



REVISIÓN

Instrumentos para la evaluación del clima de aprendizaje en residencias médicas: síntesis de la evidencia a la luz de las definiciones psicométricas



Luis Carlos Domínguez

Departamento de Cirugía, Universidad de la Sabana, Chia, Colombia

PALABRAS CLAVE

Clima de aprendizaje;
Residencias médicas;
Posgrado;
Validez;
Confiabilidad;
Instrumentos

Resumen

Introducción y objetivo: La medición del clima de aprendizaje en residencias médicas es fundamental para asegurar la calidad de la educación médica. Existen múltiples instrumentos diseñados para este propósito. El objetivo de este estudio es presentar una síntesis de estos instrumentos con énfasis en sus propiedades psicométricas.

Métodos: Se realizó una revisión sistemática de la literatura de estudios originales que evaluaran propiedades psicométricas de instrumentos para la medición del clima de aprendizaje en residencias médicas. Se describieron las características generales de cada instrumento y se revisaron sus propiedades psicométricas de acuerdo con definiciones teóricas aceptadas.

Resultados: Se incluyeron un total de 40 estudios que evalúan 11 instrumentos. El 100% de los instrumentos cumplieron con criterios de validación de contenido y el 81% de validación de constructo. Pocos instrumentos contaron con validación convergente/divergente (36%), predictiva (18,1%) y concurrente (18,1%). El 81% de los instrumentos contaron con algún tipo de análisis de confiabilidad, siendo alta en la mayoría de ellos.

Conclusiones: Existe un número importante de instrumentos disponibles para la evaluación del clima de aprendizaje en posgrado. Se dispone de suficiente evidencia que respalda su validez de contenido, constructo y confiabilidad. Se requieren nuevos estudios que aporten evidencia en cuanto a otros tipos de validez.

© 2018 Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

KEYWORDS

Learning climate;
Residency training;
Postgraduate;
Validity;
Reliability;
Measurement tools

Tools for the evaluation of learning climate in postgraduate training: Synthesis of evidence in the light of psychometric definitions

Abstract

Introduction and objective: The evaluation of the learning climate in medical residences is fundamental to ensure the quality of medical education. There are several tools available for this purpose. The aim of this study is to present a synthesis of these tools, with an emphasis on their psychometric properties.

Correo electrónico: carlosdot@unisabana.edu.co

<https://doi.org/10.1016/j.edumed.2017.10.015>

1575-1813/© 2018 Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Methods: A systematic review of the literature was conducted to evaluate original studies focused on the psychometric properties of tools for the measurement of learning climate in medical residences. The general characteristics of each tool were described, as well as their psychometric properties according to accepted theoretical definitions.

Results: A total of 40 studies evaluating 11 tools were included. All (100%) of the tools studied satisfied criteria for content (100%) and construct validity (81%). Less tools showed convergent / divergent (36%), predictive (18.1%), and concurrent (18.1%) validation. Most of them had some type of reliability analysis (81%), being high in most of them.

Conclusions: There are a significant number of tools available for the evaluation of learning climate in postgraduate training. There is sufficient evidence to support content and construct validity, and reliability. Further studies are required to provide evidence with regard to other types of validity.

© 2018 Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

Existe una necesidad de evaluar críticamente el clima de aprendizaje en residencias médicas. Esta necesidad se enmarca en la tendencia global del aseguramiento y la monitorización de la calidad de la educación médica. También es una medida útil para medir y garantizar el bienestar de los residentes en el trabajo. En este contexto, el clima de aprendizaje se define como la percepción de los residentes de las características, tanto formales como informales, de la educación^{1,2} y se refiere más específicamente a las percepciones sobre prácticas y procedimientos comunes en los departamentos clínicos³, así como a la calidad de las relaciones entre los residentes y sus profesores, supervisores y la institución⁴. En el siglo XXI, y en especial durante los últimos 5 años, se han desarrollado múltiples instrumentos para la medición de estas percepciones. Estos han sido construidos en diversos contextos culturales y se utilizan de forma rutinaria en múltiples escenarios clínicos. Igualmente, muchos de ellos han sido traducidos a varios idiomas y posteriormente validados. Como resultado, existe un aumento acelerado en el número de estudios disponibles que dan cuenta del interés y la relevancia de la evaluación continua del clima de aprendizaje para los educadores médicos, los residentes y la comunidad académica en general. Esta tendencia se inscribe además en la necesidad de obtener medidas objetivas que respalden la educación basada en resultados (outcome-based education), la cual ha sido promovida especialmente en Europa y Norteamérica, pero que cada vez avanza con más fuerza a nivel global⁵. En este contexto, la evaluación del clima de aprendizaje mediante el uso de múltiples instrumentos está ahora en la agenda de muchas facultades de medicina y programas académicos en educación para profesiones de la salud y proporciona evidencia empírica para actuar con buenos argumentos en el momento de realizar ajustes curriculares y procesos de mejoramiento continuo, entre otros⁶. Sin embargo, el creciente número de instrumentos, y de estudios que los respaldan, conlleva nuevos problemas. Uno de estos tiene que ver con la disponibilidad de un extenso número de

instrumentos que exploran múltiples atributos del clima de forma heterogénea. Igualmente, los instrumentos disponibles presentan importantes variaciones en sus propiedades psicométricas, en términos de validez y confiabilidad. Estos problemas traen dificultades en el momento de elegir el instrumento más adecuado para la evaluación del clima, en contextos particulares y poblaciones específicas de residentes. Frente al progresivo número de instrumentos y estudios empíricos al respecto, y teniendo en cuenta que las últimas revisiones de la literatura al respecto son mayores de 5 años⁷⁻⁹, el propósito de este estudio es realizar una revisión actualizada de las propiedades psicométricas de los instrumentos disponibles para la evaluación objetiva del clima de aprendizaje en residencias médicas, que facilite la elección por parte del lector en situaciones y contextos específicos. La pregunta subyacente de investigación es: ¿cuál son la validez y la confiabilidad de los instrumentos disponibles para la medición del ambiente de aprendizaje de posgrado en medicina?».

Materiales y métodos

Diseño del estudio

El objetivo primario de este estudio es describir los aspectos psicométricos relacionados con la validez (de contenido, apariencia, constructo, criterio y convergencia/divergencia) y confiabilidad de los instrumentos disponibles para la evaluación del clima de aprendizaje de posgrado en medicina. Con este propósito se llevó a cabo una revisión sistemática de la literatura, la cual tiene como objetivo «reunir toda la evidencia empírica que cumple unos criterios de elegibilidad previamente establecidos, con el fin de responder una pregunta específica de investigación. Utiliza métodos sistemáticos y explícitos, que se eligen con el fin de minimizar sesgos, aportando así resultados más fiables a partir de los cuales se puedan extraer conclusiones y tomar decisiones»¹⁰.

Criterios de inclusión y exclusión de los estudios

Los criterios de inclusión de los estudios para la revisión fueron: *a)* fuentes primarias de información (estudios originales); *b)* estudios publicados entre enero de 2000 a marzo del 2017; *c)* estudios publicados en idioma inglés o español; *d)* estudios focalizados en la medición del clima de aprendizaje en posgrado en medicina, y *e)* estudios diseñados para la medición de una o varias características psicométricas del instrumento (validez o confiabilidad). Los criterios de exclusión fueron: *a)* estudios de medición del clima de aprendizaje sin análisis psicométrico; *b)* estudios desarrollados en pregrado, enfermería u otras profesiones de la salud, y *c)* otro tipo de estudios diferentes de originales (revisión, cartas al editor, editoriales, resúmenes en conferencias, tesis, literatura gris, correspondencia).

Definiciones conceptuales y búsqueda de los estudios

Las definiciones de los tipos de validez y confiabilidad descritas en la teoría psicométrica de Streiner et al.¹¹ para la construcción de escalas e instrumentos en salud, fueron utilizados como marco conceptual para el desarrollo del presente estudio (tabla 1). La búsqueda de la literatura se llevó a cabo en las bases de datos electrónicas de Pubmed, British Nursing Index (BNI), Applied Social Sciences Index and Abstracts (ASSIA), Science Direct, Educational Resource Information Center (ERIC), BioOne, Cochrane Library y Cumulative Index to Nursing and Allied Health Literature (CINALH). También se condujeron búsquedas manuales de acuerdo con referencias cruzadas. Las palabras clave para la búsqueda de resúmenes o títulos de estudio, tanto en idioma inglés como español, fueron:

- Learning environment*, learning climate*, educational environment*, environment for training*.
- Validity*, reliability*, psychometric properties*.
- Postgraduate*, residency training*, medicine*, registrar*, medical education*, postgraduate student*.
- Survey*, test*, scale*, measure*, instrument*, inventory* or questionnaire*.
- Entorno de aprendizaje*, clima de aprendizaje*, entorno educativo*, entorno de formación*.
- Validez*, fiabilidad*, propiedades psicométricas*.
- Postgrado*, residencia*, medicina*, educación médica*, estudiante de posgrado*.
- Encuesta*, prueba*, escala*, medida*, instrumento*, inventario*, cuestionario*.

Extracción y análisis de la información

Inicialmente, el investigador principal realizó una revisión de los resúmenes de estudios potenciales. Una vez se seleccionaron los estudios a incluir, se realizó la extracción de los datos de acuerdo con las siguientes categorías: nombre del instrumento (autor, país, año de publicación e idioma del instrumento), características demográficas de los participantes (población), teoría educativa subyacente que respaldó el desarrollo del instrumento y descripción de las características del instrumento. Respecto a las propiedades

psicométricas, se identificó la presencia o no de validez de contenido, validez de apariencia, validez de constructo, validez de criterio, validez de convergencia/divergencia y confiabilidad, para cada instrumento. Igualmente se describen de forma general los métodos utilizados en dichos análisis. En la revisión se informaron el método psicométrico utilizado por estudio, estadístico o no estadístico, así como la frecuencia y el porcentaje de estudios que incluyen alguna de las mediciones de validación o confiabilidad previamente mencionadas.

Resultados

Descripción de los estudios y poblaciones

Un total de 40 estudios cumplieron los criterios de inclusión (tabla 2). La identificación y la selección de estos estudios se presenta en la figura 1. Estos estudios evalúan las propiedades psicométricas (validez o confiabilidad) de 11 instrumentos diferentes: Scan of Postgraduate Educational Environment Domains (SPEED) (n = 1), Job Evaluation Survey Tool (JEST) (n = 1), Ambulatory Care Learning Education Environment Measure (ACLEEM) (n = 4), Anesthetic Theatre Educational Environment Measure (ATEEM) (n = 1), Surgical operating theatre learning environment (STEEM) (n = 3), Operating Room Educational Environment Measure (OREEM) (n = 2), Practice Vocational Training Environment (PEEM) (n = 1), Postgraduate Hospital Educational Environment Measure (PHEEM) (n = 19), Dundee Ready Educational Environment Measure (DREEM) (n = 2), Dutch Residency Educational Climate Test (D-RECT) (n = 5), Diagnostic radiology clinical learning environment questionnaire (DR-CLE) (n = 1).

Los estudios que soportan la evidencia psicométrica sobre estos instrumentos fueron desarrollados en diferentes países y contextos culturales: Holanda (n = 6), Reino Unido (n = 6), Irán (n = 4), Arabia Saudí (n = 3), Chile (n = 4), Australia y Nueva Zelanda (n = 3), multicéntrico², Brasil (n = 2), Dinamarca (n = 1), Grecia (n = 1), India (n = 1), Irlanda (n = 1), Japón (n = 1), Pakistán (n = 1), Canadá (n = 1), Singapur (n = 1), EE. UU. (n = 1) y Alemania (n = 1). El 100% de estos instrumentos se encontraron disponibles en idioma inglés (n = 11), el 18,1% (n = 2) en español, portugués, persa y holandés, y el 9% (n = 1) en otros idiomas (danés, holandés, griego, alemán y japonés). La disponibilidad de instrumentos según el idioma fue: SPEED (inglés), JEST (inglés), ACLEEM (español, inglés, persa), ATEEM (inglés), STEEM (inglés), OREEM (inglés), PEEM (inglés), PHEEM (inglés, español, japonés, danés, holandés, persa, portugués, griego), DREEM (portugués, inglés), D-RECT (inglés, holandés, alemán) y DR-CLE (inglés).

Finalmente, la mayoría de estos estudios incluyeron diferentes poblaciones de residentes distribuidos en diversos niveles de entrenamiento y especialidades clínicas y quirúrgicas, tanto en escenarios hospitalarios (unidades de cuidado intensivo, salas de cirugía, sala de urgencias, servicios de hospitalización y consulta), así como escenarios comunitarios (tabla 2). Del total, 2 de los instrumentos fueron diseñados específicamente para evaluación del ambiente de aprendizaje en cirugía (STEEM, OREEM), uno para el ambiente de aprendizaje en anestesia (ATEEM), uno en el escenario ambulatorio (ACLEEM), uno para el ambiente

Tabla 1 Conceptos teóricos sobre validez y confiabilidad de un instrumento de medición

| Concepto | Definición | Forma de medición |
|-------------------------------------|---|--|
| Validez de contenido | Indica si el instrumento muestrea todos los contenidos relevantes, importantes o dominios del constructo teórico subyacente | Las fuentes de la validez de contenido son grupos focales, entrevistas, teoría, investigación, revisión de la literatura, opinión de expertos |
| Validez de apariencia | Indica si los elementos que aparecen en la superficie del instrumento miden lo que realmente son. Esta validez simplemente indica si, en su apariencia, el instrumento parece estar evaluando las cualidades deseadas | Se realiza por un juicio subjetivo basado en una revisión de la medida misma por uno o más expertos, y rara vez se utilizan enfoques empíricos. |
| Validez de criterio | Es la correlación de una escala con otro instrumento, idealmente, un «patrón de oro» que ha sido aceptado y ampliamente utilizado. Se divide en validación concurrente y validación predictiva. En la primera se correlaciona la nueva escala con la medida de criterio al mismo tiempo. En la validación predictiva, el criterio estará disponible en otro momento en el futuro | Correlaciones de Pearson o Spearman |
| Validez de constructo | Se refiere a la validez sobre la forma en la que se mide el constructo (en qué medida el instrumento refleja el contenido teórico subyacente) | Análisis factorial exploratorio. Análisis factorial confirmatorio |
| Validez convergente y discriminante | Indica qué tan estrechamente está relacionada la nueva escala con otras variables y otras medidas del mismo constructo al que debe relacionarse. En la práctica se habla de validez convergente cuando las mediciones del mismo rasgo realizadas con distintos métodos correlacionan entre sí, y divergente cuando las medidas de distintos aspectos o facetas del constructo utilizando el mismo instrumento muestran una baja correlación, lo cual indica que estos aspectos o facetas son independientes entre sí | Correlaciones de Pearson o Spearman. Matriz multirrasgo-multimétodo |
| Confiabilidad | Se refiere a si el instrumento está midiendo algo de una manera reproducible bajo diferentes condiciones (p. ej., que tanto las mediciones en diferentes ocasiones, o por diferentes observadores, o por pruebas similares o paralelas, producen los mismos resultados o resultados similares). El concepto es una forma de reflejar la cantidad de errores (sistemático y aleatorio), inherente a cualquier medición (fiabilidad = variabilidad del sujeto/variabilidad del sujeto + error de medición). Dado que una medida estadística de la variabilidad es la varianza, esto puede expresarse más formalmente como: $\text{confiabilidad} = \frac{\sum 2s}{\sigma 2s} + \sigma 2e$. Dos teorías subyacen: la teoría clásica (se basa en que la confiabilidad de una prueba se descompone en 2 partes: una puntuación «verdadera» (que nadie sabe realmente) y una puntuación de «error») y la teoría de la generalizabilidad (considera simultáneamente múltiples fuentes de varianza de error, lo que permite enfoques racionales para optimizar la confiabilidad) | Consistencia interna: representa el promedio de las correlaciones entre todos los elementos de la medida (alfa de Cronbach, Kuder-Richardson, o mitades divididas) Otras medidas: se basan en la estabilidad del instrumento en varias situaciones y condiciones. Por ejemplo, el grado de acuerdo entre diferentes observadores (confiabilidad entre observadores); el acuerdo entre las observaciones hechas por el mismo evaluador en 2 ocasiones (confiabilidad intraobservador); las observaciones sobre el sujeto en 2 ocasiones separadas por algún intervalo (fiabilidad test-retest) y el poder discriminante (correlaciones entre los ítems de una escala y los dominios a los cuales pertenecen o no pertenecen) |

Fuente: Streiner et al.¹¹.

Tabla 2 Características, población, teoría educativa subyacente, validez y confiabilidad de los instrumentos disponibles para la evaluación del ambiente de aprendizaje en posgrado

| Instrumento | Autor | País/año | Idioma | Características del instrumento | Población | Teoría subyacente | Validez de contenido | Validez de apariencia | Validez de constructo | Validez convergente/divergente | Validez de criterio | | Confiabilidad |
|--|--------------------------------------|-------------------|----------------|---|---|--|--|-----------------------|-----------------------|--|---------------------|------------|---|
| | | | | | | | | | | | Concurrente | Predictiva | |
| Scan of Postgraduate Educational Environment Domains | Schönrock-Adema et al. ³⁵ | Holanda, 2015 | Inglés | 15 ítems distribuidos en 3 dominios: contenido, atmósfera y organización del ambiente de aprendizaje | 223 residentes de diferentes especialidades | — | Panel de expertos (método Delphi) | ✓ | AFE | — | — | — | Chronbach Alpha-correlaciones interescala |
| Job Evaluation Survey Tool | Wall et al. ¹⁹ | Reino Unido, 2014 | Inglés | 15 ítems distribuidos en 2 dominios (no especificados) | 3.367 residentes de varias especialidades | Estándares del General Medical Council | Panel de expertos | — | AFE | — | Contra PHEEM | — | Chronbach Alpha - Teoría de generalizabilidad |
| Ambulatory Care Learning Education Environment Measure | Riquelme et al. ²⁹ | Chile, 2013 | Español/inglés | 50 ítems distribuidos en 3 dominios: soporte, enseñanza clínica, habilidades clínicas | 63 residentes de 7 especialidades | Referencia a planteamientos OMS | Grupos focales con residentes permitieron desarrollar 173 ítems que fueron validados en su contenido por residentes y profesores, mediante método Delphi | ✓ | — | — | — | — | — |
| | Riquelme et al. ³⁰ | Chile, 2015 | Español | 50 ítems distribuidos en 8 dominios: profesores, actividades clínicas y cuidado del paciente, tiempo protegido, infraestructura, habilidades clínicas, evaluación y retroalimentación, información comunicación y tecnología, supervisión clínica. Adicionalmente, en este estudio se identificó que con 15 encuestados los resultados son confiables | 441 participantes: 151 internos y 260 residentes de varios programas de especialización | — | — | ✓ | AFE | — | — | — | Chronbach alpha-teoría de generalizabilidad |
| | Parviziet al. ²⁴ | Irán, 2016 | Persa | 50 ítems distribuidos en los mismos 3 dominios de la versión original | 180 residentes | — | Traducción a persa y revisión del instrumento por panel de expertos | ✓ | AFE | Validez convergente y divergente altas, por cada dominio y globalmente | — | — | Chronbach alpha |
| | Naghizadeh et al. ³⁶ | Irán, 2015 | Persa | 50 ítems distribuidos en los mismos 8 dominios de la versión original tras solución factorial | 157 residentes, la mayoría de ginecología (91%) | — | Traducción a persa y revisión por panel de expertos | ✓ | AFE | — | — | — | Test-retest Cronbach alpha |

Tabla 2 (continuación)

| Instrumento | Autor | País/año | Idioma | Características del instrumento | Población | Teoría subyacente | Validez de contenido | Validez de apariencia | Validez de constructo | Validez convergente/divergente | Validez de criterio | | Confiabilidad | |
|---|-------------------------------|---------------------------------|--------|---|-----------------------------|-------------------|---|-----------------------|-----------------------|--------------------------------|---------------------|------------|---------------|---|
| | | | | | | | | | | | Concurrente | Predictiva | | |
| Anaesthetic Theatre Educational Environment Measure | Holt y Roff ²⁶ | Reino Unido, 2004 | Inglés | 40 ítems distribuidos en 5 dominios: autonomía, percepciones de la atmósfera, carga de trabajo/supervisión/soporte, percepciones de enseñanza y profesores, oportunidades de aprendizaje (tras solución factorial) | 218 residentes de anestesia | — | Revisión de la literatura. Grupos focales (residentes, anestesiólogos, directores de programa) que permitieron desarrollar 156 ítems que fueron validados en su contenido por residentes y anestesiólogos obteniendo un total de 43 ítems | ✓ | AFE | — | — | — | — | |
| Surgical operating theatre learning environment | Cassar ³⁷ | Reino Unido, 2004 | Inglés | 40 ítems distribuidos en 4 dominios: percepciones del profesor y entrenamiento; percepciones de las oportunidades de aprendizaje; percepciones de la atmósfera en quirófano; percepciones de supervisión, carga de trabajo y soporte | 26 residentes de cirugía | — | Revisión de la literatura, entrevistas exploratorias a cirujanos con experiencia en educación y residentes de cirugía. A partir de este proceso se derivó un instrumento de 40 ítems que fue revisado nuevamente por estudiantes de cirugía | ✓ | — | — | — | — | — | Cronbach alpha. Correlaciones interescala |
| | Mahoney et al. ¹² | Australia y Nueva Zelanda, 2010 | Inglés | 40 ítems distribuidos en 5 dominios: habilidades y comportamiento del profesor; oportunidades de aprendizaje facilitadas por el supervisor; características de la lista quirúrgica; interacción con staff no quirúrgico; distracciones del quirófano (tras AFE) | 365 residentes de cirugía | — | — | — | AFE | — | — | — | — | Chronbach alpha |
| | Binsaleh et al. ¹³ | Arabia Saudí, 2014 | Inglés | 40 ítems distribuidos en 4 dominios: percepciones del profesor y entrenamiento; percepciones de las oportunidades de aprendizaje; percepciones de la atmósfera en quirófano; percepciones de supervisión, carga de trabajo y soporte | 33 residentes de urología | — | — | — | — | — | — | — | — | Chronbach alpha |

Tabla 2 (continuación)

| Instrumento | Autor | País/año | Idioma | Características del instrumento | Población | Teoría subyacente | Validez de contenido | Validez de apariencia | Validez de constructo | Validez convergente/divergente | Validez de criterio | | Confiabilidad |
|---|-----------------------------------|---|--------|--|---|-------------------|---|-----------------------|-----------------------|--|---------------------|------------|---|
| | | | | | | | | | | | Concurrente | Predictiva | |
| Operating Room Educational Environment Measure | Kanashiro et al. ¹⁴ | Canadá, 2006 | Inglés | 40 ítems distribuidos en 4 dominios: percepciones del profesor y entrenamiento; percepciones de las oportunidades de aprendizaje; percepciones de la atmósfera en quirófano; percepciones de supervisión, carga de trabajo y soporte | 22 residentes de cirugía | — | Adaptado a contexto canadiense por investigadores principales | — | — | — | — | — | Cronbach alpha |
| | Diwadkar y Jelovsek ¹⁵ | Estados Unidos, 2010 | Inglés | | 27 residentes de ginecología | — | — | — | — | Se estableció contra la impresión global del ambiente de entrenamiento, demostrando alta validez convergente | — | — | Cronbach alpha. Correlaciones interitem |
| Practice Vocational Training Environment | Mulrooney ²⁰ | Irlanda, 2005 | Inglés | 37 ítems distribuidos en 4 dominios: práctica en el trabajo; relación con el supervisor; enseñanza y aprendizaje; interacción con otros profesionales de la salud | 48 residentes | — | Grupos focales con residentes. A partir de estos se desarrolló un instrumento de 55 ítems que fue revisado por un grupo nominal de expertos quienes redujeron el instrumento a 37 ítems | ✓ | — | — | — | — | — |
| Postgraduate Hospital Educational Environment Measure | Roff et al. ¹ | Reino Unido, 2005 | Inglés | 40 ítems distribuidos en 3 dominios: percepciones de autonomía, percepciones de enseñanza y percepciones de soporte social | Dos poblaciones: la primera de 117 residentes, y la segunda de 109; realizaron la priorización y evaluación inicial de ítems para desarrollar el cuestionario final | — | Revisión de la literatura. Teoría fundada incluyendo grupos focales, grupos nominales y método Delphi que involucró a decanos, supervisores, residentes y expertos | ✓ | — | — | — | — | Cronbach alpha |
| | Wall et al. ³⁴ | Brasil, Chile, Holanda, Reino Unido, 2009 | Varios | 40 ítems distribuidos en 3 dominios: percepciones de autonomía, percepciones de enseñanza y percepciones de soporte social (tras AFE) | 1.563 residentes de diferentes contextos | — | — | — | AFE | — | — | — | Cronbach alpha. Correlaciones interescala |

Tabla 2 (continuación)

| Instrumento | Autor | País/año | Idioma | Características del instrumento | Población | Teoría subyacente | Validez de contenido | Validez de apariencia | Validez de constructo | Validez convergente/divergente | Validez de criterio | | Confiabilidad |
|-------------|-------------------------------|-----------------|-----------|---|----------------|-------------------|--|-----------------------|-----------------------|--------------------------------|---------------------|--|---|
| | | | | | | | | | | | Concurrente | Predictiva | |
| | Shimizu et al. ¹⁷ | Japón, 2013 | Japonés | 40 ítems distribuidos en 3 dominios: percepciones de autonomía, percepciones de enseñanza y percepciones de soporte social | 206 residentes | — | — | — | — | — | — | Predictiva Predictor de desempeño académico | — |
| | Boor et al. ¹⁸ | Holanda, 2007 | Holandés | 40 ítems: unidimensional. En este estudio se determinó que con 14 sujetos por departamento la medición es confiable. Igualmente, para múltiples departamentos, 3 sujetos combinados con 10 departamentos proveen un resultado confiable para ambos grupos | 256 residentes | — | Traducción del instrumento al holandés. No evaluación de contenido | ✓ | AFE | — | — | — | Teoría de generalizabilidad |
| | Aspegren et al. ³⁸ | Dinamarca, 2007 | Danés | 40 ítems distribuidos en 3 dominios: percepciones de autonomía, percepciones de enseñanza y percepciones de soporte social (luego del AFE). Los autores indican que se pueden eliminar 7 ítems interconectados y reducir el tamaño del cuestionario | 342 residentes | — | Traducción y revisión del instrumento a danés por investigadores | ✓ | AFE | — | — | — | Cronbach alpha Correlaciones interitem |
| | Vieira ³⁹ | Brasil, 2009 | Portugués | 40 ítems distribuidos en 9 dominios (no especificados) | 306 residentes | — | Traducción al portugués y revisión por investigadores | ✓ | AFE | — | — | — | Cronbach alpha |
| | Riquelme et al. ²⁸ | Chile, 2009 | Español | 40 ítems distribuidos en 5 dominios: percepciones de enseñanza y ambiente; enseñanza y soporte social (consejería), justicia en el programa, autonomía, alojamiento, catering e instalaciones) | 125 residentes | — | Grupos focales con profesores y estudiantes luego de traducción | ✓ | AFE | — | — | — | Cronbach alpha |

Tabla 2 (continuación)

| Instrumento | Autor | País/año | Idioma | Características del instrumento | Población | Teoría subyacente | Validez de contenido | Validez de apariencia | Validez de constructo | Validez convergente/divergente | Validez de criterio | | Confiabilidad |
|---------------------------------------|----------------------|--------------------|--|--|--|-------------------|--|-----------------------|-----------------------|--|---------------------|------------|----------------|
| | | | | | | | | | | | Concurrente | Predictiva | |
| Shokoohi et al. ⁴⁰ | Irán, 2014 | Persa | 40 ítems distribuidos en 5 dominios: enseñanza; sistema educativo; facilidades; satisfacción en el trabajo; soporte social | 128 residentes | — | — | Traducción al persa y revisión por expertos | ✓ | AFE | — | — | — | Cronbach alpha |
| Gooneratne et al. ⁴¹ | Singapur, 2008 | Sri Lanka (inglés) | 34 ítems. Tras AFE investigadores describen 12 factores, sin embargo, consideran que el instrumento es unidimensional pues solo retienen el primer factor que ocupa 19,7% de la varianza | 100 residentes | — | — | Traducción y adaptación al contexto de Sri Lanka por investigadores principales | ✓ | AFE | — | — | — | Cronbach alpha |
| Al-Shiekh et al. ⁴² | Arabia Saudita, 2014 | Inglés | 40 ítems, estructura factorial similar al estudio original | 193 internos y residentes de varias especialidades | — | — | — | — | — | — | — | — | Cronbach alpha |
| Pinnock et al. ⁴³ | Australia, 2009 | Inglés | 40 ítems distribuidos en 2 factores (tras AFE): no descritos | 53 Residentes de pediatría | — | — | Modificación de la redacción de algunos ítems y ajuste al contexto | ✓ | AFE | — | — | — | Cronbach alpha |
| Jalili et al. ⁴⁴ | Irán, 2014 | Persa | 37 ítems distribuidos en 11 dominios (no especificados) (tras AFE) | 89 residentes de urgencias | — | — | Traducción, remoción de 3 ítem y adaptación | ✓ | AFE | — | — | — | Cronbach alpha |
| Sch€onrock-Adema et al. ⁴⁵ | Holanda, 2009 | Holandés | 40 ítems distribuidos en 3 dominios: contenido de aprendizaje y coaching; clima afectivo benéfico; regulación externa | 279 residentes de 8 hospitales | Tras AFE se revisaron los factores a la luz de las teorías de funciones de aprendizaje | — | Traducción, revisión y adaptación al contexto por los investigadores principales | ✓ | AFE-AFC | — | — | — | Cronbach alpha |
| Van Hell et al. ⁴⁶ | Holanda, 2009 | Holandés | 40 ítems 3 sub-escalas (igual que versión original) | 133 estudiantes de medicina | — | — | — | — | — | — | — | — | Cronbach alpha |
| Koutsogiannou et al. ⁴¹ | Grecia, 2015 | Griego | 40 ítems en 3 factores (tras AFE-AFC) | 731 residentes diversas especialidades | — | — | Previamente trasladada. Revisión luego por panel de expertos | ✓ | AFE-CFA | Establecida contra 1 ítem de satisfacción global con la rotación (alta convergencia) | — | — | Cronbach alpha |

Tabla 2 (continuación)

| Instrumento | Autor | País/año | Idioma | Características del instrumento | Población | Teoría subyacente | Validez de contenido | Validez de apariencia | Validez de constructo | Validez convergente/divergente | Validez de criterio | | Confiabilidad |
|--|--|--------------------|-----------|--|--|------------------------------|--|-----------------------|-----------------------|--|--|------------|---|
| | | | | | | | | | | | Concurrente | Predictiva | |
| | Sheikh et al. ⁴⁷ | Pakistán, 2017 | — | 40 ítems 3 dominios (similar a versión original) | 320 residentes de varias especialidades | — | Revisión de contenido de la escala y modificación de algunos de sus ítems de acuerdo conl contexto | — | — | — | — | — | Cronbach alpha |
| | Herrera et al. ³² | Chile, 2012 | Español | 40 ítems 3 dominios (similar a versión original) | 318 residentes diversas especialidades | — | — | — | — | — | — | — | Cronbach alpha/teoría de generalizabilidad |
| | Clapham et al. ⁴⁸ | Reino Unido, 2007 | Inglés | 40 ítems 3 dominios (similar a versión original) | 134 residentes cuidado intensivo | — | — | ✓ | — | — | — | — | Cronbach alpha |
| | Binsaleh et al. ⁴⁹ | Arabia Saudí, 2015 | Inglés | 40 ítems distribuidos en 3 dominios como en la versión original | 38 residentes de urología | — | — | — | — | — | — | — | Cronbach alpha |
| Dundee Ready Educational Environment Measure | De Oliveira Filho et al. ²⁵ | Brasil, 2005 | Portugués | 50 ítems distribuidos en 5 factores iguales a versión original (tras AFE) | 97 residentes de 4 programas de residencia | — | Revisión de redacción por investigadores | ✓ | AFE | Validez discriminante fue alta entre instituciones, programas y especialidades | Calidad de vida evaluada mediante e Quality of School Life Inventory | — | Cronbach alpha. Teoría de generalizabilidad |
| | Chhabra et al. ⁵⁰ | India, 2016 | Inglés | 50 ítems distribuidos en 5 factores iguales a versión original (tras AFE) | 35 residentes de medicina de urgencias | — | — | — | — | — | — | — | Cronbach alpha |
| Dutch Residency Educational Climate Test | Boor et al. ²² | Holanda, 2011 | Inglés | 50 ítems distribuidos en 11 dominios: supervisión, coaching y evaluación, trabajo en equipo, retroalimentación, colaboración, relaciones profesionales entre los profesores, trabajo adaptado al residente, papel del profesor y educación formal, papel del tutor de la especialidad, cambio de turno | 1.251 residentes. Sin embargo, ofrece medición agregada a nivel de departamentos | Marco teórico socio-cultural | Método Delphi entre expertos (revisaron cuestionario previo Dutch Residency Educational Climate Test 75 ítems) | ✓ | AFE-AFC | — | — | — | Cronbach alpha. Teoría de generalizabilidad |

Tabla 2 (continuación)

| Instrumento | Autor | País/año | Idioma | Características del instrumento | Población | Teoría subyacente | Validez de contenido | Validez de apariencia | Validez de constructo | Validez convergente/divergente | Validez de criterio | | Confiabilidad |
|--|---------------------------------|---------------------------------|--------|--|--|-------------------|---|-----------------------|-----------------------|--------------------------------|---------------------|--|---|
| | | | | | | | | | | | Concurrente | Predictiva | |
| | Lombarts et al. ³ | Holanda, 2014 | Inglés | 50 ítems distribuidos en 11 dominios (igual a versión original) | 451 residentes de 45 programas en 17 hospitales | — | — | — | — | — | — | Desempeño de los profesores evaluado mediante el System for Evaluation of Teaching Qualities | — |
| | Piek et al. ⁵¹ | Unión Europea, 2015 | Inglés | 50 ítems distribuidos en 11 dominios (igual a versión original) | 119 fellows de oncología ginecológica en Europa | — | — | — | — | — | — | — | Cronbach alpha |
| | Silkens et al. ²³ | Hola, 2016 | Inglés | 35 ítems distribuidos en 9 dominios: atmósfera educacional, papel del tutor de la especialidad, trabajo en equipo, coaching y evaluación, educación formal, colaboración entre residentes, adaptación del trabajo al nivel de competencia del residente, acceso a supervisores, cambio de turno (tras AFE) | 2.306 residentes de 291 departamentos en 48 hospitales | — | Panel de expertos y método Delphi (revisión de Dutch Residency Educational Climate Test 50 ítems) | ✓ | AFE AFC | — | — | — | Cronbach alpha adecuado, correlaciones interescala |
| | Iblher et al. ¹⁶ | Alemania, 2015 | Alemán | 50 ítems distribuidos en 11 dominios (igual a versión original) | 255 residentes de 17 hospitales | — | Se realizó traducción del instrumento original (Dutch Residency Educational Climate Test 50 ítems) | ✓ | AFC | — | — | — | Cronbach alpha correlaciones inter-ítem y poder discriminante. Correlaciones interescala Cronbach Alpha |
| Diagnostic radiology clinical learning environment questionnaire | Bloomfield et al. ³² | Australia y Nueva Zelanda, 2008 | Inglés | 24 ítems distribuidos en 5 dominios: aprendizaje basado en el trabajo; atmósfera social; entrenamiento formal; supervisión; carga de trabajo (tras AFE) | 133 residentes de radiología | — | Revisión de la literatura, adaptación de ítems de otros instrumentos y revisión por panel de expertos | ✓ | AFE | — | — | — | Cronbach Alpha |

AFE: análisis factorial exploratorio; AFC: análisis factorial confirmatorio; GMC: General Medical Council; OMS: Organización Mundial de la Salud.

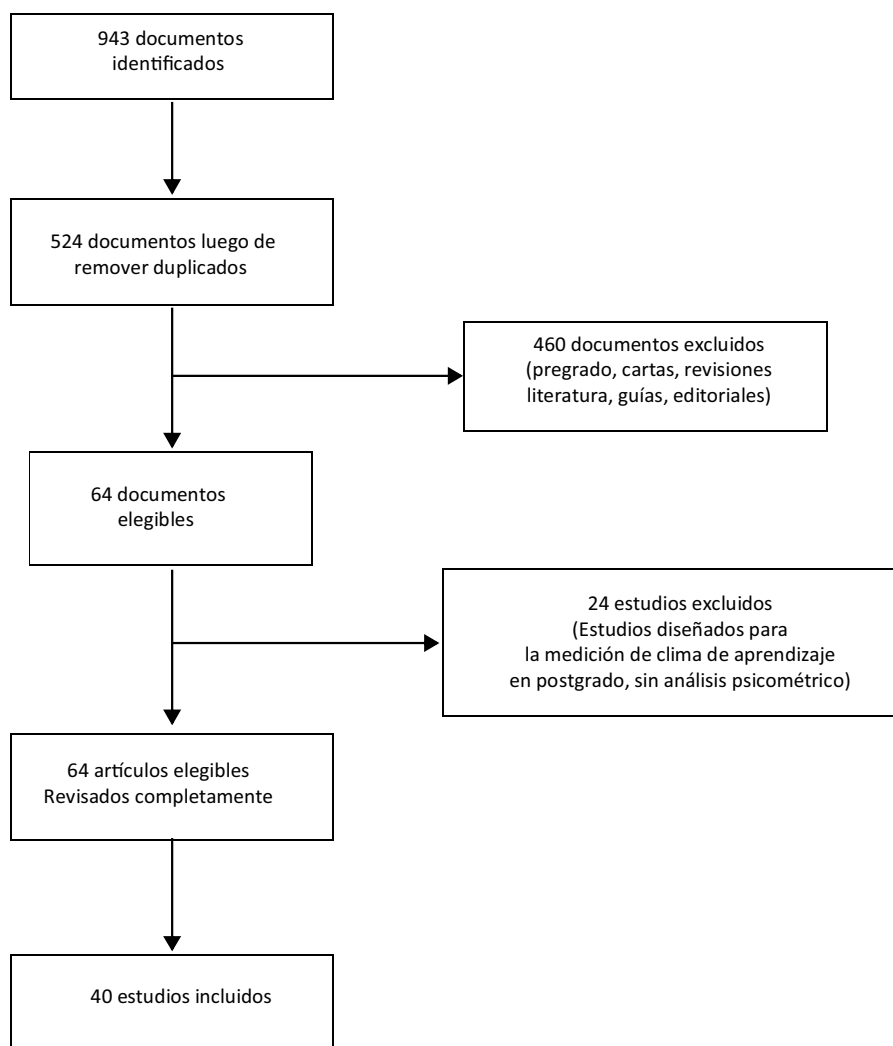


Figura 1 Diagrama de flujo para la selección e inclusión de estudios en la revisión.

de aprendizaje en radiología (DR-CLE) y el resto puede aplicarse de forma multifacética (SPEED, JEST, PHEEM, D-RECT, PEEM). Respecto a la teoría subyacente, solo D-RECT fue desarrollado con base en un marco teórico socio-cultural que explica el ambiente y el clima de aprendizaje. Sin embargo, es preciso anotar que JEST fue desarrollado con referencia a los estándares del General Medical Council (GMC) y ACLEEM con referencia a los planteamientos de la Organización Mundial de la Salud (OMS). Igualmente, la versión holandesa del PHEEM, tras el análisis factorial exploratorio (AFE), fue revisada a la luz de las teorías de funciones de aprendizaje.

Validez y confiabilidad de los instrumentos

a. *Validez de contenido y apariencia*: el 100% de los instrumentos originales cuentan con el uso, individual o combinado, de diversas técnicas de validación de contenido, como panel de expertos, método Delphi, grupos focales, entrevistas y revisión por expertos temáticos, entre otros (tabla 2). No obstante, las versiones en idioma inglés del STEEM diferentes de la original^{12,13}, OREEM^{14,15}, las

versión traducida al alemán del D-RECT¹⁶ y las versiones en japonés¹⁷ y holandés¹⁸ del PHEEM no cuentan con nuevo análisis de validez de contenido. Igualmente, 9 de los 11 instrumentos poseen algún tipo de análisis de validez de apariencia (81%). Se exceptúan de este último tipo de validez el JEST¹⁹, el OREEM^{14,15} y algunas versiones de STEEM, PHEEM y D-RECT (tabla 2).

b. *Validez de constructo*: 9 de 11 instrumentos (81%) cuentan con validez de constructo mediante AFE, exceptuando OREEM^{14,15} y PEEM²⁰. Solo 2 instrumentos (18,1%), en algunas de sus versiones, presentan validez de constructo mediante análisis factorial confirmatorio (AFC). En particular, la versión holandesa y griega de PHEEM^{18,21}, así como la versión original y traducida al alemán de D-RECT 50 ítems²² y D-RECT 35 ítems²³ cuentan con una combinación de AFE-AFC.

c. *Validez convergente/divergente*: 4 instrumentos cuentan con algún tipo de validación divergente/convergente en los estudios disponibles (36%). Se incluyen la versión persa de ACLEEM²⁴, una de las versiones en inglés de OREEM¹⁴, la versión griega de PHEEM²¹ y la versión portuguesa de DREEM²⁵.

d. *Validez de criterio*: 2 estudios cuentan con análisis de validez concurrente (18,1%) y 2 de validez predictiva (18,1%). En cuanto a los de validación concurrente, un estudio evalúa JEST contra PHEEM¹⁹ y otro DREEM contra el Quality of School Life Inventory (QSL)²⁵. Con relación a la validez predictiva, PHEEM y D-RECT ofrecen información en 2 estudios. El primero en un estudio analiza los puntajes del PHEEM como predictor de desempeño académico¹⁷ y el segundo en un estudio en el que D-RECT predice desempeño de los profesores evaluado mediante el System for Evaluation of Teaching Qualities (SETQ) a partir de regresiones lineales multivariantes³.

e. *Confiabilidad*: 9 de los 11 instrumentos disponibles (81%) cuentan con algún tipo de análisis de confiabilidad, exceptuando ATEEM²⁶ y PEEM²⁰. Los estudios que respaldan este análisis demuestran alta confiabilidad de los instrumentos (tabla 2).

Ningún instrumento cumplió todos los tipos de validez. Teniendo en cuenta la heterogeneidad de los estudios seleccionados, referida a variabilidad en los participantes, las intervenciones y los resultados estudiados, así como la heterogeneidad metodológica, referida al diseño y análisis de los resultados de cada estudio¹⁰, no se consideró pertinente el conducir un análisis estadístico combinando los datos provenientes de los estudios para determinar su heterogeneidad estadística.

Discusión

Este estudio fue diseñado para describir las propiedades psicométricas de los instrumentos disponibles para la evaluación del clima de aprendizaje en residencias médicas, a la luz de las definiciones de validez y confiabilidad de Streiner et al.¹¹. Un primer resultado indica que existen 11 instrumentos disponibles, siendo PHEEM, STEEM y D-RECT los más evaluados en la literatura. La mayor parte de estos instrumentos fueron desarrollados en Europa y todos están disponibles en idioma inglés (solo ACLEEM y PHEEM están disponibles en idioma español). Estos instrumentos evalúan escenarios quirúrgicos y clínicos, así como escenarios ambulatorios. Un segundo resultado es que la mayor parte de los instrumentos poseen algún tipo de evidencia que sustenta su validez de contenido, apariencia y constructo, aunque el uso de teoría educativa como marco conceptual es pobre. Igualmente, la evidencia disponible sobre la validez convergente y divergente, así como sobre la validez de criterio (concurrente y predictiva), es escasa. La mayoría de los instrumentos son altamente confiables. Finalmente, ningún instrumento ofrece evidencia sobre todos los tipos de validez propuestos.

El primer resultado puede explicarse por varias razones. La mayor proporción de instrumentos, así como de la evidencia que los respalda, provienen del contexto europeo. Las explicaciones pueden estar relacionadas con la tradición de la educación médica europea, presente como disciplina de conocimiento al menos durante los últimos 50 años, así como por el marco conceptual vigente orientado hacia la educación basada en resultados, enmarcada en el proceso de Bolonia de 1999 y luego en la Declaración de Bolonia²⁷. Este último en particular puede explicar

la necesidad de mayor evidencia empírica en torno al clima de aprendizaje en contextos de enseñanza clínica, orientada al mejoramiento del ambiente y a la toma de decisiones, para favorecer que los estudiantes alcancen las competencias esperadas en el marco de la declaración. También muchos de estos estudios provienen de procesos de investigación en centros académicos de educación para profesiones de la salud en Europa, en el escenario de maestrías y doctorados. No obstante, pese a esta tendencia, durante los últimos 10 años varios de estos instrumentos han sido traducidos a múltiples idiomas, dando cuenta de la relevancia de la medición del clima de aprendizaje en contextos diferentes del europeo. Un caso particular ha sido el del PHEEM, hoy en día disponible en 8 idiomas diferentes. Posiblemente, el crecimiento y la expansión global de la educación médica sean una explicación a este fenómeno. Sin embargo, en idioma español la disponibilidad de instrumentos es aún limitada. Precisamente, PHEEM ha sido el único traducido y validado²⁸ y únicamente ACLEEM ha sido desarrollado como instrumento original^{29,30}. En particular, la evidencia empírica disponible en idioma español no permite, hasta el momento, establecer comparaciones y un seguimiento longitudinal sobre el clima de aprendizaje en residencias médicas en la región latino e iberoamericana. Solo algunos estudios han evaluado la asociación entre el clima (medido por PHEEM) y otros desenlaces, como el «burnout syndrome» (síndrome de desgaste profesional)³¹, y otros han comparado el clima de aprendizaje en varias especialidades médicas dentro de una misma institución³². Se requieren nuevas investigaciones, traducciones, validaciones y adaptaciones culturales de otros instrumentos al idioma español.

El segundo resultado también tiene varias explicaciones. En primer lugar, el uso de la teoría educativa como marco conceptual para el desarrollo de estos instrumentos es pobre. Únicamente D-RECT posee un marco teórico que lo respalda, en tanto ACLEEM y JEST involucran algún tipo de referencia sobre la base de estándares y recomendaciones internacionales. En particular, el marco conceptual de D-RECT está enfocado en la evaluación a nivel colectivo (departamentos clínicos) y no solo de residentes, siendo el único instrumento empleado a este nivel. Al respecto, existe un llamado a que la investigación en educación para profesiones de la salud se apropie de la teoría educativa³³. La mayoría de los instrumentos provienen de marcos conceptuales derivados de teoría fundada, cuya generalización requiere de nueva evidencia. Por ejemplo, pocos instrumentos traducidos y validados, en contextos diferentes del original, cuentan con nuevas validaciones de contenido y adaptaciones culturales que permitan soportar la generalización de sus resultados. La pérdida de este análisis puede afectar y explicar la variabilidad de los resultados en cuanto a otros tipos de validación y en particular la estructura factorial de los instrumentos. Cuestionarios como el PHEEM han pretendido dar una solución a este problema mediante la inclusión de diversas poblaciones, a través de diseños multicéntricos en una misma medición, lo cual ha permitido definir una estructura factorial estable aportando al mejoramiento de su validez de constructo³⁴. No obstante, el resto de los instrumentos carecen de este tipo de evaluaciones. Se requieren nuevas investigaciones al respecto. Igualmente, otros tipos de validaciones

son aún limitados. Un caso tiene que ver con el escaso número de estudios focalizados en la validación de criterio, tanto predictiva como concurrente. Respecto a la predictiva, solo uno de los 40 estudios incluye algún tipo de predicción entre el clima de aprendizaje, mediante PHEEM, y el desempeño académico. Igualmente, la comparación de la validez de estos instrumentos entre sí es pobre. Por ejemplo, solo un estudio compara JEST y PHEEM, lo cual impide que puedan establecerse diferencias objetivas respecto al rendimiento de los instrumentos. Igualmente, se requieren nuevos estudios al respecto. Esta limitación también afecta a la magnitud en la que los instrumentos están relacionados con otras variables y otras medidas del mismo constructo al que debe relacionarse (validez convergente y divergente). En la práctica, 4 estudios ofrecen algún tipo de evidencia al respecto. Sin embargo, se requieren nuevos estudios que indiquen cómo la medición del clima de aprendizaje con distintos instrumentos se correlaciona entre sí (validez convergente), y como las medidas de distintos aspectos de clima de aprendizaje, utilizando el mismo instrumento, podrían demostrar una baja correlación, indicativa de aspectos independientes entre sí (validez divergente).

El presente estudio tiene fortalezas y debilidades. Dentro de las primeras, se encuentran el uso de un diseño sistemático para la revisión de la literatura, así como de un marco conceptual que permite analizar la evidencia a la luz de definiciones aceptadas. Algunas debilidades tienen que ver con la falta de análisis de las fortalezas intrínsecas de cada estudio en sus procesos de validación, en particular las que se relacionan con la valoración de la calidad propia del análisis psicométrico, así como de los aspectos propios de la heterogeneidad metodológica. Sin embargo, estos aspectos abren nuevas oportunidades de investigación. Otra debilidad está en la no inclusión de instrumentos utilizados para la evaluación del clima de aprendizaje provenientes de otras profesiones de la salud a nivel de posgrado. Si bien uno de los criterios de exclusión se refería a este específicamente, un número importante de instrumentos han sido desarrollados en otras disciplinas y podrían llegar a ser útiles a la educación médica en particular.

Igualmente este estudio ofrece oportunidades e identifica vacíos de conocimiento, que pueden ser útiles para la investigación futura. Varios han sido descritos anteriormente. Además, tiene implicaciones para la práctica. Una de estas tiene que ver con la síntesis descriptiva de los instrumentos disponibles a la fecha para la evaluación del clima de aprendizaje en posgrado, la cual puede servir como marco de referencia al momento de elegir un instrumento. También ofrece importante información descriptiva sobre los dominios que exploran estos cuestionarios, los cuales pueden ser útiles en mediciones de aspectos particulares del clima de aprendizaje. Finalmente, aporta elementos en cuanto a su validez y confiabilidad, que permiten que el lector tenga un marco comparativo e integrado de la evidencia disponible. En conclusión, el presente estudio ha explorado las características psicométricas de los instrumentos disponibles para la evaluación del clima de aprendizaje en posgrado mediante un método sistemático de análisis de la literatura, ofreciendo oportunidades para la práctica e investigación futura.

Financiación

Facultad de Medicina, Universidad de la Sabana, Colombia.

Conflicto de intereses

Ninguno.

Bibliografía

1. Roff S, McAleer S, Skinner A. Development and validation of an instrument to measure the postgraduate clinical learning and teaching educational environment for hospital-based junior doctors in the UK. *Med Teach*. 2005;27:326–31.
2. Delva MD, Kirby J, Schultz K, Godwin M. Assessing the relationship of learning approaches to workplace climate in clerkship and residency. *Acad Med*. 2004;79:1120–6.
3. Lombarts KM, Heineman MJ, Scherpbier AJ, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. *PLoS One*. 2014;9:e86512.
4. Biggs J, Tang C. *Teaching for quality learning at university*. Berkshire: Society for Research into Higher Education & Open University Press; 2007.
5. Ahmed R, Farooq A, Storie D, Hartling L, Oswald A. Building capacity for education research among clinical educators in the health professions: A BEME (Best Evidence Medical Education) Systematic Review of the outcomes of interventions: BEME Guide No. 34. *Med Teach*. 2016;38:123–36.
6. Isba R, Boor K. Creating a learning environment. In: Dornan T, Mann K, Scherpbier A, Spencer J, editor. *Medical education: theory and practice*. New York: Churchill Livingstone, Elsevier; 2011. p. 99–114.
7. Soemantri D, Herrera C, Riquelme A. Measuring the educational environment in health professions studies: A systematic review. *Med Teach*. 2010;32:947–52.
8. 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;23:445–54.
9. Genn JM. AMEE Medical Education Guide No. 23 (Part 1): Curriculum, environment, climate, quality and change in medical education—a unifying perspective. *Med Teach*. 2001;23:337–44.
10. Higgins JPT, Green S (editores). *Cochrane handbook for systematic reviews of interventions version 5.1.0* [updated March 2011] [consultado 24 Sep 2017]. The Cochrane Collaboration, 2011 Disponible en: www.cochrane-handbook.org.
11. Streiner, Norman, Cairney J, editores. *Health Measurement Scales: A practical guide to their development and use*. 5th ed. Oxford: Oxford University Press; 2015.
12. Mahoney A, Crowe PJ, Harris P. Exploring Australasian surgical trainees' satisfaction with operating theatre learning using the 'surgical theatre educational environment measure'. *ANZ J Surg*. 2010;80:884–9.
13. Binsaleh S, Babaeer A, Alkhalayal A, Madbouly K. Evaluation of the learning environment of urology residency training using the postgraduate hospital educational environment measure inventory. *Adv Med Educ Pract*. 2015;2:271–7.
14. Kanashiro J, McAleer S, Roff S. Assessing the educational environment in the operating room—a measure of resident perception at one Canadian institution. *Surgery*. 2006;139:150–8.
15. Diwadkar GB, Jelovsek JE. Measuring surgical trainee perceptions to assess the operating room educational environment. *J Surg Educ*. 2010;67:210–6.
16. 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. Doc55.
17. Shimizu T, Tsugawa Y, Tanoue Y, Konishi R, Nishizaki Y, Kishimoto M, et al. The hospital educational environment and performance of residents in the General Medicine In-Training Examination: A multicenter study in Japan. *Int J Gen Med.* 2013;29:637–40.
 18. Boor K, Scheele F, van der Vleuten CP, Scherpbier AJ, Teunissen PW, Sijtsma K. Psychometric properties of an instrument to measure the clinical learning environment. *Med Educ.* 2007;41:92–9.
 19. Wall D, Goodyear H, Singh B, Whitehouse A, Hughes E, Howes J. A new tool to evaluate postgraduate training posts: The Job Evaluation Survey Tool (JEST). *BMC Med Educ.* 2014;2:210.
 20. Mulrooney A. Development of an instrument to measure the Practice Vocational Training Environment in Ireland. *Med Teach.* 2005;27:338–42.
 21. Koutsogiannou P, Dimoliatis ID, Mavridis Bellos S, Karathanos V, Jelastopulu E. Validation of the Postgraduate Hospital Educational Environment Measure (PHEEM) in a sample of 731 Greek residents. *Res Notes.* 2015;30:734.
 22. Boor K, van der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach.* 2011;33:820–7.
 23. Silkens ME, Smirnova A, Stalmeijer RE, Arah OA, Scherpbier AJ, van der Vleuten C, et al. Revisiting the D-RECT tool: Validation of an instrument measuring residents' learning climate perceptions. *Med Teach.* 2016;38:476–81.
 24. Parvizi MM, Amini M, Dehghani MR, Jafari P, Parvizi Z. Psychometric properties of the Persian version of the Ambulatory Care Learning Educational Environment Measure (ACLEEM) questionnaire, Shiraz, Iran. *Adv Med Educ Pract.* 2016;29:559–66.
 25. De Oliveira Filho GR, Vieira JE, Schonhorst L. Psychometric properties of the Dundee Ready Educational Environment Measure (DREEM) applied to medical residents. *Med Teach.* 2005;27:343–7.
 26. Holt MC, Roff S. Development and validation of the Anaesthetic Theatre Educational Environment Measure (ATEEM). *Med Teach.* 2004;26:553–8.
 27. Patrício M, Harden RM. The Bologna Process —A global vision for the future of medical education. *Med Teach.* 2010;32:305–15.
 28. Riquelme A, Herrera C, Aranís C, Oporto J, Padilla O. Psychometric analyses and internal consistency of the PHEEM questionnaire to measure the clinical learning environment in the clerkship of a Medical School in Chile. *Med Teach.* 2009;31:e221–5.
 29. Riquelme A, Padilla O, Herrera C, Olivos T, Román JA, Sarfatis A, et al. Development of ACLEEM questionnaire, an instrument measuring residents' educational environment in postgraduate ambulatory setting. *Med Teach.* 2013;35:e861–6.
 30. Riquelme A, Padilla O, Herrera C, Olivos T, Román JA, Sarfatis A, et al. Medición del ambiente educacional en contextos ambulatorios. *Ed Med.* 2015;16:131–40.
 31. Llera J, Durante E. Correlation between the educational environment and burn-out syndrome in residency programs at a university hospital. *Arch Argent Pediatr.* 2014;112:6–11.
 32. Herrera CA, Olivos T, Román JA, Larraín A, Pizarro M, Solís N, et al. Evaluation of the educational environment in medical specialty programs. *Rev Med Chil.* 2012;140:1554–61.
 33. McMillian W. Theory in healthcare education research: The importance of worldwide. En: Cleland J, Durning S, editores. *Researching medical education.* 1st ed. Oxford: Jhon Wiley & Sons (Ed); 2015.
 34. Wall D, Clapham M, Riquelme A, Vieira J, Cartmill R, Aspegren K, et al. Is PHEEM a multi-dimensional instrument? An international perspective. *Med Teach.* 2009;31:e521–7.
 35. Schönrock-Adema J, Visscher M, Raat AN, Brand PL. Development and validation of the Scan of Postgraduate Educational Environment Domains (SPEED): A brief instrument to assess the educational environment in postgraduate medical education. *PLoS One.* 2015;10:e0137872.
 36. Naghizadeh Moogari Z, Koohpayehzadeh J, Roff S, Montazeri A, Soltani Arabshahi SK, Bigdeli S, et al. Psychometric analysis of the ambulatory care learning education environment measure (ACLEEM) in Iran. *Med J Islam Repub Iran.* 2015;21:299.
 37. Cassar K. Development of an instrument to measure the surgical operating theatre learning environment as perceived by basic surgical trainees. *Med Teach.* 2004;26:260–4.
 38. Aspegren K, Bastholt L, Bested KM, Bonnesen T, Ejlertsen E, Fog I, et al. Validation of the PHEEM instrument in a Danish hospital setting. *Med Teach.* 2007;29:498–500.
 39. Vieira JE. The postgraduate hospital educational environment measure (PHEEM) questionnaire identifies quality of instruction as a key factor predicting academic achievement. *Clinics (Sao Paulo).* 2008;63:741–6.
 40. Shokoohi S, Hossein Emami A, Mohammadi A, Ahmadi S, Mojtahezadeh R. Psychometric properties of the postgraduate hospital educational environment measure in an Iranian hospital setting. *Med Educ Online.* 2014;19:24546.
 41. Gooneratne IK, Munasinghe SR, Siriwardena C, Olupeliyawa AM, Karunathilake I. Assessment of psychometric properties of a modified PHEEM Questionnaire. *Ann Acad Med Singapore.* 2008;37:993–7.
 42. Al-Shiekh MH, Ismail MH, Al-Khater SA. Validation of the postgraduate hospital educational environment measure at a Saudi University Medical School. *Saudi Med J.* 2014;35:734–8.
 43. Pinnock R, Reed P, Wright M. The learning environment of paediatric trainees in New Zealand. *J Paediatr Child Health.* 2009;45:529–34.
 44. Jalili M, Mortaz Hejri S, Ghalandari M, Moradi-Lakeh M, Mirzazadeh A, Roff S. Validating modified PHEEM questionnaire for measuring educational environment in academic emergency departments. *Arch Iran Med.* 2014;17:372–7.
 45. Schönrock-Adema J, Heijne-Penninga M, Van Hell EA, Cohen-Schotanus J. Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Med Teach.* 2009;31:e226–32.
 46. Van Hell EA, Kuks JB, Cohen-Schotanus J. Time spent on clerkship activities by students in relation to their perceptions of learning environment quality. *Med Educ.* 2009;43:674–9.
 47. Sheikh S, Kumari B, Obaid M, Khalid N. Assessment of postgraduate educational environment in public and private hospitals of Karachi. *J Pak Med Assoc.* 2017;67:171–7.
 48. Clapham M, Wall D, Batchelor A. Educational environment in intensive care medicine —use of Postgraduate Hospital Educational Environment Measure (PHEEM). *Med Teach.* 2007;29:e184–91.
 49. Binsaleh S, Babaeer A, Rabah D, Madbouly K. J Evaluation of urology residents' perception of surgical theater educational environment. *Surg Educ.* 2015;72:73–9.
 50. Chhabra S, Misra A, Shah S, Kole T. Survey of student perception of medical education environment among emergency medicine residents of an academic medical centre in Northern India. *Int J Emerg Med.* 2016;9:5.
 51. Piek J, Bossart M, Boor K, Halaska M, Haidopoulos D, Zapardiel I, et al. The work place educational climate in gynecological oncology fellowships across Europe: the impact of accreditation. *Int J Gynecol Cancer.* 2015;25:180–90.
 52. Bloomfield L, Subramaniam R. Development of an instrument to measure the clinical learning environment in diagnostic radiology. *J Med Imaging Radiat Oncol.* 2008;52:262–8.