TROVARE
Repositorio Institucional
Instituto Universitario y Hospital Italiano de Buenos Aires

Puyol, Raúl Bernardo

Adaptación transcultural y validación del test holandés del clima de aprendizaje en residencias (D-RECT)

Maestría en Educación para Profesionales de la Salud

Tesis 2017

Cita sugerida (Vancouver): *Puyol RB. Adaptación transcultural y validación del test holandés del clima de aprendizaje en residencias (D-RECT) [tesis de maestría] [Internet]. [Buenos Aires]: Instituto universitario Hospital Italiano; 2017 [citado AAAA MM DD]. viii; 52; xivi p. Disponible en: http://trovare.hospitalitaliano.org.ar/descargas/tesisytr/20191203095540/tesis-puyol.pdf*

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.

HOSPITAL ITALIANO de Buenos Aires

CC BY NC ND

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 Test Holandés Del Clima De Aprendizaje En Residencias (D-RECT)

por

Raúl Bernardo Puyol

Director: Sergio Eduardo Giannasi

Tutor: Eduardo Durante

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
Abril de 2018

© Copyright Raúl B. Puyol, 2018

AGRADECIMIENTOS

- Al cuerpo docente de la Maestría en Educación para Profesionales de la Salud del Instituto Universitario Escuela de Medicina del Hospital Italiano de Buenos Aires y particularmente a su Director Prof. Dr. Eduardo Durante. Por su guía en el espíritu de la docencia universitaria, y su aliento en todo el camino.

- Al Dr. Sergio Eduardo Giannasi, por su acompañamiento y orientación tutorial en la confección de esta tesis.

- Al Dr. Fernando Ramón Vázquez, por su invalorable ayuda y consejo en instancias críticas de este trabajo.

- A mis compañeros de la Maestría, por el mutuo apoyo durante el curso, y por compartir la fe en la profesión docente.

- A la Facultad de Medicina de la Universidad Nacional del Nordeste, especialmente a su Decano Prof. Omar Larroza, por su aval institucional al proyecto, y su valioso estímulo.

- A mis colegas docentes que colaboraron desinteresadamente en pasos decisivos de este trabajo: Dra. Viviana Navarro (Coordinadora de Departamento Medicina de la Facultad de Medicina de la UNNE), Dr. Germán Pagno (Director de Docencia e Investigación del Hospital Escuela Gral. San Martín- Corrientes), Dr. Daniel Palma (Jefe de Servicio de Clínica Médica del Hospital Escuela Gral. San Martín- Corrientes), como así también a las autoridades de los Hospitales públicos de Corrientes y Chaco que permitieron la recolección de los datos necesarios para el presente trabajo.

- A los Médicos Residentes de los distintos Hospitales, destinatarios finales de esta labor, por su buena voluntad y disposición para someterse a la encuesta, permitiendo la recolección de los datos necesarios.

- A mi heroica familia, por su infinito cariño e incondicional aguante. Perdón por el tiempo que debí estar ausente, sin reclamo alguno de su parte.
- A Sonia, mi compañera de vida, y motor de todos mis logros.

INTRODUCCIÓN

El término “clima de aprendizaje” (CA) se refiere al modo en que el ambiente educacional es percibido o experimentado por los estudiantes (Genn, 2001). El CA comprende las distintas locaciones físicas, contextos y culturas en que los estudiantes aprenden, incluyendo las interacciones entre pares, y la organización de actividades educacionales (Bakhshialiabad et al., 2015).

Constituye un determinante de su comportamiento, contribuyendo con sus logros, satisfacción y éxito, el desarrollo de profesionalismo (West & Shanafelt (2007), y mejores resultados de aprendizaje (Lizzio et al., 2002). Por lo tanto, la calidad del CA es crucial para el aprendizaje efectivo, y los reportes de los estudiantes reflejan su efectividad (Seabrook, 2004).

Tanto en el ámbito global de la universidad (Lizzio et al., 2002), como en el contexto específico de la educación médica, el CA percibido por los alumnos tiene un papel trascendente (Riquelme Pérez et al., 2007), y es un elemento vital para la educación médica de alta calidad (Schönrock-Adema et al., 2012).

En el caso de las residencias médicas, el complejo marco de actividad y su carga de estrés (Butterfield, 1988) ocurre dentro del contexto clínico de los hospitales, y tiene un impacto decisivo en el aprendizaje (Boor et al., 2009). Las instituciones hospitalarias influyen fuertemente en el CA y la satisfacción de los residentes (Gruppen et al., 2015). Un CA saludable puede beneficiar al desarrollo del residente como profesional así como a su calidad de atención (Silkens et al., 2016).

El CA es un constructo que relaciona múltiples aspectos de formación de los residentes, y refleja el modo en que los departamentos de docencia abordan este aprendizaje, revelado a través de percepciones grupales de los aprendices sobre cuestiones como la atmósfera, la supervisión, y el estado de aprendizaje. Este entorno de aprendizaje es construido a través de las interacciones entre aprendices y los demás trabajadores de la salud, bajo influencia del diseño de cada organización y sus artefactos (Boor et al., 2011).

En las residencias médicas de nuestro país se han detectado problemas en distintos aspectos: regulación y planificación, orientación y gestión, enfoque de la formación, y modelos pedagógicos (Borrel Bentz, 2005), donde se encuentran

implicados muchos componentes del CA considerados en este trabajo. Las características del ambiente donde aprenden los residentes inciden sobre la adquisición de competencias (Hoff et al., 2004), sensación de estrés y acoso (Cohen & Patten, 2005), y aparición de burn-out (Thomas, 2004; Llera & Durante, 2014)

La percepción de este ambiente de desempeño, donde ocurren las acciones de enseñanza, puede ser recabada a través de instrumentos validados a tal fin (Riquelme Pérez et al., 2007; Schönrock-Adema et al., 2012). En tal sentido, las escuelas de medicina son organizaciones educativas que evolucionan y cambian evaluando su entorno y su currículo a través de estudios de su ambiente (Genn, 2001).

Evaluar el CA es un modo de apreciar la calidad de los programas de formación, revelando el contexto de desempeño de los residentes, y su medición puede servir como amplio indicador del funcionamiento educacional de un departamento dada su versatilidad (Boor et al., 2011).

Esta tesis trata del trabajo llevado a cabo para la adaptación transcultural del instrumento denominado D-RECT (Dutch Residency Educational Climate Test) y el proceso de verificación de sus propiedades psicométricas, para su utilización en nuestro medio.

RESUMEN

Introducción:

Los programas de residencias deben crear un ambiente adecuado de aprendizaje para los futuros especialistas. Medir el clima educacional a través de la percepción de los residentes, significa estimar la calidad de los programas para diseñar mejoras.

Objetivo:

Obtener una versión válida del cuestionario D-RECT en idioma español y aportar evidencias de validez y confiabilidad, para valorar el clima de aprendizaje en las residencias de nuestro medio.

Material y Método:

Se realizó la adaptación transcultural del cuestionario a partir del idioma original, recabando evidencia sobre validez de contenido (equivalencia lingüística y cultural, y representatividad de los ítems dentro del constructo), proceso de respuesta (pretesteo y entrevistas a grupo piloto), estructura interna y confiabilidad (alfa de Cronbach, Análisis Factorial Exploratorio y Confirmatorio, y estudio G)

Resultados:

Cumplidas las etapas de adaptación, validación de contenido y del proceso de respuesta, 403 residentes de diferentes especialidades contestaron la versión adaptada del cuestionario, ingresando 392 al estudio.

El análisis de propiedades psicométricas reveló evidencia favorable para validez y confiabilidad del instrumento.

Conclusión:

El cuestionario D-RECT resulta una herramienta válida, confiable y factible para valorar el clima de aprendizaje de las Residencias Médicas en nuestro medio. Los resultados pueden ser generalizables a los programas de residencias locales.

Palabras Clave:

1. En español:
    a. Clima educacional.
    b. Residencias.
    d. Adaptación transcultural.
    e. Validez, confiabilidad

2. En Inglés:
    a. Learning climate
    b. Residencies.
    d. Cross-cultural adaptation.
    e. Validity, reliability

TABLA DE CONTENIDOS

AGRADECIMIENTOS i
INTRODUCCIÓN iii
RESUMEN Y PALABRAS CLAVE v
TABLA DE CONTENIDOS vii
FUNDAMENTACIÓN 1
Acerca del instrumento original 7
Proceso de creación del D-RECT 5
Resultados en la elaboración del D-RECT definitivo original 9
Conclusiones del estudio original 10
Fortalezas y debilidades 10
Implicaciones en educación y futuras investigaciones 11
OBJETIVO GENERAL DEL ESTUDIO 12
Objetivos específicos 12
MATERIALES Y MÉTODO 12
Objetivo 1- Lograr una versión válida del instrumento en español 13
Objetivo 2- Obtención de evidencias de validez y confiabilidad 13
1. Evidencia basada en el Contenido del test 14
2. Evidencia basada en el Proceso de Respuesta 14
3. Evidencia basada en la estructura interna del test y confiabilidad 15
- Características del instrumento 15
- Determinación del tamaño muestral 16
- Administración del test definitivo 16
- Pruebas estadísticas 16
- Aval Institucional 18
- Conflictos de intereses 18

RESULTADOS **19**
I. Objetivo 1: Obtener una versión válida del D-RECT en español 19
- Traducción Directa 19
- Traducción inversa o Retrotraducción (del español al inglés) 19
II. Objetivo 2: Obtención de evidencia sobre validez y confiabilidad 20
1- Evidencia para la validez de contenido 20
2- Evidencia del proceso de respuesta 20
- Pre testeo sobre grupo piloto 20
- Administración del cuestionario a los residentes 21
- Resguardo del anonimato 21
3- Evidencia sobre estructura interna y confiabilidad 22
- Análisis Factorial Exploratorio 22
- Consistencia Interna- Cálculo del coeficiente Alfa de Cronbach 23
- Análisis Factorial Confirmatorio 23
- Generalizabilidad 29
DISCUSIÓN **31**
CONCLUSIONES **50**
- Limitaciones 50
- Implicaciones educativas y sanitarias futuras 51
ANEXOS ..... **ix**
BIBLIOGRAFÍA ..... **xxxiii**
CURRICULUM VITÆ DEL AUTOR ..... **xli**

FUNDAMENTACIÓN

La calidad del clima educacional resulta crucial para el aprendizaje efectivo (Seabrook, 2004) El clima de aprendizaje (CA) influye en el enfoque del estudiante (superficial o profundo) sobre el estudio, y contribuye fuertemente para todos los resultados del aprendizaje en la universidad (Lizzio, 2002). Es un elemento vital para la educación médica de alta calidad (Schönrock-Adema et al., 2011) y un factor determinante para un currículo efectivo (Abraham, 2008). Las experiencias de los estudiantes de medicina dentro de su entorno educativo guardan relación con sus logros, satisfacción y éxitos, y la bibliografía se expande sobre el rol del CA en la educación médica, como manifestación conceptual del currículo que abarca todo lo que sucede en una escuela de medicina (Abraham, 2008; Riquelme Pérez et al., 2007; Genn, 2001). El compromiso de aprender depende de la motivación y percepción de relevancia, a su vez afectados por las experiencias previas, estilos de aprendizaje, el contexto y entorno educacional (Hutchinson, 2003).

El complejo marco de actividad de las residencias médicas y su carga de estrés (Butterfield, 1988) ocurre dentro del contexto clínico de los hospitales, y tiene un impacto decisivo en el aprendizaje (Boor et al., 2009).

Los programas de residencias médicas deben crear un ambiente apropiado para que los ingresantes aprendan su tarea correctamente, en un lugar de entrenamiento diseñado a tal fin, atendiendo el contexto cultural y laboral de los hospitales, como factores que inevitablemente inciden en la adquisición de profesionalismo y otras competencias de los residentes (Hoff et al., 2004; West & Shanafelt, 2007).

El CA también es importante para los médicos de planta, dado que el desempeño del cuerpo docente es esencial para lograr alta calidad de enseñanza y asistencial, y existe evidencia de relación positiva entre ambos (Genn, 2001; Lombarts et al., 2014; van Mook et al., 2016). Enseñar significa organizar el contexto o clima de aprendizaje al igual que impartir conocimiento o compartir experticia (Hutchinson, 2003). Corresponde a los programas e instituciones que los implementan proveer la

atmósfera y condiciones de trabajo necesarios para tales propósitos (Borrel Bentz, 2005; Hoff et al., 2004). Una evaluación de la calidad del ambiente educacional podría identificar áreas problemáticas dentro del mismo, para tomar eventuales medidas de mejoramiento (Riquelme Pérez et al., 2007).

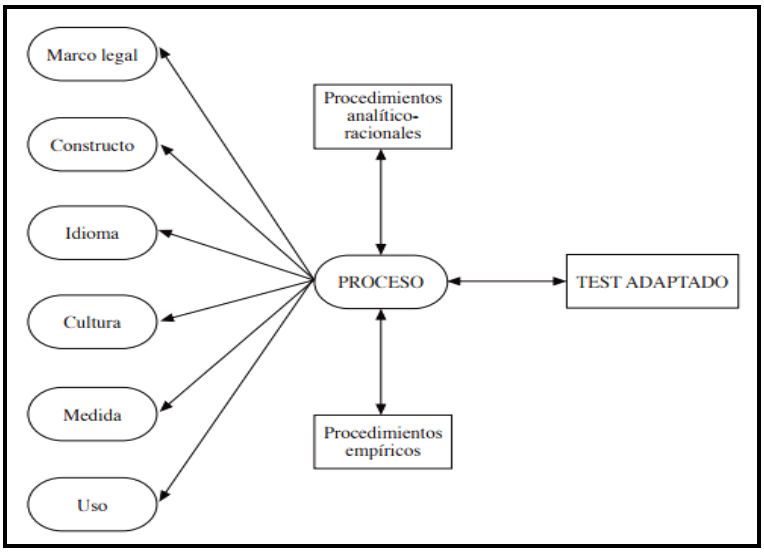
En tal sentido, al momento del presente trabajo, existen instrumentos para la valoración del CA en escuelas de medicina, la mayoría de ellos desarrollados sin base teórica (Schönrock-Adema et al., 2012). Entre los más utilizados cabe mencionar el DREEM y el PHEEM (Roff et al., 1997; 2005). El primero, aunque adaptado en varios idiomas, fue desarrollado para uso específico en pregrado. El segundo, también adaptado en otros idiomas, fue desarrollado para la etapa de posgrado denominada *Junior Training* en otros países, que precede a las residencias de especialidades. En Chile se ha desarrollado el ACLEEM (Riquelme et al., 2015), para su utilización en contexto de atención ambulatoria, y cuya población de estudio original incluía estudiantes de medicina además de residentes. Colbert-Getz et. al. (2014), analizaron 28 instrumentos de medición existentes sobre CA en Medicina (13 correspondientes a residencias) respecto de las evidencias de validez demostradas en su composición (exceptuando la validez consecuencial) y demostraron que ninguno alcanzaba el total de las 4 restantes fuentes establecidas por la APA. De esos 28 instrumentos, 17 (61%) revelaba alguna evidencia sobre contenido, 9 (32%) tenían evidencias sobre estructura interna, 6 (21%) sobre relación con otras variables y 4 (14%) sobre el proceso de respuesta. 23 de estos trabajos (82%) sugerían múltiples dominios pero sólo 5 sustentaba esa hipótesis en un análisis factorial. Por tanto, si se pretende una herramienta ideal que posea fuerte evidencia sobre las fuentes de validez determinadas por la APA, no existe un “gold standard” para medir CA a través de las percepciones de estudiantes o residentes de medicina. Esa herramienta debería además ser eficiente, amplia y rápida en su aplicación, y sensible para detectar cambios en el tiempo, ya que el constructo podría medirse regularmente.

La adaptación transcultural está orientada hacia la medición del mismo fenómeno en diferentes culturas (Hambleton et al., 2005; Arribas, 2006). Las recomendaciones de la International Test Commission en 1991, 2005 y 2013

establecen pautas que enmarcan el proceso para asegurar el cumplimiento de este propósito (Hambleton & Bollwark, 1991; Hambleton et al, 2005; Muñiz et al, 2013)

El proceso de adaptación transcultural consiste en lograr la medición confiable del mismo constructo en un contexto cultural diferente al original (Arribas, 2006). La simple traducción no implica necesariamente que el cuestionario conserve sus propiedades psicométricas, por lo que deben respetarse las recomendaciones y procedimientos establecidos para verificar la estabilidad del test (Beaton et al, 1998; Muñiz et al, 2013). Los componentes que intervienen en la adaptación de un test son representados en la figura 1:

Figura 1. Componentes en el proceso de adaptación de un test.
(Adaptado de Muñiz et al., 2013)



La interacción de elementos y procedimientos determina el resultado final del proceso de adaptación.

Citando a Malda, Van de Vijver et al. (2008), adaptar un test implica maximizar la armonización cultural y minimizar el sesgo. Superando al concepto tradicional de *traducción*, estos autores atribuyen al término *adaptación* la denominación genérica para cualquier procedimiento en que un instrumento creado para un grupo cultural es transferido para usarse en otro grupo cultural, con lo que se obtiene para ese nuevo contexto una versión equivalente (más que una mera traslación idiomática).

Las ideas abstractas representan un particular desafío en su traducción, ya que no es sencillo para los traductores describir el contenido de esos ítems en una forma

igualmente abstracta. Traducir palabras que son conceptos abstractos significa focalizarse sobre las ideas (Hambleton & Zenisky, 2010)

La tarea de adaptación transcultural comprende todas las actividades desde plantearse si un test podría medir el mismo constructo en una cultura y lenguaje diferentes, seleccionar traductores, decidir sobre apropiadas modificaciones al preparar un test para su uso en un segundo idioma, adaptarlo y chequear la equivalencia de su forma adaptada (Hambleton et al., 2005).

En ese sentido, este mismo autor organizó tres categorías para las fuentes de error de validación: a) diferencias culturales / lingüísticas, b) problemas técnicos, de diseño o método, c) interpretación de resultados. El descuido sobre alguna de estas fuentes de error produce un test que no es equivalente para ambos grupos culturales como se pretende.

Es posible enumerar cuatro nociones y prácticas comunes erróneas (Elosua et al., 2014)) que influyen negativamente sobre el producto final:

- *Hablar dos idiomas como suficiente condición para adaptar un test.*
- *Una traducción literal garantiza la equivalencia.*
- *La traducción inversa como procedimiento de verificación garantiza equivalencia entre la versión original y la adaptada.*
- *Los constructos evaluados son universales.*

Citamos: “Es habitual considerar que en una buena traducción la equivalencia entre la versión original y la versión retro-traducida generada por un traductor independiente es muy alta. Esta consideración, sin embargo, no es garantía de validez de la versión diana, es más, en una mala traducción el grado de equivalencia entre la versión original y la versión retro-traducida puede ser muy alto. La razón de esta singularidad se debe a que habitualmente las malas traducciones se apoyan en traducciones literales en lugar de en una esmerada adaptación de significados. Por ejemplo, los resultados de las traducciones automatizadas originan buenas retro-traducciones, a pesar de ello, nunca las utilizaríamos sin una evaluación y ajuste de la versión en el idioma destino.” (Muñiz et al., 2013. p. 152-153)

Con base en esta aseveración, la falta de apego a las normas podría derivar en alguna de las tres formas de sesgo relacionadas con esta tarea (Van de Vijver & Poortinga, 2005):

a) *Sesgo de constructo* (cuando el instrumento no mide el mismo concepto en ambas culturas).
b) *Sesgo de método* (causado por un aspecto metodológico p.ej. en la administración del test, la muestra o las características del instrumento).
c) *Sesgo de ítem* (problemas específicos del funcionamiento diferencial de los ítems entre distintas culturas p.ej. debido a la traducción si los individuos poseen características diferentes que les impide interpretar, del mismo modo al estudio original, el significado de una pregunta).

El sesgo no constituye una propiedad intrínseca de un instrumento, sino una resultante de la aplicación de ese instrumento en un grupo particular para un propósito determinado. Se refiere a todas las clases de factores dañinos para la interpretación de diferentes puntajes entre un grupo y otro. Desde una perspectiva de generalizabilidad puede definirse formalmente como una desigual correspondencia entre el dominio de observaciones y el universo de generalización (van de Vijver & Poortinga, 2005).

Debido a sus características, elegimos como instrumento para valorar el clima educacional en las residencias médicas al D-RECT (Dutch Residency Educational Climate Test) (Boor et al., 2011), traducido para el presente trabajo como TEST HOLANDÉS DEL CLIMA DE APRENDIZAJE EN RESIDENCIAS.

Acerca Del Instrumento Original

Los autores del test D-RECT desarrollaron dicho instrumento en base a la necesidad de los gobiernos y entes reguladores de asegurar una alta calidad de formación en las residencias médicas (Genn, 2001), tal que evaluar el clima de aprendizaje es una manera de evaluar la calidad de los programas educativos.

El Clima de Aprendizaje (CA) es un constructo referido al contexto institucional en que los residentes se forman. Refleja el enfoque del programa formativo desde los

departamentos responsables, e incorpora percepciones en cuestiones como la atmósfera, la supervisión y el estado de aprendizaje. El CA resulta de la interacción de los alumnos con los trabajadores de la salud bajo la influencia del diseño organizacional y sus artefactos (Boor et al., 2009). La medición de CA puede servir como indicador de funcionamiento de los departamentos educativos debido a la versatilidad del constructo.

No obstante, la evaluación de CA constituye una ardua labor. Los autores remarcan dos cuestiones sobre este punto:

- ✓ En primer lugar, el CA es un constructo teórico que no puede ser medido directamente. Tomando las experiencias cotidianas de los residentes que reflejan como constructo subyacente al CA, un estudio cualitativo entre 40 residentes de diversas especialidades exploró los eventos implicados en los climas de aprendizaje (Boor et al., 2009). En ese estudio, los residentes concluyeron que un óptimo CA convierte la actividad diaria en una extensión de su formación (sin limitarse a los imperativos de servicio). Así, un óptimo CA les permite tomar decisiones en su trabajo sobre sus personales necesidades de aprendizaje, resaltando la interacción con el cuerpo médico, compañeros, enfermeras y resto del personal de salud. Es decir, la actividad laboral se integra con las necesidades de aprendizaje de los residentes.
- ✓ La segunda cuestión es la dificultad para construir un instrumento con adecuadas propiedades psicométricas. Por ejemplo, el Dundee Ready Education Environment Measure (DREEM) (Roff et al., 1997) y el Postgraduate Hospital Educational Environment Measure (PHEEM) (Roff et al., 2005), dos herramientas ampliamente utilizadas, carecen de una base teórica clara, y su estructura factorial subyacente es discutida. Por ejemplo, el PHEEM revela tres subescalas diferentes en tres diferentes publicaciones (Roff et al., 2005; Boor et al., 2007; Schönrock-Adema et al., 2008). Tal controversia dificulta la retroalimentación específica para las instituciones.

Para superar estos obstáculos, y con base en el estudio cualitativo previo mencionado, el D-RECT fue desarrollado buscando evidencia de validez en el contenido y estructura interna (AERA et al., 1999). El proceso incluyó la confección de ítems y subescalas probadas mediante un panel Delphi y análisis factorial exploratorio (AFE); se realizó el cálculo del coeficiente alfa de Cronbach, y posteriormente se testeó el cuestionario con análisis factorial confirmatorio (AFC). Finalmente, se realizó un estudio de generalizabilidad para demostrar la confiabilidad de la medición.

A continuación se detalla el proceso de creación del instrumento original (Boor et al., 2011).

Proceso de creación del D-RECT

Contexto: La carrera de Medicina en Holanda tiene una duración de 6 años para obtener una graduación básica, seguido de un programa de 4 a 6 años adicionales en alguna de 27 especialidades. En dichos programas existen rotaciones dentro de un hospital universitario y un hospital general asociado, donde cada departamento encargado cuenta con un tutor especialista formalmente responsable de la educación de los residentes, incluyendo 2 evaluaciones anuales.

Etapas desarrolladas: Tomando como base el estudio cualitativo inicial (Boor et al., 2009), que consistió de 83 ítems y 21 subescalas (p.ej. feedback, colaboración entre pares, carga laboral, trabajo en equipo), se sometió esta información al análisis de un grupo médicos y un tutor especialista). Adicionalmente, cuatro residentes, dos tutores especialistas y tres educadores médicos (que no formaron parte del primer grupo de investigación) revisaron y sugirieron modificaciones sobre los ítems. Los reactivos repetidos o confusos fueron eliminados. Se obtuvo así un formulario preliminar del D-RECT de 75 ítems en una escala Likert de 5 puntos, incluyendo una opción de No Aplicable. Estos ítems fueron remitidos para ser sometidos a un panel Delphi, y a los médicos residentes en toda Holanda, en forma simultánea según se detalla a continuación.

El Procedimiento Delphi: En Abril y Mayo de 2008, 38 panelistas fueron elegidos en mérito a su rol en formación en especialidades de posgrado (incluyendo hospitales universitarios y no universitarios) para conformar el grupo encargado del procedimiento

Delphi respectivo. Luego de recibir el D-RECT preliminar, determinaron la relevancia de cada ítem en relación con el CA sobre una escala de 5 puntos (1=no relevante y 5= altamente relevante). Para el análisis, los puntajes 1, 2 y 3 se interpretaron como No Relevantes; 4 y 5 como Relevantes). Se consideró logrado el consenso cuando el 90 % de los ítems incluidos o excluidos en una ronda, permanecieran sin cambios en la ronda subsecuente.

Envío Por Correo Del Cuestionario Preliminar: En Mayo de 2008, se mandó por correo a todos los residentes médicos en Holanda (excepto en la especialidad Psiquiatría por cuestiones logísticas) el cuestionario D-RECT preliminar para ser completado en idioma del país. Algunos fueron incentivados económicamente para responder el cuestionario.

Análisis Factorial Exploratorio y Alfa de Cronbach: Para el AFE se seleccionaron al azar la mitad de los cuestionarios respondidos. El software incluía rotación ortogonal Varimax (que presupone falta de correlación entre los factores) y rotación oblicua Oblimin (que asume correlación entre los factores). La Matriz de Correlación de Componentes indicó correlaciones existentes entre dichos factores, por lo que se tomaron solo las rotaciones Oblimin. Los ítems con carga factorial débil se eliminaron, y se determinó la consistencia interna a través del cálculo de Alfa de Cronbach.

En base a los resultados de ambos procedimientos simultáneos, se continuó con la confección del D-RECT definitivo.

El D-RECT definitivo: AFC y generalizabilidad: La segunda mitad de los cuestionarios respondidos por residentes se sometió al Análisis Factorial Confirmatorio (AFC). A través del modelo de ecuaciones estructurales, se testeó el modelo multifactorial, considerando para la bondad de ajuste el índice de ajuste comparativo (CFI), el error cuadrático medio de aproximación (RMSEA) y el chi cuadrado relativo (CMIN/DF).

Este mismo grupo de cuestionarios fue tomado para determinar la confiabilidad. Se realizó un análisis de varianza para medir la contribución de todos los componentes relevantes sobre la varianza del test (en este caso, residentes, departamentos, ítems, y sus interacciones). Se efectuó análisis de generalizabilidad para el test global y cada

subescala por separado, estimando el número de residentes necesario para obtener puntajes confiables del test.

Para el AFC se utilizó software de modelo de ecuación estructural Amos, para el análisis de generalizabilidad se usó software URGENOVA, y para todos los demás se utilizó SPSS.

Resultados en la elaboración del D-RECT definitivo original

Procedimiento Delphi: En la primera ronda, se envió el D-RECT preliminar a los participantes solicitando valorar la relevancia de cada ítem, y eventualmente, agregar opiniones cualitativas.

En la segunda ronda, el cuestionario preliminar D-RECT fue enviado a todos los participantes, con los promedios de puntaje y las desviaciones estándar de todos los ítems en la primera ronda, y las opiniones en forma anónima. Se solicitó opinión a favor o en contra de leves ajustes en siete ítems. El panel nuevamente completó el cuestionario, que fue analizado en los nuevos puntajes y opiniones. Todos estuvieron de acuerdo en los siete ajustes de ítems. Más aún, el consenso se obtuvo en esta ronda: 91 % de los ítems incluidos y excluidos en la 1º y 2º ronda fueron idénticos.

El procedimiento Delphi llevó a la eliminación de 30 ítems (por su irrelevancia para CA), quedando un cuestionario de 45 ítems.

Cuestionarios por correo: De un total de 4835 residentes invitados, respondieron 1276 de diferentes especialidades (53,1 % eran mujeres). La muestra del grupo era comparable a la población total de residentes respecto de sexo y especialidades, excepto para Medicina Interna (p=0.01), Tocoginecología (p=0.01) y Rehabilitación (p=0.01). 25 cuestionarios fueron excluidos por tener más de 25 ítems sin responder. 591 marcaron *No Aplicable* en algunos ítems, pero la tasa de respuesta por ítem nunca fue menor de 94 %.

Análisis Factorial Exploratorio: Primero se analizaron 600 cuestionarios al azar usando AFE. Considerando un mínimo de 5 sujetos por ítem, se contaba con un tamaño muestral más que suficiente. Se eliminaron 14 ítems con débil carga factorial. Otros 11 ítems se eliminaron por decisión del panel Delphi. 15 ítems se incluyeron en el D-RECT porque los residentes del estudio cualitativo de base remarcaron fuertemente

su importancia específica y su carga factorial era alta, aunque estos ítems no fueron aceptados por el panel Delphi. 5 subescalas se eliminaron completamente, y 5 subescalas surgieron con otras relacionadas. *Esto llevó al definitivo cuestionario D-RECT de 11 subescalas con 50 ítems. El Alfa de Cronbach varió entre 0.64 y 0.85 para los diferentes factores.*

Análisis Factorial Confirmatorio: Se testeó el D-RECT definitivo de 50 ítems tomando los restantes 651 cuestionarios de residentes, usando AFC. Los índices de bondad de ajuste fueron: índice de ajuste comparativo (CFI)=0.89. Error cuadrático medio de aproximación (RMSEA)=0.04. CMIN/DF=2.9, Todos ellos mostraron un ajuste adecuado.

Análisis de Confiabilidad: Se examinó la generalizabilidad de los 50 ítems. Para obtener resultados confiables basados en el score global para un departamento, fue necesario un número de 3 residentes. Para cada subescala en un departamento, fue necesario un número de 11 residentes. Para resultados confiables en grupos de departamentos, 2 residentes de 3 departamentos fueron suficientes para un score total reproducible. Para resultados confiables para cada subescala, 4 residentes de 6 diferentes departamentos fueron requeridos.

Conclusiones Del Estudio Original

El D-RECT resulta una herramienta válida, confiable y viable para medir la calidad del Clima de Aprendizaje en contexto clínico. Es un instrumento que abarca importantes aspectos como supervisión, entrenamiento y evaluación, y retroalimentación, y otros menos obvios como las relaciones entre profesionales y la transferencia de pacientes. Estos puntos fueron aportados por opiniones de los residentes y confirmados por el panel Delphi, y pueden ofrecer nuevos puntos de vista sobre por qué algunos CA funcionan y otros no.

Fortalezas y Debilidades

Se consideró como fortaleza del estudio el uso de diferentes estrategias coordinadas, combinando una base teórica previa, un procedimiento Delphi, AFE y AFC, y estudio de generalizabilidad para el desarrollo y validación del instrumento.

Otra fortaleza fue que la muestra provino de residentes en 26 diferentes especialidades a diferentes niveles de formación, y de 76 diferentes hospitales, lo cual remarca la aplicabilidad del D-RECT. El número de residentes necesarios para un resultado confiable en un departamento variaba entre 3 y 11, pero la mayoría de las subescalas pudieron ser juzgadas confiablemente por 8 residentes. Esto soporta la factibilidad del instrumento. Así como en los grupos de departamentos, incluso pocos residentes (4 de 4 diferentes departamentos) arrojan una impresión confiable en la mayoría de las subescalas.

Los autores sugieren probar el D-RECT fuera de Holanda y testearlo en su forma final, como se hizo en este estudio.

Implicaciones en Educación y Futuras Investigaciones

Este cuestionario multifactorial es capaz de arrojar resultados confiables para las instituciones respecto de su CA local. Basados en estos resultados, uno o varios departamentos pueden ser informados sobre qué elementos (subescalas) muestran altos puntajes y deben sostenerse, y cuáles factores necesitan ser mejorados para ofrecer un mejor CA para los residentes en formación.

Esto puede llevar a mejorar la calidad y optimizar el CA, y ofrecer una señal de alerta para aquellos que necesiten ayuda. También es relevante para la transparencia y monitoreo a nivel estatal de la calidad educativa en medicina de posgrado.

Desde el punto de vista de la investigación, es interesante examinar el cuestionario para reforzar la base de datos existente, y si realmente su consecuencia es mejorar el CA, o sus efectos colaterales positivos y adversos. Queda pendiente su validación a nivel internacional. Finalmente, los autores dejan pendiente definir si existe correlación entre el buen CA y otros resultados en pacientes y residentes: los altos puntajes en D-RECT, o alguna de sus subescalas, ¿se correlacionan con mejor cuidado de pacientes? ¿Se obtienen mejores especialistas?

Este estudio sobre medición de CA y sus implicaciones para la educación médica sirvieron como base para el desarrollo de la presente tesis.

OBJETIVO GENERAL DEL ESTUDIO

Realizar la adaptación transcultural del cuestionario Dutch Residency Educational Climate Test (D-RECT) para su utilización en nuestro medio.

Objetivos específicos:

1. Lograr una versión válida del D-RECT en español.
2. Obtener evidencia de validez y confiabilidad.

MATERIAL Y MÉTODO

El presente trabajo se basa en un instrumento de medición educacional. Se trata del cuestionario D-RECT, desarrollado en Holanda para la valoración del clima de aprendizaje en las residencias.

La población destino de dicho instrumento está constituida por los médicos residentes de distintas especialidades de nuestro país.

Al tratarse de un diseño originalmente realizado en idioma extranjero, la tarea consiste en lograr la utilización del cuestionario en el idioma y contexto cultural de nuestro medio, conservando sus propiedades psicométricas.

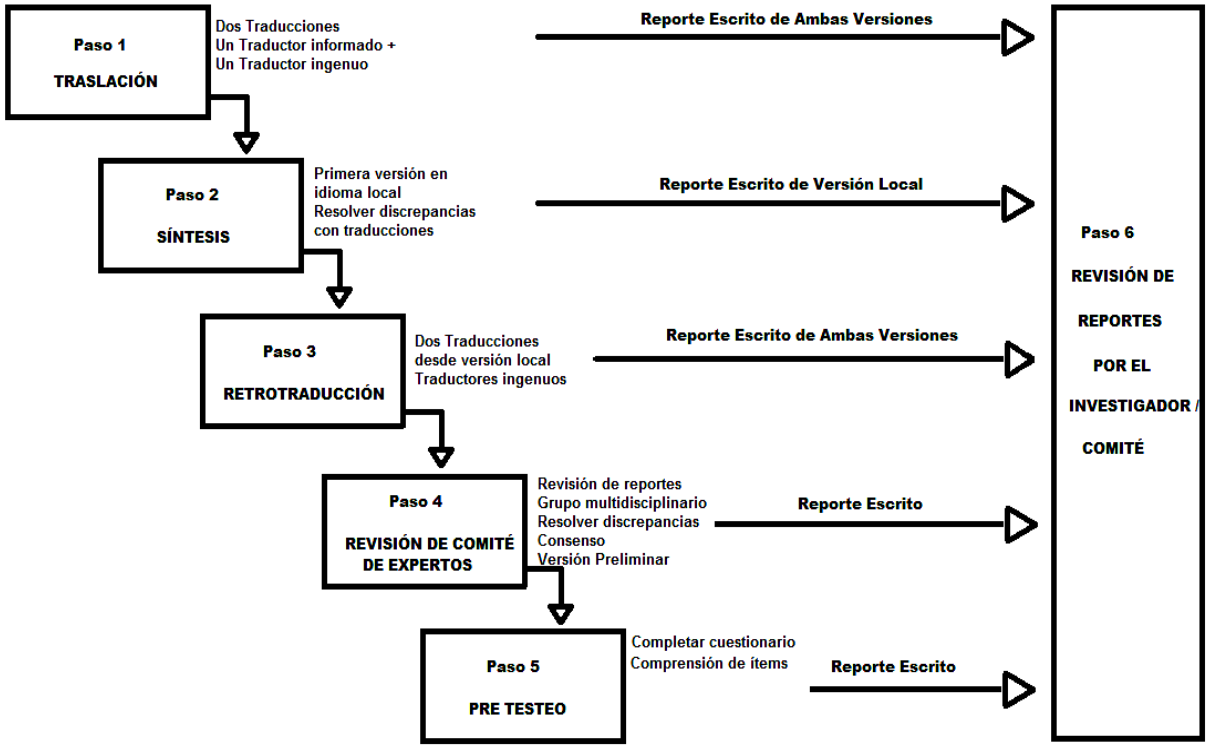
El primer paso consiste en obtener una versión válida del documento en el idioma español. Una vez logrado, el paso siguiente es someterlo a prueba en un grupo pequeño y verificar su comportamiento como testeo inicial (Beaton et al., 1998; 2000; Fernández, 2008; Arribas, 2006)

Posteriormente, corresponde aplicarlo en una muestra representativa de la población destino, para el análisis de sus propiedades psicométricas, a fin de demostrar la estabilidad de la medición luego del proceso de adaptación, obteniendo evidencias sobre su validez y confiabilidad.

• **Objetivo 1- Lograr una versión válida del instrumento en español**

Siguiendo las *Directrices para la traducción y adaptación de los tests: segunda edición* (Muñiz et al., 2013; ITC, 2017) se obtuvo autorización de uno de los autores (C. Van der Vleuten) del trabajo original (Boor et al., 2011), tras lo cual se continuó con las etapas destinadas a lograr una versión equivalente en nuestro idioma, siguiendo el esquema de tareas expuesto en la Figura 2.

Figura 2. Representación gráfica de las etapas de Adaptación Transcultural
(Adaptado de Beaton et al. 2000)



Esta secuencia describe las dos primeras fuentes de validez definidas en *Standards for educational and psychological testing* (AERA et al., 1999)

- **Objetivo 2- Obtención de evidencias de validez y confiabilidad.**

La validez implica un juicio sobre el grado en que la evidencia empírica y la teoría apoyan las interpretaciones y acciones devenidas de los puntajes dentro del propósito de un test (Messick, 1989; AERA et al., 1999)

De acuerdo con lo anterior, el proceso incluyó las siguientes instancias:

1. Evidencia basada en el Contenido del test:

Este aspecto implica demostrar que los ítems (el contenido del test) constituyen una muestra representativa del universo de posibles ítems sobre el constructo que se pretende medir. Significa establecer la pertinencia de los ítems dentro de los dominios, y de éstos en el constructo, en el contexto cultural de aplicación. El aporte de evidencia en este punto se efectuó mediante el análisis y valoración del cuestionario por un comité de expertos, integrantes de la Dirección de Docencia e Investigación del Hospital Escuela "Gral. San Martín" y del cuerpo docente de la Facultad de Medicina de la Universidad Nacional del Nordeste. Dicha tarea abarcó ambas versiones lingüísticas del documento, teniendo en cuenta la congruencia de los ítems dentro de su respectiva subescala y del constructo, como también su apropiada equivalencia y claridad en cada categoría lingüística (conceptual, técnica, lectura y comprensión) (Beaton et al., 2000; Fernández, 2008; Hambleton, 2005)

2. Evidencia basada en el Proceso de Respuesta:

En este punto, la tarea consiste en indagar sobre circunstancias inherentes a la aplicación del instrumento que pudieran afectar a los datos obtenidos. Así, se pretende evitar la ocurrencia de sesgo relativo a la administración del test.

Para ello, finalizado el proceso de traducción y validación de contenido, correspondió el testeo del cuestionario preliminar sobre un grupo piloto de 15 residentes de distintas especialidades del Hospital Escuela Gral. San Martín (Corrientes). Asimismo, se mantuvieron entrevistas individuales con ellos, enfocada sobre sus procesos de pensamiento al responder cada ítem, la adecuada redacción del cuestionario y la interpretación adecuada para el constructo a medir (Padilla & Benítez, 2014). Esta instancia incluye un aspecto de validez denominada *face validity* (*validez aparente*), vinculada con la validez de contenido. Implica el grado en que un test da la impresión que mide lo que se pretende (Abad et al., 2006. pp. 62).

Ningún residente del grupo de prueba participó en la aplicación del test definitivo.

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

Los resultados del test definitivo sobre una muestra de la población destino, fueron analizados mediante software estadístico respectivo:

- La estructura y relaciones (entre ítems y dominios del constructo), por medio de análisis factorial exploratorio.
- El funcionamiento del modelo teórico establecido a priori (según el modelo de ecuaciones estructurales), a través de análisis factorial confirmatorio.
- El análisis de consistencia interna con el cálculo del coeficiente Alfa de Cronbach, considerando confiable un valor mayor de 0.80.
- Por último, se efectuó un estudio G para estimar la capacidad de generalización de los resultados, y fuentes de variación en las mediciones.

- **Características del instrumento**

El D-RECT constituye un cuestionario estructurado en 50 ítems, distribuidos originalmente en 11 subescalas o dominios. El formulario de la encuesta se muestra en el Apéndice en el formato aplicado en el presente trabajo. Cada ítem es valorable mediante una escala Likert, asignándose un puntaje de 1 a 5 según el grado de acuerdo con el enunciado, como se muestra a continuación (tabla 1):

Tabla 1. Escala Likert del cuestionario

Valoración	Puntaje
Totalmente de acuerdo	5 puntos
Parcialmente de acuerdo	4 puntos
Neutro	3 puntos
Parcialmente en desacuerdo	2 puntos
Totalmente en desacuerdo	1 punto

- **Determinación del tamaño muestral:**

A los fines de esta investigación se consideró un mínimo de 5 sujetos por ítem para el tamaño de muestra, en base al trabajo original (Boor et al, 2011) y a la bibliografía respectiva (ver Discusión).

- **Administración del Test Definitivo:**

Se invitó a participar de esta investigación a los residentes en todas las etapas de formación, pertenecientes a los Hospitales públicos de las ciudades de Corrientes (provincia de Corrientes) y Resistencia (provincia del Chaco), con el permiso de las autoridades responsables de cada uno de ellos. Asimismo, se invitó a participar del mismo trabajo a los médicos residentes que asistieron al Congreso Nacional de Medicina Familiar y Comunitaria, realizado en la Facultad de Medicina de la ciudad de Corrientes en Octubre de 2014.

En el caso de los Hospitales se habilitó un salón con fecha y hora pactadas, donde los residentes de todas las especialidades fueron reunidos a efectos de aclarar alcances y objetivos del trabajo con el investigador principal, quien entregó el cuestionario impreso a cada participante, y recolectó las respuestas. La participación fue voluntaria, anónima y sin mediar retribución alguna para los participantes.

En el caso del Congreso Nacional de Medicina Familiar y Comunitaria, con similar estrategia se distribuyeron los formularios durante los intervalos entre actividades del mismo, con el permiso y la colaboración de los organizadores del evento.

- **Pruebas Estadísticas:**
  - Coeficiente Alfa de Cronbach

Se trata de un cálculo para evaluar la consistencia interna del conjunto de ítems de un test. Se basa en la varianza atribuida a factores comunes de los puntajes observados de los diferentes ítems, por lo que existen tres condiciones que inciden en el valor del coeficiente α: la dimensionalidad del conjunto de ítems, el grado de covariación de ítems entre sí, y la cantidad de ítems o partes que conforman la prueba en estudio (Cervantes, 2005).

- Análisis Factorial Exploratorio (AFE)

Consiste en una técnica de interdependencia que define la estructura subyacente entre las variables (Hair et al. 2010). Aplicado sobre el conjunto de reactivos de una prueba, el AFE revela la tendencia de los ítems a agruparse alrededor de *Factores*, que equivalen a constructos subyacentes o latentes (no observables directamente) deducidos de las correlaciones entre esas variables. De este modo, el contenido conceptual de los ítems agrupados sobre un mismo factor indica qué factores (constructos) explican esa correlación. El AFE contribuye con la validez de constructo de la medición, examinando la estructura del constructo que se pretende medir. Adicionalmente, aporta sobre la confiabilidad mostrando en qué medida ese constructo es unidimensional, y extrae valiosa información sobre el número de factores de la medición, el peso o carga de cada ítem y la proporción de varianza correspondiente a cada factor, y el puntaje del sujeto respecto de cada factor (Morales Vallejo, 2013)

- Análisis Factorial Confirmatorio (AFC)

Para la confirmación del modelo teórico, el AFC es un procedimiento encuadrado en los modelos de ecuaciones estructurales (*SEM, Structural Equation Models*), centrado en el estudio de los *modelos de medida*, es decir, analiza las relaciones entre variables observadas y *constructos latentes* o *factores*, establecidas *a priori* por el investigador acorde con evidencia teórica y empírica (Hair et al., 2010). Un constructo o variable latente no es observable directamente, por tanto debe ser medido mediante variables observables (Cupani, 2012). A diferencia del AFE, el investigador establece el modelo a probar previamente al AFC, especificando hipotéticas relaciones entre variables y constructos latentes. Los resultados muestran el funcionamiento del modelo (con distintos parámetros de Bondad de Ajuste), y la existencia de validez convergente y discriminante (Arias, 2008). En este caso se tomaron como indicadores de ajuste: X^2/grados de libertad, NNFI, CFI, SRMR, RMSEA, GFI, AGFI (ver Resultados)

Dada la variabilidad del X^2 respecto del tamaño muestral, y tratándose de variables ordinales sin el supuesto de normalidad multivariante, no se tomó el valor *p* derivado del valor absoluto del X^2 (Arias, 2008).

- Test de Generalizabilidad

La teoría de la Generalizabilidad (G) es una extensión de la Teoría Clásica de los Tests, que examina la contribución relativa de la variable de interés (el objeto de medición) comparada con la varianza de las fuentes de error (Bloch & Norman, 2012) a través del análisis de varianza (ANOVA). Las fuentes de variación que afectan la precisión de una prueba son consideradas facetas, y el coeficiente de Generalizabilidad (Coeficiente G) obtenido es una estimación objetiva de la estabilidad de las mediciones frente a variaciones aleatorias (Díaz et al., 2003).

La información recolectada fue registrada en una base de datos Excel, y luego procesada a través de software estadístico SPSS (IBM SPSS Statistics 19.0) para el análisis factorial exploratorio, LISREL 8.80 (Jöreskog y Sörbom. SSI Scientific Software International Inc. USA) para el análisis factorial confirmatorio, y G-String IV/urGenova para el estudio de generalizabilidad.

- **Aval Institucional**

El presente trabajo fue avalado por las autoridades de la Facultad de Medicina de la Universidad Nacional del Nordeste para su realización. En el Apéndice figura el documento respectivo.

- **Conflictos de intereses**

Se declara ausencia de conflictos de interés en cualquier aspecto de la investigación por parte del investigador.

RESULTADOS

I. Objetivo 1: Lograr una versión válida del D-RECT en idioma español:

Mediante una comunicación con uno de los autores del trabajo original (C. van der Vleuten), se obtuvo autorización para realizar la adaptación transcultural del instrumento. Además fue confirmada la ausencia hasta ese momento de otro proceso de validación en español del D-RECT.

- *Traducción Directa*

El procedimiento estuvo a cargo de dos traductores expertos de lengua nativa española a quienes se entregó una copia individual del cuestionario original. Uno de ellos fue informado de las características del estudio, mientras que el otro desconocía las mismas (traductor ingenuo). En el Apéndice constan ambas traducciones directas.

La revisión de los documentos enviados por los traductores fue realizada por un grupo de 4 expertos en educación médica, reunidos en la Dirección de Docencia e Investigación del Hospital Escuela Gral. San Martín de Corrientes. Luego del análisis exhaustivo de cada una de las traducciones, se confeccionó una versión única en español, considerada como la mejor representación del constructo a medir en nuestra lengua y contexto cultural. Asimismo, los miembros del grupo, en forma individual y luego en reunión de consenso, consideraron que los ítems eran pertinentes con respecto al sistema de formación de residentes vigente en nuestro medio.

Se guardó copia textual de cada versión del instrumento según lo escrito por cada traductor, según el plan establecido en la Tabla 1 (Beaton et al., 1998; 2000).

- *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 grupo de expertos. 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.

Nuevamente el comité de expertos, tomando en cuenta lo anterior, confeccionó el primer cuestionario definitivo en el idioma español. Específicamente como consigna dentro de la tarea, se consideró esencial que el producto final lograra equivalencia en sus cuatro variantes (conceptual, técnica, lectura y comprensión). Todas las reuniones del comité se llevaron a cabo en la Dirección de Docencia e Investigación del Hospital Escuela General San Martín de la ciudad de Corrientes.

En el apéndice figura el texto final del cuestionario obtenido por el grupo de expertos.

II. Objetivo 2: Obtención de evidencia sobre validez y confiabilidad.

1. Evidencia sobre validez de contenido:

El grupo de cuatro expertos en educación médica de postgrado en Residencias Médicas fue reunido *ad hoc* para examinar los instrumentos en ambos idiomas (inglés y español) teniendo en cuenta 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. Con especial énfasis se examinó la relación existente de cada uno de los ítems del cuestionario con el dominio correspondiente y con el constructo a medir. Los expertos consideraron que los 50 ítems correspondientes a las 11 subescalas eran un reflejo apropiado y suficiente del CA de las residencias médicas. Se estableció irrelevante agregar cuestiones eliminadas por los autores originales, ni preguntas complementarias.

2. Evidencia sobre el proceso de respuesta:

- *Pre testeo sobre grupo piloto:*

La prueba de campo o pretesteo del nuevo cuestionario implica la utilización de la versión preliminar del instrumento en un grupo de sujetos de la población destino (Beaton et al., 2000). En esta instancia, se aplicó el test a un grupo piloto de 15 residentes del Hospital Escuela Gral. San Martín, quienes luego de responder fueron citados a entrevistas individuales con el investigador, respecto del proceso de

interpretación clara del test y los ítems, comentarios y sugerencias sobre eventuales cambios o agregado de preguntas. En todos los casos la interpretación de cada reactivo fue similar o igual a la interpretación hecha por los expertos. No hubo modificaciones y/o agregados sugeridos por algunos de los entrevistados. Saturada la muestra y 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 definitivo fue utilizado en las siguientes instancias del estudio.

- *Administración del cuestionario a los residentes:*

Luego de obtener autorización de las autoridades de Investigación y Docencia de siete Hospitales Públicos Universitarios en las ciudades de Corrientes y Resistencia, en fecha y hora convenida el investigador reunió en cada Institución al conjunto de los residentes, explicando los alcances de la investigación, los objetivos del estudio y la confidencialidad de las respuestas. A continuación, se entregó a cada residente una copia del cuestionario definitivo para ser respondido. En el caso de los residentes que asistieron al Congreso Nacional de Medicina Familiar y Comunitaria, el procedimiento fue similar, llevado a cabo en los intervalos entre las actividades del evento, con la colaboración de los organizadores.

Cinco cuestionarios devueltos en blanco fueron anulados. Aunque no se especificó un plazo determinado para completar las respuestas, todos los cuestionarios fueron entregados en un lapso menor de 60 minutos.

- *Resguardo del anonimato:*

El anonimato fue preservado respecto de todos los encuestados. Los formularios de la encuesta entregados son copias iguales que carecen de cualquier señal identificatoria particular. Una vez contestado, cada residente entregó su cuestionario en una pila conjunta con el resto para evitar su individualización, excepto la institución a la que pertenecen. De esa manera se garantizaba el anonimato y la confidencialidad en sus respuestas sin percibirse presión alguna. Sólo el investigador principal tuvo acceso a los formularios, a efectos de confeccionar la base de datos respectiva. Los resultados

de las evaluaciones fueron almacenados de un modo seguro. La base de datos digitales se encuentra en poder de los investigadores.

3. Evidencia sobre estructura interna y confiabilidad

Las características de los residentes que respondieron el cuestionario D-RECT se expresan en la Tabla 2:

- **Tabla 2. Características de la muestra de residentes.**

n	403
% masculino/femenino	59,4 / 40,6
Edad promedio (años)	28,24
Nº especialidades	20

Del total de cuestionarios recolectados en la investigación, se excluyeron 11 observaciones (2,7 %) por presentar puntajes confusos o faltantes, por lo cual se analizaron 392 encuestas.

- Análisis Factorial Exploratorio

Se utilizó el análisis de componentes principales como método de extracción de factores en el AFE, y para la rotación ortogonal se usó la normalización Varimax con Kaiser. En el Apéndice se expone la matriz factorial de componentes rotados resultante.

Todos los dominios resultantes estuvieron representados al menos por tres ítems, al igual que en el trabajo original (Boor, 2011). Se consideró como significativo a un valor de carga factorial mínimo de 0.30 (Hair, 2010; Morales Vallejo, 2013). El agrupamiento de variables según su carga factorial mostró un patrón similar al estudio original, excepto en los dominios *Trabajo Adaptado a la Competencia del Residente* y *Educación Formal*, que revelaron un comportamiento en conjunto, es decir, los reactivos de ambos dominios cargan en un mismo factor (ítems 25-26-27-28 para el primero, y 37-38-39-40 para el segundo). Ello significó una reducción del número total

de dominios o subescalas respecto del original (de 11 a 10) manteniendo constante el total de ítems del test. Ese modelo se tomó como base para la realización del AFC y el cálculo del α de Cronbach.

- Consistencia Interna - Cálculo del coeficiente Alfa de Cronbach

El valor del coeficiente α de Cronbach fue de 0,972 para el test, revelando evidencia a favor de la confiabilidad de la medición global.

Para un análisis más detallado, y mayor aporte de evidencia, se analizó la consistencia interna para cada dominio, lo que arrojó un valor del coeficiente α de Cronbach mayor de 0,80 en todos los casos (ver tabla 3).

Tabla 3. Consistencia interna para cada dominio del test.

Dominio	α de Cronbach
Supervisión	**0,817**
Valoración del Desempeño	**0,896**
Retroalimentación	**0,873**
Trabajo en Equipo	**0,805**
Colaboración entre Pares	**0,814**
Relaciones entre Instructores	**0,813**
Trabajo Adaptado a Competencias + Educación Formal	**0,905**
Rol de los Instructores	**0,898**
Rol del Instructor Especialista	**0,907**
Registro de Alta del Paciente	**0,871**

- Análisis Factorial Confirmatorio

Se utilizó el software LISREL 8 (Jöreskog & Sörbom, 1996) para comprobar el ajuste de los modelos de estructuras de covarianzas. Dada la sensibilidad del X^2 al

tamaño de la muestra, se decidió emplear el valor X^2 de Satorra- Bentler dividido entre sus grados de libertad (SBχ2/gl) considerando un ajuste adecuado a un valor inferior a 2.0 (Arias, 2008)

En la tabla 4 se muestran los índices habituales para estimar la bondad de ajuste del modelo en AFC.

Tabla 4- Adaptada de Ruiz et. al (2010)

TABLA 4
ESTADÍSTICOS DE BONDAD DE AJUSTE Y CRITERIOS DE REFERENCIA

Estadístico	Abreviatura	Criterio
Ajuste absoluto		
Chi-cuadrado	χ^2	Significación > 0,05
Razón Chi-cuadrado / grados de libertad	χ^2/gl	Menor que 3
Ajuste comparativo		
Índice de bondad de ajuste comparativo	CFI	≥ 0,95
Índice de Tucker-Lewis	TLI	≥ 0,95
Índice de ajuste normalizado	NFI	≥ 0,95
Ajuste parsimonioso	NFI	
Corregido por parsimonia	PNFI	Próximo a 1
Otros		
Índice de bondad de ajuste	GFI	≥ 0,95
Índice de bondad de ajuste corregido	AGFI	≥ 0,95
Raíz del residuo cuadrático promedio	RMR	Próximo a cero
Raíz del residuo cuadrático promedio de aproximación	RMSEA	< 0,08

Respecto de los métodos de estimación en el AFC (como el de Máxima Verosimilitud o Maximum Likelihood), y considerando en este trabajo la presencia de variables con escala ordinal (en este caso Likert) y la ausencia de normalidad multivariante, el software LISREL 8.80 ofrece en su menú los métodos ULS (Unweighted Least Squares) y DWLS (Diagonally Weighted Least Squares) para tales condiciones (Arias, 2008; Morata-Ramírez et al, 2015; Li, 2014). Se decidió utilizar ambos métodos por separado para observar eventuales diferencias.

Los resultados obtenidos con ambos métodos se muestran en la tabla 5.

Tabla 5. Resultados en los índices de bondad de ajuste según el método

Método AFC	X^2/ grados de libertad *(<2)	NFI *(>0,90)	NNFI *(>0,95)	CFI *(>0,95)	SRMR *(<0,08)	RMSEA *(<0,06)	GFI *(>0,95)	AGFI *(>0,95)	N crítico
ULS	**1,90**	**0,98**	**0,99**	**0,99**	**0,044**	**0,048**	**0,99**	**0,99**	**226,2**
DWLS	**1,93**	**0,98**	**0,99**	**0,99**	**0,044**	**0,049**	**0,99**	**0,99**	**222,6**

AFC: análisis factorial confirmatorio. NFI: Normed Fit Index. NNFI: Non-Normed Fit Index
CFI: Comparative Fit Index. SRMR: Standardized Root Mean Square Residual
RMSEA: Root Mean Square Error of Approximation. GFI: Goodness-Of-Fit Index
AGFI: Adjusted Goodness-Of-Fit Index
Entre paréntesis, el valor de significación de cada índice (ver texto)

En este trabajo, todos los índices mostraron valores que revelan un adecuado ajuste para el modelo teórico.

Como parte del mismo análisis, el modelo de medida obtenido en software LISREL 8.80 confirmó el ajuste del modelo observado acorde con el teórico (respecto de las relaciones entre variables observadas y latentes). Dicho procedimiento incluyó indagar sobre la validez convergente y discriminante de los ítems respecto del constructo (Messick, 1979; Hair, 2010). Ambos aspectos surgen de los resultados del software estadístico:

- Las cargas factoriales estandarizadas para todas las variables latentes revelaron valores superiores a 0.60, y sus valores T por encima de 1.96 (tanto en método ULS como en DWLS). La Varianza Media Extractada (VME) de cada dominio arrojó un valor superior a 0,5. Todos estos resultados indican existencia de validez convergente. (Hair, 2010)

- La confiabilidad puede verse como un componente de la validez de convergencia (Hair, 2010). El cálculo de Fiabilidad Compuesta aporta evidencia acerca de la consistencia interna (como alternativa al α de Cronbach) en que todas las medidas miden consistentemente el mismo constructo. Un valor superior a 0,70 para cada variable latente brinda evidencia a favor de la confiabilidad de la medición, según puede verse en la tabla 6:

Tabla 6- Fiabilidad compuesta para cada dominio

Método	VL1	VL 2	VL 3	VL 4	VL 5	VL 6	VL 7	VL8	VL 9	VL 10
ULS	0,85	0,91	0,89	0,85	0,88	0,85	0,92	0,93	0,92	0,90
DWLS	0,86	0,91	0,93	0,78	0,89	0,85	0,92	0,93	0,93	0,88

VL: variable latente
Los valores deben ser superiores a 0,70 (Hair, 2010)

- Respecto de la validez discriminante, Fornell y Larcker (1981) proponen que cada constructo debería compartir más varianza con sus propios indicadores que con otros constructos del modelo. Por tanto, la varianza media extractada (VME) de cada constructo debe ser mayor que el cuadrado de las correlaciones con los otros (los resultados se muestran en tablas 7 y 8).

Tabla 7- Valores comparativos para validez discriminante (Fornell y Larcker, 1981) (método ULS)

	Variable latente 1	Variable latente 2	Variable latente 3	Variable latente 4	Variable latente 5	Variable latente 6	Variable latente 7	Variable latente 8	Variable latente 9	Variable latente 10
	0,66									
Variable latente 2	0,65	**0,58**								
Variable latente 3	0,47	0,65	**0,74**							
Variable latente 4	0,32	0,31	0,38	**0,60**						
Variable latente 5	0,36	0,34	0,20	0,23	**0,71**					
Variable latente 6	0,42	0,56	0,51	0,39	0,31	**0,66**				
Variable latente 7	0,56	0,60	0,50	0,42	0,36	0,65	**0,63**			
Variable latente 8	0,51	0,54	0.51	0,33	0,34	0,57	0,67	**0,66**		
Variable latente 9	0,43	0,53	0,57	0,32	0,25	0,51	0,68	0,59	**0,68**	
Variable latente 10	0,50	0,59	0,49	0,34	0,30	0,67	0,72	0,64	0,68	**0,71**

En negrita aparecen los valores de Varianza Media Extractada (VME)
En rojo aparecen los cuadrados de las correlaciones entre Variables latentes (subrayados si >VME)
V. latentes: 1- Supervisión 2- Valoración del desempeño 3- Retroalimentación 4- Trabajo en Equipo 5- Colaboración entre pares
6- Relaciones entre instructores 7- Trabajo adaptado a la competencia + Educación Formal
8- Rol de los especialistas 9- Rol de los instructores 10- Alta de los pacientes

Tabla 8- Valores comparativos para validez discriminante (Fornell y Larcker, 1981) (método DWLS)

	Variable latente 1	Variable latente 2	Variable latente 3	Variable latente 4	Variable latente 5	Variable latente 6	Variable latente 7	Variable latente 8	Variable latente 9	Variable latente 10
	0,67									
Variable latente 2	0,64	**0,59**								
Variable latente 3	0,43	0,60	**0,81**							
Variable latente 4	0,32	0,38	0,33	**0,61**						
Variable latente 5	0,33	0,33	0,17	0,22	**0,74**					
Variable latente 6	0,42	0,56	0,47	0,32	0,39	**0,66**				
Variable latente 7	0,54	0,60	0,44	0,40	0,47	0,65	**0,62**			
Variable latente 8	0,50	0,51	0.46	0,32	0,33	0,59	0,65	**0,64**		
Variable latente 9	0,43	0,51	0,32	0,24	0,24	0,51	0,68	0,59	**0,69**	
Variable latente 10	0,49	0,59	0,44	0,34	0,29	0,67	0,70	0,64	0,68	**0,71**

En negrita aparecen los valores de Varianza Media Extractada (VME)
En rojo aparecen los cuadrados de las correlaciones entre Variables latentes (subrayados si ≥VME)
V. latentes: 1- Supervisión 2- Valoración del desempeño 3- Retroalimentación 4- Trabajo en Equipo 5- Colaboración entre pares
6- Relaciones entre instructores 7- Trabajo adaptado a competencia + Educación Formal
8- Rol de los especialistas 9- Rol de los instructores 10- Alta de los pacientes

- Generalizabilidad

Para sumar evidencia sobre la confiabilidad de la medición, se efectuó un estudio G, determinando el coeficiente de generalizabilidad y los porcentajes de varianza de las facetas, para el cuestionario global y para cada dominio. El estudio se ejecutó a través de dos modelos de análisis, el primero de ellos de tipo cruzado (residentes/ítems) cuyos resultados revelaron un aporte de varianza de la faceta de diferenciación (Estudiantes) de un 45,8 %, y para la faceta de generalización (Ítems) de 0,05 %, que representan valores satisfactorios. En este modelo E/I de análisis, el cálculo del coeficiente G equivale al estudio de la consistencia interna del test. En nuestro caso, el valor obtenido (0,97) revela una alta confiabilidad para la prueba, de modo similar al α de Cronbach (ver en Anexo 13 la tabla de resultados originales de G-String/urGenova)

En este mismo modelo se analizó la confiabilidad de cada dominio por separado, obteniéndose los siguientes valores (ver tabla 9).

Tabla 9 – Coeficiente G para cada dominio

Dominio	Coeficiente G
Supervisión	**0,83**
Valoración del Desempeño	**0,88**
Retroalimentación	**0,87**
Trabajo en Equipo	**0,80**
Colaboración entre Pares	**0,81**
Relaciones entre Instructores	**0,81**
Trabajo Adaptado a Competencias + Educación Formal	**0,90**
Rol de los Instructores	**0,90**
Rol del Instructor Especialista	**0,91**
Registro de Alta del Paciente	**0,87**

Estos valores son similares a los obtenidos para α de Cronbach (ver Tabla 3), y son considerados satisfactorios para la confiabilidad de cada dominio individual

El segundo modelo de estudio G para el test global, se realizó considerando como faceta de diferenciación a los Hospitales, y con dos facetas de generalización: Residentes (anidados en Hospitales) e Ítems.

El Coeficiente de Generalizabilidad obtenido fue de 0,87. Todos estos valores fueron considerados satisfactorios para la confiabilidad del test global. En este modelo de estudio, para un resultado confiable del test se calculó un número mínimo de 16 residentes encuestados por cada Hospital.

DISCUSIÓN

El presente trabajo tuvo como objetivo general lograr la adaptación transcultural de una herramienta diseñada para valorar el CA en las Residencias, y reunir evidencia sobre sus propiedades psicométricas.

Cuando se pretende utilizar un instrumento específico de medición como una encuesta, no sólo debe asegurarse que dicho instrumento a utilizar sea confiable, sino que además sea válido en la población a investigar (Arribas, 2006). En nuestro caso, una herramienta originada en Holanda y destinada a evaluar el CA por residentes de ese país, es sometida a prueba para verificar su apropiada utilización en nuestro medio.

Si se considera a la validez como la ausencia de sesgo (Díaz et al, 2003), puede señalarse:

a) La primera forma (*sesgo de constructo*) representa una diferencia del concepto entre grupos culturales, ya sea por desigualdad en la definición del constructo o por existir una apropiación diferencial de conductas asociadas al constructo (p.ej. una habilidad que no forma parte del repertorio cultural de uno de los grupos) (van de Vijver & Poortinga, 2005*)*.

El término *Clima de Aprendizaje* (CA) se refiere a diversas locaciones físicas, contextos y culturas en que los estudiantes aprenden. Como esto puede ocurrir en una amplia variedad de escenarios (incluso fuera de una escuela y en ambientes exteriores), el término es una alternativa mejor que *salón de clase*, que tiene connotaciones más limitadas y tradicionales –una sala con filas de pupitres y pizarrón, por caso-. El término también abarca la cultura institucional o de la clase –su propio *ethos* y características, incluido el modo en que los individuos interactúan entre sí- así como la forma en que los profesores organizan el marco educacional para facilitar el aprendizaje (“Learning environment”, 2013). Entre los componentes del CA se mencionan los recursos de enseñanza disponibles, objetivos claros de aprendizaje, competencias de los instructores, metodologías de enseñanza y evaluación, horarios, apoyo al estudiante,

atmósfera grupal, entre otros (Schonrock-Adema et al, 2012; Roff & McAleer, 2001; Palés et al, 2015)

Un elemento central en este punto es definir el significado de la expresión "residente", para establecer la población destino de la presente investigación.

El término "residente" es referido en el trabajo original (Boor et al, 2011) como un profesional "junior" que se encuentra en formación de posgrado luego de obtener su título en una carrera de 6 años. El entrenamiento de posgrado prevé una duración de 4 a 6 años, donde un especialista se constituye en responsable formal de su educación, entrenamiento y evaluación.

El vocablo "residente" del trabajo original se muestra equivalente con nuestro trabajo, en cuanto a significación, duración de carrera de grado y posgrado, como también el ámbito hospitalario para el desempeño de la actividad formativa y asistencial, interactuando con pares, instructores, y resto del personal de salud.

Estos componentes de la definición de CA se revelan comunes en ambas muestras. Por tanto asumimos que no hay evidencia de sesgo en este punto respecto de la muestra poblacional.

Asimismo, el término "Clima de Aprendizaje", referente al escenario o contexto de actividades de las residencias en las especialidades, constituye un constructo teórico cuya medición no es factible directamente, sino a través de las percepciones de los residentes: el constructo subyace bajo sus experiencias cotidianas (Boor et al., 2011).

Formulando las preguntas que competen al investigador en cuanto a la validez de constructo en una adaptación transcultural, corresponde decir (Hambleton et al., 2005): ¿Es posible comparar este constructo en estas dos culturas? ¿El constructo que se pretende medir tiene similar significado en las culturas donde se compara? ¿El constructo funciona operativamente de la misma manera en ambas culturas estudiadas? En definitiva, interesa responder si ambas versiones lingüísticas del instrumento miden el mismo constructo (van de Vijver & Poortinga, 2005*)* La eventual existencia de sesgo afectaría la equivalencia de los valores obtenidos, y en ese sentido pueden delinearse cuatro niveles o jerarquías de equivalencia:

- *Ineguivalencia de constructo: hay completa falta de comparabilidad (comparar naranjas y manzanas) y determina la imposibilidad de lograr comparaciones transculturales.*
- *Equivalencia estructural o funcional: para lograr identidad de constructo entre grupos se requieren procedimientos y operacionalizaciones en el instrumento de medición. Los patrones de correlación entre variables deben ser los mismos.*
- *Equivalencia de unidad de medida: existe igual métrica (a nivel de intervalos) en las escalas de ambos grupos, aunque tienen distinto origen. Restringe la comparación de grupos entre sí, sólo permite comparar valores individuales intragrupo.*
- *Equivalencia escalar o de puntaje completo: las mediciones tienen iguales unidades de medida y origen, permitiendo superar fronteras culturales y lograr comparaciones válidas.*

La segunda categoría usualmente atañe a las adaptaciones transculturales en distinto idioma. Aquí cobra esencial relevancia el uso de análisis factorial confirmatorio y/o exploratorio seguido de rotaciones, apuntando a comparar la estructura factorial de los ítems entre muestras poblacionales, como en nuestro caso.

Dado que ambos trabajos mostraron adecuada equivalencia en ese aspecto, no se evidencian diferencias significativas que representen una fuente de sesgo. Más adelante retomaremos el punto al discutir los procedimientos estadísticos puestos en juego para la validación.

b) El *sesgo de método* (el origen del término radica en que sus componentes aparecen descriptos en la sección “Método” de los trabajos) incluye dos variantes:

- El sesgo instrumental, por las características del instrumento que, sin formar parte específica del estudio, inducen desvíos en los valores.
- El sesgo de administración del test, que incluye los desvíos originados en la aplicación de la prueba sobre la muestra (p.ej. dificultades comunicacionales, error de muestreo, y otros).

La consecuencia de este sesgo lleva a la aparición de diferencias en los valores observados que no tienen que ver con el constructo, sino con artefactos de la medición. El investigador enfrenta entonces una decisión sobre dos opciones opuestas sobre las diferencias observadas: válidas y relacionadas con el constructo vs. Sesgo de método. Un apropiado diseño investigacional y del instrumento involucra el esfuerzo por evitar esta eventualidad (van de Vijver & Poortinga, 2005)

Consideramos que ambas muestras poblacionales, tanto en el estudio original como en el presente trabajo, presentan características (socioculturales, profesión y nivel educativo) similares, y comparables en su capacidad de interpretación de los ítems del cuestionario. Asimismo el sondeo preliminar con el grupo de prueba (validez aparente) no mostró dificultades ni originó controversias en ese sentido.

En este caso, se consideró esencial la preservación de las distintas formas de equivalencia lingüística del instrumento (conceptual, técnica, lectura y comprensión) en el producto final para su administración, revelado en el análisis de expertos (como consigna dentro de la tarea) y posterior sondeo de prueba con entrevistas a los residentes.

c) El *sesgo de ítem* significa que un determinado reactivo se comporta psicométricamente diferente para diversos grupos, es decir, el funcionamiento del ítem es disímil frente a grupos comparables, esto es, grupos similares respecto al constructo o rasgo medido por el test. Dicho de otro modo, el sesgo de ítem surge si grupos igualmente capaces presentan una probabilidad distinta de responderlo con éxito o en cierta dirección, en función del grupo o cultura donde pertenecen (van de Vijver & Poortinga, 2005)

La incorrecta traducción, la ambigüedad del ítem, o connotaciones diferenciales vinculadas a las culturas suelen ser fuentes de su origen. Los ítems con alta chance de respuesta en un determinado sentido (> 90%) son poco útiles, ya que no contribuyen a discriminar entre individuos diferentes respecto de una característica. Además, tales preguntas podrían agregar un innecesario error de medición en caso de respuestas aleatorias, u otras razones no relacionadas con la respuesta verdadera (Keszei, Novak & Streiner, 2010).

El análisis por expertos en la validación de contenido, y el sondeo de prueba constituyen medios para evitar este sesgo; en nuestro caso esta condición no reveló discrepancias relacionadas con este aspecto, ni generó la necesidad de modificar significativamente el texto o la terminología empleada, para adaptarlo a este contexto cultural. La escala métrica del cuestionario coincide con la original, y su comprensión y aceptación por el grupo de expertos como por los residentes no reveló dificultad o rechazo alguno. El ordenamiento de los ítems dentro del formulario es la que aparece en el trabajo original.

En cuanto a la implementación de la prueba, si bien existen algunas diferencias particulares, en lo general fueron equivalentes. En ambas investigaciones la participación fue voluntaria. En el trabajo original, la invitación, distribución y devolución del cuestionario fue por vía de correo electrónico a todos los residentes del país a partir de una base de datos. Algunos encuestados recibieron una incentivación para responder.

En nuestro caso, la gestión y el contacto con los residentes fue personal, la participación fue voluntaria y gratuita, y el anonimato fue garantizado. Esta estrategia se utilizó a fines de lograr la mayor colaboración sin afectar la sensibilidad de los participantes. Asumimos que esta diferencia metodológica no fue generadora de sesgo al no interferir sobre la accesibilidad, la circulación, la lectura y el llenado del cuestionario, o algún otro aspecto metodológico de la investigación.

Una última forma de sesgo (por fuera de las anteriores) que merece mención es el *sesgo de muestra*. Se refiere a diferencias sistemáticas subyacentes (además de las culturales) en las características de las muestras que se comparan, e.g., poblaciones creadas por diferentes mecanismos de selección (van de Vijver & Poortinga, 2005). Por todo lo anteriormente expuesto, consideramos que no ocurre en este caso.

La conclusión respecto del primer objetivo específico del trabajo es que las pruebas aportadas son suficientes para demostrar que se atendieron todas las circunstancias inherentes a esta etapa del trabajo, y el resultado es el correcto para evitar las distintas variantes de sesgo.

La validación del test en este estudio abarcó tres de los cinco criterios definidos por la *Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education* (1999) como evidencia de validez: *Contenido, Proceso de respuesta*, y *Estructura interna y confiabilidad*. Los dos criterios restantes (*Relación con otras variables* y *Evaluación de las consecuencias del Test*) se comentarán en último lugar.

En cuanto al primer criterio (*Contenido*), el aporte de este trabajo se acopla con otras fuentes: en primer lugar, un estudio cualitativo previo de los autores (Boor et al., 2009) con 40 residentes de distintos niveles de entrenamiento y especialidades, exploró sobre los componentes del CA, la vinculación entre la labor asistencial y la formación, y sus necesidades de aprendizaje, el rol de la interacción con el resto del personal de salud, para definir los componentes de un CA positivo. Sobre esa base se diseñó la primera versión del cuestionario que en sucesivas etapas (evaluación por panel Delphi, análisis factorial, y estudio de generalizabilidad) logró el D-RECT definitivo (Boor et al., 2011)

La validez de contenido abarca dos aspectos: la *relevancia* (se refiere a definir el dominio específico en cuestión y sus límites operativos) y la *cobertura* (el proceso de selección de una muestra representativa) (Messick, 1979). Para este fin, el investigador se enfoca en la definición del constructo, el propósito planeado para el instrumento, el proceso de desarrollo y selección de los ítems, la redacción de cada ítem individual y la intervención de calificadores (Cook & Beckman, 2006)

El presente trabajo cumplió con todas las tareas previstas para cumplir con este criterio, según lo establecido en las directivas de la ITC y la bibliografía respectiva (Hambleton & Bollwark, 1991; Hambleton et al., 2005; Muñiz et al, 2013; Beaton et al., 1998, 2000; Arribas, 2006), respecto de las instancias de validación de contenido con intervención del comité de expertos sobre el cuestionario.

Respecto del segundo criterio (*Proceso de Respuesta*), la revisión de las acciones y pensamientos de los encuestados puede “iluminar el ajuste entre el constructo y la detallada naturaleza del desempeño involucrado” (AERA et al., 1999,

Cook & Beckman, 2006). En este sentido, siguiendo a Padilla & Benítez (2014, p. 137) es importante marcar que las fuentes de validez son complementarias, especialmente las de *Contenido* y *Proceso de respuesta*, tanto en sus objetivos como en las áreas que abarcan. Estos autores señalan que este enfoque complementario de ambas fuentes apunta a la detección e interpretación del funcionamiento diferencial de los ítems (ver *Sesgo de ítem* más arriba). Las condiciones para esta fuente de validez pueden enumerarse como sigue:

- El desempeño de los encuestados refleja los procesos psicológicos y operaciones cognitivas específicas del test.
- Los procesos de evaluación de los examinadores sobre los resultados del test, son consistentes con la interpretación planeada sobre los puntajes.
- Los grupos de encuestados (definidos por sus condiciones demográficas, lingüísticas u otras asociadas con el propósito del test) no difieren en la naturaleza de su desempeño o respuestas, por causas de varianza irrelevantes al constructo.

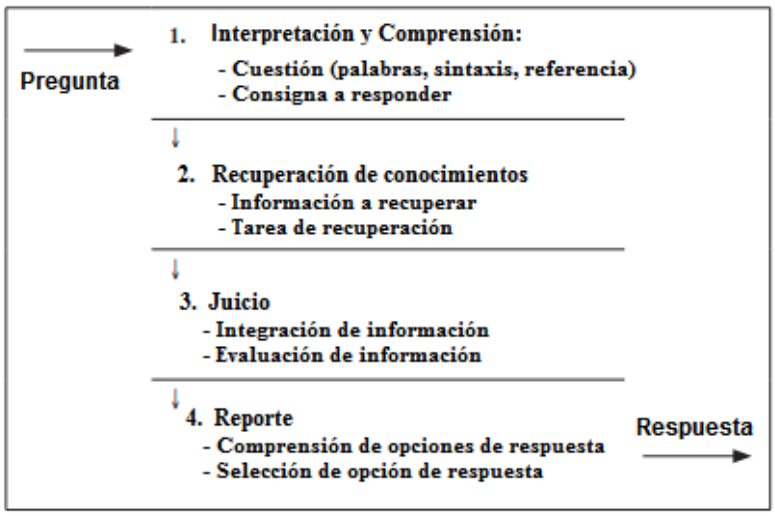
Los métodos para obtener evidencia sobre el proceso de respuesta pueden clasificarse en dos categorías (Padilla & Benítez, 2014):

- Con acceso directo a los procesos psicológicos y operaciones cognitivas de los encuestados (entrevistas, en voz alta y grupos enfocados)
- Indicadores indirectos que requieren inferencias (tiempo de respuesta y movimientos oculares)

En nuestro trabajo se utilizó el método de entrevistas, habitualmente utilizado por los investigadores para validación basada en el proceso de respuesta. Además de contemplar los procesos de pensamiento de los encuestados, su aplicación es fácil y requiere pocos recursos. Su base teórica considera una secuencia de procesos cognitivos que ocurren entre la formulación de una pregunta y la emisión de respuesta. En la figura siguiente, se muestra un modelo de esta secuencia: primero, interpretar y entender el ítem involucra al propósito, los conceptos y expresiones incluidas. Luego,

recuperar la información necesaria, hacer un juicio para integrar y evaluar esa información, y finalmente ajustar la respuesta entre las alternativas al propósito y comunicarla (ver figura 3)

Figura 3. Representación del modelo pregunta-a-respuesta (Adaptado de Padilla & Benítez, 2014)



Con esta base, corresponde someter el test a un testeo sobre un grupo piloto, lo que implica dos decisiones: quiénes y cuántos sujetos formarán dicho grupo. La decisión no es a priori una cuestión numérica. En primer lugar, el grupo piloto debe poseer similares características demográficas, lingüísticas, culturales, etc., respecto de la población destino. Dos condiciones cualitativas definen el número total de participantes de este testeo inicial: saturación teórica (continuar el testeo hasta que no aparece nueva información) y relevancia (seleccionar entrevistas por su importancia sobre los hallazgos surgidos). Las entrevistas pueden estar estructuradas en formularios que guíen el procedimiento, para asegurar la adecuada saturación muestral.

En nuestra investigación, el número de 15 residentes fue suficiente para cumplir este criterio de validación en forma satisfactoria. Por las características del test, no fue necesario asegurar capacitación de evaluadores u observadores.

El tercer criterio de validez (*Estructura Interna y Confiabilidad*) merece los siguientes comentarios respecto de su verificación en este trabajo:

Tamaño muestral

El tamaño de la muestra es un aspecto esencial en estudios que incluyen modelos de ecuaciones estructurales (SEM, en inglés), aunque la bibliografía no establece una respuesta concluyente para determinar la cantidad de casos requeridos para un adecuado análisis, sino por el contrario, la diversidad de criterios constituye una masa desarticulada de literatura que dificulta el trabajo del investigador (Vargas-Halabí & Mora Esquivel, 2017). Otros autores remarcan la influencia de ciertas características del estudio para determinar el n de muestras, como la presencia de normalidad multivariada, el número de variables latentes, indicadores y comunalidades (Hair et al., 2010; Bentler, 1989). Un estudio comparativo sobre distintos modelos de SEM, propone adoptar la sugerencia general de un mínimo de 200 observaciones, asegurando que los indicadores sean cuidadosamente elegidos y confiables, y que existe un adecuado número de indicadores por variable latente (Jackson, 2003). Westland (2010) propone que el cociente *indicadores/variables latentes* sea la base para calcular el tamaño muestral, con la fórmula de la figura 4:

Figura 4. Fórmula para calcular tamaño muestral en SEM (Adaptado de Westland, 2010)

$$n \geq 50r^2 - 450r + 1100$$

r = cociente indicadores/variables latentes

Para el presente trabajo, n= 100.

MacCallum et al. (2000) rechazan establecer una regla para calcular el tamaño muestral, que podrían resultar inválidas en el contexto del SEM.

Independientemente de la proporción de sujetos por variable, Arias (2008) recomienda un mínimo de 150 observaciones, descartando aquellas con puntuaciones extremas. En términos generales muestras más pequeñas pueden ser adecuadas si la estructura factorial es muy clara (al menos más de tres variables definiendo cada factor y ninguna variable en más de un factor).

Streiner (1994) recomienda para el AFE un mínimo de 5 sujetos por variable, con la premisa de un número de por lo menos 100. En caso de un tamaño menor, entonces la relación de sujetos/ítem debería ser 10/1.

Dado que el requerimiento de tamaño muestral para el Análisis Factorial Confirmatorio para algunos autores (Joreskog & Sorbom, 1996; Hair et al., 1995; Norman & Streiner, 2008) podría representar una limitación para este trabajo, nos decidimos a realizar AFE en primera instancia. Sin embargo, varios estudios indican que este tamaño muestral puede resultar satisfactorio para la mayoría de los casos (Li, 2016; DiStefano & Morgan, 2014; Rhemtulla et al., 2012; Savalei & Rhemtulla, 2013; Yang-Wallentin et al., 2010) sobre todo cuando hay varios indicadores por dimensión, los ítems no son binarios y las cargas factoriales son altas (Forero et al., 2009). Por lo tanto verificamos también el ajuste del modelo con AFC.

En el trabajo original de creación del D-RECT, se tomó como base mínima de 5 sujetos por ítem (Boor et al., 2011). En este trabajo, se consideró un número mínimo de 5 sujetos por ítem en base a lo anterior. Dado que la muestra total para el análisis factorial fue de 392 sujetos, los resultados fueron compatibles con la interpretación de las cargas del análisis factorial exploratorio (Hair et al., 2010). Como veremos posteriormente, también el resultado del AFC reveló la cifra de 222,68 (DWLS) y 226,20 (ULS) como número crítico de observaciones, brindando mayor soporte a las conclusiones obtenidas.

Análisis Factorial

El AFE se realizó mediante el método de componentes principales y rotación ortogonal con Kaiser, dado que se recomienda para la reducción de datos y por ser el más utilizado (Hair et al., 2010). Este análisis reveló un comportamiento de los ítems similar al estudio original (Boor et al., 2011), excepto que 8 ítems (del 25 al 28, y del 37 al 40) se agruparon en torno a un solo factor (en el original se agruparon en dos). Esta característica fue interpretada por un comité de 4 expertos como debida al similar contenido conceptual de esos ítems. Eso significó una reducción del número de dominios (de 11 a 10) sin afectar la confiabilidad de los dominios (ver α de Cronbach).

Para explicar este comportamiento, consideramos que el aspecto educativo y el asistencial no constituyen entes separados en la residencia. Incluso, en varias instancias los mismos residentes se convierten en docentes dentro de sus ámbitos. Citamos: " El proceso asistencial y educativo en las instituciones de salud implica diversos actores, uno de los cuales es el médico residente. Este grupo desarrolla simultáneamente dos roles diferentes, como alumnos de programas de especialización de las universidades y como profesionales de la salud empleados por las instituciones hospitalarias" (Méndez-López et al., 2016) En Argentina, Carballido et al. (1998) realizaron una encuesta del estado de las residencias de Cardiología. En el apartado sobre actividad docente, documentaron que la docencia depende fundamentalmente de los residentes, con menor participación de los profesores formales.

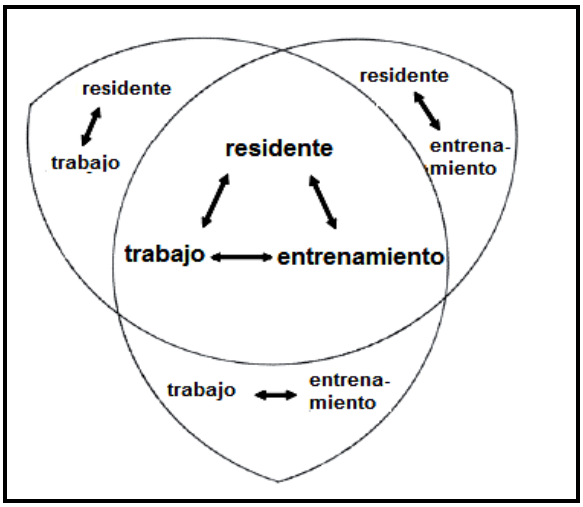
Según Rodríguez Fernández et al. (2017) la forma fundamental de organización de la enseñanza en la residencia es la educación en el trabajo, es decir, el conjunto de actividades del residente en el servicio de salud bajo supervisión docente, para adquirir habilidades, hábitos de trabajo y búsqueda de información para la resolución efectiva de los diferentes problemas de salud y el logro de los objetivos del programa de formación profesional. Este sistema de trabajo pedagógico para las especialidades ocurre en los centros de educación médica superior, y comprende actividades de los profesores, residentes y el resto del colectivo laboral, las cuales inciden directa o indirectamente en los procesos docentes, de atención médica e investigación, integrando aspectos táctico-operativos y estratégico- perspectivos, cuyos componentes no pueden considerarse de forma aislada.

Ruiz de Gauna et al. (2015) explica con mayor profundidad (citamos): "La competencia, el ser competente, es el resultado (la demostración, la comprobación y la evaluación) de un aprendizaje holístico. Para que haya aprendizaje integrado, el estudiante o residente debe afrontar situaciones complejas que tengan sentido para él. Mientras resuelve las situaciones, aprende, y cuando lo ha conseguido puede decirse que ha tenido un desempeño competente. Para desarrollar este tipo de aprendizaje es de suma importancia elaborar un banco de situaciones, porque la situación empuja a todo lo demás. No es a la inversa (primero las competencias y después las situaciones), porque caeríamos fácilmente en la pedagogía por objetivos y en los aprendizajes

desconectados entre sí. Desde este planteamiento, competencias y situaciones son lo mismo, puesto que la competencia es un conocimiento situado."

Consideramos importante remarcar que quien contesta el cuestionario **es el residente**, puesto que para los docentes que planifican y supervisan, resulta natural asumir esa disociación. Estas citas resumen que para el residente no hay confusión en asumir la tarea asistencial y la educación como una combinación de ambas. el pase de sala, el alta de los pacientes, la atención ambulatoria, la guardia médica, la entrega de guardia, la discusión diagnóstica y la atención médico quirúrgica, no sólo son prácticas asistenciales sino también actividades de aprendizaje (Rodríguez Fernández et al., 2017). Son dos caras de la misma moneda en el proceso de formación, expresadas en las preguntas de esos dos dominios, y en el contexto cultural de nuestro medio podría explicar el comportamiento factorial. En la investigación cualitativa de Boor et al. (2009), los residentes describen un óptimo CA como la combinación de trabajo y entrenamiento, ajustados a sus necesidades específicas. Los residentes sienten que su mayor aprendizaje proviene de la atención de pacientes y valoran su participación en esas instancias, aunque el CA se ve fuertemente afectado por un difícil balance entre provisión de servicio y educación. La figura 5 esquematiza la interacción entre trabajo, entrenamiento y necesidades de los residentes.

Figura 5. Interacción de los tres aspectos del CA (adaptado de Boor et. al., 2009)



El CA en la formación de especialistas es afectado por tres aspectos, que interactúan de 4 maneras diferentes. El centro de la figura muestra un óptimo CA con integración de los 3 aspectos. Las áreas externas muestran situaciones con interacción entre 2 aspectos.

Por otro lado, el cuestionario D-RECT fue recientemente testeado por algunos de sus autores originales (Silkens et al., 2015) sobre una muestra mayor de residentes. En esa revalidación, se verificó un cambio en la estructura del cuestionario: el número de ítems se redujo a 35, y el número de factores a 9, luego de un nuevo análisis factorial. No obstante, los autores remarcan la necesidad de nuevas comprobaciones en la aplicabilidad del instrumento.

En nuestro trabajo, las cargas factoriales fueron adecuadas para los ítems dentro de cada dominio (> 0,30) considerando el tamaño muestral del estudio (Hair et al., 2010). Todos los dominios se mantuvieron definidos por 3 ítems como mínimo apropiado para identificarse como factor (Morales Vallejo, 2013).

A continuación, para mayor aporte de evidencia, el ajuste del modelo se verificó también mediante AFC. Dado que las variables del estudio son ordinales, y que la literatura considera para ese caso a los métodos robustos (Li, 2016; DiStefano & Morgan, 2014; Rhemtulla et al., 2012; Savalei & Rhemtulla, 2013; Yang-Wallentin et al., 2010; Cupani, 2012), el ajuste de los modelos de estructuras de covarianzas fue verificado con el contraste de X^2 corregido de Satorra-Bentler (1994) dividido por sus grados de libertad (SBχ2/gl), asumiendo ausencia de normalidad multivariada para variables ordinales y su comportamiento robusto en tales condiciones, además de la sensibilidad del X^2 al tamaño de la muestra (Arias, 2008).

El método habitual de estimación en el AFC para variables continuas es el de Máxima Verosimilitud o Maximum Likelihood. La utilización de variables con escala ordinal y ausencia de normalidad multivariante en esta investigación demanda métodos robustos como el ULS (Unweighted Least Squares) y DWLS (Diagonally Weighted Least Squares) para tales condiciones (Arias, 2008; Morata-Ramírez et al., 2015; Li, 2014) que difieren levemente en su matriz de carga de datos y el tamaño muestral requerido. Ambos mostraron resultados similares:

Entre los distintos índices de ajuste global, el NFI (Normed Fit Index, o Índice de Ajuste Normalizado) que varía entre 0.0 y 1.0, revelando buen ajuste con valores

superiores a cercanos a 1.0. También el NNFI (Non-Normed Fit Index, o Índice de Ajuste No Normalizado) incorpora los grados de libertad en la comparación de ambos modelos evitando la subestimación del ajuste, con el eventual inconveniente de mostrar valores fuera del rango (0 a 1). Su interpretación es coincidente con la del NFI: un ajuste adecuado ocurre con resultados >0,95, mejores cuanto más cercanos a 1.

El índice CFI (Comparative Fit Index, o Índice de Ajuste Comparado) también puede usarse en caso de muestras pequeñas, con valor en el rango 0 a 1, con interpretación similar a los anteriores.

El RMR (Root Mean Square Residual o raíz cuadrada media residual) es la raíz cuadrada del cuadrado promedio de residuo de covarianzas. Junto con su alternativa estandarizada (el SRMR) estiman el grado de desviación de esos valores respecto de la predicción del modelo teórico, y como su rango es de 0 a 1, un buen ajuste mostrará valores inferiores a 0,08 (mejor cuanto más próximo a 0) (Hair et al., 2010; Arias, 2008).

El RMSEA (Root Mean Square Error of Approximation - raíz cuadrada media del error de aproximación) es un indicador de uso habitual que pretende corregir la complejidad del modelo y el tamaño de la muestra, e incluye un rango de confianza para el valor. El valor de ajuste adecuado para RMSEA es menor a 0.06 (o 0.07 si el CFI es $\geq$0.90) (Arias, 2008; Hair et al., 2010)

El GFI (Goodness of Fit Index o Índice de Bondad de Ajuste) es sensible al tamaño muestral, y su rango es de 0 a 1, con un valor satisfactorio superior a 0,95. El AGFI (Adjusted Index of Fit Index, o AGFI Ajustado) es una corrección del anterior tomando en cuenta la cantidad de parámetros y datos disponibles en función de la complejidad del modelo y el tamaño de la muestra (Hair et al., 2010)

Todos ellos mostraron valores compatibles con un ajuste adecuado del modelo respecto del teórico.

La misma conclusión puede alcanzarse con la Estrategia de Dos Índices (Hu & Bentler, 1999) donde dichos autores proponen combinar dos indicadores para mostrar la bondad de ajuste, con las siguientes posibilidades (tabla 10):

Tabla 10- Estrategia de Dos Índices (Hu y Bentler, 1999)

Combinación de Índices	Regla de combinación
NNFI y SRMR	NNFI $\geq$ 0,96 y SRMR $\leq$ 0,09
RMSEA y SRMR	RMSEA $\leq$ 0,06 y SRMR $\leq$ 0,09
CFI y SRMR	CFI $\geq$ 0.96 y SRMR $\leq$ 0,09

Los autores propusieron combinar RMSEA y SRMR para muestras >250 sujetos, y los restantes para muestras < 250. Esta estrategia no es aceptada universalmente.

NNFI: Non-Normed Fit Index SRMR: Standardized Root Mean Square Residual
CFI: Comparative Fit Index. RMSEA: Root Mean Square Error of Approximation.

A partir de estos resultados pueden establecerse otras categorías de validez (Messick, 1979; Hair et al., 2010):

- Validez convergente: se refiere a proveer evidencia de una relación coherente entre medidas del mismo constructo.
- Validez discriminante: por contrapartida, probar que la medida no posee relación significativa con mediciones de otro constructo; es decir, ese constructo mide un concepto diferente de otros constructos.

En nuestro estudio, la validez convergente fue demostrada por las cargas factoriales estandarizadas (> 0,60) y valores T (> 1,96) en ambos métodos de estimación del AFC. Del mismo modo, la Varianza Media Extractada > 0,5 para cada dominio indica similar conclusión (Hair et al., 2010)

Con respecto a la validez discriminante, se consideró aceptable ya que las Varianzas Medias Extractadas de los distintos dominios no son inferiores a la mayoría de los cuadrados de sus correlaciones con otros constructos.

Consistencia Interna

Como evidencia sobre confiabilidad, la consistencia interna estima la correlación promedio entre todos los ítems de la medición, dado que se basan en la misma dimensión subyacente. Una baja consistencia interna significa que los ítems miden

diferentes atributos o que las respuestas de los sujetos son inconsistentes. El índice α de Cronbach es sencillo de obtener, pero merece algunas observaciones (Keszei, Novak & Streiner, 2010):

- Al no tomar en cuenta variaciones en el tiempo o interobservador, es una estimación optimista de la verdadera confiabilidad del test.
- su sensibilidad no sólo depende de la consistencia interna de la escala, sino también de su extensión. Tiende a valores elevados con más de 15 ítems, independientemente de su correlación, y con escalas más largas, otros índices deben combinarse.

.

El cálculo del Coeficiente Alfa de Cronbach como medida de consistencia interna mostró resultados a favor de la confiabilidad para el test global, y para cada dominio en particular. Cabe remarcar que en el trabajo original (Boor et al., 2011) el α de tres dominios en ese estudio fue inferior a 0,70 (Supervisión=0,64; Trabajo en equipo=0,69; Trabajo adaptado a competencia del residente=0,66). En nuestro trabajo, todos los valores estuvieron por encima de 0,80.

Otro dato relevante a favor de la confiabilidad de la medición (como alternativa al α de Cronbach) es el valor de Fiabilidad Compuesta obtenido (> 0,70 para cada variable latente). Esto indica una adecuada consistencia interna donde todas las variables miden consistentemente el mismo constructo (Hair et al., 2010).

Generalizabilidad

En la teoría G, los resultados de un test son considerados como una muestra de un universo de observaciones admisibles. Este universo consiste de todas las posibles observaciones que un investigador podría recabar sustituyendo a la original. Cada oportunidad de aplicación del test posee condiciones propias (como ser, el formulario del test, los ítems, el evaluador, la ocasión de la medición) denominadas *facetas*. Es decir, el universo de observaciones admisibles está definido por todas las posibles combinaciones de estas facetas (Shavelson & Webb, 1991).

El propósito de un estudio G es obtener estimaciones de componentes de varianza dentro del universo de observaciones admisibles. (Brennan, 2001). Partiendo

de la teoría de la generalizabilidad (TG), las múltiples fuentes de variabilidad (o error) en una medición pueden estimarse por separado en un solo análisis (Shavelson & Webb, 1991) permitiendo al investigador establecer la confiabilidad y la factibilidad de generalizar los resultados de la medición al universo de generalización. Como resultado global, se obtiene un valioso indicador de confiabilidad denominado "coeficiente de generalizabilidad". Esto hace posible cuantificar y distinguir las fuentes de inconsistencias en los puntajes surgidas al replicar una medición. Los datos así obtenidos incluyen errores, que dependen de las distintas condiciones en que se efectúa la medición. En la TG, el "error" no significa una "falta" en el sentido convencional, y en el caso de un procedimiento de medición, es en parte una cuestión de definición (Brennan, 2011). La interpretación de los resultados depende del diseño establecido por el investigador para el estudio G, asignando como "faceta de diferenciación" al objeto de interés (de donde proviene la varianza verdadera), y a las condiciones de medición como "facetas de generalización" o "instrumentación" (fuentes de la varianza de error). La estimación de cada componente se realiza mediante el método ANOVA (Análisis Of Variance), y el coeficiente G es la proporción de la varianza observada atribuida a la puntuación universo, es decir, es la razón entre el valor esperado de varianza de puntuaciones universo, y el valor esperado de varianza de puntuaciones observadas (Blanco Villaseñor, 1989).

En nuestro estudio, los resultados del estudio G revelaron evidencia a favor de la confiabilidad del instrumento global y de los dominios individuales.

Corresponden algunos comentarios finales acerca de los dos últimos criterios de validez, en relación con el presente trabajo.

El criterio de *Relación con otras variables* requiere contar con estudios que midan el mismo constructo para demostrar comparativamente similares resultados, en forma simultánea (*validez concurrente*) o luego de transcurrido cierto tiempo (*validez predictiva).* Esto permitiría estimar la eficacia para predecir o pronosticar el comportamiento de los individuos respecto del constructo. Un estudio en Irlanda (Bennett et al., 2014) comparó el puntaje del D-RECT al ingreso y al año dentro del programa de formación de posgrado del Royal College of Physicians of Ireland,

observando un cambio (gap) en los puntajes globales y en las subescalas, atribuyendo importancia al test en la medición del CA para sugerir cambios en los programas educativos, y evitar el éxodo de sus profesionales.

En Holanda, van Mook et al. (2016) investigaron la relación entre CA (medido por el D-RECT) y el desempeño de los instructores docentes (medido por el Maastricht Clinical Teaching Questionnaire (MCTQ)) como indicadores de calidad del programa de formación de especialistas en Cuidados Críticos, y hallaron una correlación significativa entre ambos instrumentos.

Lombarts et al. (2014), en un estudio de mayor envergadura, investigaron la relación entre CA (medido por D-RECT) y calidad de desempeño docente (medido por el System for Evaluating Teaching Qualities (SETQ)) en 45 programas de residencia de 17 hospitales de Holanda. Una positiva asociación fue demostrada entre ambos constructos, y particularmente tres dominios del D-RECT estuvieron asociados con el mejor desempeño docente.

En el mismo país, van Vendeloo et al. (2014) demostraron significativa correlación entre el CA (medido por D-RECT), calidad de vida y Burn-out entre residentes de Ortopedia.

Silkens et al. (2016) efectuaron un seguimiento de 3 años consecutivos sobre CA en 223 programas de residencia en Holanda, mediante el D-RECT. Los autores demostraron una leve aunque significativa mejoría en los resultados, como dato relevante en ese país sobre la responsabilidad de las instituciones en garantizar un adecuado CA.

Smirnova et al. (2017) demostraron una significativa asociación entre CA (medido por D-RECT) y resultados perinatales adversos en centros de atención obstétrica en Holanda.

Otros estudios relacionaron el CA con otras variables, aunque medido con otros instrumentos: de Oliveira Filho & Vieira (2007) no pudieron demostrar una asociación significativa entre el desempeño académico, calidad de vida y CA (medido por el DREEM) en 63 residentes de Anestesiología. En nuestro país, Llera & Durante (2014) demostraron una correlación significativa entre CA (medido con PHEEM) y Burn-out (medido con el Maslach Burnout Inventory) en 92 residentes.

El criterio de *Evaluación de las consecuencias del test,* también denominada *validez consecuencial*, se basa en la vinculación entre los resultados del test con las consecuencias actuales y potenciales de su aplicación (Messick, 1993). Es decir, decidir si los efectos esperados refuerzan la validez de la prueba, o por el contrario, sus efectos inesperados la afectan. Dicho de otro modo, comprobar si se lograron efectos beneficiosos y se evitaron efectos perjudiciales con su aplicación (Cook & Beckman, 2006). El punto no sólo es asunto de debate entre distintos autores sobre su utilización en la validación de los tests, sino que además las consecuencias de los tests se han empleado en pleitos y decisiones legales (e.g., sobre justa admisión en puestos de trabajo), lo que forma parte del aspecto de *justicia* en la confección de pruebas de evaluación. Los autores del D-RECT también mencionan la cuestión, remarcando el interés de correlacionar los resultados de esta prueba con el impacto educativo y sanitario de su aplicación (Boor et al., 2011).

Tanto en el trabajo original como en esta adaptación transcultural, no se cuenta con un modelo específico para este propósito, y la creación eventual del mismo excede los objetivos de este trabajo.

CONCLUSIONES

El presente trabajo cumplió todas las instancias formales para una adecuada adaptación transcultural del Test Holandés del Clima de Aprendizaje en las Residencias (D-RECT) para su utilización en nuestro medio, verificando la preservación de las propiedades psicométricas de dicha herramienta. Quedan pendientes nuevas investigaciones que eventualmente confirmen estos resultados, y otras sobre el impacto educativo y sanitario que resulte de su aplicación dentro de los programas de formación en distintas especialidades.

- **Limitaciones:**

Cabe aclarar que el D-RECT original fue aplicado en idioma holandés en el estudio original, aunque su publicación ulterior se encuentra en idioma inglés, a partir del cual se realizó la presente adaptación transcultural, y las demás existentes en la literatura en nuestro conocimiento. Más recientemente, el D-RECT ha sido revisado y validado en su mismo país de origen por algunos de sus autores originales con modificaciones en su estructura (Silkens et al, 2015).

Existen otras adaptaciones transculturales de este instrumento:

- ✓ En Arabia Saudita, Alshomrani & AlHadi (2017) probaron el instrumento en 78 residentes de Psiquatría, con texto en inglés (sin traducción al idioma nacional), reduciendo a 44 los ítems (por su diferente diseño curricular) y tomando el α de Cronbach como evidencia de consistencia interna, sin análisis factorial o de generalizabilidad, concluyendo en la necesidad de revisar los programas educativos debido a los bajos puntajes obtenidos (media de 2,76 ± 0,55).
- ✓ En Alemania, lblher et al. (2015) aplicaron el D-RECT en 255 residentes de 4 especialidades (Anestesia, Medicina Familiar, Pediatría y Medicina Interna) traducido al idioma alemán. Su análisis de la consistencia interna

(utilizando α de Cronbach, AFE y AFC) reveló insuficientes evidencias de validez para su aplicación en ese contexto cultural.

- ✓ En Filipinas, Pacifico et al (2018) probaron el D-RECT (en la versión modificada por Silkens et al (2015) conteniendo 35 ítems en 9 dominios), en idioma inglés (segunda lengua oficial del país) sobre 843 residentes de Medicina Interna. Utilizando α de Cronbach, Análisis Factorial Confirmatorio y de Generalizabilidad, los autores modificaron el instrumento de base, resultando en un cuestionario de 28 ítems / 9 dominios para obtener suficiente evidencia de validez en ese escenario.
- ✓ En Colombia, Dominguez, Silkens & Sanabria (2018) estudiaron la aplicación del D-RECT en la versión modificada por Silkens et al (2015) de 35 ítems/9 dominios, en idioma español sobre 220 residentes de 17 especialidades. Además de aportar evidencias sobre contenido, proceso de respuesta y estructura interna, el trabajo analizó la correlación con otro instrumento de medición del mismo constructo (PHEEM) validado en español (validez concurrente). El análisis incluyó α de Cronbach, Análisis Factorial Confirmatorio y correlación de Pearson. La conclusión de los autores estima suficiente validez para esa versión en español del D-RECT, y demuestra correlación con el PHEEM como aporte de mayor evidencia de validez en ese contexto.

En nuestro trabajo, la inclusión de residentes reclutados de un congreso nacional de una especialidad médica determinada podría representar otra limitación, aunque la pertenencia a cualquier rama de la profesión no es objetivo de estudio en particular en este trabajo.

Implicaciones educativas y sanitarias futuras:

Dada la relación comprobada del CA en distintos aspectos de los programas educacionales y de la profesión, cabe esperar un impacto en esos aspectos luego de eventuales cambios institucionales basados en la medición de este constructo.

Los organismos responsables de las residencias podrían verse beneficiados en la búsqueda de mayor calidad de sus programas de residencia, al contar con este recurso, y la oportunidad de mejorar los distintos componentes del entorno de aprendizaje de sus futuros especialistas. Al respecto, más estudios sobre correlación del mismo con otros parámetros educativos podría aportar también en este sentido.

Los cambios positivos en el CA podrían también incidir favorablemente sobre el bienestar de los residentes, la calidad asistencial y la incidencia de burn-out dentro de la profesión, aspectos que formarían parte de posteriores estudios.

En el futuro, el eventual desarrollo de otros instrumentos válidos para el mismo constructo y en el mismo contexto arrojaría nuevos datos sobre la importancia del CA en el ámbito específico de la educación médica, y sobre el impacto sobre la calidad de atención.

Asimismo, la utilidad de esta herramienta podría ser comparada con otras similares en el mismo escenario.

ANEXOS

Anexo 1- Aval de la Facultad de Medicina de la Universidad Nacional del Nordeste x

Anexo 2- Cuestionario D-RECT original xii

Anexo 3- Traducción al Español (traductor inocente) xiv

Anexo 4- Traducción al Español (traductor comprometido) xvii

Anexo 5- Versión unificada en Español por experto (base para las retrotraducciones) xx

Anexo 6- Retrotraducción al Inglés N° 1 xxiii

Anexo 7- Retrotraducción al Inglés N° 2 xxv

Anexo 8- Permiso de autor para la Adaptación Transcultural del D-RECT xxvii

Anexo 9- Formulario del D-RECT aplicado a los residentes xxviii

Anexo 10- Puntajes del test global y dominios (por Hospital) xxx

Anexo 11- AFC – Diagrama path- Correlaciones entre factores (DWLS) xxxi

Anexo 12- AFC – Diagrama path- Correlaciones entre factores (ULS) xxxii

Anexo 13- Tabla de Resultados de Estudio G – Modelo R/I.....xxxiii

Anexo 14- Tabla de Resultados de Estudio G- Modelo H/R:H x I xxxiv

Anexo 1- Aval de la Facultad de Medicina de la UNNE

Universidad Nacional del Nordeste
Facultad de Medicina

M.E.S.A.- FACULTAD DE MEDICINA
10-2013-06361

NOTA Nº 2317 -Ac.
CORRIENTES, 15 de agosto de 2013.

UNIVERSIDAD NACIONAL DEL NORDESTE
FACULTAD DE MEDICINA
21 AGO 2013
EXPTE. Nº 10-13-06361

Al señor Director Ejecutivo del
Hospital Escuela "Gral. José Francisco de San Martín"
Méd. RAÚL BERNARDO PUYOL
S U D E S P A C H O

Tengo el agrado de dirigirme a Ud., con el objeto de remitirle adjunto a la presente, copia de la Resolución Nº 2300/13-D.-, por la cual esta Facultad de Medicina avala la realización de una ***Investigación Educativa sobre Clima de Aprendizaje en las Residencias de Salud del NEA,*** *augurándole el mayor de los éxitos.*

Sin otro particular, hago propicia la oportunidad, para saludarlo con atenta y distinguida consideración.-

HOSPITAL ESCUELA - CORRIENTES - ENTRO SALIO 21/8/13

Prof. GERARDO OMAR LARROZA
Secretario Académico

mjch.

Universidad Nacional del Nordeste
Facultad de Medicina

RESOLUCIÓN Nº 2300/13-C.D.-
CORRIENTES, 15 de agosto de 2013.-

VISTO:

El Expediente Nº 10-2013-05996, por el cual se tramita el aval de esta Facultad de Medicina, para la realización de una Investigación Educativa sobre Clima de Aprendizaje en las Residencias de Salud del NEA; y

CONSIDERANDO:

Que la mencionada actividad permitirá obtener información valiosa acerca del ambiente formativo de las distintas disciplinas, que redundará en un impacto significativo sobre la calidad profesional ulterior;

Lo informado por el señor Secretario de Post Grado, Profesor Jorge Alberto Costa, a fs. 03;

La importancia que representan estas actividades para la formación y perfeccionamiento de Profesionales de la Salud;

Las atribuciones conferidas al suscripto;

EL VICE DECANO A CARGO DEL DECANATO DE LA FACULTAD DE MEDICINA DE LA UNIVERSIDAD NACIONAL DEL NORDESTE
R E S U E L V E :

ARTÍCULO 1º.- Avalar, la realización de una **Investigación Educativa sobre Clima de Aprendizaje en las Residencias de Salud del NEA**, ante la solicitud efectuada por el Director Ejecutivo del Hospital Escuela "Gral. José Francisco de San Martín", Méd. Raúl Bernardo Puyol y por las razones enunciadas precedentemente.

ARTÍCULO 2º.- Regístrese, comuníquese y archívese.-

Prof. GERARDO OMAR LARROZA
Secretario Académico

Prof. CARLOS ALBERTO MARKOWSKY
Vice Decano a/c. Decanato

ES COPIA

CORDELIA AUCHTER de SANTILLAN
Directora General
de Coordinación Administrativa
FACULTAD DE MEDICINA - U.N.N.E.

Anexo 2
Cuestionario D-RECT original (Boor, 2011)

Table 1. D-RECT questionnaire.

	Items	*N*	Mean	SD	α	Delphi
	Subscale: Supervision				0.64	
1	The guidelines clearly outline when to request input from a supervisor	600	3.77	1.06		A
2	The amount of supervision I receive is appropriate for my level of experience	600	4.09	0.91		A
3	It is clear which attending supervises me	600	4.26	0.95		A
	Subscale: Coaching and assessment				0.80	
4	I am asked on a regular basis to provide a rationale for my management decisions and actions	600	3.88	0.88		A
5	My attendings coach me on how to communicate with difficult patients	600	3.42	1.03		NA
6	My attendings take the initiative to explain their actions	600	3.46	0.98		NA
7	My attendings take the initiative to evaluate my performance	600	3.08	1.06		NA
8	My attendings take the initiative to evaluate difficult situations I have been involved in	600	3.01	1.04		A
9	My attendings evaluate whether my performance in patient care is commensurate with my level of training	600	3.40	1.09		A
10	My attendings occasionally observe me taking a history	600	2.85	1.22		A
11	My attendings assess not only my medical expertise but also other skills such as teamwork, organization or professional behavior	600	3.67	1.12		A
	Subscale: Feedback				0.75	
12	My attendings give regular feedback on my strengths and weaknesses	600	3.37	1.14		A
13	Observation forms (i.e., Mini-CEX) are used to structure feedback	600	3.81	1.19		NA
14	Observation forms (i.e., Mini-CEX) are used periodically to monitor my progress	600	3.42	1.27		NA
	Subscale: Teamwork				0.69	
15	Attendings, nursing staff, other allied health professionals and residents work together as a team	600	3.82	1.00		NA
16	Nursing staff and other allied health professionals make a positive contribution to my training	600	3.66	1.09		NA
17	Nursing staff and other allied health professionals are willing to reflect with me on the delivery of patient care	600	3.76	1.02		NA
18	Teamwork is an integral part of my training	600	3.00	1.04		A
	Subscale: Peer collaboration				0.76	
19	Residents work well together	600	4.48	0.79		A
20	Residents, as a group, make sure the day's work gets done	600	4.14	1.04		NA
21	Within our group of residents it is easy to find someone to cover or exchange a call	600	4.26	0.91		NA
	Subscale: Professional relations between attendings				0.77	
22	Continuity of care is not affected by differences of opinion between attendings	600	3.76	1.12		A
23	Differences of opinion between attendings about patient management are discussed in such a manner that is instructive to others present	600	3.50	1.14		A
24	Differences of opinion are not such that they have a negative impact on the work climate	600	3.62	1.16		A
	Subscale: Work is adapted to residents' competence				0.66	
25	The work I am doing is commensurate with my level of experience	600	4.12	0.84		A
26	The work I am doing suits my learning objectives at this stage of my training	600	3.97	0.92		A
27	It is possible to do follow up with patients	600	3.96	1.08		A
28	There is enough time in the schedule for me to learn new skills	600	3.24	1.08		A
	Subscale: Attendings' role				0.85	
29	My attendings take time to explain things when asked for advice	600	4.12	0.80		A
30	My attendings are happy to discuss patient care	600	4.16	0.82		A
31	There is (are) NO attending physician(s) who have a negative impact on the educational climate	600	3.34	1.27		NA
32	My attendings treat me as an individual	600	3.47	1.18		NA
33	My attendings treat me with respect	600	4.46	0.87		A
34	My attendings are all in their own way positive role models	600	3.48	1.04		NA
35	When I need a attending, I can always contact one	600	4.55	0.69		A
36	When I need to consult a attending, they are readily available	600	4.52	0.72		A
	Subscale: Formal education				0.75	
37	Residents are generally able to attend scheduled educational activities	600	3.60	1.06		A
38	Educational activities take place as scheduled	600	3.89	0.98		A
39	Attendings contribute actively to the delivery of high-quality formal education	600	3.63	1.14		A
40	Formal education and training activities are appropriate to my needs	600	3.61	1.06		NA
	Subscale: Role of the specialty tutor	600			0.78	
41	The specialty tutor monitors the progress of my training	600	4.49	0.84		A
42	The specialty tutor provides guidance to other attendings when needed	600	3.46	1.26		A
43	The specialty tutor is actively involved in improving the quality of education and training	600	4.12	1.06		A
44	In this rotation evaluations are useful discussions about my performance	600	3.65	1.11		A
45	My plans for the future are part of the discussion	600	3.61	1.16		NA

	Items	*N*	Mean	SD	α	Delphi
46	During evaluations, input from several attendings is considered	600	3.83	1.17		A
	Subscale: Patient sign out				0.75	
47	When there is criticism of a management plan I have developed in consultation with my attending physician, I know the attending physician will back me up	600	3.74	1.13		A
48	Sign out takes place in a safe climate	600	4.02	1.04		A
49	Sign out is used as a teaching opportunity	600	3.96	1.06		A
50	Attendings encourage residents to join in the discussion during sign out	600	3.69	1.15		A

Anexo 3
Traducción al Español (traductor inocente)

Tabla 1- D-RECT. Cuestionario

Sub-escala: Supervisión

1- Las directrices establecen claramente el momento en el cual solicitar la intervención de un supervisor.
2- El asesoramiento que recibo es el adecuado para mi nivel de formación
3- Está claro quién me supervisa.

Sub-escala: Valoración de desempeño

4- Me solicitan fundamentar mis acciones y decisiones
5- Los profesionales a cargo me asesoran sobre el modo de comunicarme con los pacientes difíciles.
6- Mis tutores explican su accionar.
7- Mis tutores evalúan mi desempeño.
8- Mis tutores evalúan mi intervención en situaciones complejas.
9- Mis tutores evalúan si mi procedimiento con el paciente hospitalizado condice con mi nivel de entrenamiento.
10- Mis tutores me observan la elaboración de la historia clínica.
11- Mis tutores evalúan mis competencias y otras habilidades médicas como el trabajo en equipo, la organización o el comportamiento profesional.

Sub-escala: Retroalimentación

12- Los tutores me retroalimentan en mis puntos débiles y fuertes.
13- Se utilizan formularios de observación estructurada de la práctica profesional (Mini-CEX) para la posterior provisión de feed-back.
14- Se utilizan formularios de observación de la práctica clínica (Mini-CEX) para monitorear mi progreso.

Sub-escala: Trabajo en equipo

15- Los tutores, el personal de enfermería, otros profesionales de la salud y los residentes trabajan en equipo.
16- El personal de enfermería y otros profesionales de la salud contribuyen positivamente a mi formación.
17- El personal de enfermería y otros profesionales de la salud están dispuestos a reflexionar conmigo sobre la atención y cuidado al paciente.
18- El trabajo en equipo es una parte integral de mi formación.

Sub-escala: Colaboración entre pares

19- Los residentes trabajan bien juntos.
20- Los residentes, como grupo, se aseguran de que el trabajo del día se realice.
21- Dentro de nuestro grupo de residentes, es fácil encontrar a alguien para cubrir o intercambiar un turno.

Sub-escala: Relaciones profesionales entre los tutores

22- La continuidad asistencial no se ve afectada por las diferencias de opinión entre tutores.
23- El intercambio de opiniones entre los tutores sobre el manejo del paciente constituye un rol instructivo.
24- Las opiniones de los tutores no tienen un impacto negativo en el clima de trabajo.

Sub-escala: La actividad está adecuada a las competencias de los residentes

25- Las actividades asignadas son proporcionales a mi nivel de experiencia.
26- Las actividades que realizo condicen con mis objetivos de aprendizaje en este nivel de entrenamiento.
27- Es posible hacer un seguimiento de los pacientes.
28- El periodo de entrenamiento programado me permite aprender nuevas habilidades.

Sub-escala: El rol de los tutores

29- Los tutores se dan tiempo para responder a los interrogantes que surgen.
30- Los tutores están dispuestos a intercambiar opiniones sobre la atención del paciente.
31- Todos los tutores contribuyen positivamente a la formación del residente.
32- Los tutores me tratan como a su par.
33- Los tutores me tratan con respeto.
34- Los tutores muestran un modelo de conducta a seguir.
35- Los tutores están a nuestra disposición.
36- Siempre puedo contactar a un tutor ante un interrogante.

Sub-escala: Educación formal

37- Los residentes, en general, pueden asistir a las actividades educativas programadas.
38- Las actividades educativas se llevan a cabo según lo programado.
39- Los tutores contribuyen activamente con la calidad de la educación formal.

40- La educación formal y actividades de entrenamiento son adecuadas para mis necesidades.

Sub-escala: rol del tutor de la especialidad

41- El tutor especialista supervisa mi progreso de entrenamiento.
42- El tutor especialista orienta a otros tutores cuando surge la necesidad.
43- El tutor de la especialidad es un participante activo en la mejora de la calidad de educación y formación.
44- En la rotación, las evaluaciones son intercambios útiles de mi desempeño.
45- Mis planes para el futuro son parte del intercambio.
46- Durante la evaluación, se consideran el aporte de otros tutores.

Sub-escala: Registro del alta del paciente

47- Cuando existe una crítica del plan de gestión desarrollado en conjunto con mi tutor, sé que mi tutor me apoyará.
48- El alta de los pacientes se lleva a cabo en un ambiente seguro.
49- El alta de los pacientes se utiliza como una oportunidad de enseñanza.
50- Los tutores alientan a los residentes a participar de los intercambios durante el alta de los pacientes.

Anexo 4
Traducción al Español (traductor comprometido)

Instrumento de medición del Clima de Aprendizaje de los Residentes de Holanda D-RECT

Tabla 1- Cuestionario D-RECT

Sub-escala: Supervisión

1- Los lineamientos establecen claramente cuándo pedir información a un supervisor.
2- El tipo de supervisión que recibo es apropiado para mi nivel de experiencia
3- Está claro quién me supervisa.

Sub-escala: Asesoramiento y tutoría

4- Regularmente me piden que dé razones para justificar mis decisiones y mi actuar.
5- Los tutores me asesoran en cómo comunicarme con pacientes difíciles.
6- Los tutores toman la iniciativa para explicar su actuar.
7- Los tutores toman la iniciativa para evaluar mi desempeño.
8- Los tutores toman la iniciativa para evaluar las situaciones difíciles en las que estuve involucrado/a.
9- Los tutores evalúan si mi desempeño en el paciente hospitalizado está acorde con mi nivel de entrenamiento.
10- De vez en cuando mis tutores me observan realizar una historia clínica.
11- Los tutores evalúan mis competencias y otras habilidades médicas tales como el trabajo en equipo, la organización o el comportamiento profesional.

Sub-escala: Retroalimentación

12- Los tutores me brindan retroalimentación sobre mis debilidades y mis fortalezas.
13- Se utilizan los formularios de observación estructurada de la práctica profesional (Mini-CEX) para brindar retroalimentación.
14- Se utilizan los formularios de observación (Mini-CEX) para monitorear mi progreso.

Sub-escala: Trabajo en equipo

15- Los tutores, el personal de enfermería, otros profesionales de la salud y los residentes trabajan juntos como un equipo.
16- El personal de enfermería y otros profesionales de la salud hacen una contribución positiva para mi entrenamiento.
17- El personal de enfermería y otros profesionales de la salud están dispuestos a reflexionar conmigo sobre el cuidado del paciente.
18- El trabajo en equipo es una parte integral de mi entrenamiento.

Sub-escala: Colaboración entre pares

19- Los residentes trabajan bien juntos.
20- Los residentes, como grupo, se aseguran de que el trabajo del día se lleve a cabo.
21- Dentro de nuestro grupo de residentes, es fácil encontrar a alguien para cubrir o cambiar un turno.

Sub-escala: Relaciones profesionales entre los tutores

22- La continuidad asistencial no se ve afectada por las diferencias de opinión entre los tutores.
23- Las diferencias de opinión entre los tutores sobre el manejo del paciente es una instancia de instrucción para los presentes.
24- Las opiniones de los tutores no son tan marcadas como para crear un impacto negativo en el clima de trabajo.

Sub-escala: La actividad está adecuada a las competencias de los residentes

25- Las actividades asignadas son proporcional a mi nivel de experiencia.
26- Las actividades asignadas se adecúa a mis objetivos de aprendizaje en este estadio de entrenamiento.
27- Es posible hacer un seguimiento a los pacientes.
28- El periodo de entrenamiento es suficiente para aprender nuevas habilidades.

Sub-escala: El rol de los tutores

29- Los tutores se dan tiempo para responder cada vez que se les pide consejo.
30- Los tutores están dispuestos a intercambiar opiniones sobre el cuidado del paciente.
31- No hay médicos tutores que tengan un impacto negativo en el clima educativo.
32- Los tutores me tratan como a un par.
33- Los tutores me tratan con respeto.
34- Los tutores, cada uno a su manera, son modelos positivos a seguir.
35- Cuando necesito a un tutor siempre puedo contactarme con alguno de ellos.
36- Cuando necesito consultar a un tutor, ellos están siempre disponibles.

Sub-escala: Educación formal

37- Los residentes, en general, pueden asistir a las actividades educativas programadas.
38- Las actividades educativas se llevan a cabo según fueron programadas.
39- Los tutores contribuyen activamente a brindar una educación formal de alta calidad.
40- La educación formal y actividades de entrenamiento son apropiadas para mi formación.

Sub-escala: rol del tutor especialista.

41- El tutor especialista monitorea el progreso de mi entrenamiento.
42- El tutor especialista guía a los otros tutores cuando lo necesitan.
43- El tutor especialista está involucrado en mejorar la calidad de la educación y del entrenamiento.
44- En la rotación, las evaluaciones son intercambios útiles sobre mi desempeño.
45- Mis planes sobre el futuro son parte de estos intercambios.
46- Durante la evaluación, se consideran las opiniones de los otros tutores.

Sub-escala: Registro del alta del paciente

47- Cuando hay alguna crítica sobre un plan de gestión desarrollado bajo la supervisión de mi tutor, sé que mi tutor me apoyará.
48- El alta de los pacientes tiene lugar en un clima seguro.
49- El alta de los pacientes se utiliza como una herramienta de enseñanza.
50- Los tutores alientan a los residentes a participar en los intercambios durante el alta de los pacientes.

Anexo 5
Versión unificada por grupo de expertos del D-RECT en Español
(base para las retrotraducciones)

D-RECT- Dutch Residency Educational Climate Test

Sub-escala *Supervisión*

1- Las normas establecen claramente el momento en el cual solicitar intervención de instructor.
2- El grado de supervisión que recibo de mi instructor es apropiado para mi nivel de experiencia.
3- Está claro quién me supervisa.

Sub-escala *Valoración del desempeño*

4- Regularmente, me piden que fundamente mis decisiones y mis acciones.
5- Mis instructores me asesoran sobre el modo de comunicarme con pacientes difíciles.
6- Mis instructores explican espontáneamente su accionar.
7- Mis instructores espontáneamente evalúan mi desempeño.
8- Mis instructores espontáneamente evalúan las situaciones difíciles en las que estuve involucrado.
9- Mis instructores evalúan si mi desempeño con el paciente hospitalizado es acorde con mi nivel de entrenamiento.
10- Ocasionalmente, mis instructores me observan al realizar una historia clínica.
11- Mis instructores evalúan no solamente mi experticia médica, sino también otras habilidades como trabajo en equipo, organización o conducta profesional.

Sub-escala *Retroalimentación*

12- Mis instructores me brindan regularmente retroalimentación sobre mis debilidades y fortalezas.
13- Se utilizan herramientas de observación de la práctica profesional (por ejemplo Mini-CEX) para brindar retroalimentación estructurada.
14- Se utilizan herramientas de observación (p. ej. Mini-CEX) para monitorear mi progreso.

Sub-escala *Trabajo en Equipo*

15- Los instructores, el personal de enfermería, otros profesionales de la salud y los residentes, trabajan en equipo.
16- El personal de enfermería y otros profesionales de la salud hacen una contribución positiva para mi formación.
17- El personal de enfermería y otros profesionales de la salud están dispuestos a reflexionar conmigo sobre el cuidado de los pacientes.
18- El trabajo en equipo es una parte integral de mi formación.

Sub-escala *Colaboración entre pares*

19- Los residentes trabajan bien en forma conjunta.
20- Los residentes como grupo, se aseguran que el trabajo del día se lleve a cabo.

21- Dentro de nuestro grupo de residentes, es fácil encontrar a alguien para reemplazar una tarea.

Sub-escala *Relaciones profesionales entre instructores*

22- La continuidad asistencial no se ve afectada por las diferencias de opiniones entre los instructores.
23- Las diferencias de opiniones entre los instructores acerca del manejo del paciente son discutidas de tal manera que son educativas para los demás presentes.
24- Las diferencias de opiniones no producen un impacto negativo en el clima de trabajo.

Sub-escala *El trabajo está adaptado a la competencia del residente*

25- El trabajo que yo hago es proporcional a mi nivel de experiencia.
26- El trabajo que yo hago condice con los objetivos de aprendizaje en este nivel de formación.
27- Es posible hacer seguimiento de los pacientes.
28- Hay tiempo suficiente programado para que yo aprenda nuevas habilidades.

Sub-escala *El rol de los instructores*

29- Los instructores se dan tiempo para responder cada vez que se les pide consejo.
30- Los instructores están dispuestos a intercambiar opiniones sobre el cuidado del paciente.
31- No hay instructores que tengan un impacto negativo en el clima educativo.
32- Mis instructores me tratan como persona individual.
33- Los instructores me tratan con respeto.
34- Los instructores, cada uno a su manera, son modelos positivos a seguir.
35- Cuando necesito a un instructor, siempre puedo contactarme con alguno de ellos.
36- Cuando necesito consultar a un instructor, ellos están siempre disponibles.

Sub-escala *Educación Formal*

37- Los residentes, en general, pueden asistir a las actividades educativas programadas.
38- Las actividades educativas se llevan a cabo según lo programado.
39- Los instructores contribuyen activamente a brindar una educación formal de alta calidad.
40- La educación formal y las actividades de entrenamiento son adecuadas para mis necesidades.

Sub-escala *El rol del instructor especialista*

41- El instructor especialista monitorea el progreso de mi entrenamiento.
42- El instructor especialista guía a los otros instructores cuando lo necesitan.
43- El instructor especialista está involucrado en mejorar la calidad de la educación y del entrenamiento.
44- En la rotación, las evaluaciones constituyen discusiones útiles sobre mi desempeño.
45- Mis planes sobre el futuro son parte de la discusión.
46- Durante las evaluaciones, se consideran las opiniones de los otros instructores.

Sub-escala *Registro del alta del paciente*

47- Cuando hay una crítica sobre el manejo que he desarrollado en conjunto con mi instructor, sé que mi instructor me apoyará.
48- El alta de los pacientes tiene lugar en un clima de seguridad.
49- El alta de los pacientes se utiliza como una oportunidad de enseñanza.
50- Los instructores alientan a los residentes a participar en la discusión durante el alta de los pacientes.

Anexo 6
Retrotraducción al Inglés Nº 1

D-RECT

DUTCH RESIDENCY EDUCATIONAL CLIMATE TEST

Subscale *Supervision*

1. Standards clearly state the moment a supervisor intervention is asked for.
2. Supervision rate I get from the coach is appropriate for my experience level.
3. It is obvious who I am supervised from.

Subscale *Performance Assessment.*

4. I am usually asked to found my decisions and my actions.
5. My coaches point me out the way to contact with unruly patients.
6. My coaches explain me their actions spontaneously.
7. My coaches assess my performance spontaneously.
8. My coaches assess difficult situations where I was involved in a spontaneously way.
9. My coaches assess if my performance with patient care agree with my training level.
10. Occasionally my coaches watch me out while making a clinical history.
11. My coaches assess not only my clinical experience but other abilities such as team work, organization or professional behavior.

Subscale *Feedback*

12. My coaches usually provide me with feedback about weaknesses and strengths.
13. Observation tools in professional practice are used (i.e. Mini CEX) to provide a structured feedback.
14. Observation tools (for instance Mini CEX) are usually used for monitoring my progress.

Subscale *Teamwork*

15. Coaches, nurses, health workers and residents work altogether as a team.
16. Nursing staff and other health professionals make a positive contribution to my training.
17. Nursing staff and other health professionals are well disposed to think over patient care with me.
18. Teamwork is an integral part in my training.

Subscale *Peer collaboration*

19. Residents work well altogether.
20. Residents as a group, make sure to do the daily work.
21. Within our group of residents, it is easy to find someone to replace for a task.

Subscale *Professional relationships among Coaches*.

22. Care continuity is not affected by difference in opinions of coaches.
23. Differences in opinions of coaches about patient handling are discussed so that they become educative for others.
24. Differences in opinions of coaches do not produce a negative impact in work climate.

Subscale *Work is adapted to residents' abilities.*

25. The work I do is in proportion with my experience level.

26. The work I do agree with learning aims in this formative level.
27. It is possible to monitoring patients.
28. There is plenty of planned time so that I can learn new abilities.

Subscale *Role of Coaches*

29. Coaches take time to answer every time they are asked for.
30. Coaches are disposed to make interchange of opinions about patients care.
31. There are no coaches who have a negative impact in educative climate.
32. My coaches treat me as an individual person.
33. Coaches treat me with respect.
34. Coaches, each of them in their way, are positive models to follow to.
35. Every time I need a coach I can contact with one of them.
36. Whenever I need a coach, they are always disposed to.

Subscale *Formal education*

37. Residents generally attend planned educative activities.
38. Educative activities are held according planned schedules.
39. Coaches contribute actively to provide a high quality formal education.
40. Formal education and training activities are adequate for my needs.

Subscale *Role of Specialist Coach*

41. Specialist coach monitors progress in my training.
42. Specialist coach guides other coaches when they need to.
43. Specialist coach is involved in getting better education and training quality.
44. In turns, evaluations become useful discussions about my performance.
45. My future plans are part of discussions.
46. Opinions of other coaches are considered during evaluations.

Subscale *Patient sign out.*

47. When a criticism about my management plan that I have developed together with my coach exists, I know he will support me.
48. Patient sign out takes place in a secure climate.
49. Patient sign out is used as a teaching opportunity.
50. Coaches encourage residents to take part of discussion during patient sign out.

Anexo 7
Retrotraducción al Inglés Nº 2

D-RECT

DUTCH RESIDENCY EDUCATIONAL CLIMATE TEST

Subscale *Supervision*

1. The work scheme clearly state the moment a supervisor intervention is asked for.
2. Supervision amount I get from the coach is appropriate for my experience level.
3. It is evident which coach I am supervised from.

Subscale *Performance Assessment.*

4. Periodically I am asked to found my decisions and my actions.
5. My coaches explain me the way to communicate with difficult patients.
6. My coaches explain me their actions spontaneously.
7. My coaches assess my performance spontaneously.
8. In a spontaneous way, my coaches assess difficult situations where I was involved.
9. My coaches assess if my performance with patient care agree with my training level.
10. Occasionally, I am observed by my coaches when I make a clinical history.
11. My coaches assess not only my clinical experience but other abilities such as team work, organization or professional behavior.

Subscale *Feedback*

12. My coaches usually provide me with feedback about my own weaknesses and strengths.
13. Observation tools in professional practice are used (i.e. Mini CEX) to provide a structured feedback.
14. Observation tools (for instance Mini CEX) are used for monitoring my progress on a regular basis.

Subscale *Teamwork*

15. Coaches, nurses, health workers and residents work altogether as a team.
16. Nursing staff and other health professionals make a positive contribution to my training.
17. Nursing staff and other health professionals are willing to reflect about patient care with me.
18. Teamwork is an integral part in my training.

Subscale *Peer collaboration*

19. Residents work well altogether.
20. Residents as a group, make sure to do the daily work.
21. Within our group of residents, it is easy to find somebody to replace for a task.

Subscale *Professional relationships among Coaches*.

22. Care continuity is not affected by differences of opinions between coaches.
23. Differences of opinions between coaches about patient handling are discussed so that they become educative for others.
24. Differences of opinions between coaches do not produce a negative impact in work climate.

Subscale *Work is adapted to residents' skills.*

25. The work I do agree with my experience level.
26. The work I do agree with learning aims in this formative level.
27. It is possible to monitoring patients.
28. There is enough time in the work scheme so that I can learn new skills.

Subscale *Role of Coaches*

29. Coaches spend enough time to explain things when they are asked for advice.
30. Coaches are willing to discuss about patients care.
31. There are no coaches who have a negative impact in educational climate.
32. My coaches treat me as an individual person.
33. Coaches treat me with respect.
34. Coaches, each of them in their way, are positive role models.
35. Every time I need a coach I can contact with one of them.
36. Whenever I need a coach, they are always available.

Subscale *Formal education*

37. Residents generally can attend planned educative activities.
38. Educative activities are met as planned.
39. Coaches contribute actively to provide a high-quality formal education.
40. Formal education and training activities are adequate for my needs.

Subscale *Role of Specialty Coach*

41. Specialty coach monitors progress in my training.
42. Specialty coach guides other coaches when they need to.
43. Specialty coach is involved in getting better educational and training quality.
44. In this rotation, evaluations become useful discussions about my performance.
45. My future plans are part of discussions.
46. Opinions of other coaches are considered during evaluations.

Subscale *Patient sign out.*

47. Whenever a criticism about the management plan that I have developed together with my coach exists, I know he will support me.
48. Patient sign out takes place in a secure climate.
49. Patient sign out is used as a teaching opportunity.
50. Coaches encourage residents to join in the discussion during patient sign out.

Anexo 8
Permiso de autor para la Adaptación transcultural del D-RECT

Raul Puyol 16 hours ago

Good evening.
My name is Raúl Puyol. I am an Argentinian physician, and I'm working in a postgraduate thesis about Transcultural Adaptation of the questionnaire D-RECT.
The objective is to get a valid and reliable instrument to be applied in my country (Argentina) and possibly in Latin America.
I ask for your permission to complete this task, in order to get my master's degree in Medical Education in the Instituto Universitario Hospital Italiano de Buenos Aires (Argentina).
If you agree, this will be a valuable contribution to medical education, in addition to being particularly important to me.
Any comment or opinion about this work will be very appreciated.
Thank you very much for your attention, and your time.

Waiting for your answer
Gratefully

Dr. Raúl Puyol

Cees Van der Vleuten to you 11 hours ago

Feel free to use the questionnaire. Good luck with your Master thesis.

Cees

Raul Puyol 6 minutes ago

Thank you very much.
Raúl

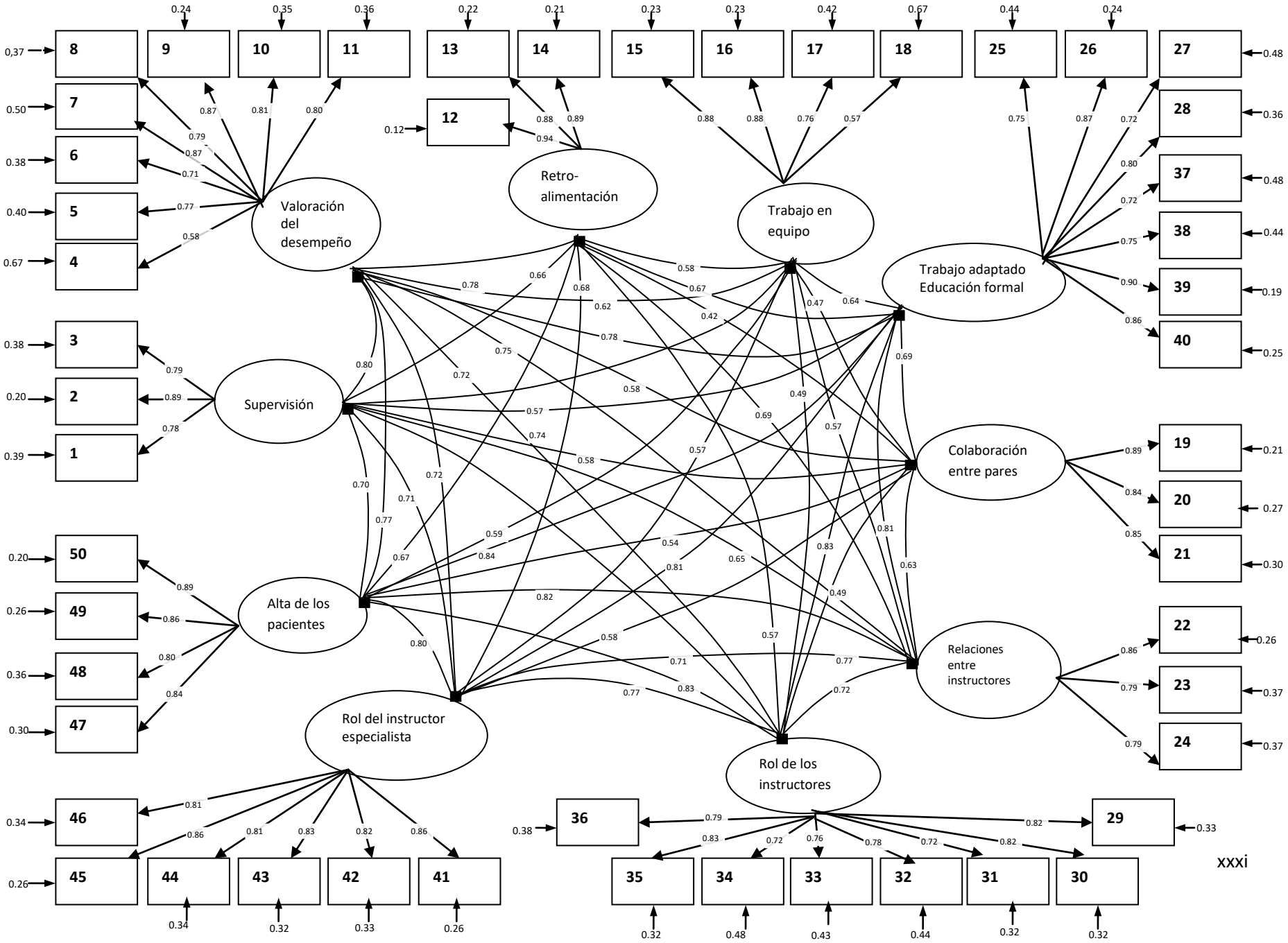
Anexo 9- Formulario del D-RECT aplicado a los residentes

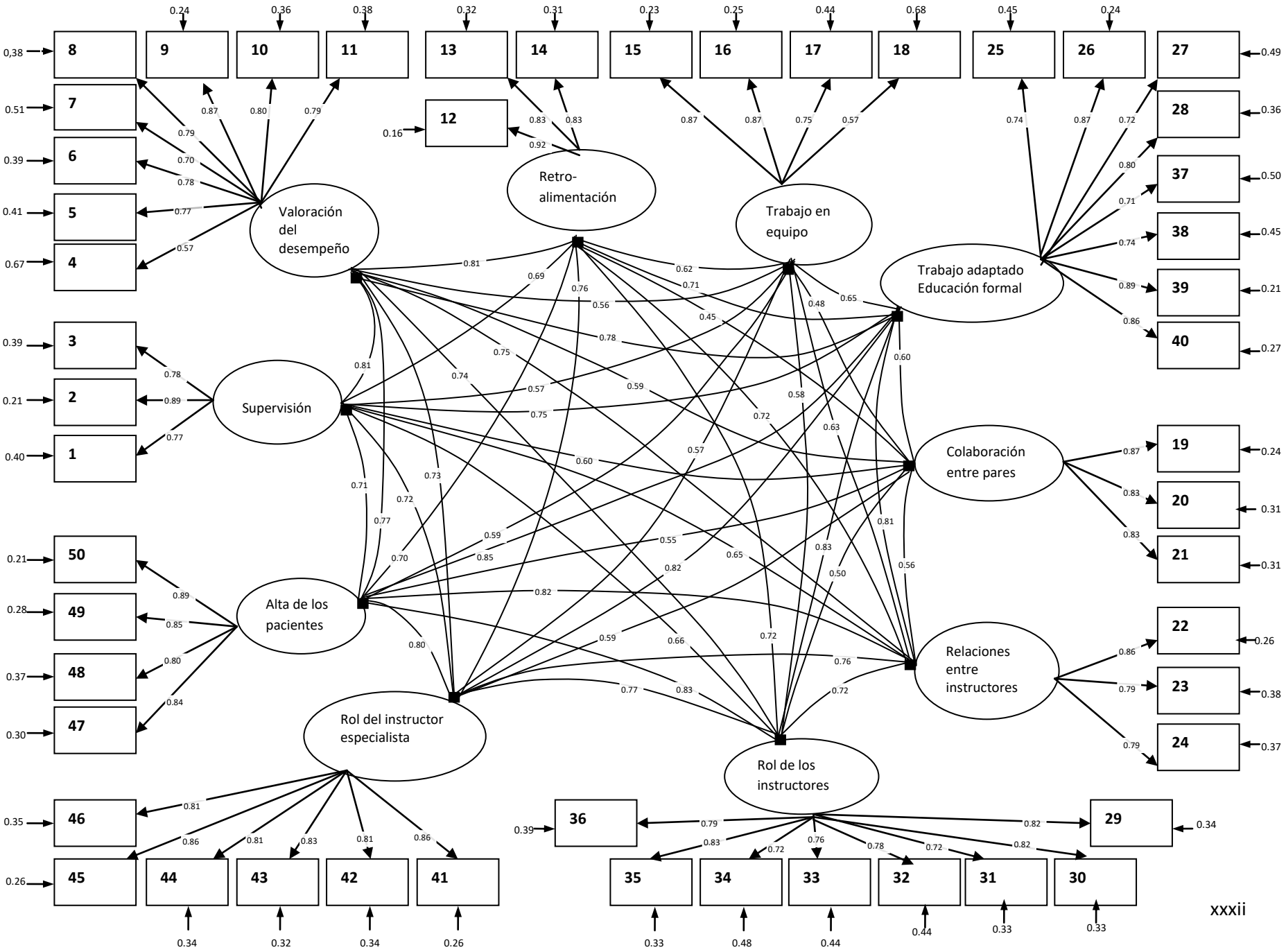
CLIMA DE APRENDIZAJE EN LAS RESIDENCIAS	Muy de acuerdo	Bastante de acuerdo	Neutro	Bastante en desacuerdo	Muy en desacuerdo
Sub-escala *SUPERVISION*					
1- Las normas establecen claramente el momento en el cual solicitar intervención de instructor.					
2- El grado de supervisión que recibo de mi instructor es apropiado para mi nivel de experiencia.					
3- Está claro quién me supervisa.					
Sub-escala *VALORACION DEL DESEMPEÑO*					
4- Regularmente, me piden que fundamente mis decisiones y mis acciones.					
5- Mis instructores me asesoran sobre el modo de comunicarme con pacientes difíciles.					
6-Mis instructores explican espontáneamente su accionar.					
7- Mis instructores espontáneamente evalúan mi desempeño.					
8- Mis instructores espontáneamente evalúan las situaciones difíciles en las que estuve involucrado.					
9- Mis instructores evalúan si mi desempeño con el paciente hospitalizado es acorde con mi nivel de entrenamiento.					
10- Ocasionalmente, mis instructores me observan al realizar una historia clínica.					
11- Mis instructores evalúan no solamente mi experticia médica, sino también otras habilidades como trabajo en equipo, organización o conducta profesional.					
Sub-escala *RETROALIMENTACION*					
12- Mis instructores me brindan regularmente retroalimentación sobre mis debilidades y fortalezas.					
13- Se utilizan herramientas de observación de la práctica profesional (por ejemplo Mini-CEX) para brindar retroalimentación estructurada.					
14- Se utilizan herramientas de observación (por ejemplo Mini-CEX) para monitorear mi progreso.					
Sub-escala *TRABAJO EN EQUIPO*					
15- Los instructores, el personal de enfermería, otros profesionales de la salud y los residentes, trabajan en equipo.					
16- El personal de enfermería y otros profesionales de la salud hacen una contribución positiva para mi formación.					
17- El personal de enfermería y otros profesionales de la salud están dispuestos a reflexionar conmigo sobre el cuidado de los pacientes.					
18- El trabajo en equipo es una parte integral de mi formación.					
Sub-escala *COLABORACION ENTRE PARES*					
19- Los residentes trabajan bien en forma conjunta.					
20- Los residentes como grupo, se aseguran que el trabajo del día se lleve a cabo.					
21- Dentro de nuestro grupo de residentes, es fácil encontrar a alguien para reemplazar una tarea.					
Sub-escala *RELACIONES PROFESIONALES ENTRE INSTRUCTORES*					
22- La continuidad asistencial no se ve afectada por las diferencias de opiniones entre los instructores.					
23- Las diferencias de opiniones entre los instructores acerca del manejo del paciente son discutidas de tal manera que son educativas para los demás presentes.					
24- Las diferencias de opiniones no producen un impacto negativo en el clima de trabajo.					

Sub-escala *EL TRABAJO ESTÁ ADAPTADO A LA COMPETENCIA DEL RESIDENTE*	Muy de acuerdo	Bastante de acuerdo	Neutro	Bastante en desacuerdo	Muy en desacuerdo
25- El trabajo que yo hago es proporcional a mi nivel de experiencia.					
26- El trabajo que yo hago condice con los objetivos de aprendizaje en este nivel de formación.					
27- Es posible hacer seguimiento de los pacientes.					
28- Hay tiempo suficiente programado para que yo aprenda nuevas habilidades.					
Sub-escala *EL ROL DE LOS INSTRUCTORES*					
29- Los instructores se dan tiempo para responder cada vez que se les pide consejo.					
30- Los instructores están dispuestos a intercambiar opiniones sobre el cuidado del paciente.					
31- No hay instructores que tengan un impacto negativo en el clima educativo.					
32- Mis instructores me tratan como persona individual.					
33- Los instructores me tratan con respeto.					
34- Los instructores, cada uno a su manera, son modelos positivos a seguir.					
35- Cuando necesito a un instructor, siempre puedo contactarme con alguno de ellos.					
36- Cuando necesito consultar a un instructor, ellos están siempre disponibles.					
Sub-escala *EDUCACION FORMAL*					
37- Los residentes, en general, pueden asistir a las actividades educativas programadas.					
38- Las actividades educativas se llevan a cabo según lo programado.					
39- Los instructores contribuyen activamente a brindar una educación formal de alta calidad.					
40- La educación formal y las actividades de entrenamiento son adecuadas para mis necesidades.					
Sub-escala *EL ROL DEL INSTRUCTOR ESPECIALISTA*					
41- El instructor especialista monitorea el progreso de mi entrenamiento.					
42- El instructor especialista guía a los otros instructores cuando lo necesitan.					
43- El instructor especialista está involucrado en mejorar la calidad de la educación y del entrenamiento.					
44- En la rotación, las evaluaciones constituyen discusiones útiles sobre mi desempeño.					
45- Mis planes sobre el futuro son parte de la discusión.					
46- Durante las evaluaciones, se consideran las opiniones de los otros instructores.					
Sub-escala *REGISTRO DEL ALTA DEL PACIENTE*					
47- Cuando hay una crítica sobre el manejo que he desarrollado en conjunto con mi instructor, sé que mi instructor me apoyará.					
48- El alta de los pacientes tiene lugar en un clima de seguridad.					
49- El alta de los pacientes se utiliza como una oportunidad de enseñanza.					
50- Los instructores alientan a los residentes a participar en la discusión durante el alta de los pacientes.					

Anexo 10
Puntajes promedio por Dominio y Global
(por Hospital)

Hospital	Supervisión	Valoración del desempeño	Retroali-mentación	Trabajo en equipo	Colaboración entre pares	Relaciones entre instructores	T. adaptado + Ed. formal	Rol de los instructores	Rol del instructor especialista	Registro del alta de pacientes	Global
J.R. Vidal	3,58	3,62	3,21	3,69	4,01	3,47	3,62	3,89	3,53	3,88	3,67
Llano	3,76	3,31	2,73	3,54	4,09	3,30	3,56	4,10	3,25	3,78	3,56
San Francisco de Asís	3,23	3,41	2,69	3,92	3,56	2,82	2,76	3,42	2,98	3,11	3,19
Juan Pablo II	3,44	3,65	2,54	3,70	4,33	3,34	3,64	3,67	2,97	3,33	3,49
Gral. San Martín	3,71	3,67	3,12	3,71	4,19	3,51	3,77	3,86	3,68	3,83	3,73
J.C. Perrando	4,15	4,05	3,41	3,94	4,42	3,83	3,95	4,13	3,78	4,14	3,99
Castelán	4,19	3,93	3,64	3,72	4,45	3,19	4,02	4,27	3,50	3,86	3,91


Valoración del desempeño
Retro-alimentación
Trabajo en equipo
Trabajo adaptado Educación formal
Supervisión
Colaboración entre pares
Alta de los pacientes
Relaciones entre instructores
Rol del instructor especialista
Rol de los instructores
8
9
10
11
13
14
15
16
17
18
25
26
27
7
12
28
6
37
5
38
4
39
3
40
2
19
1
20
50
21
49
22
48
23
47
24
46
36
29
45
44
43
42
41
35
34
33
32
31
30


Valoración del desempeño
Retro-alimentación
Trabajo en equipo
Trabajo adaptado Educación formal
Supervisión
Colaboración entre pares
Alta de los pacientes
Relaciones entre instructores
Rol del instructor especialista
Rol de los instructores
8
9
10
11
13
14
15
16
17
18
25
26
27
7
12
28
6
37
5
38
4
39
3
40
2
19
1
20
50
21
49
22
48
23
47
24
46
36
29
45
44
43
42
41
35
34
33
32
31
30

ANEXO 13

Tabla de Resultados de Estudio G- Modelo R/I

```
Effect            df            T           SS           MS           VC
-----------------------------------------------------------------------------
R                391   9216.12812   9216.12811     23.57066      0.45839
I                 49   1150.80720   1150.80719     23.48586      0.05825
RI             19159  22846.22810  12479.29279      0.65135      0.65135
-----------------------------------------------------------------------------
Mean                       0.00001
-----------------------------------------------------------------------------
Total          19599                22846.22808
-----------------------------------------------------------------------------
        Grand Mean:        0.00002

Date and time at beginning of Run 1:  Wed Nov 22 20:49:54 2017
Processor time for run: 0 seconds

The calculated grand mean = 3,8713
This value has been subtracted from the actual scores for the calculations.
While this improves the apparent accuracy, it does not affect the calculated variances.

Facets with adjusted sample sizes according to Brennan
'R'     Differentiation     392,00
'I'              Random      50,00
-------------------------------------------------------------------------------------

Pattern   Var. Comp.    Levels      Signature      Rule
-------------------------------------------------------------------------------------
R              0,458     1.00                d     tau only
I              0,001     50,00               r     Delta only
R|I            0,013     50,00               dr    Delta and delta

RESULTS:

s2(T)       = 0,458
s2(D)       = 0,014
s2(d)       = 0,013
Er2         = 0,972
Phi         = 0,970
```

ANEXO 14

Tabla de Resultados de Estudio G- Modelo H/R:HxI

Effect	df	T	SS	MS	VC
H	7	1722.47073	1722.47072	246.06725	0.09665
R:H	384	9216.12812	7493.65738	19.51473	0.37786
I	49	1150.80720	1150.80719	23.48586	0.05230
HI	343	3655.98146	782.70353	2.28193	0.03568
RI:H	18816	22846.22810	11696.58926	0.62163	0.62163
Mean		0.00001			
Total	19599		22846.22808		
Grand Mean:		0.00002			

Date and time at beginning of Run 1: Wed Nov 22 20:25:36 2017
Processor time for run: 0 seconds

The calculated grand mean = 3,8713
This value has been subtracted from the actual scores for the calculations.
While this improves the apparent accuracy, it does not affect the calculated variances.

Facets with adjusted sample sizes according to Brennan

'H'	Differentiation	8,00
'R'	Random	30,62
'I'	Random	50,00

Pattern	Var. Comp.	Levels	Signature	Rule
H	0,097	1.00	d	tau only
R:H	0,012	30,62	dr	Delta and delta
I	0,001	50,00	r	Delta only
H|I	0,001	50,00	dr	Delta and delta
R:H|I	0,000	30,62 * 50,00	dr	Delta and delta

RESULTS:

s2(T) = 0,097
s2(D) = 0,015
s2(d) = 0,013
Er2 = 0,878
Phi = 0,870

BIBLIOGRAFÍA

Abad, F., Garrido, J., Olea, J., & Ponsoda, V. (2006). *Introducción a la Psicometría, Teoría Clásica de los Tests y Teoría de Respuesta al Ítem*. Universidad Autónoma de Madrid.

Abraham, R., Ramnarayan, K., Vinod, P., & Torke, S. (2008). Students' perceptions of learning environment in an Indian medical school. *BMC Med Educ* 8:20. Disponible en http://www.biomedcentral.com/1472-6920/8/20

Alshomrani AT, & AlHadi AH. Learning environment of the Saudi psychiatry board training program (2017). *Saudi Med J; Vol. 38 (6): 629-635. doi:10.15537/smj.2017.6.18164*

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). Standards for educational and psychological testing. Washington, D.C.: American Educational Research Association.

Arias Martínez, B. (2008) Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS y SAS. En Verdugo, M.A., Crespo, M., Badía, M., & Arias, B. (coord.), *Metodología en la investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales*. VI Seminario Científico SAID. Salamanca: Publicaciones del INICO. Colección Actas. pp. 75-120. Disponible en http://sid.usal.es/idocs/F8/FDO20749/said_2008.pdf#page=76

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.

Bakhshialiabad, H., Bakhshi, M., & Hassanshahi, G. (2015). Students' perceptions of the academic learning environment in seven medical sciences courses based on DREEM. *Adv Med Educ Pract*, 6, 195–203. doi: 10.2147/AMEP.S60570

Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M.B. (1998). Recommendations for the Cross-Cultural Adaptation of Health Status Measures. American Academy of Orthopaedic Surgeons, Institute for Work & Health, USA. Disponible en *https://www.ortho.umn.edu/sites/ortho.umn.edu/files/recommendations-cultural.pdf*

Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M.B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25:3186-3191.

Bennett, D., Dornan, T., Bergin, C., & Horgan, M. (2013) Postgraduate training in Ireland: expectations and experience. *Ir J Med Sci*, 183, 611-620. doi:10.1007/s11845-013-1060-5

Bentler, P. (1989). EQS 6 Structural Equations Program Manual. Multivariate Software, Inc. Encino, California USA.

Blanco Villaseñor, A. (1989). Fiabilidad y generalización de la observación conductual. *Anuario de Psicología*, 43, 6-32.

Bloch, R. & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide Nº 68. *Med Teach*, 34, 960-992.

Boor, K., Scheele, F., van der Vleuten, C., Scherpbier ,A, Teunissen, P., & Sijtsma, K . (2007). Psychometric properties of an instrument to measure the clinical learning environment. *Med Educ* 41 (1): 92-99.

Boor, K., Scherpbier, A., Teunissen, P., van der Vleuten, C., & Scheele, F. (2009). The postgraduate learning climate: integration of work, training, and residents' needs. En Boor, K.: *The clinical learning climate*. Amsterdam pp. 65-78.

Boor, K., Van der Vleuten, C., Teunissen, P., Scherpbier, A., & Scheele, F. (2011). Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach* 33: 820–827.

Borrell Bentz, R.M. (2005). La educación médica de posgrado en Argentina: el desafío de una nueva práctica educativa – 1ª ed. - Buenos Aires: Organización Panamericana de la Salud – OPS.

Brennan, R. L. (2001). Generalizability Theory. New York: Springer-Verlag

Brennan, R. L. (2011). Generalizability Theory and Classical Test Theory. *Applied measurement in education*, 24, 1–21. doi: 10.1080/08957347.2011.532417

Butterfield P.S. (1988). The Stress of Residency. A Review of the Literature. *Arch Intern Med.* 148(6):1428–1435. doi:10.1001/archinte.1988.00380060192034

Carballido, R., Ayala, F.P., & Amarilla, A. (1998). Estado Actual de las residencias de Cardiología. Encuesta Nacional de Residentes 1997. *Rev Argent Cardiol* , 66, 3, 269-276.

Cervantes, V. (2005). Interpretaciones del coeficiente Alfa de Cronbach. *Avances en Medición,* 3(1), 9-28.

Cohen, J. S., & Patten, S. (2005). Well-being in residency training: a survey examining resident physician satisfaction both within and outside of residency training and mental health in Alberta. *BMC Medical Education*, *5*, 21. http://doi.org/10.1186/1472-6920-5-21

Cook, D.A., & Beckman, T.J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*, 119, 166.e7-166.e16. doi:10.1016/j.amjmed.2005.10.036

Colbert-Getz, J., Sooyoun, K., Goode, V., Schochet, R., & Wright, S. (2014). Assessing Medical Students' and Residents' Perceptions of the Learning Environment: Exploring Validity Evidence for the Interpretation of Scores From Existing Tools. *Acad Med*, 89(12):1687-93. doi: 10.1097/ACM.0000000000000433

Cupani, M. (2012). Análisis de Ecuaciones Estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación. *Tesis*, 2(1):186-199.

Díaz, C., Batanero, C., & Cobo, B. (2003). Fiabilidad y generalizabilidad. Aplicaciones en evaluación educativa. *Números*, 54, 3-21.

DiStefano, C., & Morgan, G.B. (2014). A comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21, 1–14.

Domínguez, L. C., Silkens, M. E., & Sanabria, A. (2018) The D-RECT in Spanish: validation of a Tool evaluating Residents' Learning Climate. *New Perspectives on Learning Climates in Postgraduate Medical Education*, 61.

Elosua, P., Mujika, J., Almeida, L.S., & Hermosilla, D. (2014). Procedimientos analítico-racionales en la adaptación de tests. Adaptación al español de la batería de pruebas de razonamiento. *Rev Latinoam Psicol,* 46, 2, 117-126

Fernandez, A. (2008). Adaptación de Tests a Otras Culturas. En Tornimbeni, S., Perez, E., & Olaz, F. (eds.), *Introducción a la Psicometría* (191-205). 1ª Ed. Buenos Aires: Paidós.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16, 625-641. doi: 10.1080/10705510903203573

Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18: 39-50.

Genn, J.M. (2001). AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education—A unifying perspective. *Med Teach* 23(5):445–454.

Gruppen, L., Stansfield, B., Zhao, Z., & Sen, S. (2015). Institution and specialty contribute to resident satisfaction with their learning environment and workload. *Acad Med*, 90, S77–S82. doi: 10.1097/ACM.0000000000000898

Hair, J.F., Black W.C., Babin, B.J., & Anderson, R.E. (eds.)(2010) *Multivariate Data Analysis*.(7th Ed.). Pearson Prentice Hall. S.

Hambleton, R.K., & Bollwark, J. (1991). Adapting tests for use in different cultures: Technical issues and methods. *Bulletin of the International Test Commission*, *18*, 3-32.

Hambleton, R.K. (2005) Issues, Designs and Technical Guidelines for Adapting Tests Into Multiple Languages and Cultures. En Hambleton, R.K., Merenda, P. F., & Spielberger, C. D. (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ:Lawrence Erlbaum Associates.

Hambleton, R., & Zenisky, A. (2010). Translating and Adapting Tests for Cross-Cultural Assessments. En Matsumoto, D., & Van de Vijver, F.(Eds.), *Cross-Cultural Research Methods in Psychology* (Culture and Psychology, pp. 46-70). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511779381.004

Hoff, T.J., Pohl, H., & Bartfield, J. (2004). Creating a learning environment to produce competent residents: the roles of culture and context. *Acad Med* 79 (6): 532-540.

Hutchinson, L. (2003). Educational environment. *BMJ*, 326, 810-812. doi: 10.1136/bmj.326.7393.810

Iblher, P., Zupanic, M., & Ostermann, T. (2015) The Questionnaire D-RECT German: Adaptation and test theoretical properties of an instrument for evaluation of the learning climate in medical specialist training. *GMS Z Med Ausbild*.16; 32(5): Doc 55. doi: 10.3205/zma000997

International Test Commission. (2017). The ITC Guidelines for Translating and Adapting Tests (Second edition). [www.InTestCom.org]

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.

Keszei, A.P., Novak, M., & Streiner, D.L. (2010). Introduction to health measurement scales. *Journal of Psychosomatic Research*, 68, 319–323. doi:10.1016/j.jpsychores.2010.01.006

Learning Environment (Agosto 29, 2013). In S. Abbott (Ed.), The glossary of education reform. Recuperado de http://edglossary.org/learning-environment/

Li, C-H. (2016). The performance of MLR, USLMV, and WLSMV estimation in structural regression models with ordinal variables. *Psychological Methods*, 21, 3, 369-387. doi: http://dx.doi.org/10.1037/met0000093

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: implications for theory and practice. *Studies in Higher Education*, 27 (1): 27-52. doi: 10.1080/03075070120099359

Llera, J., & Durante, E. (2014). Correlación entre el clima educacional y el síndrome de desgaste profesional en los programas de residencia de un hospital universitario. *Arch Argent Pediatr*, 112, 1, e6-e11. http://dx.doi.org/10.5546/aap.2014.e6

Lombarts, K.M.J.M.H., Heineman, M.J., Scherpbier, A.J.J.A., & Arah, O.A. (2014) Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. *PLoS ONE* 9(1): e86512. doi:10.1371/journal.pone.0086512

MacCallum, R.C., & Austin, J.T. (2000). Aplications of structural equation modeling in psychological research. *Annu Rev Psychol*, 51, 201-226.

Malda, M., van de Vijver, F.J.R., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*, 50, 451-468.

Méndez-López, J.F., Mendoza-Espinosa, H., Torruco-García, U. & Sánchez-Mendiola, M. (2016). El médico residente como educador. *Inv Ed Med*, 2, 7, 154-161.

Messick, S. (1979). Test validity and the ethics of assessment. *ETS Research Report Series*, 1979: i–43. doi:10.1002/j.2333-8504.1979.tb01178.x

Messick, S. (1993). Foundations of validity: meaning and consequences in psychological assessment. *ETS Research Report Series*, 1993: i–18. doi:10.1002/j.2333-8504.1993.tb01562.x

Morales Vallejo, P. (2010). El análisis factorial en la construcción e interpretación de tests, escalas y cuestionarios. Disponible en *http://www.upcomillas.es/personal/peter/investigacion/AnalisisFactorial.pdf* (Última revisión, 15 de Septiembre de 2013).

Morata-Ramírez, M. A., Holgado-Tello, F.P., Barbero-García, I. & Mendez, G. (2015). Análisis Factorial Confirmatorio. Recomendaciones sobre Mínimos Cuadrados no Ponderados en función del error Tipo I de Ji-cuadrado y RMSEA (Confirmatory Factory Analysis. Recomendations for Underweighted Least Squares method related to Chi-square And RMSEA Type I error). *Acción Psicológica,* 12(1), 19-90. doi: http://dx.doi.org/10.5944/ap.12.1.14362

Muñiz J., Elosua P., & Hambleton R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema, 25*(2), 151-157. doi: 10.7334/psicothema2013.24

Norman, G.R., & Streiner, D.L. (2008). Biostatistics. The Bare Essentials. 3rd edition. Hamilton: B.C.Decker Inc.

Pacifico. J.L., van der Vleuten, C.P.M., Muijtjens, A.M.M., Sana, E.A., & Heeneman, S. (2018).Cross-validation of a learning climate instrument in a non-western postgraduate clinical environment. MC Med Educ. 2018 Jan 25; 18 (1): 22. doi: https://doi.org/10.1186/s12909-018-1127-0

Padilla, J.L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26, 1, 136-44. doi: 10.7334/psicothema2013.259

Palés J., Gual, A., Escanero, J., Tomás, I., Rodríguez-de Castro, F., Elorduy, M., Virumbrales, M., Rodríguez, G., & Arce, V. (2015) Educational climate perception by preclinical and clinical medical students in five Spanish medical schools. *Int J Med Educ*, 8; 6, 65-75. doi: 10.5116/ijme.5557.25f9

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17,* 354–373. doi: 10.1037/a0029315

Riquelme Pérez, A., Fuentes, G., Jeria, Á., Méndez, I., Aranís, C., Larios, G., & Oporto Caroca, J. (2007). Ambiente educacional y calidad de la docencia en la Escuela de Medicina. *Ars Médica* 15 (15): 125-39. doi:http://dx.doi.org/10.11565/arsmed.v36i2.149

Riquelme, A., Padilla, O., Herrera, C. Olivo, T, Roman, J.A., Sarfatis, A., Solís, N., Pizarro, M., Díaz, L.A., Torres, P., & Roff, S. (2015). Measuring the educational environment in ambulatory settings. *Educ Med* 16 (2): 131-140

Rodríguez Fernández, Z., Rizo Rodríguez, R., Mirabal Fariñas, A., Nazario Dolz, A.M., & García Céspedes, M.E. (2017) Educación en el trabajo en la enseñanza médica superior. *MEDISAN*, 21, 7, 913-925. Disponible en http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192017000700017&lng=es&tlng=es

Roff, S. McAleer, S., Harden, R., Al-Qahtani, M., Ahmed, A., Deza, H., Groenen, G., & Primparyon, P. (1997). Development and validation of the Dundee Ready Education Environment Measure (DREEM).*Med Teach* 19:295-299.

Roff, S., & McAleer, S. (2001). What is educational climate? *Med Teach*, 23, 333-334.

Roff, S., McAleer, S, & Skinner, A. (2005). Development and validation of an instrument to measure postgraduate clinical learning and teaching educational environment for hospital based junior doctors in the UK. *Med Teach* 27 (4) 326-331

Ruiz de Gauna, P., González Moro, V., & Morán-Barrios, J. (2015). Diez claves pedagógicas para promover buenas prácticas en la formación médica basada en competencias en el grado y en la especialización. *Educ Med*, 16, 1, 34-42.

Savalei, V. & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, 66, 2, 201-223. doi: 10111/j2044-8317201202049x

Schönrock-Adema, J., Bouwkamp-Timmer, T., van Hell, E.A., & Cohen-Schotanus J. (2012). Key elements in assessing the educational environment: where is the theory?. *Adv in Health Sci Educ*, 17, 727–742. doi:10.1007/s10459-011-9346-8.

Seabrook, M. A. (2004), Clinical students' initial reports of the educational climate in a single medical school. *Med Educ* 38: 659–669.

Shavelson, R., & Webb, N. (1991). *Generalizability theory*. Beverly Hills, CA: Sage.

Silkens, M. E. W. M., Smirnova, A., Stalmeijer, R.E., Arah, O.A., Scherpbier, A.J.J.A., van der Vleuten, C.P.M., Lombarts, K.M.J.M.H. (2015). Revisiting the D-RECT tool: Validation of an instrument measuring residents' learning climate perceptions. *Med Teach*, 38, 5, 476-481. http://dx.doi.org/10.3109/0142159X.2015.1060300

Silkens, M. E. W. M., Arah, O. A., Scherpbier, A. J. J. A., Heineman, M. J., & Lombarts, K. M. J. M. H. (2016). Focus on quality: investigating residents' learning climate perceptions. *PLoS ONE* 11(1): e0147108. doi:10.1371/journal.pone.0147108

Streiner, D.L. (1994). Figuring out factors: the use and misuse of factor analysis. *Can J Psychiatry,* 39, 3, 135-140.

Van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. En R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 39-63). Mahwah, NJ:Lawrence Erlbaum Associates.

Van Mook, W., Stalmeijer, R., Muijtjens, A., & Scherpbier, A. (2016). A preliminary study on the quality of the intensive care medicine training programmes in the Netherlands. *Netherlands Journal of Critical Care*, *24*, 7, 10-15.

Van Vendeloo, S. N., Brand, P. L. P., & Verheyen, C. C. P. M. (2014). Burnout and quality of life among orthopaedic trainees in a modern educational programme. *Bone Joint J*, 96-B, 8, 1133-1138. doi: 10.1302/0301-620X.96B8.33609

Vargas Halabí, T., & Mora-Esquivel, R. (2017). Tamaño de la muestra en modelos de ecuaciones estructurales con constructos latentes: Un método práctico. *Actualidades Investigativas en Educación*, *17*, 1, 25-60. https://dx.doi.org/10.15517/aie.v17i1.27294

West, C.P., & Shanafelt, T.D. (2007). The influence of personal and environmental factors on professionalism in medical education. *BMC Med Educ,* 7, 29. doi:10.1186/1472-6920-7-29

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, PII:S1567-4223(10)00054-2, DOI: 10.1016/j.elerap.2010.07.003

Yang-Wallentin F., Jöreskog K. G., & Luo H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, 17, 392–423. doi: 10.1080/10705511.2010.489003

CURRICULUM VITÆ

Nombre: **RAÚL BERNARDO PUYOL**
Lugar y Fecha de Nacimiento: Corrientes- 13 de Febrero de 1965
Títulos obtenidos:

1- Médico Cirujano- Facultad de Medicina de la UNNE- 1988
2- Especialista en Terapia Intensiva –1994
3- Especialista en Neumonología – 2000
4- Especialista en Clínica Médica - 2011
5- Especialista en Docencia y Gestión Universitaria-Orientación en Ciencias de la Salud (2008)

Antecedentes Académicos

FACULTAD DE MEDICINA – UNIVERSIDAD NACIONAL DEL NORDESTE

1) Cátedra de Parasitología- Auxiliar de Docencia 2a categoría- (1985- 1989)
   - Jefe de Trabajos Prácticos (1989 – 1990)

2) Cátedra III de Medicina – Auxiliar de Docencia de 1ª categoría (1993 -2001)
   - Instructor de Internado Rotatorio 1994- 1997)
   - Jefe de Trabajos Prácticos (2001-2003)
   - Profesor Adjunto (2003-2015)

3) Cátedra de Medicina Respiratoria
   - Coordinador Docente (2003-2008)

4) Cátedra de Medicina Basada en la Evidencia
   - Docente Adscripto (2007-20015)

FACULTAD DE CIENCIAS EXACTAS- UNIVERSIDAD NACIONAL DEL NORDESTE

5) Cátedra de Fisiopatología- Carrera de Bioquímica
   - Jefe de Trabajos Prácticos (1994-2013)
   - Profesor Responsable Interino (2014-2015)

6) Asignatura Urgencias Hospitalarias- Carrera de Bioquímica
   - Profesor Responsable (2008- 2015)

FACULTAD DE INGENIERÍA- UNIVERSIDAD DE LA CUENCA DEL PLATA

7) Carrera de Licenciatura en Nutrición – Asignatura: Fisiopatología del Adulto
   Profesor Titular (2009-2010)