Validación transcultural del Cuestionario de Enseñanza Clínica de Maastricht

Transcultural validation of the Maastricht Clinical Teaching Questionnaire

Dr. Sergio E. Giannasi^a, Dr. Eduardo Durante^a, Dr. Fernando R. Vázquez^a,
Dra. Claudia E. Kecskes^a, Dra. Roberta Ladenheim^b y Dr. Carlos A. Brailovsky^c

RESUMEN

La evaluación del desempeño del docente clínico permite proporcionar *feedback* para motivarlo a mejorar sus habilidades docentes.

Se realizó la validación transcultural del Cuestionario de Enseñanza Clínica de Maastricht, mediante la guía internacional para la adaptación de los cuestionarios. Se investigó validez de contenido, proceso de respuesta, constructo y confiabilidad. Luego de la adaptación, residentes de dos hospitales universitarios evaluaron a 187 docentes clínicos.

La validación del contenido y del proceso de respuesta fue adecuada. En el análisis factorial confirmatorio, todos los índices y criterios se ajustaron bien con el modelo de 5 factores y 14 ítems. El alfa de Cronbach fue de 0,80. El coeficiente G fue > 0,70, con bajos valores de varianza del error absoluto. Fueron necesarias 6 evaluaciones por docente para una evaluación confiable en cada dominio y en el desempeño global.

La versión en castellano del Cuestionario es válida y confiable para evaluar a los docentes.

Palabras clave: educación médica, adaptación transcultural, reproducibilidad de los resultados, encuestas y cuestionarios.

ABSTRACT

The evaluation of the clinical teacher's performance provides feedback to motivate them to improve their teaching skills.

To perform the cross-cultural adaptation of the Maastricht Clinical Teaching Questionnaire, the International Guide for the Adaptation of the Questionnaires was followed. The validity of content, response process, construct and reliability were investigated. After cross-cultural adaption, residents of two University hospitals evaluated 187 clinical teachers.

Content and answering process were validated. In the confirmatory factor analysis, all indexes and criteria for a good fit suited the 5 factors and 14 items model. The Cronbach's alpha coefficient was 0.80. The G coefficient was > 0.70, with low variance of the absolute error. Every clinical teacher should receive at least 6 evaluations to achieve a reliable evaluation of every domain and of their global performance.

The Spanish version of the questionnaire is a valid, reliable instrument for medical residents to evaluate teachers.

Key words: medical education, cross-cultural adaptation, reproducibility of results, surveys and questionnaires.

<http://dx.doi.org/10.5546/aap.2019.e425>

Cómo citar: Giannasi SE, Durante E, Vázquez F, Kecskes CE et al. Validación transcultural del Cuestionario de Enseñanza Clínica de Maastricht. *Arch Argent Pediatr* 2019;117(4):e425-e432.

INTRODUCCIÓN

La formación de los docentes clínicos (DC) es de larga data, pero se desarrolló más plenamente a partir de la década de los años ochenta y noventa. La raíz de este movimiento fue "enseñar a enseñar", y su foco estuvo, más bien, en la didáctica (enseñanza con pequeños y grandes grupos, observación y *feedback*, enseñanza en distintos lugares de práctica, entre otros). Sin embargo, la mayoría de los programas de formación en enseñanza clínica no han tenido un marco teórico que los fundamente ni evaluaciones adecuadas del desempeño docente.¹

La teoría del aprendizaje cognitivo desarrollada por Collins y col., en 1989² incluye conceptos relacionados con la educación y el aprendizaje de los adultos que pueden aplicarse a la enseñanza clínica.^{3,4} Este modelo propone seis métodos de enseñanza: "modelado, *coaching* y andamiaje" son el corazón del modelo y facilitan la adquisición de habilidades cognitivas y metacognitivas; "articulación y reflexión" son estrategias que los profesores utilizan para exteriorizar los procesos tácitos que subyacen a su pensamiento y a sus acciones en la práctica, y para mostrar las estrategias que usan para resolver los problemas.⁵ Finalmente, mediante

a. Hospital Italiano de Buenos Aires.

b. Centro de Educación Médica e Investigaciones Clínicas.

Ciudad Autónoma de Buenos Aires, Argentina.

c. Universidad de Laval, Quebec, Canadá.

Correspondencia:

Dr. Sergio E. Giannasi:
sergio.giannasi@hospitalitaliano.org.ar

Financiamiento:

Ninguno.

Conflicto de intereses:

Ninguno que declarar.

Recibido: 24-7-2018

Aceptado: 20-3-2019

la “exploración”, el docente fomenta que los estudiantes formulen y persigan sus propias metas de aprendizaje.⁶

Además, la evaluación de la calidad docente permite proporcionar *feedback* a los profesores para motivarlos a mejorar sus habilidades de enseñanza.^{7,8} Un *feedback* adecuado requiere que la evaluación sea realizada con instrumentos válidos, confiables⁹ y que cuenten con evidencias de validez de múltiples fuentes.^{10,11}

El Cuestionario de Enseñanza Clínica de Maastricht (*Maastricht Clinical Teaching Questionnaire*, MCTQ) fue construido sobre la base de la teoría del aprendizaje cognitivo y ha mostrado ser válido y confiable para evaluar el desempeño de los DC.¹² Los ítems utilizados en la versión final del instrumento se validaron en 4 de los 6 métodos descritos originalmente: modelado, *coaching*, articulación y exploración, a los que se le sumó la evaluación de “clima de aprendizaje seguro”, fundamental en todo proceso educativo.¹³ La validez y confiabilidad del MCTQ fue probada con estudiantes de Medicina durante sus rotaciones clínicas, en dos hospitales universitarios de Holanda.¹³ Además, demostró ser válido y confiable para evaluar a DC de estudiantes de Veterinaria durante rotaciones cortas de hasta un día.¹⁴

El MCTQ también fue utilizado para evaluar el desempeño de los DC en la residencia de Terapia Intensiva en el Hospital Universitario de Maastricht,¹⁵ de Psiquiatría en Nueva Zelanda¹⁶ y de todas las residencias en la Universidad Javeriana, en Bogotá.¹⁷

Considerando el excelente desempeño del MCTQ y la falta de instrumentos de evaluación del DC en idioma español, el objetivo de la investigación fue realizar su validación transcultural.

MATERIAL Y MÉTODO

Se realizó la adaptación del MCTQ siguiendo las directrices de la Comisión Internacional para la Adaptación de Test.¹⁸ Luego de realizar la traducción al español, se exploró la evidencia de validez de cuatro de las cinco fuentes descritas en los *Estándares para pruebas educativas y psicológicas*:¹⁹ 1ª De contenido mediante un panel de expertos. 2ª Del proceso de respuesta, mediante una prueba piloto. 3ª La estructura interna del test, validez de constructo y confiabilidad, mediante un análisis factorial confirmatorio (AFC), determinación de la validez convergente y discriminante, la varianza media extractada

(VME) y la confiabilidad compuesta. Además, se la evaluó con el coeficiente alfa de Cronbach. Para la evaluación de diferentes aspectos de la varianza total, se realizó un estudio de generalizabilidad (estudio G). 4ª Las relaciones con otras variables se midieron mediante un modelo de ecuaciones estructurales (MEE), la relación entre el puntaje recibido por los DC en cada dominio y el juicio general.

Las consecuencias de la aplicación del instrumento (quinta fuente de validez) no se incluyeron como objetivo en este estudio.

Marco, observadores y docentes clínicos

Médicos residentes de dos hospitales universitarios de la Ciudad Autónoma de Buenos Aires (CABA), Argentina, participaron como observadores del desempeño de los DC después de completar 4 semanas de rotación clínica. La participación de los residentes y de los DC fue voluntaria y anónima; en todos los casos, se explicaron los objetivos y alcances de la investigación.

Para el cálculo del N de observaciones, se tuvo en cuenta la necesidad impuesta por el AFC y la existencia o no de normalidad multivariada de los datos.²⁰ Dado que un análisis de una muestra parcial reveló que los datos no presentaban normalidad multivariada, se utilizó el método de análisis de *robust maximum likelihood*,²¹ y quedó establecido el N de DC para evaluar en mayor de 120.

Pruebas estadísticas

Para el AFC y el MEE, se utilizaron los siguientes indicadores de buen ajuste: valor de p del estadístico χ^2 ajustado, χ^2 /grados de libertad, *root mean square error of approximation* (RMSEA), *normed fit index* (NFI), *non normed fit index* (NNFI), *comparative fit index* (CFI), *root mean square residual* (RMR), *goodness of fit index* (GFI), *adjusted goodness of fit index* (AGFI). Se utilizó el programa Lisrel 9.10.²² Se evaluó la validez convergente y discriminante. Para estimar la confiabilidad de cada dominio, se calculó la VME y la confiabilidad compuesta. Se calculó el coeficiente alfa de Cronbach por tratarse de ítems no dicotómicos, y se lo consideró como significativo si era mayor de 0,80. Mediante un MEE, se estableció la relación entre los factores latentes y el rendimiento general, y se compararon los resultados con los del estudio de Stalmeijer y col.¹³ Para determinar si los resultados de la evaluación de los DC por residentes que aplicaban

el MCTQ podrían generalizarse, se realizó un estudio G.²³ La faceta de diferenciación fueron los residentes observadores (O) anidados en los DC (O : DC) y la faceta de generalización, los ítems (I). Para el estudio G, se utilizó el *software* EduG 6.1 y se consideró un coeficiente G satisfactorio igual a 0,70 o mayor. Finalmente, mediante un proceso de optimización, se determinó la menor cantidad de evaluaciones necesarias para evaluar a un DC en cada una de las actividades docentes (modelado, *coaching*, articulación, exploración y establecer un clima seguro) y para el juicio global.

El protocolo fue aprobado por el Comité de Ética de Investigación del Hospital Italiano de Buenos Aires, N° 1776.

RESULTADOS

Dos traductores, nativos de habla hispana, independientes de la investigación, realizaron una traducción individual del MCTQ del inglés al español. Un panel de 7 expertos en educación médica produjo un documento único en español, que representó de la mejor manera posible el constructo de la teoría del aprendizaje cognitivo y la pertinencia de los ítems a cada uno de los modos de enseñanza. Otros dos traductores de habla nativa inglesa, independientes de la investigación, realizaron la traducción inversa al inglés. Fueron analizadas por Renee Stalmeijer (autora del estudio original), quien eligió la versión más adecuada y sugirió algunas correcciones. Ese documento fue la base del primer cuestionario definitivo en español (*Anexo*).

Aporte de evidencia de validez y confiabilidad del Cuestionario de Enseñanza Clínica de Maastricht en español

1. Aporte de evidencia para la validez de contenido: los expertos en educación médica constataron la pertinencia de los ítems en español con los dominios y de estos con el constructo. Consideraron que la estructura gramatical era correcta y que los 14 ítems correspondientes a los 4 modos de enseñanza de la teoría del aprendizaje cognitivo y al clima del aprendizaje eran claros y adecuados para la nueva población.
2. Aporte de evidencia del proceso de respuesta: mediante la técnica de sondeo,⁷ se aplicó el MCTQ a un grupo de residentes de los hospitales intervinientes que no participarían de la investigación final. La comprensión del instrumento y la forma de respuesta fue adecuada y no requirió cambios.

3. Aporte de evidencia de la estructura interna del test y confiabilidad

- Recopilación de datos. Entre febrero de 2013 y mayo de 2014, se recibió un total de 1078 evaluaciones correspondientes a 187 DC de los dos hospitales intervinientes. Luego de eliminar las respuestas incompletas y las correspondientes a DC con menos de 4 evaluaciones, la muestra final incluyó a 151 DC.
- Validez de constructo. En el AFC, se probaron 3 modelos con 1, 3 o 5 factores latentes (*Tabla 1*). El modelo de cinco factores mostró el mejor ajuste. Para asegurar la validez del constructo y la estabilidad del modelo, se aplicó un AFC en dos muestras aleatorias de 120 evaluaciones cada una. En ambos casos, los índices mostraron un aceptable ajuste a los datos.
- La validez convergente fue garantizada, ya que las cargas factoriales estandarizadas en el modelo de 5 factores latentes fueron elevadas ($> 0,7$) y significativas (con valores $t > 1,96$).¹⁸ La validez discriminante también se confirmó fuertemente, ya que la VME (en negrita, en la *Tabla 2*) superó el cuadrado de la correlación de cada uno de esos dominios con el resto.²⁰
- El índice de ajuste para el MEE también fue adecuado (véase la *Tabla 1*). Las pruebas en el MEE, con los cinco factores latentes, produjeron un resultado adecuado (consúltese la 7ª línea en la *Tabla 1*). Los “coeficientes *path*” mostraron fuertes analogías con los resultados de la investigación original, con algunas diferencias (véase la *Figura 1*). En nuestro caso, “*coaching*” estuvo menos influenciado por “modelado” y más por “clima de aprendizaje seguro”. Además, “*coaching*” tuvo una mayor influencia en “*articulación*” y menos en “*exploración*”. Finalmente, “*articulación*” tuvo una incidencia aún mayor en “*exploración*” que en el estudio original.
- La confiabilidad de cada dominio fue adecuada, ya que la confiabilidad compuesta y la VME excedieron el 0,7 y el 0,5, respectivamente (véase la *Tabla 2*).²⁰
- Los resultados del coeficiente alfa de Cronbach fueron adecuados en todos los dominios (véase la *Tabla 2*).
- Para evaluar las fuentes de varianza, se diseñó un estudio G con tres modelos

TABLA 1. Los 3 modelos de índices de ajuste del análisis factorial confirmatorio con 1, 3 y 5 factores latentes y las dos muestras de 120 de las 151 evaluaciones. Además, en la 7^{ma} línea, están los valores de los índices de ajuste del modelo de ecuaciones estructurales. Método de estimación en todos los casos: robust maximum likelihood

Distintos modelos	Chi ² /g de libertad ^a (< 2)	Valor de p ajustado ^b (> 0,05)(*)	RMSEA ^c (< 0,08 y no mayor de 0,1) ^c	NFI (> 0,90)	NNFI ^e (> 0,90)	CFI ^f (> 0,90)	RMR st. ^g (< 0,08)	GFI ^h (> 0,90)
1 FL	241,32 / 74 = 3,26	0,0000	0,152	0,96	0,96	0,97	0,049	0,74
3 FL	153,22 / 71 = 2,15	0,0028	0,116	0,98	0,98	0,99	0,043	0,83
5 FL(**)	95,24 / 64 = 1,49	0,0642	0,079	0,98	0,99	0,99	0,03	0,902
Primeros 120 al azar (5 FL)	98,97 / 64 = 1,55	0,0582	0,09	0,98	0,99	0,99	0,04	0,88
Segundos 120 al azar (5 FL)	97,56 / 64 = 1,52	0,0632	0,09	0,98	0,99	0,99	0,03	0,88
MEE(**)	118,13 / 75 = 1,57	0,0303	0,08	0,98	0,99	0,99	0,0398	0,885

Encabezando la *Tabla 1*, está abreviado cada índice, y se especifican los valores que debería tener para un buen ajuste. FL: factor latente. ^a Chi² sobre grados de libertad. ^b Valor de p del estadístico chi² ajustado (Satorra-Bentler). ^c RMSEA: error de aproximación cuadrático medio. ^d NFI: índice de ajuste normativo. ^e NNFI: índice ajustado no normado. ^f CFI: índice de ajuste comparado. ^g RMR st.: residuo cuadrático medio estandarizado. ^h GFI: índice de la bondad del ajuste.

(*) Con una muestra de tamaño menor de 250 y teniendo entre 12 y 30 indicadores, como es este caso, la p puede ser significativa y tener un buen ajuste [18].

(**) Todos los errores de medida presentan valores de t mayores de 1,96, como recomienda la literatura. También se aclara que se aceptaron las sugerencias del programa Lisrel acerca de correlacionar los errores de medida de las variables (o ítems) 9 y 10, 9 y 11, y 14 y 16, por considerar que, probablemente, se repetirán en futuros estudios, ya que son razonables desde el punto de vista conceptual (véase el apéndice. Cuestionario definitivo usado para la validación).

FIGURA 1. Relaciones causales entre los diferentes modos de enseñanza (factores latentes) del Cuestionario de Enseñanza Clínica de Maastricht y la evaluación global

Los valores adjuntos a las líneas que relacionan los diferentes factores corresponden a los coeficientes de regresión estandarizados; los valores que se encuentran recuadrados corresponden al coeficiente del trabajo local, y el otro, al del trabajo original. El esquema o imagen se divide con línea de puntos, de la siguiente forma: "facilitado por el docente", con los dominios "modelado" y "clima de aprendizaje seguro"; "interacción entre el docente y el alumno", con el dominio "coaching", y "estimulación de la autorregulación del alumno", con los dominios "articulación" y "exploración".

El coeficiente de regresión estándar es significativo al nivel 0,01 (también indicado por líneas oscuras). La única diferencia es que, en el estudio local, hay una correlación significativa entre clima de aprendizaje y coaching, algo que no se evidenció en el trabajo original. Si bien la correlación entre articulación y evaluación global es significativa en ambas investigaciones, en la investigación local, esa correlación es pobre.

diferentes de análisis del instrumento siguiendo el plan O : DC /I.

- Análisis del instrumento, que incluyó el juicio global (15 ítems): en la *Tabla 3*, se representan los componentes de la varianza y el error absoluto para cada faceta. La varianza DC fue del 13,5 % y la de observadores anidados en los DC fue del 10,9 %. El coeficiente G absoluto fue 0,83.
- Análisis del instrumento sin el juicio global (14 ítems): el coeficiente absoluto G para los 14 ítems fue de 0,92.
- Análisis de cada dominio: el coeficiente absoluto G para cada dominio independiente fue superior a 0,7 en todos los casos.

Para un G de, al menos, 0,70, se necesitarían seis o más respuestas para evaluar la competencia de los DC en exploración, cinco para articulación y establecer un ambiente de aprendizaje seguro, y cuatro para modelado, *coaching* y obtener un juicio del desempeño global.

4. Evidencia relacionada con la relación con otras variables: todas las correlaciones entre los factores latentes y el juicio global fueron altas y estadísticamente significativas (véase la *Tabla 4*).

DISCUSIÓN

La adaptación transcultural de un test exige no solo una correcta traducción, sino también una adaptación y validación adecuadas. En la traducción, se buscó mantener las equivalencias lingüísticas, considerando las más importantes la conceptual (significado y contenido), la técnica (sintaxis y gramática) y la lectura y comprensión.^{24,25} Luego, se buscó evitar los sesgos más frecuentes que aparecían durante la adaptación.

El sesgo de muestreo se produce cuando las características socioculturales o de educación de los que responden el test son distintas y hacen que los ítems sean interpretados de modo diferente a como ocurrió en la investigación original.²⁶

Se consideró que, en ambos estudios, como los observadores de los DC eran médicos residentes (en nuestro caso) o estudiantes de Medicina (en el estudio original), la capacidad de comprensión de las preguntas era similar. Respecto a los DC evaluados y sus características, tanto de Holanda como de la Argentina, cumplieron tareas clínicas y docentes similares, y, en ambos casos, el nivel de formación docente era heterogéneo.²⁷

TABLA 2. Validez discriminante, convergente y confiabilidad

	Modelado	Coaching	Articulación	Exploración	Clima de aprendizaje seguro	Fiabilidad compuesta	Alfa de Cronbach
Modelado	0,781(*)	0,617(a)	0,558(a)	0,585(a)	0,576(a)	0,92	0,91
Coaching	0,617(a)	0,713(*)	0,654(a)	0,649(a)	0,664(a)	0,88	0,88
Articulación	0,558(a)	0,654(a)	0,731(*)	0,729(a)	0,466(a)	0,89	0,91
Exploración	0,585(a)	0,649(a)	0,729(a)	0,893(*)	0,465(a)	0,95	0,95
Clima de aprendizaje seguro	0,576(a)	0,664(a)	0,466(a)	0,465(a)	0,716(*)	0,87	0,90

Los valores de esta tabla corresponden al modelo de 5 factores latentes (véase la *Tabla 1*).

(*) Valores de VME por dominio.

(a) Cuadrado de las correlaciones entre los dominios.

TABLA 3. Resultados del estudio G. 15 ítems

Facetas	% de la varianza total	% de contribución al error absoluto (I + DI + O : DI)
DC	13,5	
O : DC	10,9	
Ítems	32,1	42,4
DC_I	4,1	5,4
O : DC_I	39,5	52,2

D = docentes clínicos; O : D = residentes observadores anidados en los docentes; I = ítems;

DC_I = interacción entre los docentes y los ítems; O : DC_I = interacción entre los observadores anidados en los docentes y los ítems.

El sesgo de muestreo también puede aparecer por un uso inadecuado del instrumento.²⁸ Mediante una prueba piloto en un grupo de residentes, se comprobó que todos los ítems se consideraban fáciles de entender y el proceso de respuesta era simple. Además, el proceso de implementación final del MCTQ fue el mismo que en el estudio original; se mantuvo la confidencialidad y la participación fue voluntaria.

El sesgo del instrumento aparece cuando las características del test, a causa de las diferencias transculturales, no refleja del mismo modo el constructo por evaluar.²⁶ Se evitó este sesgo mediante la validación del contenido realizado por un panel de expertos en enseñanza clínica y se indicó que todos los elementos incluidos en la versión final del MCTQ eran relevantes para la población que se evaluaba.

Después de completar la adaptación transcultural, se verificaron las propiedades psicométricas del test y se proporcionó evidencia de validez¹⁰ de cuatro de las cinco fuentes propuestas en *Estándares para pruebas educativas y psicológicas*.¹⁹ Se encontró que la versión en español del cuestionario era adecuada para aplicar en nuestra población. La validez de constructo del MCTQ se estableció mediante el AFC, siguiendo el modelo de cinco factores propuesto por Stalmeijer et al.¹³ La validez convergente y discriminatoria también se confirmó en nuestro análisis. La confiabilidad se calculó utilizando el VME, la confiabilidad compuesta y el coeficiente alfa de Cronbach para la consistencia interna, con resultados muy alentadores.

La interacción entre los factores latentes y el juicio global, evaluada con un MEE, fue muy

similar a la reportada en el estudio original, con pequeñas variaciones. En nuestro medio, “coaching” tuvo una mayor influencia sobre “articulación” y menos sobre “exploración”. Además, “articulación” tuvo una incidencia aún mayor en “exploración” que en el estudio original. Este hecho podría explicarse por el mayor nivel de competencia de los residentes en comparación con los estudiantes y su necesidad de traducirlo en la práctica (es decir, cuidado del paciente). Esta es solo una hipótesis que puede valer la pena explorar en futuros estudios. Se encontró una relación adecuada entre cada factor latente y el juicio global sobre el desempeño del docente.

Finalmente, el estudio G demostró que la evaluación del DC realizada por los observadores residentes era generalizable, ya que los coeficientes G para el juicio general eran 0,83. Las fuentes de varianza para el error absoluto, para el DC y para los residentes fueron bajas. Se determinó que el 42,4 % de la varianza del error se relacionó con los ítems y el 52,2 %, con la interacción de los estudiantes anidados en los docentes y los ítems; la interacción de los DC con los ítems agregó solo el 5,4 % al error absoluto.

Con estos resultados, se puede suponer que la variabilidad observada deriva de diferencias reales entre los DC y que la variabilidad (confiabilidad interobservador) que aporta el residente en su forma de evaluar es muy baja. La poca variabilidad asociada con los ítems, probablemente, se relaciona con la buena correlación del aprendizaje cognitivo.

El coeficiente G sin el juicio global fue de 0,92 y, para cada uno de los métodos de enseñanza, mayor de 0,70. Se puede concluir que, con, al

TABLA 4. Correlaciones entre cada uno de los factores y la evaluación global (N = 151)

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

** La correlación es significativa al nivel 0,01 (bilateral). CC: coeficiente de correlación de Pearson.

menos, seis observadores, se puede evaluar, en forma confiable, cada uno de los métodos de enseñanza y que, eliminado el ítem de juicio global, los resultados son aún más generalizables.

Se considera que, tanto con los resultados obtenidos con el análisis de la consistencia interna como con los del estudio G, se puede afirmar que esta investigación aporta evidencia de la confiabilidad del MCTQ en español.

Una limitación de este trabajo es haber validado el cuestionario en dos hospitales universitarios de la CABA de características similares. Dado que la enseñanza clínica está contextualizada por diferentes medios en donde se realiza, para aportar mayor fuente de validez y evitar sesgo de muestreo, es necesario que el MCTQ sea aplicado en otros hospitales de nuestro país y de otros países de habla hispana.

Futuras líneas de investigación podrían estar destinadas a aportar evidencia sobre el impacto de la utilización de este instrumento sobre el desempeño docente luego de brindar *feedback* basado en los resultados de evaluaciones hechas con el MCTQ.

Se concluye que esta investigación cumplió con todos los requisitos formales para una correcta adaptación transcultural del instrumento de evaluación del desempeño del DC y que aportó validez basada en el contenido, en el proceso de respuesta, en la consistencia interna y en la confiabilidad, así como en las relaciones con otras variables. ■

REFERENCIAS

- Steinert Y. Commentary: faculty development: the road less traveled. *Acad Med*. 2011; 86(4):409-11.
- Collins A, Brown J, Newman S. Cognitive apprenticeship: Teaching the crafts of Reading, writing, and mathematics. In Resnick L (ed.). *Knowing, Learning, and Instruction: Essays in honor of Robert Glaser*. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1989. Págs.453-94.
- Kaufman DM, Mann KV. Teaching and learning in medical education: how theory can inform practice. In: Swanwick T (ed.). *Understanding Medical Education: Evidence, Theory and Practice*. 2nd ed. London, UK: Wiley Blackwell; 2014. Págs.7-30.
- Morris C, Blaney D. Work-Based Learning. In Swanwick T (ed.). *Understanding Medical Education: evidence, theory and practice*. 2nd ed. London, UK: Wiley-Blackwell; 2014. Págs.97-109.
- Collins A, Brown J, Holum A. Cognitive Apprenticeship: Making Thinking Visible. *American Educator*. 1991; 6(11):38-46.
- Stalmeijer RE, Dolmans DH, Wolfhagen IH, Scherpbier AJ. Cognitive apprenticeship in clinical practice: can it stimulate learning in the opinion of students? *Adv Health Sci Educ Theory Pract*. 2009; 14(4):535-46.
- Litzelman DK, Stratos GA, Marriott DJ, Lazaridis EN, et al. Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances. *Acad Med*. 1998; 73(3):324-32.
- Snell L, Tallett S, Haist S, Hays R, et al. A review of the evaluation of clinical teaching: new perspectives and challenges. *Med Educ*. 2000; 34(10):862-70.
- Bowden J, Marton F. Quality and qualities. In Bowden J (ed.). *The University of Learning: Beyond Quality and Competence*. London & New York: Taylor & Francis; 2005. Págs.212-37.
- Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med*. 2005; 20(12):1159-64.
- Colliver JA, Conlee MJ, Verhulst SJ. From test validity to construct validity... and back? *Med Educ*. 2012; 46(4):366-71.
- Stalmeijer RE, Dolmans DH, Wolfhagen IH, Muijtjens AM, et al. The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity. *Med Teach*. 2008; 30(8):e272-7.
- Stalmeijer RE, Dolmans DH, Wolfhagen IH, Muijtjens AM, et al. The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med*. 2010; 85(11):1732-8.
- Boerboom TB, Dolmans DH, Jaarsma AD, Muijtjens AM, et al. Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations. *Med Teach*. 2011; 33(2):e84-91.
- Van Mook WNKA, Stalmeijer RE, Muijtjens AMM, Scherpbier AJA. A preliminary study on the quality of the intensive care medicine training programmes in the Netherlands. *Neth J Crit Care*. 2016; 24(7):10-5.
- Stephan A, Cheung G. Clinical teaching and supervision in postgraduate psychiatry training: the trainee perspective. *Australasian Psychiatry*. 2017; 25(2):191-7.
- Olmos-Vega F, Dolmans D, Donkers J, Stalmeijer RE. Understanding how residents' preferences for supervisory methods change throughout residency training: a mixed-methods study. *BMC Med Educ*. 2015; 15(1):177.
- Muñiz J, Elosua P, Hambleton RK, International Test Commission. Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*. 2013; 25(2):151-7.
- American Education Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington: American Psychological Association; 2014.
- Hair JFJ, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 8th ed. Upper Saddle River, NJ: Prentice Hall; 2018.
- Li CH. The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol Methods*. 2016; 21(3):369-87.
- Jöreskog KG, Sörbom D, Inc S. PRELIS 2 User's Reference Guide: a program for multivariate data screening and data summarization: a preprocessor for LISREL. 3rd ed. Chicago: Scientific Software International; 1996.
- Wiley EW, Webb NM, Shavelson RJ. The Generalizability of test scores. In: Kurt F. Geisinger P (ed.). *APA Handbook of Testing and Assessment in Psychology*. New York: American Psychological Association; 2013. Págs.43-60.
- Gudmundsson E. Guidelines for translating and adapting psychological instruments. *Nordic Psychology*. 2009; 61(2):29-45.
- Fernández A. Adaptación de Tests a Otras Culturas. En: Tornimbeni S, Pérez E, Olaz F (ed.). *Introducción a la Psicometría*. Buenos Aires: Paidós; 2008. Págs.191-205.
- Van de Vijver FJR, Hofer J, Chasiotis A. Methodological aspects of cross-cultural developmental studies. In: Bornstein MH (ed.). *Handbook of cross-cultural developmental science*. Mahwah: Erlbaum; 2010. Págs.21-37.
- Centeno AM. The programs and context of medical education in Argentina. *Acad Med*. 2006; 81(12):1081-4.
- Hambleton RK, Merenda PF, Spielberger CD. Adapting educational and psychological tests for cross-cultural assessment: Mahwah, NJ: Psychology Press; 2004.

ANEXO

Cuestionario de Maastricht para la Enseñanza Clínica

Complete el cuestionario luego de haber finalizado su rotación.

Nota: Complete un cuestionario por cada docente clínico.

Nombre y apellido del docente clínico:

1. Especialidad del residente:

2. Especialidad de la rotación:

Este cuestionario tiene como objetivo evaluar la calidad de la enseñanza de su docente clínico.

Por favor, responda a cada afirmación marcando lo que considere apropiado.

Los docentes clínicos recibirán devolución (*feedback*) anónima respecto a su desempeño docente, luego de haber sido evaluados por 6 o más estudiantes

1 = totalmente en desacuerdo.

2 = en desacuerdo.

3 = ni de acuerdo ni en desacuerdo.

4 = de acuerdo.

5 = totalmente de acuerdo.

3. El docente clínico:

	1	2	3	4	5
Me demostró cómo realizar las habilidades clínicas de manera consistente.	☐	☐	☐	☐	☐
Creó suficientes oportunidades para que yo pudiera observarlo en su desempeño.	☐	☐	☐	☐	☐
Sirvió como modelo para seguir del tipo de médico en el que me gustaría convertirme.	☐	☐	☐	☐	☐
Me dio una devolución (<i>feedback</i>) útil inmediatamente después de observarme en contacto directo con mis pacientes o durante eso.	☐	☐	☐	☐	☐
Ajustó sus actividades de enseñanza a mi nivel de experiencia.	☐	☐	☐	☐	☐
Me ofreció suficientes oportunidades para realizar actividades en forma independiente.	☐	☐	☐	☐	☐
Me pidió que fundamentara mis acciones.	☐	☐	☐	☐	☐
Me formuló preguntas dirigidas a aumentar mi nivel de comprensión.	☐	☐	☐	☐	☐
Me incentivó a explorar mis fortalezas y debilidades.	☐	☐	☐	☐	☐
Me incentivó a formular objetivos de aprendizaje.	☐	☐	☐	☐	☐
Me incentivó a cumplir mis objetivos de aprendizaje.	☐	☐	☐	☐	☐
Generó un clima de aprendizaje favorable (cooperativo, no hostil y saludable).	☐	☐	☐	☐	☐
Estuvo interesado en mí como estudiante.	☐	☐	☐	☐	☐
Me demostró respeto.	☐	☐	☐	☐	☐

4. Evalúe el desempeño general de su supervisor como docente.

Escriba el n.º del 1 al 10. ()

Tenga en cuenta que una calificación menor de 6 significa insatisfactorio y 10 es excelente.

5. ¿Cuáles son las fortalezas de este docente clínico?

6. ¿En qué aspectos podría mejorar este docente clínico?

Gracias por completar el cuestionario.

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