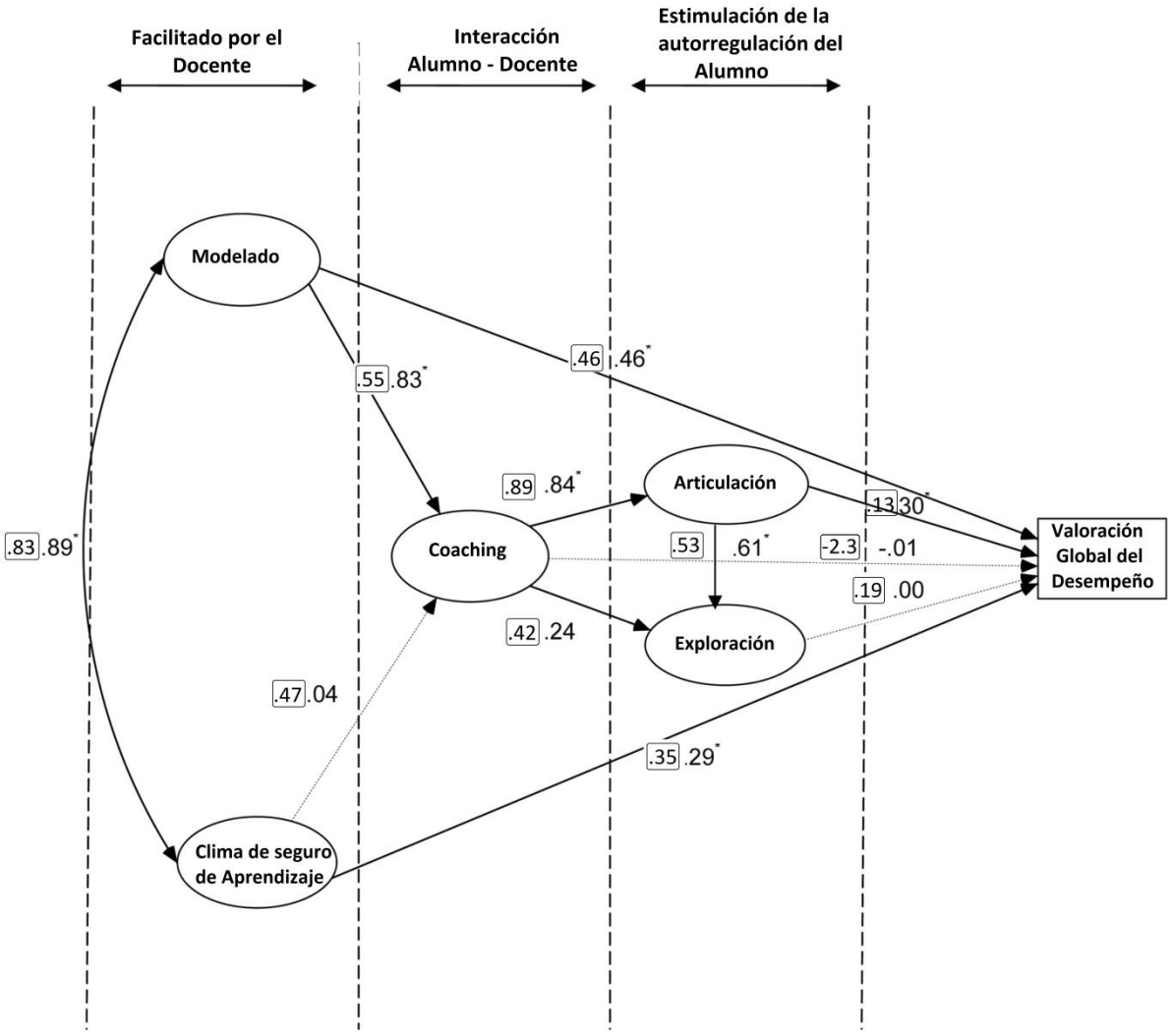
Los coeficientes Path mostraron fuertes analogías y fundamentalmente una diferencia, ver Fig. 2, referente a la incidencia del Factor latente "Clima de aprendizaje" sobre el factor latente "Coaching", casi inexistente en el trabajo original (Ver más detalles en Discusión).

Tabla 8. Resultados de ajuste del AFC y MEE

Modelos	*Chi2/g. de liber. (≤ 3)*	*RMSEA (< 0.05)*	*NFI (>0.95)*	*NNFI (>0.95)*	*CFI (>0.95)*	*RMR (< 0.08)*	*GFI (>0.95)*	*AGFI (>0.95)*
MEE	202.24/79: **2.56**	0.0269	0.972	0.977	0.982	0.0356	0.998	0.996
AFC	179.62/67: **2.68**	0.04	0.97	0.975	0.98	0.03	0.998	0.997

MEE: Modelo de ecuaciones estructurales, AFC: Análisis Factorial confirmatorio. En todos los casos los valores de ajuste son adecuados.

Figura 2. Relaciones causales entre los diferentes modos de enseñanza (factores latentes) del Instrumento de Evaluación del Desempeño del Docente Clínico de Maastricht y la valoración global.



Los valores adjuntos a las líneas que relacionan los diferentes factores corresponden a los coeficientes de regresión estandarizados, los valores que se encuentran recuadrados corresponden al coeficiente del trabajo local y el otro al del trabajo original. El esquema o imagen se divide con línea de puntos, de la siguiente forma: "Facilitado por el docente", con los dominios "Modelado" y "Clima de aprendizaje seguro"; "Interacción entre el docente y el alumno", con el dominio "Coaching", y "Estimulación de la auto-regulación del alumno", con los dominios "Articulación" y "Exploración".
El coeficiente de regresión estándar es significativa al nivel 0,01 (también indicada por líneas oscuras). La única diferencia es que el estudio local hay una correlación significativa entre clima de aprendizaje y coaching, algo que no se evidenció en el trabajo original. Si bien la correlación entre articulación y valoración global es significativa en ambas investigaciones, en la investigación local esa correlación es pobre.

Determinación de la confiabilidad

- **Consistencia Interna - Cálculo del coeficiente Alfa de Cronbach.**

Del mismo modo a lo realizado en el estudio original, al tener una sola muestra de datos, estimamos la confiabilidad mediante el método de consistencia interna. Elegimos el método de correlación entre los ítems usando el coeficiente Alfa de Cronbach por tratarse de ítems no dicotómicos, específicamente una escala tipo Likert (Aiken, Salinas et al. 2003). Realizamos el cálculo mediante el programa SPSS, subprocedimiento: ¨Análisis de la confiabilidad¨, opción: ¨Escalas¨. Se extrajo el ítem 15 (valoración global) por no estar relacionado al constructo que se buscó medir. El valor Alfa para cada uno de los dominios evaluados, de los DC de ambos hospitales, fue en todos los casos mayor a 0.90. Ver tabla 9.

De modo complementario, valoramos el comportamiento de cada uno de los ítems, incluido el 15, en ambos hospitales como muestras separadas. Como se puede observar en las tablas 10 y 11, la correlación de cada ítem con el resto es la más alta y la del 15 la más baja.

Tabla 9. Análisis de la confiablidad de la muestra mediante el Alfa de Cronbach para cada uno de los factores

Ítems	*Factores*	*Alfa de Cronbach*
1,2,3	Modelado	0.9119
4,5,6	Coaching	0.8766
7,8,9	Articulación	0.9069
10,11	Exploración	0.9462
12,13,14	Clima de aprendizaje seguro	0.9012

Como se puede observar todos los coeficientes son elevados, demostrando la adecuada consistencia interna del test.

Tabla 10. Cronbach: Estadísticos de todos los elementos. Muestra del HIBA

Ítem	*Promedio de la escala, sin eliminar elementos*	*Varianza de la escala si se elimina el ítem*	*Correlación de los elementos completos*	*Cuadrado de la Correlación Múltiple*	*El alfa de Cronbach para la supresión del Ítem*
P01	57,81	138,974	0,715	0,627	0,877
P02	57,85	139,522	0,687	0,606	0,878
P03	58,1	137,811	0,652	0,499	0,878
P04	58,16	137,531	0,628	0,45	0,879
P05	57,83	138,917	0,677	0,53	0,878
P06	57,83	138,089	0,663	0,524	0,878
P07	57,91	139,519	0,653	0,576	0,879
P08	57,82	137,918	0,694	0,598	0,877
P09	58,23	135,973	0,729	0,666	0,876
P10*	58,31	136,066	0,704	0,73	0,876
P11*	58,24	136,178	0,723	0,712	0,876
P12	57,56	139,674	0,629	0,594	0,88
P13	57,8	137,802	0,697	0,582	0,877
P14	57,28	144,261	0,542	0,512	0,883
P15	55,18	123,197	0,315	0,14	0,936

Tabla de correlaciones del análisis de consistencia interna de la muestra tomada en el HIBA
* Los ítems 10 y 11 son los que muestran mayor correlación; ambos forman el dominio: ¨exploración¨. ** El ítem 15: ¨valoración global del DC: es el que menos correlaciona y que aporta mayor carga al valor final del Alfa si es removido

Tabla 11. Cronbach: Estadísticos de todos los elementos. Muestra del CEMIC

Ítem	*Promedio de la escala, sin eliminar elementos*	*Varianza de la escala si se elimina el ítem*	*Correlación de los elementos completos*	*Cuadrado de la Correlación Múltiple*	*El alfa de Cronbach para la supresión del Ítem*
P01	59,6	91,655	0,646	0,484	0,862
P02	59,6	93,402	0,529	0,406	0,866
P03	59,82	90,104	0,63	0,493	0,861
P04	60,05	88,195	0,642	0,495	0,86
P05	59,57	92,2	0,662	0,491	0,862
P06	59,63	94,681	0,466	0,31	0,869
P07	59,72	93,405	0,503	0,408	0,867
P08	59,59	92,501	0,556	0,444	0,865
P09	60,08	89,563	0,677	0,599	0,86
P10	60,11	89,791	0,647	0,66	0,861
P11	60,02	88,836	0,71	0,673	0,858
P12	59,42	92,445	0,572	0,532	0,864
P13	59,49	89,848	0,7	0,576	0,859
P14	59,04	96,498	0,478	0,444	0,869
P15	56,13	81,57	0,339	0,179	0,909

Tabla de correlaciones del análisis de consistencia interna de la muestra tomada en el CEMIC
* Los ítems 10 y 11 son los que muestran mayor correlación; ambos forman el dominio: ¨exploración¨
** El ítem 15: ¨valoración global del DC: es el que menos correlaciona y que aporta mayor carga al valor final del Alfa si es removido

- **Análisis de Generalizabilidad.**

Realizamos un estudio de generalizabilidad referido a criterios, por tal motivo en todos los casos tuvimos en cuenta, como fuente de error, todos los componentes de varianza, excepto el objeto de medida (Zúñiga-Brenes, 2007). Las estrategias de análisis buscaron responder 3 preguntas que aportaran evidencias sobre la confiablidad del IEDDC y, determinar las diferentes fuentes de error.

1. ***¿En qué medida se pueden generalizar los resultados de la evaluación del desempeño de los DC obtenidos a través de una valoración hecha por los residentes usando el IEDDC?***

Utilizamos el mismo plan de observación y de estimación que se utilizó en el estudio original {Stalmeijer, 2010 #17}. Usamos tres facetas: la de diferenciación formada por los DC (D), cruzada y de universo infinito, y dos de generalización, la de los residentes (O) anidados en los docentes (O:D), de universo infinito, y la de los ítems (I), cruzada y de universo infinito. Se incluyeron los 151 DC de ambos centros con al menos 4 evaluaciones. El diseño de medición usado fue **D/I(O:D)** que permitió la estimación de los componente de varianza de las siguientes fuentes: (1) las diferencias entre los DC, (2) la varianza del error absoluto aportado por los residentes anidados en los DC y 3) la interacción entre los residentes, los ítems y el error general. El diseño de observación y estimación se redujo a los 14 ítems que responden a los dominios de la teoría del aprendizaje cognitivo, es decir, se quitó del análisis al juicio global. La varianza del puntaje universo, objeto de medición (DC), fue de 0.29. El efecto constante determinado por los observadores anidados en los docentes representa el 77% de la varianza del error absoluto. Las inconsistencias determinadas por las interacciones entre los docentes y los ítems es baja, repesentando el 17.2% de la varianza del error absoluto (ver tabla 12). El Coeficiente G absoluto fue de 0,79. En la tabla 10 se pueden apreciar los componentes de la varianza y del error absoluto de cada una de las fuentes. En el apéndice 9 se presenta el análisis completo.

Tabla 12. Tabla del Estudio G correspondiente al Diseño de Medición: D/OI.

Fuente de varianza	*Varianza de diferenciación*	*Fuente de varianza*	*Varianza del Error Relativo*	*% relativo*	*Varianza del Error Absoluto*	*% absoluto*
D	0.29241		.....		.....	
	.....	O:D	0.06171	83.0	0.06171	77.1
	.....	I	.....		0.00572	7.1
	.....	DI	0.00452	6.1	0.00452	5.7
	.....	OI:D	0.00809	10.9	0.00809	10.1
Suma de varianzas	0.29241		0.07432	100%	0.08004	100%
Desviación standar	0.54075		ES Relativo:	0.27261	ES Absoluto:	0.28291
Coef_G relativo	0.80					
Coef_G absoluto	0.79					
	Gran media para los niveles usados:			3.96381		
	Varianza de error de la media para los niveles usados :			0.00815		
	Error Standar de la gran media:			0.09027		

D= Docentes Clínicos; O:D= Residentes Observadores anidados en los docentes; I= Ítems; DI= interacción entre los docentes y los ítems; OI:D= interacción entre los observadores anidados en los docentes y los ítems. ES**:** error estándar de la medición (expresado en las mismas unidades de la escala)

2. ¿En qué medida las evaluaciones hechas por los residentes son generalizables a través del uso de los ítems del IEDDC?

Se configuró un plan de observación con 4 facetas: Hospitales (H), Grupos de residentes como observadores anidados en Hospitales (G:H), Residentes como observadores anidados en los grupos y en los hospitales (O:G:H) y los Ítems (I), todas de universo infinito. El plan de medición fue O:G:H/I. Los grupos de residentes como observadores (G) estuvieron conformados por 4 residentes para cada DC a ser observado. Para este análisis se tuvieron en cuenta los resultados de la evaluación de ambos hospitales por separado. Como fue comentado anteriormente, el programa EduG requiere muestras balanceadas para llevar adelante el análisis. Por lo tanto, se balanceó la cantidad de DC evaluados y de evaluaciones en cada Hospital en forma aleatorizada. Como resultado, la muestra quedó conformada por 48 docentes por hospital con cuatro valoraciones cada uno. Aquí el Coeficiente G absoluto fue de 0,79 y al eliminar el ítem ¨desempeño global¨ el Coeficiente G fue de 0.9. Los resultados de este análisis mostraron evidentes diferencias al tener en cuenta un modelo

con 15 ítems de otro con 14 ítems (este excluye la valoración global). De acuerdo con la tabla de ANOVA, la varianza de la faceta de diferenciación (diferencias sistemáticas entre los observadores anidados en los grupos y hospitales) fue similar tanto para el modelo de 14 como el de 15 ítems (0.43 y 0.39 respectivamente). El efecto constante del universo de los ítems, determinado por una varianza del error absoluto, fue muy distinto según se consideren 14 o 15 ítems (de 0.08 (71% del total) vs 0.006 (15.2% del total) respectivamente). Como consecuencia de los resultados de la varianza de los ítems, las interacciones representaron el 30% de la varianza del error absoluto en el modelo de 14 ítems y casi el 84% al considerar los 15 ítems. En la tabla 13 se muestran las diferencias en los resultados del estudio G al considerar el modelo de 14 o 15 ítems.

En el apéndice 10 se presentan los resultados de los estudios G completos de ambos modelos con la estrategia de medición O:G:H/I.

Tabla 13. Diferencias en los resultados del estudio G al considerar el modelo de 14 o 15 ítems. Estrategia de medición O:G:H/I..

Fuente de la varianza	Varianza de Diferenciación		Fuente de la varianza	Varianza del Error Absoluto	% absoluto	Varianza del Error Absoluto	% absoluto
Modelo	**15 items**	**14 ítems**		**15 ítems**		**14 ítems**	
H	(0.000)	(0.000)			.....		
G:H	0.166	0.149			.....		
S:G:H	0.265	0.250			.....		
		.....	**I**	**0.079**	**70.7**	**0.006**	**15,2**
		.....	**HI**	(0.000)	0.0	(0.000)	0.0
		.....	**GI:H**	0.003	3.2	0.003	8,7
		.....	**SI:G:H**	**0.029**	**26.2**	**0.020**	**76,1**
Suma de Varianzas	0.432			**0,112**	100%	0,039	100%
Desviación Estándar	0,657			**ES absoluto**:	**0,335**	**ES absoluto:**	**0,199**
Coef_G absoluto			***0,79***	***0,91***			
Gran Media para los niveles usados			4,323	4,05			
Varianza del error para la media			0,082	0,008			
ES de la Gran Media			**0,286**	**0,091**			

H= Hospitales, corresponde a los resultados obtenidos de la evaluación realizada en el HIBA y el CEMIC. G:H= grupos de evaluadores anidados en Hospitales, corresponde a los 48 grupos formados por 4 evaluadores en cada centro; O:G:H= Residentes Observadores anidados en los Grupos y Hospitales; I= Ítems; HI= interacción entre los Hospitales y los ítems; GI:H= interacción entre los grupos anidados en los hospitales y los ítems; OI:G:H= interacción entre los residentes observadores anidados en los grupos y hospitales con los ítems

Las dos columnas tituladas "Varianza del error absoluto" muestran cómo se dividen las fuentes de varianza en función del diseño de medición ya sea para el modelo de los ítems completos o extrayendo el correspondiente a la valoración global. Se puede al gran peso aportado a la varianza del error del ítem 15. También se pueden apreciar las diferencias existentes en los valores del error absoluto (expresado en la misma escala que la utilizada en el cuestionario) y en los coeficientes G.

.

3. ***¿Cuál es el número de residentes necesarios para una evaluación confiable de los DC en cada unos de los dominios del IEDDC y del desempeño global, para proveerles un feedback adecuado?***

Mediante un diseño de optimización de los resultados con el plan diseñado D/I(O:D) se propusieron diferentes escenarios de observación con el propósito de determinar la cantidad mínima necesaria de evaluadores para una confiable medición de cada dominio y la valoración global. Los resultados demostraron que para obtener evaluaciones de cada DC con coeficientes de confiablidad mayores a 0.7 en todos los dominios, son necesarias por lo menos 6 evaluaciones por docente. Ver tabla 14.

Tabla 14. Coeficientes G absolutos para cada uno de los Dominios y el Desempeño global

Evaluadores	*4*	*5*	*6*	*8*	*10*	*12*
Modelado	0.72*	0.75	0.78	0.81	0.83	0.85
Coaching	0.70*	0.73	0.76	0.80	0.82	0.83
Articulación	0.69	0.72*	0.75	0.79	0.81	0.82
Exploración	0.34	0.68	0.71*	0.76	0.80	0.82
Clima seguro	0.68	0.72*	0.74	0.77	0.79	0.81
Valoración Global del desempeño	0.77*	0.81	0.84	0.87	0.89	0.91

En la columna de cuatro evaluadores se presentan los coeficientes G que resultaron del análisis general. A continuación los resultados de la optimización con creciente número de evaluadores.
Considerando un valor 0.70 los asteriscos señalan el número mínimo de evaluadores para cada dominio

4. Evidencia basada en la Relaciones con otras variables. Análisis de las correlaciones entre de cada uno de los factores y el juicio sobre el desempeño global

Se analizaron las correlaciones de cada uno de los factores y la puntuación del desempeño global de docente. Todas los índices de las correlaciones fueron altas y significativas aportando mayor solidez al testeo de la validez del instrumento. Realizado con SPSS. Ver tabla 15.

Tabla 15. Correlaciones entre cada uno de los factores y la valoración global

Factores		*Valorac Global*	*Coaching*	*Modelado*	*Articulación*	*Exploración*	*Clima Seguro*
V. Global	CC	1	,720**	,772**	,734**	,715**	,695**
	p		,000	,000	,000	,000	,000
Coaching	CC	,720**	1	,786**	,809**	,806**	,815**
	p	,000		,000	,000	,000	,000
Modelado	CC	,772**	,786**	1	,747**	,765**	,759**
	p	,000	,000		,000	,000	,000
Articulación	CC	,734**	,809**	,747**	1	,854**	,683**
	p	,000	,000	,000		,000	,000
Exploración	CC	,715**	,806**	,765**	,854**	1	,682**
	p	,000	,000	,000	,000		,000
Clima seguro	CC	,695**	,815**	,759**	,683**	,682**	1
	p	,000	,000	,000	,000	,000	

**. La correlación es significativa al nivel 0,01 (bilateral).El N analizado en todos los casos fue de 151 docentes. CC: coeficiente de correlación de Pearson

Discusión

El objetivo de nuestra investigación fue realizar la adaptación transcultural de un instrumento de evaluación de la competencia de los docentes clínicos y su posterior validación para su uso en nuestro medio.

La adaptación transcultural de un cuestionario es un proceso más complejo que la simple traducción de un texto(Arribas 2006; Fernandez 2008). Para un correcto proceso de traducción y adaptación de mismo nos basamos en las ¨Directrices para la traducción y adaptación de pruebas¨ de ¨The International Test Commission¨(Commission 2005). Buscamos asegurar la adecuada conservación de las equivalencias lingüísticas entre las culturas (anglosajona e hispana), considerando las más importantes a la conceptual (significado y contenido), la técnica (sintaxis y gramática) y la de lectura y comprensión.

Sin una apropiada adecuación y posterior validación del cuestionario, es posible incurrir en alguno de los sesgos más frecuentes(van de Vijver and Poortinga 1997). Particularmente, tuvimos en cuenta evitar el sesgo de *constructo*, que en nuestro caso significa la capacidad del docente de favorecer el “aprendizaje cognitivo”. Esto fue evaluado mediante un AFC, al que me referiré luego.

El sesgo metodológico se presenta durante la realización de la investigación propiamente dicha y puede estar dado por tres elementos fundamentales: las características de la muestra elegida para hacer la investigación, las características del instrumento en sí mismo, que pueden provocar errores de interpretación; y la implementación de la prueba.

El sesgo dependiente de la muestra elegida para la investigación, aparece cuando los individuos poseen características diferentes que les impida interpretar, del mismo modo al estudio original, el significado de la pregunta. En nuestro caso buscamos determinar si el instrumento creado para ser utilizado por estudiantes para evaluar DC de Holanda podría mantener la validez y confiabilidad al ser usado por médicos residentes para evaluar DC de habla hispana. En general, los problemas provocados por las diferencias entre las individuos que utilizan un cuestionario surge cuando los mismos difieren en sus capacidades de comunicación, o lectura y comprensión de texto, o respecto de

propiedades relacionadas a cuestiones sociales o culturales que puedan alterar la interpretación de las preguntas(Streiner and Norman 2008). Este sesgo se presenta especialmente cuando los que responden son pacientes, entre quienes es más probable que existan diferencias en el nivel de educación(van de Vijver, Hofer et al. 2010). En nuestra investigación, los responsables de completar los cuestionarios poseen características socio culturales y nivel de educación similares. En la investigación de la Universidad de Maastricht(Stalmeijer, Dolmans et al. 2010) los evaluadores fueron alumnos avanzados de la carrera de medicina, en cambio en nuestro estudio los evaluadores fueron médicos residentes en diferentes etapas de su formación. Consideramos que ambas poblaciones poseen capacidades similares para interpretar las preguntas del cuestionario y por lo tanto, esta no parece ser una fuente de sesgo. Además, al evaluar la validez aparente entre un grupo de residentes, ninguno de los ítems se prestó a confusión.

Respecto al objeto de evaluación, los DC, no parecen existir diferencias significativas. Los DC de Holanda cumplen con tareas asistenciales y docentes en hospitales universitarios similares a las que cumplen los DC en los que se realizó el estudio en nuestro país. Por otro lado, del mismo modo a lo que ocurre en nuestro país, los médicos holandeses no poseen una formación docente formal homogénea(Fluit C. R, Scherpbier J. A et al. 1997).

El sesgo de ítem, se refiere a que alguna de las preguntas no corresponda en la nueva cultura a investigar (Arribas 2006; van de Vijver, Hofer et al. 2010), situación que fue controlada a través de la validación de contenido llevada a cabo por los expertos en docencia del CIEM que calificaron a todos los ítems del instrumento como relevantes para la población en la cual se realizaría la prueba.

La escala de puntuación también fue rápidamente aceptada y comprendida, tanto por los integrantes del comité de expertos como por los residentes.

Con relación a las características del instrumento, el mismo fue confeccionado manteniendo la misma estructura conceptual y gráfica del original.

Respecto a la implementación de la prueba, si bien existen algunas diferencias particulares, en lo general fueron equivalentes. En Holanda se distribuyeron los

cuestionarios entre alumnos avanzados de la carrera. Los nombres de los alumnos fueron escogidos de una base de datos de la facultad y a cada uno de ellos se le envió un cuestionario con una nota explicativa de los alcances de la investigación. En nuestro caso, la elección de los evaluadores fue similar, sin embargo el contacto con los mimos fue personal. En ambas investigaciones la participación fue voluntaria, los evaluadores elegían los DC a evaluar, y en nuestro caso los sobres con las respuestas fueron entregadas en forma anónima en lugar designado por los propios participantes. Consideramos que nuestra estrategia facilitó el entendimiento del propósito de la investigación y por lo tanto el compromiso de los evaluadores con la misma, disminuyendo de este modo la pérdida de observadores. Creemos que esta diferencia metodológica no es una causante de sesgo ya que no hubo ninguna interferencia durante el llenado del cuestionario.

Concluimos que las pruebas aportadas respecto a la traducción del instrumento son suficientes para demostrar que el proceso fue el apropiado y, el resultado es el correcto para evitar los sesgos mencionados.

Una vez finalizada la adaptación transcultural, procedimos a verificar las propiedades psicométricas del cuestionario y a compararlo con el original. Se buscaron evidencias para aportar validez de cuatro de los cinco criterios definidos por la Asociación Americana de Investigación en Psicológica y Educación (American-Education-Research-Association, American-Psychological-Association et al. 1999).

Nuestra investigación buscó sumar validez de contenido a la ya aportada por otras dos fuentes. Una investigación previa de Stalmeijer y col. aportó validez de contenido del instrumento, en la enseñanza clínica en pregrado(Stalmeijer, Dolmans et al. 2008; Stalmeijer, Dolmans et al. 2009). En ese estudio, realizado en la facultad de medicina de Maastricht, estudiantes avanzados trabajando en grupos focales consideraron que los métodos de enseñanza del aprendizaje cognitivo eran observables y viables durante las prácticas clínicas. Además, todos los involucrados en la DC, incluidos los médicos, educadores y estudiantes de medicina calificaron del modo más alto la relevancia de los ítems del instrumento. Más recientemente estudiantes y docentes de la facultad

de veterinaria de la misma universidad (Boerboom, Dolmans et al. 2011), coincidieron que los ítems que evalúan la competencia de los DC eran útiles y pertinentes del aprendizaje cognitivo, durante las diferentes rotaciones de corta duración en la escuela de veterinaria. En nuestro caso, tanto los expertos del CIEM como los residentes entrevistados, valoraron a los ítems y a los diferentes modos de aprendizaje como adecuados y, a las conductas a observar simples y apropiadas para medir la competencia de los DC respecto de la teoría cognitiva del aprendizaje.

Para aportar evidencia a la validez de constructo realizamos un AFC con los datos de los 151 DC evaluados al menos por 4 residentes durante sus rotaciones, en dos hospitales universitarios de la ciudad de Buenos Aires. A diferencia de lo que ocurrió en el estudio original(Stalmeijer, Dolmans et al. 2010), nuestros datos no mostraron normalidad multivariada, sin embargo, es común que esto ocurra cuando se trabaja con variables ordinales que utilizan escalas de Likert. Esto no acarrea dificultad para el análisis ya que métodos como Diagonally Weighted Least Squares (DWLS) y el Unweighted Least Squares (ULS), no necesitan que se cumpla este supuesto(Arias 2008). Este método trabaja en forma estable con poco n, incluso mejor que los métodos actualmente más usados, tales como Maximum Likelihood ó Máxima Verosimilitud (ML) (Ximénez M.C 2005; Forero, Maydeu-Olivares et al. 2009). También utilizamos de modo complementario el método Unweighted Least Squares (ULS). Este método no es usado habitualmente como método ya que puede cambiar los resultados al variar la escala de medida (Forero, Maydeu-Olivares et al. 2009). En el otro estudio de validación realizado sobre el mismo instrumento (Boerboom, Dolmans et al. 2011), los investigadores asumieron directamente a los datos como no normales y eligieron como método estadístico ULS.

Determinamos los índices de ajuste según los ítems correspondan a uno, tres o cinco factores latentes del modelo teórico del aprendizaje cognitivo, con el método DWLS. El modelo con cinco factores sugerido en la validación original(Stalmeijer, Dolmans et al. 2009), mostró valores de ajuste correctos en

todos los índices. Del mismo modo a lo que sucedió en nuestro caso, el modelo de 5 factores latentes también obtuvo la mejor evidencia de contenido durante la confección del instrumento(Stalmeijer, Dolmans et al. 2008) y mejor evidencia de validez de constructo en dos estudios anteriores(Stalmeijer, Dolmans et al. 2010; Boerboom, Dolmans et al. 2011). De un modo diferente a la estrategia realizada en nuestra investigación, Boerboom y col. inicialmente buscaron validar el modelo original de 8 factores latentes y 30 ítems; de igual manera a lo que sucedió en el estudio de Stalmeijer y col., el AFC mostró muy mal ajuste a la mayoría de los índices. Luego, de eliminar los ítems sobrantes, el análisis con el modelo de 5 factores latentes y 15 ítems mostró excelentes índices de ajuste.

Nuestros resultados sumados a los de las dos investigaciones previas refuerzan el valor del constructo de este modelo.

Es de destacar también que los Coeficientes estandarizados (las cargas) de cada indicador sobre las variables latentes fue bastante superior a lo mínimo recomendado por la literatura, reflejando la consistencia de la estructura factorial. Por otro lado los errores de medida de los indicadores (la parte de la varianza no explicada por el modelo de cada indicador), no alcanzó significancia en varios casos, pudiendo dar cuenta de cierta inestabilidad.

Para agregar mayor confianza en la estabilidad "de hecho" del Modelo, ejecutamos otros dos análisis. Realizamos AFC dividiendo la base de datos en dos mitades, de dos maneras diferentes, primero en pares e impares, y luego en primera mitad y segunda mitad. Este tipo de análisis, analizando diferentes partes de la base de datos, también fue llevado a cabo en el estudio original, pero dividiendo la base de una sola manera. El análisis lo realizamos con el método ULS y en todos los casos los índices mostraron un buen ajuste a los datos, y los valores resultaron ser estables en las diferentes mitades.

También se puede evaluar, como ya dijimos en la Sección: "Validez de constructo mediante un análisis factorial confirmatorio", la consistencia en la estabilidad del modelo, observando los resultados congruentes de los índices de ajuste, al usar sobre la base completa, dos métodos que trabajan analíticamente de modo diferente (DWLS y ULS).

Finalmente, y en forma complementaria, efectuamos un estudio de las relaciones entre los diferentes factores latentes y el juicio global, mediante un Modelo de Ecuaciones Estructurales. De un modo similar a lo que ocurre en el estudio original, la valoración global del desempeño del DC se ve afectada por el "Modelado", el "Clima del aprendizaje" y la "Articulación", aunque este último lo hace en menor grado que en el estudio original. Así como en el trabajo original de Stalmeijer (2010), ni el "Coaching" ni la "Exploración" repercuten fuertemente en el juicio global. La "Articulación" incide sobre la "Exploración" de manera muy parecida en ambos estudios. En nuestro estudio, el dominio "Clima del aprendizaje" actúa sobre "Coaching" de forma significativa, a diferencia del trabajo original que lo hace mínimamente. Por otro lado, en nuestro estudio el "Modelado" afecta a "Coaching" aunque menos que en el estudio original. El "Modelado" y "El clima del aprendizaje" correlacionan muy análogamente en ambos estudios. Estas semejanzas y diferencias deberían ser objetivo de futuros estudios.

También, valoramos la correlación entre cada uno de los factores latentes y la valoración global del desempeño docente (Juicio Global), demostrándose en todos los casos buenos valores de Correlación.

Dados los resultados obtenidos fundamentalmente en el AFC, pero también en el estudio de Correlaciones y en el de Ecuaciones Estructurales, damos por justificado el aporte de evidencia de constructo y de relaciones con otras variables para el IEDDC en español.

Evaluamos la confiabilidad mediante el método de consistencia interna, a través de la correlación entre los ítems, con el coeficiente Alfa de Cronbach. Las correlaciones entre cada uno de los ítems y el total fue alta en cada uno de los dominios; siendo la más baja el ítem que valora el juicio global y las más altos las correspondientes al dominio articulación. Para el caso del juicio global, es esperable que la correlación sea baja dado que mide un constructo diferente; cuando el mismo es retirado del análisis el valor alfa en las dos muestras de hospitales analizados alcanza su valor más alto. De un modo diferente, los ítems del dominio exploración, no solo forman un dominio del constructo, sino que parecen tener un poder discriminativo mayor que el resto.

Coincidentemente, los ítems del dominio ¨exploración¨ además de tener el valor α mas elevado (0.94) son aquellos en los cuales los DC obtuvieron una menor puntuación promedio (ver puntajes promedio en el Apéndice 11). Si bien es necesaria una investigación particular para fundamentar estos resultados, una posible explicación es el hecho que este modo de enseñanza requiere de una experticia muy elevada del DC. La exploración, implica haber preparado al alumno a reconocer sus propios problemas a resolver y retirar en forma progresiva el apoyo antes brindado, por lo cual, este puede ser un marcador de deficiencias en los procesos que le preceden.

En todos los dominios los valores α de Cronbach fueron significativos y concordantes con los resultados obtenidos en las otras dos validaciones este mismo instrumento.(Stalmeijer, Dolmans et al. 2010; Boerboom, Dolmans et al. 2011)

Como última y muy importante consecuencia de esta investigación; los resultados obtenidos del estudio de Generalizabilidad, aportan evidencia definitiva sobre la confiabilidad del IEDDC. Mediante dos estrategias de diseño de análisis de los datos, respondimos las preguntas respecto de la medida de generalización de nuestros resultados a las fuentes de error y al número necesario de evaluadores residentes para una evaluación confiable del desempeño del DC. Pudimos demostrar que los resultados de la evaluación del desempeño de los DC obtenidas a través de una valoración hecha por los residentes usando el IEDDC tiene un adecuado nivel de exactitud ya que el coeficiente G absoluto fue de 0.79 y la varianza de los DC fue del 29%. Con estos resultados podemos considerar que la variabilidad existente en las observaciones responde a verdaderas diferencias entre los DC y que la variabilidad dependiente del error es muy baja, en su mayoría dependiente del modo en el cual los residentes evalúan la competencia docente de sus DC.

La pequeña variabilidad dependiente de los ítems posiblemente sea consecuencia de que el constructo del aprendizaje cognitivo este muy bien representado por los ítems elegidos. La varianza del error absoluto relacionada con la interacción entre docentes e ítems y entre observadores e ítems fue baja (15% del total).

Respecto a la medida en la cual se pueden generalizar las evaluaciones llevadas a cabo por los residentes a través del uso de los ítems del IEDDC, también es alta. Sin embargo, hallamos fuertes diferencias entre el modelo que incluía los 15 ítems frente al de 14 (es decir, sin la valoración global). En ambos casos la varianza aportada por la fuente de diferenciación (O:G:H) fue baja y muy similar, siendo consecuencia de escasas diferencias entre los diferentes hospitales, grupos y observadores que constituyeron este análisis. En cambio, las diferencias en la varianza del error absoluto estuvieron relacionadas a la diferencias en la faceta ítem sola y en todas sus interacciones. Este fenómeno se puede explicar por el hecho de que el ítem 15 no está relacionado con el constructo de 5 dominios del modelo cognitivo y claramente disminuye la confiabilidad al ser incluido en el cálculo del coeficiente G {Norman, 2008 #27}.

En cuanto a la cantidad de evaluadores que son necesarios para obtener una valoración confiable del DC, los resultados de la optimización del estudio G nos indican que tan solo 4 observadores son suficientes para evaluar el modelado, coaching y la valoración global, 5 para la articulación y 6 para exploración. Estos resultados son diferentes a los observadores necesarios calculados en el estudio de {Stalmeijer, 2010 #17} con estudiantes de medicina, o en el de Boerboom, Dolmans et al. (2011) con estudiantes de veterinaria durante cortas rotaciones. En estas investigaciones, el número de evaluadores necesarios era mucho más alto, llegando a 14 para el dominio de exploración. Es posible que estas diferencias se deban a una mayor competencia de los residentes para evaluar a sus DC (Beckman, 2004).

Concluimos que esta investigación cumplió con todos los requisitos formales para realizar una correcta adaptación transcultural del instrumento de evaluación del desempeño del docente clínico y que aporta validez basada en el contenido, en el proceso de respuesta, en la consistencia interna y en la confiabilidad y, en las relaciones con otras variables. Futuras líneas de investigación podrían estar destinadas a aportar evidencia sobre el impacto de la utilización de este instrumento sobre el desempeño docente luego de brindar feedback basado en los resultados de evaluaciones hechas con el IEDDC.

Limitaciones:

Una consideración merece el hecho de haber realizado la adaptación transcultural de un instrumento publicado en inglés y no desde el idioma utilizado en el estudio original. En la universidad de Maastricht, si bien en una de las carreras de medicina el idioma oficial es el inglés, el instrumento se usó en holandés. Otra de las investigaciones realizada en la misma universidad, entre alumnos de la carrera de veterinaria, también fue publicada en inglés y tomada en holandés. Hasta donde conocemos, por haber revisado la bibliografía y consultado a la investigadora del instrumento original, la Dra. Renée Stalmaijer, no existen publicaciones en idioma holandés de este instrumento.

Otra de las limitaciones es no haber podido balancear las muestras de los DC y de los residentes como observadores antes del análisis de los datos. La eliminación de los DC con menos de cuatro evaluaciones y la eliminación de las evaluaciones que excedían a cuatro por DC fueron elevadas. Sin embargo, vale aclarar que la eliminación de los DC no fue causante de sesgo para nuestra investigación ya que los mismos se encontraban repartidos en igual proporción entre ambos hospitales, y entre los servicios de los mismos. En relación a las evaluaciones eliminadas, las mismas se realizaron de modo aleatorizado como sugiere la bibliografía, por lo tanto no lo consideramos una causa de sesgo de selección.

Finalmente, si bien pudimos aportar algunas evidencias a la validez, se requiere poder correlacionar los resultados del IEDDC con otros instrumentos válidos y que midan el mismo constructo. Futuras investigaciones podrían relacionar los resultados obtenidos con este instrumento con otros similares como el cuestionario de Stanford el de la Cleveland Clinic.

Aplicación educativa de los resultados de la tesis:

Es sabido que los DC de nuestro medio cuentan con escasas herramientas docentes; en la mayoría de los casos apelan al modelado y a la provisión de información; y si bien para muchos residentes esto es más que suficiente para aprender, esta realidad podría mejorar significativamente. Con una buena estrategia docente, muchos estudiantes podrían aprender de un modo superior, reflexionado sobre su acción, buscando objetivos de aprendizaje propios y comprendiendo que no hay una única manera de llegar a la solución de los problemas de los pacientes.

Esta herramienta de evaluación, al favorecer un feedback específico, podría estimular a los DC a mejorar su desempeño y lograr que los futuros especialistas sean mejores profesionales.

Contents

FUNDAMENTACIÓN: 1

OBJETIVO DEL ESTUDIO 16

REALIZAR UNA VALIDACIÓN TRANSCULTURAL DEL INSTRUMENTO DE EVALUACIÓN DEL DESEMPEÑO DEL DOCENTE CLÍNICO DE MAASTRICHT AL IDIOMA ESPAÑOL.....16

OBJETIVOS ESPECÍFICOS:.....16

- ✓ OBTENER UNA VERSIÓN VÁLIDA DEL IEDDC EN ESPAÑOL 16
- ✓ APORTAR EVIDENCIA PARA LA VALIDEZ Y DETERMINAR LA CONFIABILIDAD. 16

MATERIALES Y MÉTODO. 16

- DE LA ADAPTACIÓN TRANSCULTURAL DEL INSTRUMENTO. 16
- DEL APORTE DE EVIDENCIA SOBRE LA VALIDEZ Y DETERMINACIÓN DE LA CONFIABILIDAD DEL INSTRUMENTO. 17
  1. ***Evidencia basada en el Contenido del test:*** *.....17*
  2. ***Evidencia basada en el Proceso de Respuesta:*** *.....18*
  3. ***Evidencia basada en la estructura interna del test y confiabilidad:*** *.....18*
  4. ***Evidencia basada en la Relaciones con otras variables:*** *.....19*
- RESPECTO DE LAS CARACTERÍSTICAS DEL INSTRUMENTO DE EVALUACIÓN. 19
- CON RELACIÓN A LOS MÉDICOS RESIDENTES (EVALUADORES): 20
- CARACTERÍSTICAS DE LOS DOCENTES (EVALUADOS): 20
- RESPECTO DE LA PRESERVACIÓN DE LA INTEGRIDAD DEL ANONIMATO.....21
- DETERMINACIÓN DE N DE EVALUACIONES. 21
- **COMITÉ DE INVESTIGACIÓN EN EDUCACIÓN MÉDICA (CIEM)**:.....22
- PRUEBAS ESTADÍSTICAS:.....22
- COMITÉ DE ÉTICA DE PROTOCOLOS DE INVESTIGACIÓN: 23

RESULTADOS: 24

I. OBJETIVO 1: OBTENER UNA VERSIÓN VÁLIDA DEL IEDDC EN ESPAÑOL:.....24

- *Traducción Directa.....24*
- ***Traducción inversa o Retrotraducción*** *(del español al inglés).....25*

II. OBJETIVO 2: APORTE DE EVIDENCIA SOBRE LA VALIDEZ Y DETERMINACIÓN DE LA CONFIABILIDAD DEL IEDDC EN NUESTRO MEDIO.....26

1. *Aporte de evidencia para la validez de contenido:.....26*
2. *Aporte de evidencia del proceso de respuesta.....26*

- ***Características de la administración de la prueba:*** *.....27*
- *Recolección de la muestra: 27*

3. *Aporte de validez basada en la estructura interna del test y de la confiabilidad. 28*

- *Validez de constructo mediante un análisis factorial confirmatorio. 28*
- *Análisis de las relaciones causales mediante el modelo de ecuaciones estructurales. 31*

Determinación de la confiabilidad.....33

- *Consistencia Interna - Cálculo del coeficiente Alfa de Cronbach. 33*
- *Análisis de Generalizabilidad. 36*

4. *Evidencia basada en la Relaciones con otras variables. Análisis de las correlaciones entre de cada uno de los factores y el juicio sobre el desempeño global.....41*

DISCUSIÓN **42**

LIMITACIONES: **50**

APLICACIÓN EDUCATIVA DE LOS RESULTADOS DE LA TESIS: **51**

Tabla 1. Resultados del análisis factorial confirmatorio del Instrumento de evaluación Original de Maastricht con uno a cinco factores latentes. 13
Tabla 2. Componentes de la varianza del cuestionario IEDDC para cada uno de los factores y para el Juicio global * 14
Tabla 3. Guía orientadora para el proceso de adaptación transcultural de instrumentos 17
Tabla 4. Recomendaciones sobre el tamaño muestral mínimo cuando hay Normalidad Multivariante 22
Tabla 5. Test de valoración de la normalidad multivariada para datos continuos 28
Tabla 6. Índices de ajuste del Análisis Factorial Confirmatorio con 1, 3 y 5 factores latentes (método de estimación: DWLS). 29
Tabla 7. Índices de ajuste del Análisis Factorial Confirmatorio con modelos alternativos (método de estimación ULS) 31
Tabla 8. Resultados de ajuste del AFC y MEE 31
Tabla 9. Análisis de la confiablidad de la muestra mediante el Alfa de Cronbach para cada uno de los factores 33
Tabla 10. Cronbach: Estadísticos de todos los elementos. Muestra del HIBA 34
Tabla 11. Cronbach: Estadísticos de todos los elementos. Muestra del CEMIC 35
Tabla 12. Tabla del Estudio G correspondiente al Diseño de Medición: D/OI. 37
Tabla 13. Diferencias en los resultados del estudio G al considerar el modelo de 14 o 15 ítems. Estrategia de medición O:G:H/I.. 39
Tabla 14. Coeficientes G absolutos para cada uno de los Dominios y el Desempeño global 40
Tabla 15. Correlaciones entre cada uno de los factores y la valoración global 41

Figura 1. Análisis de la relaciones entre factores latentes mediante modelo de ecuaciones estructurales (trabajo original)(Stalmeijer, Dolmans et al. 2010) 15
Figura 2. Relaciones causales entre los diferentes modos de enseñanza (factores latentes) del Instrumento de Evaluación del Desempeño del Docente Clínico de Maastricht y la valoración global. 32

APÉNDICES

APÉNDICE 1: CUESTIONARIO DE EVALUACIÓN DEL DESEMPEÑO DEL DOCENTE CLÍNICO DE MAASTRICHT ...II

APÉNDICE 2. CARTA DE PRESENTACIÓN A LOS MÉDICOS RESIDENTES III

APÉNDICE 3. TRADUCCIONES DIRECTAS DEL INGLÉS AL ESPAÑOL. IV

TRADUCCIÓN DIRECTA REALIZADA POR TRADUCTOR ¨COMPROMETIDO¨ IV
TRADUCCIÓN DIRECTA REALIZADA POR TRADUCTOR ¨INOCENTE¨ V
DOCUMENTO ENVIADO A LOS TRADUCTORES VI

APÉNDICE 4. TRADUCCIÓN DIRECTA UNIFICADA Y CONSENSUADA POR EL CIEM SOBRE LA QUE SE HICIERON LAS RETROTRADUCCIONES. VII

APÉNDICE 5. RETROTRADUCCIONES REALIZADAS POR TRADUCTORES DE LENGUA NATIVA INGLESA..... VIII

RETROTRADUCCIÓN 1 VIII
RETROTRADUCCIÓN 2 IX

APÉNDICE 6. DOCUMENTO FINAL CONFECCIONADO A PARTIR DE LA RETROTRADUCCIÓN ELEGIDA POR LA AUTORA DEL TRABAJO Y CONSENSUADO POR EL COMITÉ DE EXPERTOS X

APÉNDICE 7. CUESTIONARIO DEFINITIVO USADO PARA LA VALIDACIÓN XI

APÉNDICE 8. GRÁFICOS CON LA DISTRIBUCIÓN DE EVALUADORES Y DC SEGÚN ESPECIALIDAD XII

GRÁFICO 1: Nº DE EVALUACIONES SEGÚN LA ESPECIALIDAD DEL EVALUADOR XII
GRÁFICO: Nº 2. DE EVALUACIONES SEGÚN LA ESPECIALIDAD DEL DOCENTE XIII

APÉNDICE 9. RESULTADOS DEL ESTUDIO G CORRESPONDIENTE AL DISEÑO DE MEDICIÓN D/I(OD) XIV

APÉNDICE 10. RESULTADOS DEL ESTUDIO G CORRESPONDIENTE AL DISEÑO DE MEDICIÓN HGO/I XV

CON TODOS LOS ÍTEMS INCLUIDOS EN EL IEDDC. XV
CON TODOS LOS ÍTEMS EXCEPTO EL CORRESPONDIENTE A LA VALORACIÓN GLOBAL DEL DESEMPEÑO. XVI

APÉNDICE 11. PUNTAJES PROMEDIO DE TODAS LAS VALORACIONES (AMBOS HOSPITALES) XVII

Apéndice 1: Cuestionario de Evaluación del Desempeño del Docente Clínico de Maastricht

Table 1. Mean ratings and SDs, on a Likert scale (1 = highly irrelevant; 5 = highly relevant), of the relevance of the items relating to eight supervision methods by stakeholder group and for the three stakeholder groups combined and overall group ratings per supervision method.

	Educationalists		Doctors		Students		Total item score
Supervision methods	M	SD	M	SD	M	SD	M
Modelling							
The clinical teacher:							
1. demonstrated how different tasks should be performed.	4.5	0.53	4.5	0.63	3.6	0.80	4.2
2. explained, while performing a task, which aspects were important and why.	4.4	0.84	4.7	0.48	4.2	1.00	4.4
3. created sufficient opportunities for me to observe him or her.	4.4	0.70	4.4	0.51	3.8	0.87	4.2
4. was a role model for me.	4.3	1.10	4.3	0.68	3.2	1.00	3.9
Overall ratings on modelling	4.4	0.41	4.5	0.45	3.7	0.47	
Coaching							
The clinical teacher:							
5. observed me while I was performing a task.	4.9	0.32	4.5	0.63	4.2	0.83	4.5
6. provided me with constructive and concrete feedback during or following direct observation.	4.6	0.70	4.7	0.48	4.5	0.70	4.6
7.* provided me with constructive and concrete feedback during my rotation.	4.3	1.10	4.5	0.90	4.7	0.50	4.5
8. gave me a better insight into aspects of my performance that needed improvement.	4.3	0.80	4.6	0.63	4.0	0.85	4.3
Overall ratings on coaching	4.5	0.50	4.6	0.60	4.3	0.50	
Scaffolding							
The clinical teacher:							
9. adjusted his/her teaching activities to my level of experience and competence.	4.1	1.00	4.2	0.75	3.8	0.62	4.0
10. allowed me to perform tasks that fit my level of experience and competence.	4.3	1.10	4.4	0.63	4.2	1.10	4.3
11. was supportive when I experienced difficulties with a task.	4.0	1.20	4.5	0.73	4.1	0.51	4.2
12. gradually decreased the amount of guidance in order to bolster my independence.	3.4	1.10	4.5	0.52	4.3	0.75	4.1
Overall ratings on scaffolding	4.0	0.66	4.4	0.47	4.1	0.44	
Articulation							
The clinical teacher:							
13. asked me to explain my reasoning and arguments.	4.6	0.70	4.6	0.63	4.5	0.52	4.6
14. alerted me to gaps in my knowledge and skills.	4.2	0.63	4.4	0.62	3.9	1.10	4.2
15. asked questions to increase my understanding.	4.3	0.71	4.6	0.62	4.5	0.70	4.5
16. stimulated me to ask questions to increase my understanding.	4.5	0.71	4.6	0.50	4.0	1.10	4.4
Overall ratings on articulation	4.0	0.50	4.5	0.48	4.4	0.57	
Reflection							
The clinical teacher:							
17. stimulated me to think about my own strengths and weaknesses.	4.4	0.84	4.4	0.62	3.4	0.90	4.1
18. stimulated me to think about how to improve my own strengths and weaknesses.	4.5	0.71	4.6	0.63	3.4	1.10	4.2
19.* stimulated me to compare my own approach to that of a more experienced doctor.	3.9	1.20	3.7	1.10	2.8	0.87	3.4
20.* stimulated me to reflect on the profession of a medical doctor.	3.5	0.85	3.8	1.00	3.0	1.30	3.4
Overall ratings on reflection	4.4	0.66	4.1	0.70	3.1	0.63	
Exploration							
The clinical teacher:							
21. stimulated me to formulate my own goals.	4.7	0.50	4.1	0.85	3.5	0.67	4.1
22. stimulated me to achieve my own goals.	4.8	0.42	4.3	0.45	3.8	0.75	4.2
23. challenged me to explore new tasks and possibilities.	4.5	0.85	3.9	0.72	4.3	0.75	4.2
24.* challenged me to extend my boundaries.	3.6	1.50	3.8	0.83	4.1	0.90	3.8
Overall ratings on exploration	4.5	0.41	4.0	0.59	3.9	0.54	
General Learning Climate							
The clinical teacher:							
25. established a safe-learning environment.	4.4	0.84	4.8	0.45	4.1	0.67	4.4
26. showed an interest in me as a student.	4.4	0.70	4.7	0.48	4.2	0.83	4.4
27. treated me with respect.	4.7	0.50	4.9	0.34	4.3	0.62	4.6
** *took enough time to supervise me.*							
Overall ratings on general learning climate	4.3	0.50	4.8	0.38	4.2	0.56	
Miscellaneous							
28. Give an overall mark (1–10) for this doctor as a clinical teacher.	3.7	1.60	4.2	0.83	3.6	1.20	3.9
29. What are the strengths of this clinical teacher? (open-ended question)	4.7	0.49	4.6	0.62	4.4	0.52	4.6
30. Which aspects of the performance of this clinical teacher can be improved? (open-ended question)	4.7	0.52	4.8	0.40	4.3	0.67	4.6
Overall ratings on miscellaneous	4.3	0.85	4.5	0.36	4.1	0.49	

*= Eliminated item; ** = added item.

En esta tabla se presentan los 30 ítems originales evaluados por DC, estudiantes y educadores con sus correspondientes puntuaciones expresadas como media y DS (ver texto). Resaltados los ítems eliminados (*) y ítem agregado (**)

Apéndice 2. Carta de presentación a los médicos residentes

Bs. As., xx/xxxx

Estimados Doctores:

Nos dirigimos a ustedes para solicitarles su colaboración y participación en el estudio multicéntrico titulado "Validación de un cuestionario para evaluar **Docentes durante la Enseñanza clínica**". El cuestionario original fue creado y validado en Maastricht (Holanda) y actualmente diferentes grupos están trabajando en la validación en otros idiomas.

El alcance del presente estudio es validar el cuestionario en nuestra población, *sin realizar análisis de los resultados ni tomar conductas respecto a ellos*. Los datos obtenidos se manejarán en forma confidencial y cada **Docente** será codificado para garantizar su anonimato. A su vez, el residente "evaluador" (quién completa el cuestionario) lo realizará en forma **anónima** (no deberá colocar su nombre ni su firma en los cuestionarios completados).

. Los invitamos a participar completando **un cuestionario** por cada **docente/tutor (médico)** en la "enseñanza clínica": la que realizan durante la atención del paciente, en la sala y/o consultorio (Aclaración: no se trata de "Enseñanza en aula, clases expositivas, ej. Ateneos, cursos, etc."). Para la elección de los **Docentes** a evaluar solo se deberá tener en cuenta el **tiempo compartido** con ellos, que deberá ser **igual o mayor a 4 semanas** y al menos**, 4 hs diarias**.

Cada residente podrá evaluar un máximo de 2 (dos) **Docentes por rotación**; respondiendo las preguntas volcadas dentro del cuestionario que tratan de explorar el desempeño del docente, calificando cada área con un número del **1 al 5** (1= en **"total desacuerdo",** progresando hasta el nº 5, en**" total acuerdo"**).

Luego de completar los cuestionarios les pedimos que lo devuelvan dentro del sobre cerrado.

Muchas Gracias.

Los saludamos Atte.,

Investigadores

Apéndice 3. Traducciones directas del Inglés al Español.

Se presentan los trabajos originales enviados por los traductores.

Traducción Directa realizada por traductor ¨comprometido¨

Modelado
Demuestra en forma consistente poder llevar a cabo adecuadamente distintas destrezas clínicas.
Generó una suficiente cantidad de oportunidades para que pueda observarlo en su desempeño.
Representa el modelo de médico que yo quisiera ser.

Entrenamiento
Me aportó una devolución constructiva durante o inmediatamente después de observarme en contacto directo con mis pacientes.
Adaptó sus actividades docentes a mi nivel de experiencia.
Me ofreció suficientes oportunidades para realizar actividades en forma independiente.

Implementación
Me estimuló a justificar racionalmente mis acciones.
Me realizó preguntas destinadas a aumentar mi nivel de comprensión.
Me estimuló para que explore mis fortalezas y debilidades.

Exploración
Me incentivó a formular objetivos de aprendizaje.
Me incentivó para cumplir mis objetivos de aprendizaje.

Ambiente de aprendizaje seguro
Generó un ambiente de aprendizaje seguro.
Estaba genuinamente interesado en mí como estudiante.
Demostró que me respetaba.

Puntuación global de la enseñanza médica (puntuación del 1 al 10)

Traducción directa realizada por traductor ¨inocente¨

Modelado

-Demostró consistentemente cómo ejecutar las habilidades clínicas.

-Creó suficientes oportunidades para que yo pueda observarla/lo.

-Sirvió como modelo a seguir del tipo de médico en el que me gustaría convertirme.

Entrenamiento

-Ofreció feedback útil durante o inmediatamente después de observar mis encuentros con pacientes.

-Ajustó sus actividades de enseñanza a mi nivel de experiencia.

-Me ofreció suficientes oportunidades para realizar actividades de forma independiente.

Articulación

-Pidió que explicara el razonamiento de mis acciones.

-Me hizo preguntas focalizadas hacia aumentar mi comprensión.

-Me incentivó a explorar mis fortalezas y debilidades.

Exploración

-Me incentivó a formular mis metas de aprendizaje.

-Me incentivó a alcanzar mis metas de aprendizaje.

Entorno de aprendizaje

-Creó un entorno de aprendizaje seguro.

-Demostró interés genuino en mí como estudiante.

-Demostró que me respetaba

Documento enviado a los traductores

Appendix 1

Mean Score (Scale 1 = Fully Disagree, 5 = Fully Agree) and Corresponding Standard Deviation (SD) per Item of the Maastricht Clinical Teaching Questionnaire and for the Overall Judgment (Scale 1–10)*

Questionnaire item—The clinical teacher …	Mean (1–5)	SD	Alpha-coefficient (α)
Modeling			**.86**
Consistently demonstrated how to perform clinical skills.	3.9	0.45	
Created sufficient opportunities for me to observe him/her.	4.1	0.45	
Served as a role model as to the kind of doctor I would like to become.	3.9	0.57	
Coaching			**.83**
Gave useful feedback during or immediately after direct observation of my patient encounters.	3.8	0.56	
Adjusted his/her teaching activities to my level of experience.	3.9	0.42	
Offered me sufficient opportunities to perform activities independently.	4.2	0.50	
Articulation			**.89**
Asked me to provide a rationale for my actions.	3.8	0.47	
Asked me questions aimed at increasing my understanding.	4.0	0.50	
Stimulated me to explore my strengths and weaknesses.	3.7	0.46	
Exploration			**.94**
Encouraged me to formulate learning goals.	3.5	0.52	
Encouraged me to pursue my learning goals.	3.5	0.54	
Safe learning environment			**.96**
Created a safe learning environment.	4.3	0.46	
Was genuinely interested in me as a student.	4.3	0.56	
Showed that he/she respected me.	4.4	0.46	
Overall judgment of clinical teaching (scale 1–10)	**7.8**	**0.71**	

* The dataset consisted of four or more anonymous ratings for 126 individual clinical teachers.

Apéndice 4. Traducción directa unificada y consensuada por el CIEM sobre la que se hicieron las retrotraducciones.

Elementos del Cuestionario - El profesor o docente clínico

- **Como modelo de rol**

✓ Demostró cómo ejecutar las habilidades clínicas de manera consistente.
✓ Creó suficientes oportunidades para que yo pueda observarla/lo en su desempeño.
✓ Sirvió como modelo a seguir del tipo de médico en el que me gustaría convertirme.

- **Como tutor o guía**

✓ Me aportó una devolución útil durante o inmediatamente después de observarme en contacto directo con mis pacientes.
✓ Ajustó sus actividades de enseñanza a mi nivel de experiencia.
✓ Me ofreció suficientes oportunidades para realizar actividades en forma independiente.

- **Implementación/*Articulación***

✓ Me pidió que explicara el fundamento o motivos de mis acciones
✓ Me realizó preguntas dirigidas a aumentar mi nivel de comprensión.
✓ Me incentivó a explorar mis fortalezas y debilidades.

- **Exploración**

✓ Me incentivó a formular objetivos de aprendizaje.
✓ Me incentivó a cumplir mis objetivos de aprendizaje.

- **Ambiente de aprendizaje seguro**

✓ Generó un ambiente de aprendizaje seguro.
✓ Estuvo genuinamente interesado en mí como estudiante.
✓ Me demostró respeto

Puntuación global de la enseñanza clínica (puntuación del 1 al 10)

Apéndice 5. Retrotraducciones realizadas por traductores de lengua nativa Inglesa.

Retrotraducción 1

Doctor in charge of training - Evaluation Form

Modeling
He/she
Showed me how to carry out clinical tasks consistently.
Provided me with opportunities to observe him/her in their role.
Was a good role model in the area that was a good role model in terms of the kind of doctor that I would like to become

As a Tutor
He/ she
Gave me useful feedback whilst or immediately after observing me attending my patients.
Made sure that my training was at the appropriate level for my experience.
Provided me with enough opportunities to work independently.

Articulation
He/she
Asked me to explain my reasons behind decisions that I took regarding practice.
Asked me specific questions to improve my understanding of the profession.
Encouraged me to explore my own strengths and weaknesses.

Exploration
He/She
Encouraged me to come up with short term and long term learning objectives.
Encouraged me to fulfill these objectives.

Safe Learning Environment
He/she
Provided a safe learning environment.
Was genuinely interested in me as a student.
Treated me with respect.

Global mark for clinical teaching (on a scale from 1-10)

Retrotraducción 2

Elements of the Questionnaire – The Professor or Clinical Teacher

As role model

Demonstrated how to execute clinical skills in a consistent manner

Created sufficient opportunities so that I could observe your performance

Served as a model to follow of the type of doctor that I would like to become

As tutor or guide

Give me useful feedback during or immediately after observing me in direct contact with patients.

Adjusted teaching activities to my level of experience

Offered me sufficient opportunities to do independent activities

Implementation/Articulation

Asked me to explain the background or motives for my actions

Asked me directed questions to increase my level of understanding

Encouraged me to explore my strengths and weaknesses

Exploration

Encouraged me to formulate learning objectives

Encouraged me to meet my learning objectives

Safe Learning Environment

Generated a safe learning environment

Were genuinely interested in me as a student

Showed me respect

Overall rating of the teaching clinic (Rating from 1 to 10)

Apéndice 6. Documento final confeccionado a partir de la retrotraducción elegida por la autora del trabajo y consensuado por el comité de expertos

El Docente Clínico:

… me demostró cómo realizar las habilidades clínicas de manera consistente.

… creó suficientes oportunidades para que yo pueda observarla/lo en su desempeño.

… sirvió como modelo a seguir del tipo de médico en el que me gustaría convertirme.

… me dio una devolución (feedback) útil durante o inmediatamente después de observarme en contacto directo con mis pacientes.

… ajustó sus actividades de enseñanza a mi nivel de experiencia.

… me ofreció suficientes oportunidades para realizar actividades en forma independiente.

… me pidió que fundamentara mis acciones.

… me formuló preguntas dirigidas a aumentar mi nivel de comprensión.

… me incentivó a explorar mis fortalezas y debilidades.

… me incentivó a formular objetivos de aprendizaje.

… me incentivó a cumplir mis objetivos de aprendizaje.

… generó un clima de aprendizaje favorable (cooperativo, no hostil y saludable).

… estuvo interesado en mí como estudiante.

… me demostró respeto

Evalúe el desempeño general de su supervisor como docente. Escriba el n° del 1 a 10.

¿Cuáles son las Fortalezas de este Docente Clínico?

¿En qué aspectos podría mejorar este Docente Clínico?

Apéndice 7. Cuestionario definitivo usado para la validación

Cuestionario de Maastricht para la Enseñanza Clínica (CMEC)

Complete el cuestionario luego de haber finalizado su rotación
Nota: Complete un cuestionario por cada docente clínico.

Nombre y Apellido del Docente Clínico: .....

1-Especialidad del Residente: ..... /2- Especialidad de la Rotación: .....

Este cuestionario tiene como objetivo evaluar la calidad de la enseñanza de su docente clínico. Por favor, responda a cada afirmación, marcando lo que considere apropiado.
Los docentes clínicos recibirán devolución (feedback) anónimo respecto a su desempeño docente, luego de haber sido evaluado por 6 o más estudiantes.

1 = totalmente en desacuerdo,
2 = en desacuerdo,
3 = ni de acuerdo ni en desacuerdo,
4 = de acuerdo y
5 = totalmente de acuerdo

El Docente Clínico:

		1	2	3	4	5
3.	... me demostró cómo realizar las habilidades clínicas de manera consistente.	☐	☐	☐	☐	☐
4.	... creó suficientes oportunidades para que yo pueda observarla/lo en su desempeño.	☐	☐	☐	☐	☐
5.	... sirvió como modelo a seguir del tipo de médico en el que me gustaría convertirme.	☐	☐	☐	☐	☐
6.	... me dio una devolución (feedback) útil durante o inmediatamente después de observarme en contacto directo con mis pacientes.	☐	☐	☐	☐	☐
7.	... ajustó sus actividades de enseñanza a mi nivel de experiencia.	☐	☐	☐	☐	☐
8.	... me ofreció suficientes oportunidades para realizar actividades en forma independiente.	☐	☐	☐	☐	☐
9.	... me pidió que fundamentara mis acciones.	☐	☐	☐	☐	☐
10.	... me formuló preguntas dirigidas a aumentar mi nivel de comprensión.	☐	☐	☐	☐	☐
11.	... me incentivó a explorar mis fortalezas y debilidades.	☐	☐	☐	☐	☐
12.	... me incentivó a formular objetivos de aprendizaje.	☐	☐	☐	☐	☐
13.	... me incentivó a cumplir mis objetivos de aprendizaje.	☐	☐	☐	☐	☐
14.	... generó un clima de aprendizaje favorable (cooperativo, no hostil y saludable).	☐	☐	☐	☐	☐
15.	... estuvo interesado en mí como estudiante.	☐	☐	☐	☐	☐
16.	... me demostró respeto	☐	☐	☐	☐	☐

17. Evalúe el desempeño general de su supervisor como docente. Escriba el n° del 1 a 10. ( )
Tenga en cuenta que una calificación menor a 6 significa insatisfactorio y 10 es excelente.

18. ¿Cuáles son las ***Fortalezas*** de este Docente Clínico?

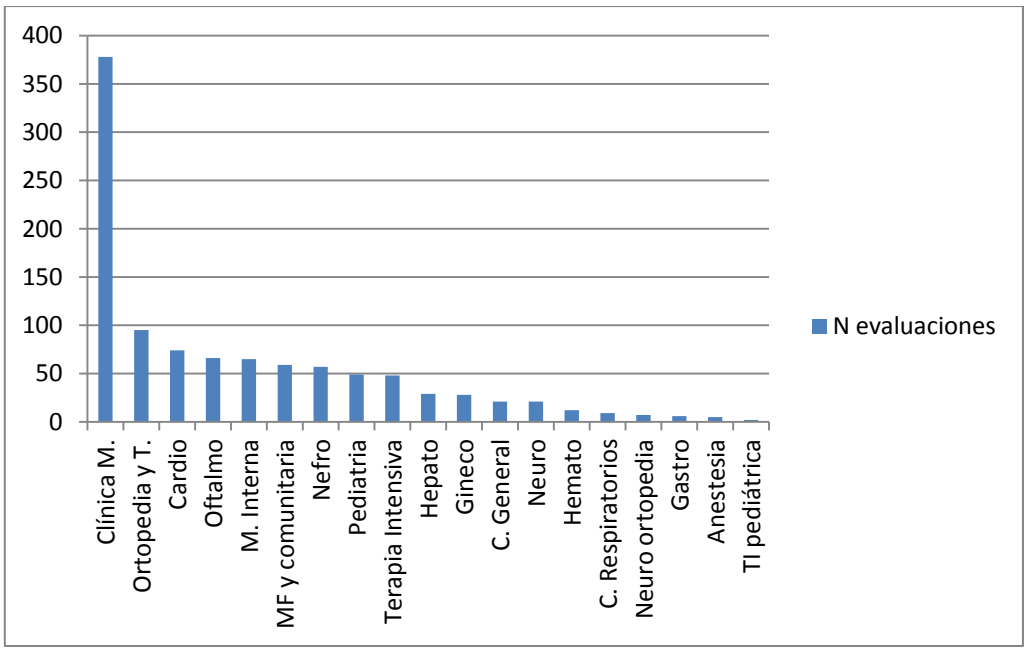
19. ¿En qué ***aspectos podría mejorar*** este Docente Clínico?

Gracias por completar el Cuestionario

Stalmeijer, R. E., et. Maastricht Clinical Teaching Questionnaire (MCTQ) as a Valid and Reliable Instrument for the Evaluation of Clinical Teachers. Academic Medicine 2010; 85(11), 1732-1738.

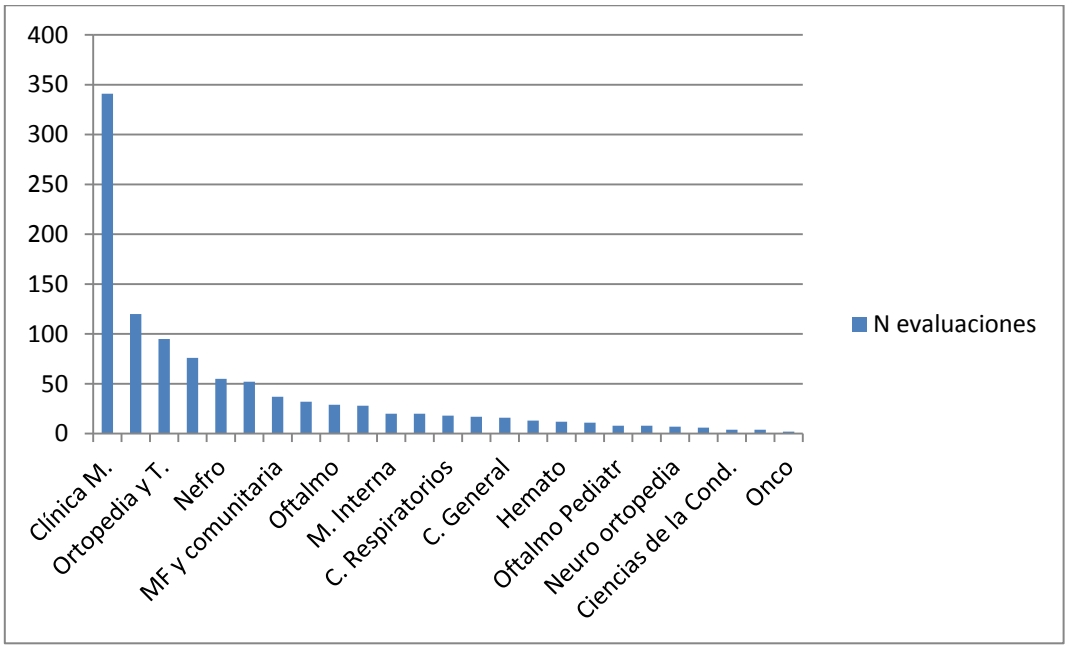
Apéndice 8. Gráficos con la distribución de evaluadores y DC según especialidad

Gráfico 1: nº de evaluaciones según la especialidad del evaluador



En el gráfico se vé la cantidad de evaluaciones realizadas según el servicio del residente evaluador.

Gráfico: nº 2. de evaluaciones según la especialidad del docente



En el gráfico se ve la cantidad de evaluaciones realizadas según el servicio del docente evaluado.

Apéndice 9. Resultados del Estudio G correspondiente al diseño de medición D/I(OD)

Todos los items menos el 15

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
DC	D	151	INF	
RES	O:D	4	INF	
I	I	15	INF	15

Analysis of variance

				Components				
Source	SS	df	MS	Random	Mixed	Corrected	%	SE
D	3080.53382	150	20.53689	0.29241	0.29241	0.29241	25.8	0.04232
O:D	1770.53571	453	3.90847	0.24683	0.24683	0.24683	21.7	0.01852
I	637.90681	13	49.06975	0.08007	0.08007	0.08007	7.1	0.02967
DI	1376.98605	1950	0.70615	0.06332	0.06332	0.06332	5.6	0.00602
OI:D	2666.96429	5889	0.45287	0.45287	0.45287	0.45287	39.9	0.00834
Total	9532.92668	8455					100%	

G Study Table
(Measurement design D/OI)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
D	0.29241		.....		.....	
	.....	O:D	0.06171	83.0	0.06171	77.1
	.....	I	.....		0.00572	7.1
	.....	DI	0.00452	6.1	0.00452	5.7
	.....	OI:D	0.00809	10.9	0.00809	10.1
Sum of variances	0.29241		0.07432	100%	0.08004	100%
Standard deviation	0.54075		Relative SE: 0.27261		Absolute SE: 0.28291	
Coef_G relative	0.80					
Coef_G absolute	0.79					

Grand mean for levels used: 3.96381
Variance error of the mean for levels used: 0.00815
Standard error of the grand mean: 0.09027

Apéndice 10. Resultados del Estudio G correspondiente al diseño de medición HGO/I

con todos los ítems incluidos en el IEDDC.

Análisis con todos los ítems incluidos en el IEDDC.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Hospitales	H	2	INF	
grupos	G:H	48	INF	
Supervisores	S:G:H	4	INF	
items	I	15	INF	

Analysis of variance

				Components				
Source	SS	df	MS	Random	Mixed	Corrected	%	SE
H	6.66944	1	6.66944	-0.00276	-0.00276	-0.00276	0.0	0.00203
G:H	1375.72431	94	14.63536	0.16663	0.16663	0.16663	7.8	0.03574
S:G:H	1273.90000	288	4.42326	0.26537	0.26537	0.26537	12.5	0.02450
I	6438.70729	14	459.90766	1.19600	1.19600	1.19600	56.3	0.42344
HI	9.02326	14	0.64452	-0.00006	-0.00006	-0.00006	0.0	0.00119
GI:H	864.36944	1316	0.65682	0.05352	0.05352	0.05352	2.5	0.00685
SI:G:H	1785.10000	4032	0.44273	0.44273	0.44273	0.44273	20.8	0.00986
Total	11753.49375	5759					100%	

G Study Table
(Measurement design HGS/I)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
H	(0.00000)		.....		.....	
G:H	0.16663		.....		.....	
S:G:H	0.26537		.....		.....	
	.....	I	.....		0.07973	70.7
	.....	HI	(0.00000)	0.0	(0.00000)	0.0
	.....	GI:H	0.00357	10.8	0.00357	3.2
	.....	SI:G:H	0.02952	89.2	0.02952	26.2
Sum of variances	0.43200		0.03308	100%	0.11282	100%
Standard deviation	0.65727		Relative SE: 0.18189		Absolute SE: 0.33588	
Coef_G relative	0.93					
Coef_G absolute	0.79					

Grand mean for levels used: 4.32396
Variance error of the mean for levels used: 0.08227
Standard error of the grand mean: 0.28683

con todos los ítems excepto el correspondiente a la valoración global del desempeño.

con 14

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Hospitales	H	2	2	
grupos	G:H	48	INF	
Observadores	O:G:H	4	INF	
items	I	15	INF	15

Analysis of variance

				Components				
Source	SS	df	MS	Random	Mixed	Corrected	%	SE
H	8.84003	1	8.84003	-0.00128	-0.00128	-0.00064	0.0	0.00277
G:H	1173.23065	94	12.48118	0.14908	0.14908	0.14908	15.6	0.03270
O:G:H	1134.25000	288	3.93837	0.25092	0.25092	0.25092	26.2	0.02337
I	429.08185	13	33.00630	0.08483	0.08434	0.08434	8.8	0.03139
HI	5.59226	13	0.43017	-0.00099	-0.00099	-0.00099	0.0	0.00083
GI:H	757.79018	1222	0.62012	0.04864	0.04864	0.04864	5.1	0.00673
OI:G:H	1593.25000	3744	0.42555	0.42555	0.42555	0.42555	44.4	0.00983
Total	5102.03497	5375					100%	

G Study Table
(Measurement design HGO/I)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
H	(0.00000)		.....		.....	
G:H	0.14908		.....		.....	
O:G:H	0.25092		.....		.....	
	.....	I	.....		0.00602	15.1
	.....	HI	(0.00000)	0.0	(0.00000)	0.0
	.....	GI:H	0.00347	10.3	0.00347	8.7
	.....	OI:G:H	0.03040	89.7	0.03040	76.2
Sum of variances	0.39999		0.03387	100%	0.03990	100%
Standard deviation	0.63245		Relative SE: 0.18404		Absolute SE: 0.19974	
Coef_G relative	0.92					
Coef_G absolute	0.91					

Grand mean for levels used: 4.05097
Variance error of the mean for levels used: 0.00835
Standard error of the grand mean: 0.09136

Apéndice 11. Puntajes promedio de todas las valoraciones (ambos hospitales)

Ítem del cuestionario IEDDC	*χ*	*DS*
Modelado: el DC		
...me demostró cómo realizar las habilidades clínicas de manera consistente.	*4.06*	*0.91*
… creó suficientes oportunidades para que yo pueda observarla/lo en su desempeño	*4.03*	*0.96*
… sirvió como modelo a seguir del tipo de médico en el que me gustaría convertirme.	*3.80*	*1.10*
Coaching		
… me dio una devolución (feedback) útil durante o inmediatamente después de observarme en contacto directo con mis pacientes.	*3.67*	*1.18*
… ajustó sus actividades de enseñanza a mi nivel de experiencia.	*4.05*	*0.97*
… me ofreció suficientes oportunidades para realizar actividades en forma independiente.	*4.03*	*1.03*
Articulación		
… me pidió que fundamentara mis acciones.	*3.95*	*1.00*
… me formuló preguntas dirigidas a aumentar mi nivel de comprensión	*4.05*	*1.03*
… me incentivó a explorar mis fortalezas y debilidades.	*3.62*	*1.09*
Exploración		
… me incentivó a formular objetivos de aprendizaje.	*3.55*	*1.12*
… me incentivó a cumplir mis objetivos de aprendizaje	*3.63*	*1.09*
Clima Seguro de Aprendizaje		
… generó un clima de aprendizaje favorable (cooperativo, no hostil y saludable).	*4.28*	*1.01*
… estuvo interesado en mí como estudiante.	*4.10*	*1.03*
… me demostró respeto	*4.60*	*0.82*
Valoración Global del desempeño del DC (1 a 10)	*6.98*	*3.05*

Puntajes obtenidos por todos los docentes de ambos hospitales expresados como promedio y DS

Bibliografía

Aiken, L. R., M. E. O. Salinas, et al. (2003). Confiabilidad y Validez. Tests psicológicos y evaluación, Pearson Educación.

Alhaqwi, A. I., J. Kuntze, et al. (2014). "Development of the clinical learning evaluation questionnaire for undergraduate clinical education: factor structure, validity, and reliability study." BMC Med Educ **14**(1): 44.

American-Education-Research-Association, American-Psychological-Association, et al. (1999). "Standards for Educational and Psychological Testing. ." The American Psychological Association.

Arias, B. (2008). Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS y SAS. Metodología en la investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales. M. C. M. A. Verdugo, M. Badía, & B. Arias. Salamanca, INICO.

Arribas, A. (2006). "Adaptación transcultural de Instrumentos. Guía para el proceso de Validación de instrumentos tipo encuestas." Revista Científica de la AMBB **16**(3): 74-82.

Bandura A (1986). Social Foundations of Thought and Action. A Social Cognitive Theory. Englewood Cliffs. New Jersey, Prentice - Hall.

Beaton, D. E., C. Bombardier, et al. (2000). "Guidelines for the process of cross-cultural adaptation of self-report measures." Spine (Phila Pa 1976) **25**(24): 3186-3191.

Beckman, T. J., A. K. Ghosh, et al. (2004). "How reliable are assessments of clinical teaching? A review of the published instruments." J Gen Intern Med **19**(9): 971-977.

Boerboom, T. B., D. H. Dolmans, et al. (2011). "Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations." Med Teach **33**(2): e84-91.

Bowden, J. and F. Marton (2003). The University of Learning: Beyond Quality and Competence, Taylor & Francis.

Candy PC (1991). Self - direction in Lifelong Learning. San Francisco, CA . Jossey - Bass.

Carline, J. D. (2004). "Funding medical education research: opportunities and issues." Acad Med **79**(10): 918-924.

Carroll, W. R. and A. Bandura (1987). "Translating cognition into action: the role of visual guidance in observational learning." J Mot Behav **19**(3): 385-398.

Collins A and Stevens A. L (1983). A cognitive theory of interactive teaching. Instructional design theories and models: An overview. Reigeluth C. M. Hillsdale, NJ:, Erlbaum.

Collins, A., J. Brown, et al. (1989). Cognitive apprenticeship: Teaching the crafts of Reading, writing, and mathematics. . Knowing,Learning, and Instruction. L. Resnick. Hillsdale, New Jersey, Lawrence Erlbaum Associates, Inc ., Publishers**:** 453-494.

Commission, I. T. (2005) "International Guidelines on Test Adaptation. ."

Copeland, H. L. and M. G. Hewson (2000). "Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center." Acad Med **75**(2): 161-166.

Cottrell, D., S. Kilminster, et al. (2002). "What is effective supervision and how does it happen? A critical incident study." Med Educ **36**(11): 1042-1049.

Deketelaere, A., G. Kelchtermans, et al. (2006). "Disentangling clinical learning experiences: an exploratory study on the dynamic tensions in internship." Med Educ **40**(9): 908-915.

Dolmans, D. H., I. H. Wolfhagen, et al. (2002). "The impacts of supervision, patient mix, and numbers of students on the effectiveness of clinical rotations." Acad Med **77**(4): 332-335.

Downing, S. M. (2003). "Validity: on meaningful interpretation of assessment data." Med Educ **37**(9): 830-837.
Elzubeir, M. A. and D. E. Rizk (2001). "Identifying characteristics that students, interns and residents look for in their role models." Med Educ **35**(3): 272-277.
Ericsson, K. A. (2004). "Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains." Acad Med **79**(10 Suppl): S70-81.
Fernandez, A. (2008). Adaptación de Tests a Otras Culturas. Introducción a la Psicometría. S. Tornimbeni, E. Perez and F. Olaz. Ciudad de Buenos Aires, Paidos Ed.**:** 191-205.
Fluit C. R, Scherpbier J. A, et al. (1997). Educational Training of Medical Teachers: Recent Developments in the Netherlands. Advances in Medical Education. A. J. J. A. Scherpbier, C. P. M. van der Vleuten, J. J. Rethans and A. F. W. van der Steeg, Springer Netherlands**:** 213-215.
Forero, C. G., A. Maydeu-Olivares, et al. (2009). "Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation." Structural Equation Modeling: A Multidisciplinary Journal **16**(4): 625-641.
Fornellsa, J. M., X. Juliàb, et al. (2008). "Feedback en educación médica." Educación Médica **11**(1): 7-12.
Frisch, S. R., F. G. Boucher, et al. (1984). "Increasing the effectiveness of clinical supervision." Can Med Assoc J **131**(6): 569-572.
Grant, J., S. Kilminster, et al. (2003). "Clinical supervision of SpRs: where does it happen, when does it happen and is it effective? Specialist registrars." Med Educ **37**(2): 140-148.
Hair, J. F., R. L. Tatham, et al. (2006). Multivariate data analysis, Pearson Prentice Hall Upper Saddle River, NJ.
Hargreaves D.H., e. a. (1997). On-the-job Training for Physicians., London: Royal Society of Medicine Press.
Henson, R. K. (2006). "Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice." Educational and Psychological Measurement **66**(June): 393-416.
Hoffman, K. G. and J. F. Donaldson (2004). "Contextual tensions of the clinical environment and their influence on teaching and learning." Med Educ **38**(4): 448-454.
Jöreskog, K. G., D. Sörbom, et al. (1996). PRELIS 2 User's Reference Guide: A Program for Multivariate Data Screening and Data Summarization : a Preprocessor for LISREL, Scientific Software International.
Kaufman D.M. and Mann K.V. (2010). Teaching and learning in medical e ducation: how t heory can inform practice. Understanding Medical Education Evidence, Theory and Practice. T. Swanwick. London, UK, The Association for the Study of Medical Education.
Kilminster, S. M. and B. C. Jolly (2000). "Effective supervision in clinical practice settings: a literature review." Med Educ **34**(10): 827-840.
Knowles M. (1973). The Adult Learner: a neglected species. Houston , TX . Gulf Publishing,.
Knowles MS (1980). The Modern Practice of Adult Education: from pedagogy to andragogy (2e). New York . Cambridge Books , .
Kolb DA (1984). Experiential Learning: experience as the source of learning and development. Englewood Cliffs,NJ, Prentice Hall.
LaCombe, M. A. (1997). "On bedside teaching." Ann Intern Med **126**(3): 217-220.
Lave J. and Wenger E. (1991). Situated Learning: legitimate peripheral participation New York, Cambridge University Press.
Litzelman, D. K., G. A. Stratos, et al. (1998). "Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances." Acad Med **73**(3): 324-332.
Litzelman, D. K., G. A. Stratos, et al. (1998). "Factorial validation of a widely disseminated educational framework for evaluating clinical teachers." Acad Med **73**(6): 688-695.
McLellan H (1996). Situated Learning Perspectives. Englewood Cliffs, New Jersey, Educational Technology Publications.

McLeod, P. J. (1986). "A successful formula for ward rounds." CMAJ **134**(8): 902-904.
Mezirow J (1994). "Understanding transformation theory." Adult Education Quarterly **44**(4): 222-244.
Morris, C., Blaney, D. (2010). Work-Based Learning. Understanding Medical Education: Evidence, Theory and Practice. T. Swanwick. Oxford, UK., Wiley-Blackwell.
Moulaert, V., M. G. Verwijnen, et al. (2004). "The effects of deliberate practice in undergraduate medical education." Med Educ **38**(10): 1044-1052.
Norman, G. R. and D. L. Streiner (2008). Biostatistics: The Bare Essentials, B.C. Decker.
Pathak, A., N. Pathak, et al. (2000). "Ideal ward round making in neurosurgical practice." Neurol India **48**(3): 216-222.
Ramani, S. and S. Leinster (2008). "AMEE Guide no. 34: Teaching in the clinical environment." Med Teach **30**(4): 347-364.
Ramani, S., J. D. Orlander, et al. (2003). "Whither bedside teaching? A focus-group study of clinical teachers." Acad Med **78**(4): 384-390.
Rees, J. (1987). "Take a teaching ward round." Br Med J (Clin Res Ed) **295**(6595): 424-425.
Romá, V. G. and S. L. Segura (1994). Análisis factorial confirmatorio: fundamentos y aplicaciones mediante LISREL, Vicente González Romá.
Schön DA (1983). The Reflective Practitioner: how professionals think in action. New York, Basic Books.
Seabrook, M. A. (2003). "Medical teachers' concerns about the clinical teaching context." Med Educ **37**(3): 213-222.
Shankel, S. W. and E. L. Mazzaferri (1986). "Teaching the resident in internal medicine. Present practices and suggestions for the future." JAMA **256**(6): 725-729.
Snell, L., S. Tallett, et al. (2000). "A review of the evaluation of clinical teaching: new perspectives and challenges." Med Educ **34**(10): 862-870.
Souster V and Jackson N (2003). Apprenticeship systems and work based learning . Work based Learning in Primary Care. Burton J and Jackson N, Radcliffe Publishing, Oxford**:** 59-72.
Spencer, J. (2003). "Learning and teaching in the clinical environment." BMJ **326**(7389): 591-594.
Stalmeijer, R. E., D. H. Dolmans, et al. (2008). "The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity." Med Teach **30**(8): e272-277.
Stalmeijer, R. E., D. H. Dolmans, et al. (2010). "The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers." Acad Med **85**(11): 1732-1738.
Stalmeijer, R. E., D. H. Dolmans, et al. (2009). "Cognitive apprenticeship in clinical practice: can it stimulate learning in the opinion of students?" Adv Health Sci Educ Theory Pract **14**(4): 535-546.
Stanley, P. (1998). "Structuring ward rounds for learning: can opportunities be created?" Med Educ **32**(3): 239-243.
Streiner, D. L. and G. R. Norman (2008). Health Measurement Scales: A practical guide to their development and use, Oxford University Press.
Thibault, G. E. (1997). "Bedside rounds revisited." N Engl J Med **336**(16): 1174-1175.
van de Vijver, F. J. R., J. Hofer, et al. (2010). Methodological aspects of cross-cultural developmental studies. Handbook of cross-cultural developmental science. Bornstein. Mahwah, Erlbaum**:** 21-37.
van de Vijver, F. J. R. and Y. H. Poortinga (1997). " Towards an integrated analysis of bias in cross-cultural assessment." Eur J Psychol Assess **13**: 21-29.
Weinstein, C. E. and R. E. Mayer (1985). The Taching of Larning Etrategies. Handbook of research on teaching. M. C. WITTROCK. New York, MacMillan Publishing Company.
Wenger E (1998). Communities of Practice: learning, meaning, and identity. New York, Cambridge University Press.

Whitman, N. and T. Schwenk (1997). The physician as teacher, Whitman Associates.
Ximénez M.C (2005). "Comparación de los métodos de estimación de máxima verosimilitud y mínimos cuadrados no ponderados en el análisis factorial confirmatorio mediante simulación Monte Carlo. ." Psicothema **17**(3): 528-535.
Zúñiga-Brenes, M. E. and E. Montero-Rojas (2007). "Teoría G: un futuro paradigma para el análisis de pruebas psicométricas." Actualidades en Psicología **21**: 117-144.
21: 117-144.