

Clima de aprendizaje en residencias: adaptación transcultural y validación del cuestionario D-RECT al español

Learning climate in residencies: transcultural adaptation and validation of D-RECT questionnaire into Spanish

Raúl Bernardo Puyol^a, Sergio Giannasi^{b,c}, Eduardo Durante^{d,c}

Resumen

Introducción. Medir el clima de aprendizaje es un aspecto relevante para estimar la calidad de los programas educativos. El Dutch Residency Educational Climate Test (D-RECT) es un instrumento ampliamente reconocido para ese propósito.

Objetivo. Realizar la adaptación transcultural y validación del D-RECT al español para su utilización en Argentina.

Métodos. A partir del cuestionario original, se realizó el proceso de traducción, y posterior demostración de 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 transcultural, validación de contenido y del proceso de respuesta, 403 residentes de diferentes especialidades contestaron la versión en español del cuestionario e ingresaron al estudio. El análisis de propiedades psicométricas se realizó con los 392 cuestionarios completos, revelando evidencia favorable sobre la validez y confiabilidad del instrumento.

Conclusión. Se realizó la adaptación transcultural del cuestionario D-RECT y se confirmó su adecuada validez y confiabilidad para evaluar el clima de aprendizaje en residencias de Argentina.

Abstract

Introduction. Measuring the learning climate is a relevant aspect to estimate the quality of educational programmes. The Dutch Residency Educational Climate Test (D-RECT) is a widely recognised instrument for that purpose.

Objective. To perform the cross-cultural adaptation and validation of the D-RECT into Spanish to be used in Argentina.

Methods. Based on the original questionnaire, the translation process was carried out, and evidence of content validity (linguistic and cultural equivalence, and representativeness of the items within the construct), response process (pretesting and pilot group interviews), internal structure and reliability (Cronbach's alpha, exploratory and confirmatory factor analysis, and G study) was subsequently demonstrated.

Results. Once the cross-cultural adaptation, content validation and response process validation stages were completed, 403 residents from different specialties answered the adapted version of the questionnaire and entered the study. The analysis of psychometric properties was performed with 392 complete questionnaires, revealing favourable evidence for the instrument's validity and reliability.

Conclusion. The cross-cultural adaptation of the D-RECT questionnaire was carried out and its adequate validity and to assess the learning climate in residences in Argentina was confirmed.

Palabras clave: Clima de aprendizaje, Residencias. Keywords: Learning climate, Residencies.

Puyol RB, Giannasi S, Durante E. Clima de aprendizaje en residencias: adaptación transcultural y validación del cuestionario D-RECT al español. Evid Actual Pract Ambul. 2021;24(2):e002104.

Introducción

El término clima de aprendizaje (CA) se refiere al modo en que el ambiente educacional es percibido por los estudiantes¹ y comprende distintas locaciones, contextos y culturas donde éstos aprenden, incluyendo interacciones entre pares, y organización de actividades educacionales². En las residencias médicas, el complejo marco de actividad y su carga de estrés³ tiene impacto decisivo en el aprendizaje⁴ influyendo fuertemente en el CA y la satisfacción de los residentes⁵.

El CA es un constructo que relaciona múltiples aspectos de formación y refleja el abordaje de los departamentos de docencia, a través de las percepciones de los aprendices sobre cuestiones como la atmósfera, la supervisión, y el estado del aprendi-

zaje. Este entorno es construido por las interacciones entre los aprendices y los demás trabajadores de la salud, bajo influencia de cada organización y sus artefactos⁶. Un CA saludable puede beneficiar el desarrollo del residente como profesional y la calidad de atención que brinda a sus pacientes⁷.

En las residencias médicas de Argentina se han detectado problemas en distintos aspectos: la regulación y la planificación, la orientación y la gestión, el enfoque de la formación, los modelos pedagógicos⁸. Varios componentes del CA están implicados en la adquisición de competencias⁹, la sensación de estrés y acoso¹⁰, y la aparición de *burn-out*¹¹.

Evaluar el CA es relevante para estimar la calidad de los programas de formación y el funcionamiento educacional de un

^a Hospital Escuela "José de San Martín", Rivadavia 1250, Corrientes, 3400, Argentina. raulpuyol42@yahoo.com.ar

^b Servicio de Terapia Intensiva, Hospital Italiano de Buenos Aires, Pres. Tte. Gral. Juan Domingo Perón 4190, Buenos Aires, Argentina.

^c Departamento de Investigación, Centro de Investigación en Educación para Profesionales de la Salud, Instituto Universitario Hospital Italiano de Buenos Aires. sergio.giannasi@hospitalitaliano.org.ar, eduardo.durante@hospitalitaliano.org.ar

^d Servicio de Medicina Familiar y Comunitaria, Hospital Italiano de Buenos Aires.

departamento, dada su versatilidad⁶. Los programas de residencias médicas deben crear un ambiente apropiado al contexto cultural y laboral de los hospitales, como factores para la adquisición de profesionalismo y otras competencias⁹⁻¹³. El CA interesa al cuerpo docente para lograr alta calidad de enseñanza y asistencial, y existe evidencia de relación positiva entre ambos^{1,14,15}. Enseñar significa organizar el CA al igual que impartir conocimiento o compartir experticia¹⁶. Corresponde a los programas e instituciones proveer la atmósfera y las condiciones de trabajo para tales propósitos^{8,9}, e identificar las áreas problemáticas para tomar eventuales medidas que permitan mejoras¹³.

Las escuelas de medicina evolucionan y cambian evaluando su entorno y su currículo a través de estudios de su ambiente¹ por instrumentos válidos^{17,18}. Por su versatilidad y facilidad de aplicación, se seleccionó el D-RECT (*Dutch Residency Educational Climate Test*)⁶, cuestionario desarrollado en Holanda para la valoración del CA en las residencias, para su aplicación en Argentina. Este instrumento también fue investigado para ese propósito en otros países¹⁹⁻²².

El objetivo general de este estudio fue realizar la adaptación transcultural y validación del D-RECT para su utilización en Argentina. Como objetivos específicos, nos propusimos obtener una versión del instrumento en español y aportar evidencias sobre su validez y confiabilidad.

Material y métodos

El primer paso consistió en obtener una versión válida del D-RECT en idioma español, para luego aplicarlo como testeo inicial en un grupo pequeño de individuos²³⁻²⁵.

Posteriormente, se lo aplicó en una muestra representativa de la población destino, para analizar 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- Obtener una versión del instrumento en español

Previo autorización de uno de los autores del trabajo original (C. Van der Vleuten), el cuestionario original fue traducido al español por dos traductores independientes bilingües. Luego, un comité de expertos de la Dirección de Docencia e Investigación del Hospital Escuela "Gral. San Martín" de la Facultad de Medicina de la Universidad Nacional del Nordeste, conformado por dos profesores titulares y dos profesores adjuntos de la carrera de Medicina, expertos en educación médica, unificó la versión considerada como mejor representación del constructo en nuestra lengua y contexto cultural. Ésta se tradujo nuevamente a su idioma original por otros dos traductores de lengua inglesa nativa que desconocían el instrumento original y los objetivos de la investigación. El comité de expertos unificó ambas traslaciones en un cuestionario preliminar²³⁻²⁸.

Objetivo 2- Obtención de evidencias de validez y confiabilidad.

- Evidencia basada en el contenido de la prueba:** implica demostrar que los ítems (contenido del instrumento) constituyen una muestra representativa del universo de posibles ítems sobre el constructo a medir, estableciendo su pertinencia dentro de los dominios y el constructo, en el contexto cultural de aplicación. El aporte de evidencia se logró mediante el comité de expertos, considerando la congruencia de los ítems en cada subescala con el constructo, su claridad lingüística y su equivalencia en el contexto cultural^{23,24}.
- Evidencia basada en el proceso de respuesta:** significa indagar circunstancias inherentes a la aplicación del instrumento y eventual sesgo relativo a la administración de la prueba. Se efectuó el testeo del cuestionario preliminar

sobre un grupo piloto de 15 residentes de distintas especialidades, con entrevistas individuales enfocadas sobre sus procesos de pensamiento al responder cada ítem, la adecuada redacción e interpretación para el constructo a medir²⁹. Esta instancia incluye un aspecto denominado validez aparente (*face validity*), vinculada con la validez de contenido³⁰. Ningún residente del grupo piloto participó en la aplicación de la prueba definitiva.

3. Evidencia basada en la estructura interna de la prueba y confiabilidad.

Tras aplicar la prueba definitiva sobre una muestra de la población destino, se analizaron: su estructura y las relaciones entre ítems y dominios, por medio de análisis factorial exploratorio (AFE); el funcionamiento del modelo teórico establecido a priori, a través de análisis factorial confirmatorio (AFC); y la consistencia interna, con el cálculo del coeficiente Alfa (α) de Cronbach (considerando confiable un valor mayor de 0,80)³¹. Por último, se efectuó un estudio de Generalizabilidad (G) para estimar la capacidad de generalización de los resultados y fuentes de variación en las mediciones.

AFE. Es una técnica que define la estructura subyacente entre las variables³² y la tendencia de los ítems a agruparse alrededor de factores, que equivalen a constructos latentes. El contenido conceptual de los ítems agrupados sobre un mismo factor indica qué constructos explican esa correlación. El AFE contribuye con la validez de constructo de la medición. Adicionalmente, aporta sobre 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 factor³³. En este trabajo se utilizó el análisis de componentes principales como método de extracción de factores en el AFE, y rotación ortogonal con normalización Varimax con Kaiser. Se consideró significativo un valor de carga factorial mínimo de 0,30^{32,33}.

AFC. Analiza las relaciones entre variables observadas y constructos latentes o factores, establecidas a priori por el investigador acorde con evidencia teórica y empírica³². Los resultados muestran el funcionamiento del modelo con distintos parámetros de bondad de ajuste, y la existencia de validez convergente y discriminante³⁴. En este caso se tomaron como indicadores de ajuste: X^2 /grados de libertad, NNFI (Non-Normed Fit Index o Índice de ajuste no normalizado), CFI (Comparative Fit Index o Índice de ajuste comparativo), SRMR (Standardized Root Mean Square Residual o Residual cuadrático medio estandarizado de la raíz), RMSEA (Root Mean Square Error of Approximation o Error cuadrático medio de aproximación), GFI (Goodness-Of-Fit Index o Índice de bondad de ajuste), AGFI (Adjusted Goodness-Of-Fit Index o Índice de bondad de ajuste ajustado). 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 . Para comprobar el ajuste de los modelos de estructuras de covarianzas, dada la sensibilidad del X^2 al tamaño muestral, se decidió emplear el valor X^2 de Satorra-Bentler dividido entre sus grados de libertad ($SB\chi^2/gl$) considerando ajuste adecuado a un valor inferior a 2,0^{34,35}. Respecto de los métodos de estimación en el AFC, y considerando en este trabajo la presencia de variables con escala ordinal y la ausencia de normalidad multivariante, se decidieron aplicar los métodos ULS (Unweighted Least Squares, o mínimos cuadrados no ponderados) y DWLS (Diagonally Weighted Least Squares, o mínimos cuadrados ponderados en diagonal) por separado para observar eventuales diferencias^{34,36,37}. El procedimiento incluyó indagar sobre validez convergente y discriminante de los ítems respecto del constructo³². La confiabilidad puede verse como un

componente de la validez de convergencia³². El cálculo de fiabilidad fopuesta aporta evidencia acerca de la consistencia interna (como alternativa al α de Cronbach). Se consideró que un valor superior a 0,70 para cada variable latente brinda evidencia a favor de la confiabilidad de la medición. Respecto de la validez discriminante, cada constructo debería compartir más varianza con sus propios indicadores que con otros constructos del modelo³⁸. Por tanto, la varianza media extractada (VME) de cada constructo debe ser mayor que el cuadrado de las correlaciones con los otros.

Estudio G. La teoría de la Generalizabilidad examina la contribución relativa de la variable de interés (el objeto de medición) comparada con la varianza de las fuentes de error³⁹ a través del análisis de varianza (ANOVA, por sus iniciales en inglés). Se determinó el coeficiente de generalizabilidad (estimación de la estabilidad de las mediciones frente a variaciones aleatorias⁴⁰) y los porcentajes de varianza de las facetas (como se denomina a las fuentes de variación que afectan la precisión de una prueba), para el cuestionario global y para cada dominio.

Características del instrumento

El D-RECT es un cuestionario de 50 ítems, distribuidos originalmente en 11 subescalas o dominios. Cada ítem es valorable mediante una escala Likert, asignándose un puntaje de 1 a 5 según el grado de acuerdo con el enunciado: totalmente en desacuerdo: 1 punto; parcialmente en desacuerdo: 2 puntos; neutro: 3 puntos; parcialmente de acuerdo: 4 puntos; totalmente de acuerdo: 5 puntos.

Muestra para la administración de la prueba definitiva.

Se estimó necesario un mínimo de 5 sujetos por ítem, en base al trabajo original⁶. Se realizó un muestreo no probabilístico por conveniencia, invitando verbalmente a participar a los residentes en todas las etapas de formación, pertenecientes a hospitales públicos de las ciudades de Corrientes (provincia de Corrientes) y Resistencia (provincia del Chaco), previa autorización de las respectivas autoridades institucionales. Asimismo, se invitó a los médicos residentes que asistieron al Congreso Nacional de Medicina Familiar y Comunitaria (Facultad de Medicina de Corrientes, Octubre de 2014). La invitación a participar fue realizada en forma personal por el investigador principal (RBP), explicando a los grupos de residentes convocados los alcances de la investigación, contestando preguntas y dudas, y garantizando el carácter voluntario y anónimo de la encuesta, que se completó en soporte de papel.

Análisis estadístico

La información recolectada fue registrada en una base de datos MS Excel, y luego procesada mediante software SPSS para el AFE, LISREL 8.80 para el AFC, y G-String IV/urGenova para el estudio G. Las variables categóricas se expresaron en valores absolutos y porcentajes, y las variables continuas, en media y desvío estándar o mediana e intervalo intercuartilo, según su distribución.

Aval Institucional

El presente trabajo fue avalado por las autoridades de la Facultad de Medicina de la Universidad Nacional del Nordeste para su realización.

Resultados

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

El cuestionario obtenido se muestra como **anexo** en el material suplementario.

Objetivo 2: Obtención de evidencia sobre validez y confiabilidad

Evidencia sobre validez de contenido

El grupo de expertos trabajó teniendo en cuenta la calidad formal de la herramienta respecto de la relación de cada ítem con el dominio correspondiente y el constructo a medir, la estructura gramatical, claridad y adecuación a la población destino.

Evidencia sobre el proceso de respuesta

- **Pre-testeo sobre grupo piloto.** Se aplicó el instrumento a un grupo de 15 residentes, quienes fueron entrevistados individualmente por el investigador, respecto del proceso de interpretación de la prueba y los ítems, comentarios y sugerencias sobre eventuales cambios o agregados. En todos los casos la interpretación de cada reactivo fue similar al comité experto. Saturada la muestra sin modificaciones, se implementó el cuestionario definitivo.
- **Administración de la prueba.** Bajo supervisión personal del autor principal (RBP), respondieron en forma anónima los residentes de siete Hospitales Públicos Universitarios de Corrientes y Resistencia, y los residentes asistentes al Congreso mencionado. Se obtuvieron respuestas de 403 participantes de 20 especialidades diferentes. La edad promedio fue de 28,2 años (desvío estándar 2,74); 239 (59,4 %) fueron de sexo masculino y 164 (49,6 %), de sexo femenino.

Evidencia sobre la estructura interna y la confiabilidad

Tras la exclusión de 11 encuestas (2,7%) por puntajes confusos o datos faltantes, se analizaron 392 cuestionarios completos.

- **Análisis Factorial Exploratorio.** Todos los dominios resultantes contuvieron al menos tres ítems, como en el trabajo original⁶. El agrupamiento de variables mostró un patrón similar al estudio original, excepto en los dominios Trabajo Adaptado a la Competencia del Residente y Educación Formal, cuyos reactivos se agruparon en un mismo factor, reduciendo el número de dominios (de 11 a 10) y manteniendo constante el total de ítems del instrumento. Ese modelo se tomó como base para el AFC y el cálculo del α de Cronbach.
- **Consistencia Interna.** El valor del coeficiente α de Cronbach fue de 0,972 para el instrumento, revelando evidencia a favor de la confiabilidad de la medición global. Para un análisis más detallado, 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 1).
- **Análisis Factorial Confirmatorio.** Todos los índices revelaron un adecuado ajuste para el modelo teórico (ver Tabla 2). Las cargas factoriales estandarizadas para todas las variables latentes revelaron valores superiores a 0,60, y sus valores T por encima de 1,96 en ambos métodos. La VME de cada dominio arrojó un valor superior a 0,5. Todos estos resultados indican existencia de validez convergente³². Los resultados obtenidos al evaluar la fiabilidad compuesta de cada dominio se presentan en la Tabla 3, mientras que los vinculados en la evaluación de la validez discriminante se resumen en la Tabla 4 (método de los mínimos cuadrados no ponderados) y la Tabla 5 (método de los mínimos cuadrados ponderados en diagonal).
- **Generalizabilidad.** El estudio incluyó dos modelos de análisis: el primero de tipo cruzado (residentes/ítems) cuyos resultados revelaron un aporte de varianza de la faceta de diferenciación (estudiantes) del 45,8 %, y para

la faceta de generalización (ítems) de 0,05 %, que representan valores satisfactorios. En este modelo E/I, el cálculo del coeficiente G equivale al estudio de la consistencia interna de la prueba. En nuestro caso, el valor obtenido (0,97) reveló una alta confiabilidad para el instrumento, de modo similar al α de Cronbach. Analizando la confiabilidad de cada dominio por separado (ver Tabla 6) los valores fueron considerados satisfactorios. El segundo modelo de estudio G para el instrumento en forma 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 G obtenido fue de 0,87. Todos estos valores fueron considerados satisfactorios para la confiabilidad del test global. Para un resultado confiable de la prueba se calculó un número mínimo de 16 residentes encuestados por cada hospital.

Tabla 1. Consistencia interna para cada dominio del cuestionario

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 competencia + Educación Formal	0,905
Rol de los instructores	0,898
Rol del instructor especialista	0,907
Registro del alta del paciente	0,871

Tabla 2. Resultados en los índices de bondad de ajuste según el método de análisis factorial confirmatorio

Método	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
Mínimos cuadrados no ponderados (ULS)	1,90	0,98	0,99	0,99	0,044	0,048	0,99	0,99	226,2
Mínimos cuadrados ponderados en diagonal (DWLS)	1,93	0,98	0,99	0,99	0,044	0,049	0,99	0,99	222,6

NFI: Normed Fit Index o Índice de ajuste normalizado; NNFI: Non-Normed Fit Index o Índice de ajuste no normalizado; CFI: Comparative Fit Index o Índice de ajuste comparativo; SRMR: Standardized Root Mean Square Residual o Residual cuadrático medio estandarizado de la raíz; RMSEA: Root Mean Square Error of Approximation o Error cuadrático medio de aproximación; GFI: Goodness-Of-Fit Index o Índice de bondad de ajuste; AGFI: Adjusted Goodness-Of-Fit Index o Índice de bondad de ajuste ajustado.

*Entre paréntesis, el valor de significación de cada índice

Tabla 3. Fiabilidad compuesta para cada dominio en el análisis factorial confirmatorio.

Método	VL 1	VL 2	VL 3	VL 4	VL 5	VL 6	VL 7	VL 8	VL 9	VL 10
Mínimos cuadrados no ponderados (ULS)	0,85	0,91	0,89	0,85	0,88	0,85	0,92	0,93	0,92	0,90
Mínimos cuadrados ponderados en diagonal (DWLS)	0,86	0,91	0,93	0,78	0,89	0,85	0,92	0,93	0,93	0,88

Nota: Los valores deben ser superiores a 0,70 (Hair et al, 2010)

Variables latentes: VL 1: Supervisión; VL 2: Valoración del desempeño; VL 3: Retroalimentación; VL 4: Trabajo en Equipo; VL 5: Colaboración entre pares; VL 6: Relaciones entre instructores; VL 7: Trabajo adaptado a competencia + Educación Formal; VL 8: Rol de los especialistas; VL 9: Rol de los instructores; VL 10: Alta de los pacientes.

Discusión

Nuestro trabajo consistió en adaptar y validar un instrumento desde su versión original para su apropiada utilización en nuestro medio. Adaptar un cuestionario implica maximizar la armonización cultural y minimizar el sesgo⁴¹⁻⁴³. Respecto del primer objetivo específico, se obtuvo una versión apropiada del D-RECT tras cumplir el proceso de adaptación transcultural previsto.

La validación abarcó tres de los cinco criterios⁴³: contenido, proceso de respuesta, y estructura interna y confiabilidad. La validez implica un juicio sobre el grado en que la evidencia empírica y la teoría apoyan las interpretaciones y acciones derivadas de los puntajes dentro del propósito de una prueba^{42,43}.

La validez de contenido abarca dos aspectos: la relevancia (definir el dominio específico y sus límites operativos) y la cober-

Tabla 4. Valores comparativos para validez discriminante estimada mediante el método de los mínimos cuadrados no ponderados (ULS) (Fornell y Larcker, 1981)

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

Los valores representan los cuadrados de las correlaciones entre variables latentes.

Resaltados en negrita aparecen los valores de Varianza Media Extractada.

Variables latentes: VL 1: Supervisión; VL 2: Valoración del desempeño; VL 3: Retroalimentación; VL 4: Trabajo en Equipo; VL 5: Colaboración entre pares; VL 6: Relaciones entre instructores; VL 7: Trabajo adaptado a competencia + Educación Formal; VL 8: Rol de los especialistas; VL 9: Rol de los instructores; VL 10: Alta de los pacientes.

Tabla 5. Valores comparativos para la validez discriminante estimada mediante el método de los mínimos cuadrados ponderados en diagonal (DWLS) (Fornell y Larcker, 1981)

	VL 1	VL 2	VL 3	VL 4	VL 5	VL 6	VL 7	VL 8	VL 9	VL 10
VL 1	0,67									
VL 2	0,64	0,59								
VL 3	0,43	0,60	0,81							
VL 4	0,32	0,38	0,33	0,61						
VL 5	0,33	0,33	0,17	0,22	0,74					
VL 6	0,42	0,56	0,47	0,32	0,39	0,66				
VL 7	0,54	0,60	0,44	0,40	0,47	0,65	0,62			
VL 8	0,50	0,51	0,46	0,32	0,33	0,59	0,65	0,64		
VL 9	0,43	0,51	0,32	0,24	0,24	0,51	0,68	0,59	0,69	
VL 10	0,49	0,59	0,44	0,34	0,29	0,67	0,70	0,64	0,68	0,71

Los valores representan los cuadrados de las correlaciones entre variables latentes.

Resaltados en negrita aparecen los valores de la Varianza Media Extractada.

Variables latentes: VL 1: Supervisión; VL 2: Valoración del desempeño; VL 3: Retroalimentación; VL 4: Trabajo en Equipo; VL 5: Colaboración entre pares; VL 6: Relaciones entre instructores; VL 7: Trabajo adaptado a competencia + Educación Formal; VL 8: Rol de los especialistas; VL 9: Rol de los instructores; VL 10: Alta de los pacientes.

Tabla 6. Coeficiente G para cada dominio del D-RECT adaptado al español

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

tura (una muestra representativa de ítems⁴⁴). Para este fin, el investigador se enfoca en la definición del constructo, el propósito del instrumento, el proceso de desarrollo y selección de los ítems, la redacción de cada ítem individual y la intervención de calificadores⁴⁵. El presente trabajo cumplió todas las tareas previstas según las directivas de la ITC y la bibliografía de referencia^{23,25-27,44} para validación de contenido con intervención de un comité de expertos sobre el cuestionario. La escala métrica y el ordenamiento de los ítems son del trabajo original, y las muestras poblacionales presentan características (socioculturales, profesión y nivel educativo) similares, comparables en su capacidad de interpretación de los ítems. Asimismo, el sondeo con el grupo piloto no mostró dificultades ni controversias sobre la versión preliminar.

Respecto del proceso de respuesta, la revisión de las acciones y pensamientos de los encuestados puede aclarar el ajuste entre el constructo y la detallada naturaleza del desempeño involucrado^{43,45}. Las fuentes de validez son complementarias, especialmente las de contenido y del proceso de respuesta²⁹, apuntando a la detección e interpretación del funcionamiento diferencial de los ítems. El método de entrevistas contempla los procesos de pensamiento de los encuestados, es sencillo de aplicar y requiere pocos recursos. Su base teórica considera una secuencia de procesos cognitivos entre la formulación de una pregunta y la emisión de respuesta: 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. Finalmente ajustar la respuesta entre las alternativas al propósito y comunicarla. El testeo sobre grupo piloto implica dos decisiones: quiénes y cuántos sujetos formarán parte de este grupo. La decisión no es a priori una cuestión numérica. 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 del testeo inicial: la saturación teórica (continuar hasta que no aparece nueva información) y la relevancia (seleccionar entrevistas por su importancia sobre los hallazgos surgidos)²⁹.

Con respecto al tercer criterio de validez (la evaluación de la estructura interna y la confiabilidad) consideramos apropiado destacar ciertos aspectos relacionados con el tamaño muestral y los métodos empleados para el análisis estadístico. El tamaño muestral es un aspecto esencial en estudios que incluyen modelos de ecuaciones estructurales (MEE), aunque la bibliografía no establece una respuesta concluyente para determinar la cantidad de casos requeridos para un adecuado análisis. Por el contrario, la diversidad de criterios constituye una masa desarticulada de literatura que dificulta el trabajo del investigador^{33,46,47}. Ciertas características influyen para determinar el tamaño muestral, como la presencia de normalidad multivariada, el número de variables latentes, los indicadores y las comunalidades^{32,47,48} (se denomina comunalidad a la proporción de la varianza explicada por los factores comunes en una variable). Un estudio comparativo sobre distintos modelos de MEE, propone un mínimo de 200 observaciones, asegurando indicadores cuidadosamente elegidos y confiables, y un adecuado número de indicadores por variable latente⁴⁸. Tomando como base la fórmula propuesta por Westland⁴⁹ para nuestro estudio, en la que $n = 50r^2 - 450r + 1100$, donde $r =$ cociente indicadores/variables latentes, correspondería un mínimo de 100 observaciones.

Otros autores rechazan establecer reglas para calcular el tamaño muestral⁵⁰, sosteniendo que podrían resultar inválidas en el contexto del MEE. Independientemente de la proporción de sujetos por variable, Arias³⁴ propuso un mínimo de 150 observaciones, descartando aquellas con puntuaciones extremas. En tér-

minos 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). Para el AFE se propuso un mínimo de cinco sujetos por variable, con una base de al menos 100 observaciones; en caso de un tamaño menor, la relación debería ser 10/1⁵¹. Dado que el tamaño muestral para el AFC para algunos autores^{37,52,53} podría representar una limitación para este trabajo, decidimos realizar un AFE en primera instancia. Sin embargo, varios estudios indican que nuestro tamaño muestral puede resultar satisfactorio para la mayoría de los casos⁵⁴⁻⁵⁸, especialmente cuando hay varios indicadores por dimensión, los ítems no son binarios y las cargas factoriales son altas⁵⁹. Por lo tanto verificamos también el ajuste del modelo con AFC. En el trabajo original la base mínima fue de cinco sujetos por ítem⁶. Dado que la muestra total fue de 392 sujetos, los resultados fueron compatibles con la interpretación de las cargas del AFE³², y también el resultado del AFC reveló la cifra de 222,68 (método DWLS) y 226,20 (método ULS) como número crítico de observaciones, brindando mayor soporte a las conclusiones obtenidas.

El método de componentes principales y rotación ortogonal con Kaiser se recomienda para la reducción de datos y por ser el más utilizado en AFE³². El agrupamiento de ocho ítems en torno a un solo factor (con reducción del número de dominios, sin afectar la confiabilidad) fue atribuido por el comité de expertos al contenido conceptual de esos ítems: el aspecto educativo y el asistencial no constituyen entes separados en la residencia. Los mismos residentes se tornan docentes esenciales dentro de su ámbito⁵⁹⁻⁶¹ donde la educación ocurre en trabajo bajo supervisión para adquirir habilidades, hábitos de trabajo y búsqueda de información; el logro de los objetivos del programa de educación médica superior involucra directa o indirectamente profesores, residentes y el resto del colectivo laboral en los procesos docentes, de atención médica e investigación⁶². El ser competente resulta de un aprendizaje integrado, donde el residente afronta situaciones complejas que tengan sentido para él. Al resolverlas, aprende, y logra un desempeño competente impulsado por las situaciones: competencias y situaciones son lo mismo, puesto que la competencia es un conocimiento situado⁶². Para los docentes que planifican y supervisan, resulta natural asumir esa disociación. El pase de sala, el alta de 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^{61,62}. La tarea asistencial y la educación son dos caras de la misma moneda en el proceso de formación, expresadas en las preguntas de esos dominios, y en el contexto cultural de nuestro medio podría explicar el comportamiento factorial. 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 educación⁴. Tras un testeo del D-RECT sobre una muestra mayor de residentes, el número de ítems se redujo a 35 y el número de factores, a 9⁶³. 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 (mayores a 0,30) considerando el tamaño muestral del estudio³². Cada dominio mantuvo 3 ítems como mínimo apropiado para identificarse como factor³³.

Para el AFC, como las variables del estudio son ordinales, y la literatura considera para ese caso a los métodos robustos^{35,54-59}, el ajuste de los modelos de estructuras de covarianzas fue verificado con el contraste de X^2 corregido de Satorra-Bentler⁶⁴ divi-

dido por sus grados de libertad ($SB\chi^2/gl$), asumiendo ausencia de normalidad multivariada para variables ordinales y su comportamiento robusto en tales condiciones, además de la sensibilidad del χ^2 al tamaño de la muestra³⁴. En tal condición, ULS y DWLS como métodos robustos^{34,36,37} difieren levemente en sus matri-

ces de carga de datos y en el tamaño muestral requerido, pero ambos mostraron resultados similares. La misma conclusión puede alcanzarse con la Estrategia de Dos Índices⁶⁵ que propone combinar dos indicadores para demostrar la bondad de ajuste, con diferentes posibilidades (Tabla 7).

Tabla 7. Estrategia de Dos Índices. Fuente: Keszei AP, et al. J Psychosom Res . 2010 Apr;68(4):319-23.

Combinación de Índices	Regla de Combinación
Índice de ajuste no normalizado (NNFI) y Residual cuadrático medio estandarizado de la raíz (SRMR)	NNFI > 0,96 y SRMR < 0,09
Error cuadrático medio de aproximación (RMSEA) y Residual cuadrático medio estandarizado de la raíz (SRMR)	RMSEA < 0,06 y SRMR < 0,09
Índice de ajuste comparativo (CFI) y Residual cuadrático medio estandarizado de la raíz (SRMR)	CFI > 0,96 y SRMR < 0,09

Nota: Los autores propusieron combinar RMSEA y SRMR para muestras mayores de 250 sujetos, y los restantes para muestras menores de 250. Esta estrategia no es aceptada universalmente.

A partir de estos resultados pueden establecerse otras categorías de validez^{32,44}: la validez convergente (evidencia de una relación coherente entre medidas del mismo constructo) y la 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 (mayores de 0,60) y valores T (mayores de 1,96) en ambos métodos de estimación del AFC. Del mismo modo, la VME mayor de 0,5 para cada dominio indica similar conclusión³². Con respecto a la validez discriminante, esta se consideró aceptable, ya que las VME de los distintos dominios no son inferiores a la mayoría de los cuadrados de sus correlaciones con otros constructos.

Por otro lado, 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 observaciones⁶⁶: al no tomar en cuenta variaciones en el tiempo o interobservador, es una estimación optimista de la verdadera confiabilidad de la prueba, y su sensibilidad se ve afectada por la extensión del instrumento, tendiendo a valores elevados con más de 15 ítems, independientemente de su correlación. Por este motivo, con escalas más largas, deben combinarse otros índices. El α de Cronbach mostró resultados a favor de la confiabilidad para el test global, y para cada dominio en particular. Cabe remarcar que, en el trabajo original⁶, el α de tres dominios fue inferior a 0,70. En nuestro trabajo, todos los valores estuvieron por encima de 0,80. Otro dato relevante a favor de la confiabilidad es el valor de la fiabilidad compuesta (mayor de 0,70 para cada variable latente³²).

Adicionalmente, obtuvimos evidencia a favor de la confiabilidad del instrumento global y de los dominios individuales mediante el estudio G. En la teoría G, los resultados de una prueba son considerados como una muestra de un universo de observacio-

nes admisibles, y cada oportunidad de aplicación de la prueba posee condiciones propias (como por ej., el formulario utilizado, 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 esas facetas⁶⁶.

Entre las limitaciones de este estudio podemos destacar que el D-RECT 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. Más recientemente, el D-RECT fue revisado y validado en su país de origen por algunos de sus autores originales con modificaciones en su estructura⁷.

Como implicaciones educativas y sanitarias futuras de este trabajo, los autores consideramos que cabe esperar un impacto positivo en los programas de formación profesional luego de eventuales cambios institucionales basados en la medición del CA. Las residencias podrían verse beneficiadas al contar con este recurso, por la oportunidad de detectar oportunidades de mejora. Destacamos la importancia de que se desarrollen futuras investigaciones que eventualmente confirmen estos resultados, y otras sobre el impacto educativo y sanitario de su aplicación dentro de los programas de formación en distintas especialidades médicas.

Conclusiones

Se realizó la adaptación transcultural del D-RECT y se confirmó su adecuada validez y confiabilidad para su utilización en Argentina.

Fuentes de financiamiento y Conflicto de interés de los autores: Este trabajo no recibió financiamiento de fuente alguna. Se declara ausencia de conflictos de intereses en cualquier aspecto por parte de los investigadores.

Recibido el 02/11/2021 Aceptado el 31/03/2021
y Publicado el 28/04/2021.

Agradecimientos

- Al Dr. Fernando Ramón Vázquez, por su invaluable ayuda y consejo en instancias críticas de este trabajo.
- 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 las autoridades de los Hospitales públicos de Corrientes y Chaco que permitieron la recolección de los datos necesarios para el presente trabajo.
- A los colegas docentes que colaboraron en pasos decisivos de este trabajo.
- A los médicos residentes de los distintos hospitales participantes, destinatarios finales de esta labor, por su buena voluntad y disposición.
- A nuestras familias.

A los miembros del jurado de tesis de la Maestría de Educación para Profesionales de la Salud del Instituto Universitario Hospital Italiano de Buenos Aires, Dr. Roberto Cherkovsky, Dr. Marcelo García Diéguez y Mg. Marta del Valle.

A Sergio Ogueta Baza y Alfredo Eymann por sus aportes como revisores de pares de este manuscrito.

Referencias

1. Genn JM. AMEE Medical Education Guide No. 23 Part 2): Curriculum, environment, climate, quality and change in medical education - A unifying perspective. *Med Teach*. 2001;23(5):445–454. Available from: 10.1080/01421590120075661.
2. Bakhshialiabad H, Bakhshi M, Hassanshahi G. Students' perceptions of the academic learning environment in seven medical sciences courses based on DREEM. *Adv Med Educ Pract*. 2015;6:195–203. Available from: 10.2147/AMEP.S60570.
3. Butterfield PS. The stress of residency. A review of the literature. *Arch Intern Med*. 1988;148(6):1428–1435. Available from: 10.1001/archinte.148.6.1428;https://dx.doi.org/10.1001/archinte.148.6.1428.
4. Boor K, Scherpier A, Teunissen P, et al. The postgraduate learning climate: integration of work, training, and residents' needs. In: *The clinical learning climate*. Amsterdam; 2009. p. 65–78. Available from: http://scholar.google.com.ar/scholar_url?url=https://research.vumc.nl/ws/files/518234/8834.pdf&hl=es&sa=X&ei=g9-eX-uRK4_0mQHX4Y6YQC&scisig=AAGBfm2aUnT115PMCxPzxNZDjSGNOo7gQ&nossl=1&oi=scholar.
5. Gruppen LD, Stansfield RB, Zhao Z, et al. Institution and Specialty Contribute to Resident Satisfaction With Their Learning Environment and Workload. *Acad Med*. 2015;90(11 Supl):S77–S82. Available from: 10.1097/acm.0000000000000898;https://dx.doi.org/10.1097/acm.0000000000000898.
6. Boor K, Van-Der-Vleuten C, Teunissen P, et al. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach*. 2011;33(10):820–827. Available from: 10.3109/0142159x.2010.541533;https://dx.doi.org/10.3109/0142159x.2010.541533.
7. Silkens MEWM, Smirnova A, Stalmeijer RE, et al. Revisiting the D-RECT tool: Validation of an instrument measuring residents' learning climate perceptions. *Med Teach*. 2016;38(5):476–481. Available from: 10.3109/0142159x.2015.1060300;https://dx.doi.org/10.3109/0142159x.2015.1060300.
8. Borrell-Bentz RM. La educación médica de posgrado en Argentina: el desafío de una nueva práctica educativa. Buenos Aires: Organización Panamericana de la Salud; 2005. Available from: https://iris.paho.org/bitstream/handle/10665.2/6244/2005-ARG-educacion-medica-postgrado.pdf?sequence=4&isAllowed=y.
9. Hoff TJ, Pohl H, Bartfield J. Creating a Learning Environment to Produce Competent Residents: The Roles of Culture and Context. *Acad Med*. 2004;79(6):532–540. Available from: 10.1097/00001888-200406000-00007;https://dx.doi.org/10.1097/00001888-200406000-00007.
10. Cohen JS, Patten S. Well-being in residency training: a survey examining resident physician satisfaction both within and outside of residency training and mental health in Alberta. *BMC Med Educ*. 2005;5(1):21–21. Available from: 10.1186/1472-6920-5-21;https://dx.doi.org/10.1186/1472-6920-5-21.
11. Thomas NK. Resident burnout. *JAMA*. 2004;292(23):2880–2889. Available from: 10.1001/jama.292.23.2880.
12. Llera J, Durante E. Correlación entre el clima educacional y el síndrome de desgaste profesional en los programas de residencia de un hospital universitario. *Arch Argent Pediatr*. 2014;112(1):6–11. Available from: 10.5546/aap.2014.6.
13. West CP, Shanafelt TD. The influence of personal and environmental factors on professionalism in medical education. *BMC Med Educ*. 2007;7(1):29–29. Available from: 10.1186/1472-6920-7-29;https://dx.doi.org/10.1186/1472-6920-7-29.
14. Lombarts KMJM, Heineman MJ, Scherpier AJJA, et al. Effect of the Learning Climate of Residency Programs on Faculty's Teaching Performance as Evaluated by Residents. *PLoS ONE*. 2014;9(1):e86512–e86512. Available from: 10.1371/journal.pone.0086512;https://dx.doi.org/10.1371/journal.pone.0086512.
15. Mook WV, Stalmeijer R, Muijtens A, et al. A preliminary study on the quality of the intensive care medicine training programmes in the Netherlands. *Neth J Crit Care*. 2016;24(7):10–15. Available from: https://nvc.nl/node/2136.
16. Hutchinson L. Educational environment. *BMJ*. 2003;326(7393):810–812. Available from: 10.1136/bmj.326.7393.810.
17. Riquelme-Pérez A, Fuentes G, Jeria A, et al. Ambiente educacional y calidad de la docencia en la escuela de medicina. *ARS MEDICA Revista de Ciencias Médicas*. 2007;36(2):95–95. Available from: 10.11565/arsmed.v36i2.149;https://dx.doi.org/10.11565/arsmed.v36i2.149.
18. Schönrock-Adema J, Bouwkamp-Timmer T, van Hell EA, et al. Key elements in assessing the educational environment: where is the theory? *Adv Health Sci Educ Theory Pract*. 2012;17(5):727–742. Available from: 10.1007/s10459-011-9346-8;https://dx.doi.org/10.1007/s10459-011-9346-8.
19. Alshomrani A, AlHadi A. Learning environment of the Saudi psychiatry board training program. *Saudi Med J*. 2017;38(6):629–635. Available from: 10.15537/smj.2017.6.18164;https://dx.doi.org/10.15537/smj.2017.6.18164.
20. Iblher P, Zupanic M, Ostermann T. The Questionnaire D-RECT German: Adaptation and testtheoretical properties of an instrument for evaluation of the learning climate in medical specialist training. *GMS Z Med Ausbild*. 2015;32(5):55–55. Available from: 10.3205/zma000997.
21. Pacifico JL, van-der Vleuten CPM, Muijtens AMM, et al. Cross-validation of a learning climate instrument in a non-western postgraduate clinical education. *BMC Med Educ*. 2018;18(1):22–22. Available from: 10.1186/s12909-018-1127-0;https://dx.doi.org/10.1186/s12909-018-1127-0.
22. Dominguez LC, Silkens M, Sanabria A. The Dutch residency educational climate test: construct and concurrent validation in Spanish language. *Int J Med Educ*. 2019;10:138–148. Available from: 10.5116/ijme.5d0c.bff7;https://dx.doi.org/10.5116/ijme.5d0c.bff7.
23. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 1998;25(24):3186–91. Available from: 10.1097/00007632-200012150-00014.
24. A F. Adaptación de Tests a Otras Culturas. In: Tornimbeni S, Perez E, Olaz F, editors. *Introducción a la Psicometría*. Buenos Aires: Paidós; 2008. p. 191–205.
25. Arribas A. Adaptación transcultural de Instrumentos. Guía para el proceso de Validación de instrumentos tipo encuestas. *Rev Asoc Med Bahía Blanca*. 2006;16(3):74–82. Available from: https://pesquisa.bvsalud.org/portal/resource/pt/biblio-1007566.
26. Muñoz J, Elosua P, Hambleton RK, et al. Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*. 2013;25(2):151–157. Available from: 10.7334/psicothema2013.24.
27. Hambleton RK. Issues, Designs and Technical Guidelines for Adapting Tests Into Multiple Languages and Cultures. In: Hambleton RK, Merenda PF, Spielberger CD, editors. *Adapting educational and psychological tests for crosscultural assessment*. Mahwah, NJ: Lawrence Erlbaum Associates; 2005. p. 3–38.
28. ITC Guidelines for Translating and Adapting Tests (Second Edition). *International Journal of Testing*. 2018;18(2):101–134. Available from: 10.1080/15305058.2017.1398166.
29. Padilla JL, Benítez I. Validity evidence based on response processes. *Psicothema*. 2014;26(1):136–180. Available from: 10.7334/psicothema2013.259;http://www.psicothema.com/pdf/4171.pdf.
30. Abad F, Garrido J, Olea J, et al. Introducción a la Psicometría, Teoría Clásica de los Tests y Teoría de la Respuesta al Ítem. Madrid: Universidad Autónoma de Madrid. Facultad de Psicología; 2006. Available from: http://aprendeenlinea.udea.edu.co/lms/investigacion/file.php/39/ARCHIVOS_2010/PDF/IntPsicometria_aristidesvara_1_.pdf.
31. Cervantes V. Interpretaciones del coeficiente alpha de Cronbach. *Avances en Medición*. 2005;3(1):9–28. Available from: http://www.humanas.unal.edu.co/psicometria/files/1113/8574/8604/Articulo_1_Alfa_de_Cronbach_9-28_2.pdf.
32. Hair JF, Black WC, Babin B, et al., editors. *Multivariate Data Analysis*. 7th ed. Pearson Prentice Hall S; 2010.
33. Morales-Vallejo P. El análisis factorial en la construcción e interpretación de tests, escalas y cuestionarios. Madrid: Universidad Pontificia Comillas. Facultad de Ciencias Humanas y Sociales; 2010. Available from: http://www.upcomillas.es/personal/peter/investigacion/AnalisisFactorial.pdf.

34. Arias-Martínez B. Metodología en la investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales. VI Seminario Científico SAID. Verdugo MA, Crespo MA, Badía M, et al., editors. Colección Actas. Salamanca: Publicaciones del INICO; 2008. Available from: https://sid.usal.es/idos/F8/FDO20749/said_2008.pdf#page=76.

35. Satorra A, Bentler PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye A, Clogg C, editors. Latent variables analysis: Applications for developmental research. 1994 (pp. 399-419). Thousand Oaks, CA: Sage. Thousand Oaks, CA: Sage; 1994. p. 399-419.

36. Morata-Ramírez MA, Holgado-Tello FP, Barbero-García MI, et al. Análisis factorial confirmatorio. Recomendaciones sobre mínimos cuadrados no ponderados en función del error Tipo I de Ji-Cuadrado y RMSEA [Confirmatory factor analysis. Recommendations for unweighted least squares method related to Chi-Square and RMSEA]. *Acción Psicológica*. 2015;12:79-79. Available from: 10.5944/ap.12.1.14362;<https://dx.doi.org/10.5944/ap.12.1.14362>.

37. Li C. The performance of MLR, USLMV, and WLSMV estimation in structural regression models with ordinal variables; 2014. Available from: https://d.lib.msu.edu/islandora/object/etd:3268/datastream/OBJ/download/The_performance_of_MLR_USLMV_and_WLSMV_estimation_in_structural_regression_models_with_ordinal_variables.pdf.

38. Fornell C, Larcker D. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*. 1981;18(1):39-50. Available from: 10.1177/002224378101800104;<https://doi.org/10.1177/002224378101800104>.

39. Bloch R, Norman G. Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Med Teach*. 2012;34(11):960-992. Available from: 10.3109/0142159x.2012.703791;<https://dx.doi.org/10.3109/0142159x.2012.703791>.

40. Díaz C, Batanero C, Cobo B. Fiabilidad y generalizabilidad. Aplicaciones en evaluación educativa. *Revista Números*. 2003;p. 3-21. Available from: http://scholar.google.com.ar/scholar_url?url=http://www.academia.edu/download/46762960/fiabilidad.pdf&hl=es&sa=X&ei=DB6fX9LFENzLsQLW6ZfWDA&scisig=AAGBfm1nTLQmo-3NB5C4bU532qBbl4yobw&nossl=1&oi=scholarr.

41. Malda M, Van-De-Vijver FJR, Srinivasan K, et al. Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*. 2008;50(4):451-468. Available from: http://journaldatabase.info/articles/adapting_cognitive_test_for_different.html.

42. Messick S. Validity. In: Linn RL, editor. Educational measurement. New York, NY: Macmillan; 1989. p. 13-103.

43. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing. Washington, D.C.: American Educational Research Association; 2014.

44. Messick S. Test validity and the ethics of assessment. ETS Research Report Series. 1979;p. i-43. Available from: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2333-8504.1979.tb01178.x>.

45. Cook DA, Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *Am J Med*. 2006;119(2):166.e7-166.e16. Available from: 10.1016/j.amjmed.2005.10.036;<https://dx.doi.org/10.1016/j.amjmed.2005.10.036>.

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

47. Jackson DL. Revisiting Sample Size and Number of Parameter Estimates: Some Support for the N:q Hypothesis. *Structural Equation Modeling*. 2003;10(1):128-141. Available from: https://doi.org/10.1207/S15328007SEM1001_6.

48. Bentler P. EQS 6 Structural Equations Program Manual. Encino, California USA: Multivariate Software, Inc; 1989. Available from: <http://www.econ.upf.edu/~satorra/CourseSEMVienna2010/EQSManual.pdf>.

49. MacCallum RC, Austin JT. Applications of Structural Equation Modeling in Psychological Research. *Annu Rev Psychol*. 2000;51(1):201-226. Available from: 10.1146/annurev.psych.51.1.201;<https://dx.doi.org/10.1146/annurev.psych.51.1.201>.

50. Streiner DL. Figuring Out Factors: The Use and Misuse of Factor Analysis. *The Canadian Journal of Psychiatry*. 1994;39(3):135-140. Available from: 10.1177/070674379403900303;<https://dx.doi.org/10.1177/070674379403900303>.

51. Jöreskog KG, Sörbom D. PRELIS 2 User's Reference Guide: A Program for Multivariate Data Screening and Data Summarization: a Preprocessor for LISREL. Scientific Software International; 1996.

52. Ferrando PJ, Anguiano-Carrasco C. El análisis factorial como técnica de investigación en Psicología. *Papeles del Psicólogo*. 2010;31(1):18-33. Available from: <https://www.redalyc.org/pdf/778/77812441003.pdf>.

53. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. 3rd ed. Canada, Ontario: B.C. Decker Inc; 2008.

54. DiStefano C, Morgan GB. A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data. *Structural Equation Modeling*. 2014;21(3):425-438. Available from: 10.1080/10705511.2014.915373;<https://dx.doi.org/10.1080/10705511.2014.915373>.

55. Rhemtulla M, Brosseau-Liard PE, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods*. 2012;17(3):354-373. Available from: 10.1037/a0029315;<https://dx.doi.org/10.1037/a0029315>.

56. Savalei V, Rhemtulla M. The performance of robust test statistics with categorical data. *Br J Math Stat Psychol*. 2013;66(2):201-223. Available from: 10.1111/j.2044-8317.2012.02049.x;<https://dx.doi.org/10.1111/j.2044-8317.2012.02049.x>.

57. Yang-Wallentin F, Joreskog K, Luo H. Confirmatory Factor Analysis of Ordinal Variables With Misspecified Models. *Structural Equation Modeling*. 2010;17(3):392-423. Available from: 10.1080/10705511.2010.489003;<https://dx.doi.org/10.1080/10705511.2010.489003>.

58. Forero CG, Maydeu-Olivares A, Gallardo-Pujol D. Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling*. 2009;16(4):625-641. Available from: 10.1080/10705510903203573;<https://dx.doi.org/10.1080/10705510903203573>.

59. Méndez-López JF, Mendoza-Espinosa H, Torruco-García U, et al. El médico residente como educador. *Inv Ed Med*. 2016;2(7):154-161. Available from: 10.1016/S2007-5057(13)72705-4.

60. Carballido R, Ayala FP, Amarilla A. Estado Actual de las residencias de Cardiología. Encuesta Nacional de Residentes. *Rev Argent Cardiol*. 1997;66(3):269-276. Available from: <https://www.redalyc.org/articulo.oa?id=305326931014>.

61. Rodríguez-Fernández Z, Rizo-Rodríguez R, Fariñas AM, et al. Educación en el trabajo en la enseñanza médica superior. *MEDISAN*. 2017;21(7). Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192017000700017.

62. de Gauna PR, González-Moro V, Morán-Barrios J. Diez claves pedagógicas para promover buenas prácticas en la formación médica basada en competencias en el grado y en la especialización. *Educación Médica*. 2015;16(1):34-42. Available from: 10.1016/j.edumed.2015.04.005;<https://dx.doi.org/10.1016/j.edumed.2015.04.005>.

63. Cupani M. Análisis de Ecuaciones Estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación. Tesis. 2012;2(1):186-199. Available from: <https://revistas.unc.edu.ar/index.php/tesis/article/view/2884>.

64. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999;6(1):1-55. Available from: 10.1080/10705519909540118;<https://dx.doi.org/10.1080/10705519909540118>.

65. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. *J Psychosom Res*. 2010;68(4):319-323. Available from: 10.1016/j.jpsychores.2010.01.006;<https://dx.doi.org/10.1016/j.jpsychores.2010.01.006>.

66. Shavelson RJ, Webb NM, Rowley GL. Generalizability theory. *American Psychologist*. 1989;44(6):922-932. Available from: 10.1037/0003-066x.44.6.922;<https://dx.doi.org/10.1037/0003-066x.44.6.922>.