



CienciAmérica: Revista de Divulgación Científica de la
Universidad Tecnológica Indoamérica

ISSN: 1390-9592

ISSN: 1390-681X

cienciamerica@uti.edu.ec

Universidad Tecnológica Indoamérica
Ecuador

Moreno, Jaime; Puertas, Luis

ESTUDIO DE VALIDEZ DE CRITERIO DE UN INSTRUMENTO PARA
MEDIR ESTILOS CONDUCTUALES EN POBLACIÓN ECUATORIANA

CienciAmérica: Revista de Divulgación Científica de la Universidad Tecnológica
Indoamérica, vol. 11, núm. 1, <http://dx.doi.org/10.33210/ca.v11i1.385>, 2022, Enero-Junio
Universidad Tecnológica Indoamérica
Ecuador

- Número completo
- Más información del artículo
- Página de la revista en redalyc.org



ESTUDIO DE VALIDEZ DE CRITERIO DE UN INSTRUMENTO PARA MEDIR ESTILOS CONDUCTUALES EN POBLACIÓN ECUATORIANA

Criterion validity of an instrument to measure behavioral styles in Ecuadorian population

Estudo de validade de critério de um instrumento para medir estilos comportamentais na população equatoriana

Jaime Moreno¹  & Luis Puertas² 

¹ Carrera de Psicología Organizacional, Facultad de Psicología, Pontificia Universidad Católica del Ecuador, Quito – Ecuador. Correo: jamorenov@puce.edu.ec

² Investigador independiente. Correo: lpuertas37@gmail.com

Fecha de recepción: 06 de marzo de 2022.
Fecha de aceptación: 06 de junio de 2022.

RESUMEN

INTRODUCCIÓN: Kudert es la prueba de mayor uso en el mercado de evaluaciones psicológicas en el ámbito organizacional en el Ecuador. Mide estilos conductuales DISC, capacidades cognitivas e inteligencia emocional. A pesar de su amplia difusión la información sobre las propiedades psicométricas del instrumento es limitada. **OBJETIVO:** analizar la validez de criterio concurrente del instrumento Kudert considerando que uno de sus usos es la selección de personal. **MÉTODO:** participaron 164 vendedores de una empresa de consumo masivo quienes aparte de rendir la prueba fueron evaluados en tres criterios de desempeño identificados con el análisis del trabajo. Después de efectuar las pruebas de normalidad y la revisión de puntuaciones extremas de los datos, se correlacionaron todas las escalas del Kudert como predictores y los tres criterios de desempeño identificados, aplicando la corrección de la atenuación debida a la falta de fiabilidad de las medidas. **RESULTADOS:** Las correlaciones mostraron que no hay relación con significación estadística entre las escalas del test Kudert y las variables criterio en la muestra. **DISCUSIÓN Y CONCLUSIONES:** se presentan algunas recomendaciones respecto a los modelos y procedimientos psicométricos que se deberían emplear considerando la naturaleza ipsativa de algunas escalas de la prueba, así como el tipo de información que debería estar disponible para los usuarios.



Moreno & Puertas. Estudio de validez de criterio de un instrumento para medir estilos conductuales en población ecuatoriana.
Enero – Junio 2022

<http://dx.doi.org/10.33210/ca.v11i1.385>



Palabras clave: test Kudert; DISC; validez de criterio; validez concurrente; personalidad en vendedores.

ABSTRACT

INTRODUCTION: Kudert is the most widely used test in the psychological assessment market in the organizational field in Ecuador. It measures DISC behavioral styles, cognitive abilities, and emotional intelligence. Despite its wide use, information on the psychometric properties of the instrument is limited. **OBJECTIVE:** to analyze the criterion validity of Kudert, considering that one of its uses is personnel selection. **METHOD:** 164 salespeople from a mass consumption company participated, who, apart from taking the test, were evaluated on the three performance criteria identified with work analysis. After conducting normality tests and outlier analysis, all the Kudert scales were correlated as predictors with the three identified performance criteria, correcting for attenuation due to unreliability of measures. **RESULTS:** The correlations showed no statistically significant relationship between the Kudert test scales and the criterion variables in the sample. **DISCUSSION AND CONCLUSIONS:** Some recommendations are offered regarding the psychometric models and procedures that should be used considering the ipsative nature of some of the test scales as well as the type of information that should be available to the users.

Keywords: Kudert test; DISC; criterion validity; concurrent validity; salespersons personality.

RESUMO

INTRODUÇÃO: Kudert é o teste mais utilizado no mercado de avaliação psicológica no campo organizacional no Equador. Mede estilos comportamentais DISC, habilidades cognitivas e inteligência emocional. Apesar de sua ampla divulgação, as informações sobre as propriedades psicométricas do instrumento são limitadas. **OBJETIVO:** analisar a validade de critério concorrente do instrumento de Kudert, considerando que um de seus usos é a seleção de pessoal. **MÉTODO:** Participaram 164 vendedores de uma empresa de consumo de massa que, além de realizarem o teste, foram avaliados em três critérios de desempenho identificados com a análise de trabalho. Após a realização dos testes de normalidade e revisão das pontuações extremas dos dados, todas as escalas de Kudert foram correlacionadas como preditoras e, os três critérios de desempenho identificados, aplicando-se a correção da atenuação, devido à falta de confiabilidade das medidas. **RESULTADOS:** As correlações mostraram que não há relação estatisticamente significativa entre as escalas do teste de Kudert e as variáveis critério na amostra. **DISCUSSÃO E CONCLUSÕES:** São apresentadas algumas recomendações quanto aos modelos e procedimentos psicométricos que devem ser utilizados considerando a natureza ipsativa de algumas escalas de teste, bem como o tipo de informação que deve ser disponível aos usuários.

Palavras-chave: teste Kudert; DISC; validade do critério; validade concorrente; personalidade nos vendedores.

INTRODUCCIÓN

El test Kudert es una prueba comercial diseñada en el Ecuador para usos en el ámbito organizacional como, por ejemplo: selección de personal, desarrollo de líderes, coaching y equipos de alto desempeño [1].

Información comercial sugiere que Kudert es la prueba que tiene la mayor participación en el mercado de pruebas psicológicas del Ecuador de uso organizacional con alrededor del 61%. Se reporta además que al menos 200.000 personas han sido evaluadas lo que comprende más de 150 empresas y su uso en seis países latinoamericanos [2].

Kudert se compone de tres dimensiones: conductual (basada en el modelo DISC); cognitiva (razonamiento verbal, orientación espacial, razonamiento lógico, habilidad numérica y razonamiento abstracto) y emocional (autocontrol) [2]. Los resultados de la evaluación se presentan en un reporte que contiene las puntuaciones de las escalas DISC divididas en un perfil natural y un perfil adaptado; las puntuaciones en las escalas cognitivas y el “índice emocional” [1]. Adicional se reportan estilos de liderazgo, estilos comerciales y competencias. Hay diversos reportes que se orientan a distintos usuarios: evaluados, empresa, equipos, etc.

La aplicación de la prueba es de 40 minutos. Las escalas DISC evalúan cuatro dimensiones de personalidad o estilos conductuales que son dominancia, influencia, solidez y cumplimiento y que serían la manifestación de diferencias individuales en los patrones de pensamiento, sentimiento y conducta de acuerdo con la teoría original de Marston propuesta en 1928 [3]. Los ítems de las escalas DISC del Kudert son de respuesta forzada. Es decir, los evaluados deben seleccionar distintos adjetivos de acuerdo con su percepción sobre cómo se caracterizan ciertos comportamientos; estos adjetivos son limitados de modo que si se asigna una puntuación a una opción se descartan otras opciones. La prueba presenta dos listas de adjetivos que conforman el modelo DISC. En la primera lista se pregunta al evaluado sobre los adjetivos que mejor lo describan; esto configura el perfil natural y en la segunda lista se pregunta los adjetivos que usarían los demás para describirlo, lo que configura el perfil adaptado. Este formato de reactivo se denomina escala ipsativa que es común en las herramientas DISC [3].

La segunda parte de Kudert mide aptitudes cognitivas por medio de reactivos por tiempo: verbal, espacial, lógico, numérico y abstracto. Estas aptitudes se integran en un “índice cognitivo”. En la sección final Kudert mide inteligencia emocional; para esto presenta una serie de imágenes de personas con expresiones faciales en la que se solicita a los evaluados identificar la emoción de la imagen con la que se reaccionaría frente a diversas situaciones. Al igual que la dimensión conductual esta parte resulta en un puntaje natural y uno adaptado.

El número de ítems de la escala original de las dimensiones DISC son: 78 reactivos que conforman las escalas de dominancia, influencia, solidez y cumplimiento [4]. La dimensión cognitiva, integrada por las escalas verbal, espacial, lógico, numérico

y abstracto, comprende 183 ítems. Finalmente, la escala emocional se presenta con cinco reactivos.

Ortiz et al. [4] reportan los coeficientes alfa ordinales para las escalas DISC: dominancia = 0.80; influencia = 0.81; solidez = 0.78 y cumplimiento = 0.82. Los coeficientes alfa de las escalas cognitivas y el “indicador emocional” se presentan en la Ficha Técnica [1] con el siguiente detalle: verbal = 0.90; espacial = 0.82; lógico; 0.87; numérico = 0.85 y abstracto = 0.74; indicador emocional = 0.81.

En otra publicación [2] se describe el procedimiento empleado para establecer la validez de contenido que consistió en la identificación de ítems que “tenían un impacto negativo en el autoconcepto de los participantes por razones culturales” [SIC] [2, p. 21] y ciertos ítems que no discriminaban por su excesiva escogencia entre los participantes, lo que se atribuyó a la deseabilidad social. Respecto a la validez de constructo Ortiz, et al. [1] reportan solamente las intercorrelaciones de la dimensión conductual y emocional y las intercorrelaciones de las escalas cognitivas.

En un estudio más reciente Ortiz, et al. [4] analizan la estructura interna de las escalas DISC de la prueba para lo que efectúan un análisis factorial exploratorio con matriz de correlaciones tetracóricas y en una segunda etapa de análisis, la aplicación de un modelo de ecuaciones estructurales exploratorias (ESEM) con rotación oblicua semi-especificada, evaluando la bondad de ajuste con el índice de ajuste comparativo (CFI), el índice Tucker-Lewis (TLI), el error cuadrático medio de aproximación (RMSEA) y la raíz cuadrática media residual ponderada (WRMR). La fiabilidad de las escalas resultantes se determinó con el Alfa ordinal.

El resultado del análisis fue la reducción sustancial del número de ítems de 78 a 41 con una estructura factorial más robusta y parsimoniosa. Destaca que el mejor ajuste se logró con el uso del modelo de ecuaciones estructurales exploratorias (ESEM) y no con el análisis factorial confirmatorio (AFC) por considerar que sus supuestos y restricciones no eran adecuados según el tipo de datos y formato de respuesta que usa la prueba (ítems dicotómicos de elección forzada).

Entre las principales conclusiones del estudio se afirma que con estos resultados se puede disminuir la longitud del instrumento con una estructura factorial optimizada, reduciendo los tiempos de aplicación, lo que resulta conveniente considerando el contexto en el que se lo usa (ámbito organizacional) en el que usualmente los evaluados disponen de poco tiempo por tener que hacer varias evaluaciones. También se menciona la necesidad de estudiar la validez predictiva con relación a indicadores de éxito laboral como el desempeño laboral, la conducta de ciudadanía organizacional, etc. Esta última recomendación es particularmente importante considerando que la prueba fue diseñada entre otros fines, para la selección de personal [1].

Si bien la optimización de la estructura interna y la mejora de la confiabilidad de las escalas es un paso necesario en el desarrollo de un instrumento [5], no es menos cierto que este análisis por sí solo es insuficiente para establecer la validez de constructo del instrumento. Mientras no se establezcan las correlaciones del instrumento con constructos medidos por pruebas diferentes y otros criterios, no se puede llegar a

conclusiones significativas sobre los constructos medidos. Aun cuando Ortiz et al. [4], mencionan algunos estudios en que los instrumentos DISC se han relacionado con otros cuestionarios de personalidad como el Big Five y el 16PF, no se menciona en la discusión la necesidad de efectuar estos estudios con las escalas DISC de la herramienta Kudert.

Con el fin de identificar los constructos medidos por la prueba y considerando que uno de sus objetivos es la selección de personal, se planteó un estudio de validez de criterio para evaluar la capacidad predictiva de la prueba. Así, la investigación se realizó dentro del área comercial de una empresa de consumo masivo tomando como unidad de análisis el cargo de vendedor. En el Ecuador “el sector comercio no es solo el más grande en el país, sino también el que emplea el mayor número de trabajadores en condiciones adecuadas” [6, p. 16]. Este sector emplea el 20% de la fuerza laboral del país lo que hace necesario, desde la gestión de talento, contar con herramientas que permitan una evaluación válida y confiable de esta fuerza laboral. Por esta razón

La evaluación de personal en las empresas comerciales es un gran reto. “La multiplicidad de especificaciones o categorías del hecho de la venta implica la dificultad de definir al buen vendedor” [7., p.5].

Por otra parte, los trabajos comerciales son competitivos ya que en la mayoría de los casos se establecen metas retadoras para maximizar las ventas. Desde la perspectiva de la gestión de talento esto se traduce en la necesidad de reclutar y contratar personal con las competencias comerciales y la capacidad de aprender rápida y correctamente el método comercial para desempeñar un trabajo que, aunque suele tener metas claras, se efectúa en un entorno dinámico y volátil en el que factores externos, sobre los que se tiene poco control, afectan el logro de los objetivos [8].

En la evaluación con fines de selección, la validez de criterio se erige como la más relevante [9]. Si la prueba no predice el desempeño, su utilidad en una situación de selección es cuestionable. La validación de criterio consiste en correlacionar las puntuaciones del test o predictor con las puntuaciones del criterio [10]. Si la correlación es significativa y está dentro de cierta magnitud, se dispone de evidencia que demuestra la utilidad de la prueba para la selección. La validación de criterio se la puede indagar en base a dos diseños: predictivo y concurrente [10]. En el predictivo las puntuaciones del predictor se recogen en una situación real de selección mientras que las del criterio se las obtiene luego de un período de tiempo cuando se puede medir el desempeño de las personas seleccionadas con la prueba. Cuando se dispone de ambos conjuntos de medidas se procede a calcular la correlación que es sometida a dos correcciones estadísticas: in fiabilidad de las medidas y restricción de rango [11]. En el diseño concurrente ambas medidas (predictor y criterio) suelen obtenerse en el mismo período de tiempo. Un factor relevante es que las personas evaluadas no son objeto de una selección, sino que ya son parte de la organización y en consecuencia es más fácil obtener las puntuaciones del predictor y el criterio. El principal ajuste estadístico posible en este diseño es la corrección por in fiabilidad de las medidas [11]. Ambos diseños tienen ventajas y desventajas, pero el diseño concurrente suele ser más factible en la práctica [10].

En las labores comerciales el criterio a predecir debe derivarse de un cuidadoso análisis del trabajo que permita dilucidar cuáles son los desempeños por pronosticar y a partir de qué predictores [14]. En el actual estudio todas las dimensiones y facetas del Kudert serían los potenciales predictores, considerando que uno de los propósitos de la prueba es la selección de personal [1]. Sin embargo, como afirman Highhouse, et al. [13] la hipótesis predictiva debe plantear unos predictores que tengan sentido conceptual respecto al criterio. Llevar a cabo una correlación con muchas variables para “ver que sale” habla de una aproximación empírica sin sustento teórico. No obstante, en el presente caso, se determinó la casi total ausencia de estudios previos que examinaran la validez de criterio del Kudert¹, por lo que se decidió hacer un análisis exploratorio con todas las dimensiones de la herramienta (conductual, cognitiva y emocional) para determinar su grado de asociación con los criterios de desempeño identificados con el análisis del trabajo.

El análisis del trabajo se hizo con el cargo de “vendedor” que es desempeñado por 164 personas. El análisis resultó en la identificación de los siguientes criterios: cumplimiento del presupuesto, cumplimiento de visitas a clientes y competencias comerciales específicas del cargo.

Considerando la definición de desempeño laboral, los tres criterios corresponden al llamado desempeño en la tarea [15]. Los dos primeros criterios son objetivos y se refieren a los resultados más importantes que se esperan del ocupante del cargo mientras que el tercer criterio, más subjetivo, evalúa ciertos comportamientos laborales (competencias), a través del juicio de los supervisores sobre el desempeño de los vendedores.

Con estos antecedentes el presente estudio se enfocó en determinar la validez de criterio de la prueba Kudert con relación al desempeño laboral y de manera particular la selección de personal comercial.

MÉTODO

Participantes

Participó la población de vendedores de una empresa comercializadora de Ecuador (N = 175). La muestra estaba repartida en las siguientes ciudades: Ambato, Guayaquil, Ibarra, Lago Agrio, Quito y Santo Domingo. Se seleccionó aquellas evaluaciones que mantuvieron niveles de fiabilidad altos según lo especificado por el instrumento y se tomaron en cuenta a las personas con un año o más en la organización como criterio de inclusión. Esto llevó a la eliminación de 11 evaluaciones quedando la muestra final en 164 participantes. El 81% de los casos (n = 133) son hombres. En cuanto a la variable edad, el 44% (n = 72) se encuentran en el rango de 30 a 40 años.

¹ Se encontró una disertación de grado no publicada en la que se concluye que el Kudert: “No evidencia las características psicométricas necesarias (...) y tampoco se comprueba que tiene la utilidad de predecir una conducta futura que asegure un buen desempeño laboral de cada candidato [14, p. 41].”

El 34 % (n = 56) pertenecen al rango de 40 años en adelante. El último rango que es de 20 a 30 años representa el 22% (n = 36). En la variable escolaridad, el 90% (n = 148) son bachilleres. El 10% (n = 16) tienen escolaridad superior. En la variable antigüedad el 37% (n = 61) pertenecen al rango de 1 a 2 años. El 29% (n = 48) pertenecen al grupo de 2 a 4 años. El 10% (n = 16) pertenecen al rango de 4 a 6 años. El 24% (n = 39) pertenece al rango 6 años en adelante. En cuanto a la variable región el 48% de los casos (n = 79) pertenecen a la región sierra. El 42% (n = 69) pertenecen a la región costa mientras que el 10% restante (n = 16) son de la región Oriente.

Normas éticas de investigación

Los participantes fueron informados sobre el objetivo del estudio destacando que la participación era voluntaria, que no había ninguna consecuencia por participar o no y que la información recolectada era confidencial.

Instrumentos

Los predictores están representados por todas las dimensiones de la prueba Kudert (conductual, cognitiva y emocional), cuyas propiedades psicométricas (principalmente confiabilidad) ya fueron descritas en la sección anterior.

Los criterios se los identificó mediante el análisis del trabajo [12] del cargo de vendedor. A continuación, la descripción de los criterios.

Cumplimiento de presupuesto de ventas: se mide en base al resultado de venta neta dividido para un presupuesto, ambos expresados en dólares que cada vendedor debe cumplir en el periodo de un mes. Este presupuesto se configura de las necesidades de venta de los distintos proveedores en base a un histórico de zonas geográficas y consumo de clientes. Si bien es un criterio importante para la organización, posiblemente tiene contaminación porque su consecución no depende únicamente del desempeño del vendedor, sino que puede estar afectado por condiciones y factores externos como: condiciones de mercado, tipo de clientes, zonas, etc.

Cumplimiento de visitas a clientes: es la medida de las visitas que realiza el vendedor a clientes sobre una planificación diaria de clientes por visitar llamada maestro de clientes. Este maestro es revisado y depurado constantemente por el supervisor de ventas para garantizar un número óptimo de clientes por ruta de visita. Las visitas son registradas por el sistema SCMI versión 3.2, que registra la gestión de campo del vendedor. Cuando el vendedor visita a un cliente debe abrir el dispositivo celular, ejecutar la *app* SCMI, registrar la visita y su resultado que puede ser un pedido o motivo de no venta. Esta información la ingresa en el sistema adjuntando las coordenadas geográficas, hora de la visita y el detalle de la gestión. Las visitas son un indicador conductual que depende principalmente de la gestión que realiza el vendedor y es registrado con precisión por el dispositivo.

Competencias comerciales (modelo PODER): acrónimo de Planifica la venta, Observa inventario, Desarrolla la venta, Establece pedido, Registra y cobra. Se trata de comportamientos necesarios para vender. Estas competencias se distribuyen en los diferentes momentos de la venta que son: previo a la venta, durante y después de la venta. El modelo de competencias fue realizado con la asesoría de una empresa externa, con un panel de expertos comerciales y del Área de talento, para asegurar su validez de contenido.

La evaluación se realiza mensualmente sobre la base de al menos dos clientes del vendedor. El supervisor de ventas es el responsable de evaluar, brindar retroalimentación inmediata y cargar las evaluaciones en el sistema. La empresa realiza capacitaciones semestrales para reforzar la metodología con la fuerza de ventas.

Técnicas de análisis de datos

En el análisis de los descriptivos se encontró que las puntuaciones de las variables cognitivas (excepto el índice cognitivo) no correspondían con las propiedades de una escala de intervalo, sino con una escala ordinal. Con el fin de correlacionar las variables cognitivas con los criterios fue necesario transformar estos últimos a una escala ordinal para aplicar la correlación de rangos de Spearman.

Se verificaron los supuestos de normalidad de todas las variables efectuando la prueba de Shapiro-Wilk junto con otras pruebas relacionadas. Luego se identificaron los valores extremos (*outliers*) de los criterios y los predictores empleando la distancia de Mahalanobis con un nivel $p < 0.001$ en la distribución Chi cuadrado [16]. Se eliminaron dos casos extremos en los criterios quedando el tamaño de la muestra en $N = 162$. Se efectuó el mismo procedimiento para los diversos predictores de modo que el tamaño de la muestra tuvo ligeras variaciones dependiendo del cruce del conjunto de predictores con los criterios. En varios casos, cuando la mayoría de las variables no asumían la distribución normal ($W_i; p < 0.05$) se calculó la correlación de rangos de Spearman que también es apropiada cuando no hay evidencia de normalidad entre las variables [17].

Eliminados los valores extremos se calcularon las correlaciones asegurando el cruce de todos los predictores y criterios aplicando además la corrección por falta de fiabilidad [5, 10, 11] al tratarse de un diseño concurrente [10, 11].

Sheskin [18] sugiere efectuar regresión lineal cuando en un conjunto de predictores hay por lo menos una correlación que muestre significación con el criterio, siempre que se cumplan los supuestos de la regresión. Esta recomendación se la tuvo presente en el análisis.

En el análisis correlacional con los predictores, las competencias del modelo PODER fueron analizadas de dos maneras: (a) como índice global (una puntuación global de todas las competencias) y (b) con puntuaciones específicas correspondientes a cada una de las cinco competencias (planifica, observa, desarrolla, establece y

registra). Para evitar la multicolinealidad, el índice global y las puntuaciones específicas fueron analizadas por separado con respecto a los predictores.

Procedimiento

Las fuentes de recolección primarias fueron los vendedores que decidieron participar a quienes se les aplicó la prueba Kudert completa y los procedimientos para medir los tres criterios de desempeño. Los datos recopilados de los predictores y los criterios correspondieron a un período aproximado de 12 meses, tiempo en el cual las evaluaciones Kudert se hicieron mediante la adquisición de las licencias necesarias.

Se capacitó a los especialistas del Área de talento de cada región sobre el procedimiento de aplicación de la prueba y se comunicó a los vendedores sobre el proceso de evaluación y su objetivo. Las pruebas se aplicaron en línea. Los vendedores que tuvieron dificultad asistieron a las agencias para realizar la evaluación de forma presencial con el soporte de especialistas del Área de talento. Las evaluaciones obtenidas con la herramienta Kudert fueron exportadas en formato Excel para su depuración y transformación. Las evaluaciones PODER fueron ingresadas por los supervisores en el sistema Knack BV3-1644 que también fueron exportadas en formato Excel.

El cumplimiento de visitas se recopiló mediante consulta en MySQL 3.23.19 por ODBC 5.43 del sistema SCMI al sistema Qlik Sense Enterprise 13, en el que fueron transformados para su posterior exportación a Excel. El cumplimiento de presupuesto de ventas fue recopilado del sistema Qlik Sense Enterprise 13 y exportado posteriormente a Excel.

Los datos depurados y almacenados en Excel fueron transformados para su tratamiento en SPSS v. 23 [19].

RESULTADOS

Con el fin de analizar la validez criterial se correlacionaron todas las escalas de la prueba con los tres criterios de desempeño del cargo de vendedor. Se presentan los resultados para cada una de las dimensiones de la prueba: conductual (DISC), cognitivo y emocional.

Tabla 1. Correlaciones de los criterios con las escalas DISC

Predictores	Criterios de desempeño			Corrección por in fiabilidad ²		
	PODER	Ventas	Visitas	PODER	Ventas	Visitas
Dominancia	-.018	-.082	-0.107	-.024	-.092	-.12
Influencia	-.021	.079	.035	-.027	.088	.039
Solidez	-.066	-.098	-.023	-.087	-.111	-.026
Cumplimiento	.092	.075	.074	.119	.083	.082
Dominancia ¹	-.034	-.122	.022	-.044	-.136	.025
Influencia ¹	-.071	-.047	-.038	-.092	-.052	-.042
Solidez ¹	-.071	-.04	-.058	-.094	-.045	-.066

Cumplimiento¹ .05 .014 .01 .065 .015 .011

Nota. N = 158. * $p < .05$; ** $p < .01$; 1 = escalas adaptadas; 2 = alfas de las escalas DISC, dominancia = 0.8; influencia = 0.81; solidez = 0.78; cumplimiento = 0.82 [29]. En el caso de PODER también se corrigió por infiabilidad del criterio con alfa = 0.73. En los restantes dos criterios solo se corrigió por infiabilidad del test.

No se encuentran correlaciones significativas.

Tabla 2. Correlaciones de los criterios con las escalas cognitivas

Predictores	Criterios de desempeño ¹		
	PODER	Ventas	Visitas
Verbal	.015	-.09	-.088
Espacial	.11	-.031	.015
Lógico	-.033	.072	-.001
Numérico	-.015	-.010	.074
Abstracto	-.036	-.004	-.003

Nota. N = 162. * $p < .05$; ** $p < .01$; 1 dado que las puntuaciones de los predictores estaban en una escala ordinal, los criterios también fueron transformados al mismo tipo de escala aplicando la correlación de rangos de Spearman.

No se encuentran correlaciones significativas.

Tabla 3. Correlaciones de los criterios con el índice cognitivo y la dimensión emoción

Predictores	Criterios de desempeño		
	PODER	Ventas	Visitas
Índice cognitivo	.001	.005	.044
Emoción natural	.083	.014	.053
Emoción adaptada	.122	.012	-.027

Nota. N = 162. * $p < .05$; ** $p < .01$

No se encuentran correlaciones significativas.

El criterio PODER es una puntuación combinada que integra cinco subescalas que corresponden a las competencias comerciales específicas del cargo. Estas subescalas también fueron correlacionadas con todas las escalas del Kudert. Por motivos de espacio se presentan a continuación solamente las conclusiones de los resultados obtenidos.

No se encontraron correlaciones significativas entre las competencias específicas y las escalas DISC (puntuaciones naturales y adaptadas). Total de correlaciones = 40.

No se encontraron correlaciones significativas entre las competencias específicas y los siete estilos comerciales (proactivo, resolutivo, planificador, comunicador, experto, asesor, soporte). Total de correlaciones = 35.

No se encontró correlación entre las competencias comerciales PODER y el índice cognitivo. Total de correlaciones = 5.

La correlación entre las competencias comerciales específicas (PODER) y el predictor disposición general para la venta (DGV) resultó en un total de cinco correlaciones de las cuales solamente una mostró indicios de significación: DGV – desarrolla ($r = 0.17$; $p < 0.05$), usando la correlación de Spearman dada la evidencia de no normalidad de la DGV ($W = 0.90$, $p < .001$). El coeficiente de determinación indica que apenas un 2.9% de la varianza en la habilidad de desarrollar la venta se explica por la DGV.

Respecto a las competencias comerciales específicas (PODER) y las puntuaciones en las escalas de emoción natural y emoción adaptada se detectaron dos correlaciones con indicios de significación que se muestran en la siguiente tabla.

Tabla 4. Correlaciones de las competencias comerciales PODER y los factores de la dimensión emocional

Criterios	Predictores	
	Emoción natural	Emoción adaptada
Planifica	-.021	-.121
Observa	.147	.186*
Desarrolla	-.025	.004
Establece	.131	-.027
Registra	.184*	-.039

Nota. N = 158. * $p < .05$; se usó correlación de rangos de Spearman ($W = 0.75$, $p < .001$).

Los coeficientes de determinación de estas correlaciones explican el 3,4% y el 3,5% de la varianza en las competencias: Registra ($r = .18$, $p < .05$) y Observa ($r = .19$, $p < .05$) respectivamente. No se efectuó regresión lineal de estos criterios y predictores por la evidencia de no normalidad de uno de los predictores (emoción natural, $W = 0.95$, $p < .001$) y de algunos criterios.

DISCUSIÓN Y CONCLUSIONES

La presente investigación tuvo como objetivo evaluar la validez de criterio concurrente del test Kudert para aportar evidencia de las propiedades psicométricas de uno de los tests de alta difusión en Ecuador. Se efectuaron un total de 225 correlaciones entre predictores y criterios encontrando indicios de significación en apenas tres casos que representan el 1.3% de todas las correlaciones efectuadas. El conjunto de resultados de los distintos análisis muestra que el test no predice los criterios establecidos y consecuentemente carece de utilidad práctica para la organización en la que se efectuó el estudio. La corrección por falta de fiabilidad en las escalas del DISC y algunos criterios no fueron suficientes para mejorar la magnitud de las correlaciones y alcanzar significación estadística como se observa en la tabla 1.

Si bien los criterios son específicos de la organización en la que se efectuó el estudio, las categorías que estos criterios representan (ventas efectuadas, visitas a

clientes y competencias comerciales) son comunes y de uso frecuente en los trabajos comerciales [20].

Se detectaron tres correlaciones no paramétricas con indicios de significación estadística, a saber: disposición general para la venta (predictor) con la competencia de desarrollar la venta (criterio); emoción natural (predictor) con la competencia registra la venta (criterio) y emoción adaptada con la competencia observa (criterio), (estas dos últimas en la tabla 4). El máximo porcentaje de varianza explicada de estos tres casos llega apenas al 3.5% en uno de los criterios, lo que no tiene relevancia práctica [21].

Si bien estas tres correlaciones mostraron un nivel fronterizo de significación estadística ($p < .05$), el resultado no debe ser interpretado como evidencia sólida de validez por estas razones: (a) la varianza explicada es mínima, sin consecuencia práctica, (b) fueron correlaciones no paramétricas (Spearman) que tienen menor poder estadístico que una correlación paramétrica (Pearson) lo que aumenta la probabilidad de errores tipo I [18]. (c) La baja potencia estadística de la muestra. (d) la naturaleza subjetiva de la calificación del criterio, (e) la aparente falta de lógica conceptual entre las variables relacionadas y (f) porque el valor p asociado a estas correlaciones ($p < 0.05$) es objeto de cuestionamientos en los últimos años [22] por la elevada tasa de errores tipo I o falsos positivos que genera [23].

No se encontró ninguna relación significativa con las escalas cognitivas (tabla 2). Por su parte el “índice cognitivo” (tabla 3), presumiblemente una combinación de los factores cognitivos tampoco correlacionó con los criterios. Esto llama la atención pues diversos estudios han demostrado que las medidas cognitivas suelen relacionar con diversos criterios de desempeño [24], incluyendo los trabajos comerciales [20]. Además, las escalas cognitivas del Kudert no son ipsativas quedando exentas de las limitaciones inherentes a este tipo de datos, como es una peor validez de criterio [25].

Los creadores de la prueba han dado algunos pasos para mejorar sus propiedades psicométricas, en especial las escalas DISC, reportando buenos niveles de confiabilidad y una mejor estructura factorial [4]. Sin embargo, contrario a lo que afirman los creadores estos hallazgos son evidencia parcial de la validez de constructo del test pues como han destacado varios autores e instituciones especializadas [5, 10, 26, 27] dicha validez se juzga a partir del conjunto de evidencias recolectadas que incluyen la correlación con otros constructos y con variables criterio, sin mencionar el principio de que la validación de constructo es un esfuerzo de investigación permanente [28].

A continuación se ofrecen algunas sugerencias para contribuir con el mejoramiento de la prueba.

Primeramente, es necesario repensar los múltiples usos asignados a la prueba. “Si el autor de un test afirma que la prueba se puede usar para cierto propósito, es esencial que presente pruebas empíricas que sustenten el uso de la puntuación del test para dicho propósito” [29, p. 933]. Según la ficha técnica [1] la prueba tiene cuatro áreas de aplicación (selección, desarrollo de líderes, coaching y equipos de alto rendimiento). El asignar varios usos a una misma prueba es cuestionable para algunos autores porque dispersa el esfuerzo y los recursos en el desarrollo del instrumento elevando los requerimientos de validez de la prueba [30].

Repasar los estándares psicométricos generalmente aceptados internacionalmente y que han sido conjuntamente desarrollados y aprobados por tres instituciones especializadas en evaluación educativa y psicológica [26, 27] es imprescindible para asegurar el desarrollo de un instrumento confiable y válido.

Una publicación que enfoca los esfuerzos de validación en el ámbito laboral, son los llamados “Principios” [10], que tomando como punto de partida los Estándares, ofrecen lineamientos para conducir estudios de validación en ámbitos aplicados y situaciones de selección de personal.

La norma UNE-ISO 10667 (2013) dirigida a proveedores de servicios de evaluación de personas en entornos organizacionales y sus clientes, especifica las guías y procedimientos para asegurar un proceso de evaluación de alta calidad que proteja los derechos y deberes de las personas evaluadas, considerando la trascendencia y riesgo que tienen las decisiones tomadas a partir de los resultados de un test [31].

Una guía más reciente y muy clara para la construcción de un test se encuentra en [32] que detalla los “diez pasos para la construcción de un test” que ofrece además los criterios necesarios para mejorar una prueba existente.

La naturaleza ipsativa de las escalas DISC [3] tiene en el presente caso algunas implicaciones. Estas escalas afectan la independencia de los errores de medida [33]. Calderón & Ximénez [34] explican: “El atributo de un individuo se mide con relación a sus propias puntuaciones en otros atributos, lo que hace que la suma de las puntuaciones de las variables para cada individuo sea constante” (p. 26). Brown, & Maydeu-Olivares [35] han señalado que los cuestionarios de elección forzada si bien ayudan a reducir varios sesgos de respuesta, generan errores de medida no independientes porque las respuestas que un individuo brinda a un ítem dependen de las respuestas que da a los demás ítems del bloque contrariando el supuesto de independencia y afectando con ello la confiabilidad, la validez de criterio y la de constructo. Esto hace que los supuestos y procedimientos de la Teoría clásica de los tests, no se ajusten bien a este tipo de ítems y escalas. Brown, & Maydeu-Olivares [35] proponen que las escalas ipsativas se traten con los supuestos y procedimientos de la Teoría de respuesta al ítem [36] basando el análisis en la llamada “ley del juicio comparativo de Thurstone” [35]. Calderón & Ximénez [34] han demostrado empíricamente que los estimadores clásicos de confiabilidad como el Alfa y el test-retest no deberían ser empleados con los cuestionarios de elección forzada por la violación del supuesto de independencia. Por tanto, las escalas ipsativas del Kudert, al ser de respuesta forzada, deberían basar su construcción psicométrica en la Teoría de respuesta al ítem y no en los supuestos de la Teoría clásica de los tests [34, 35, 37]. Esta conclusión queda reforzada con la reciente publicación de Abad, et al. [25] que examina los nuevos desarrollos metodológicos y tecnológicos para resolver los problemas psicométricos de los tests de elección forzada.

Otro aspecto de mejora se relaciona con el contenido de la Ficha técnica [1] que proporciona información básica pero incompleta sobre la herramienta si se compara con los criterios para evaluar la calidad de los tests [38]. Por ejemplo, la empresa que autorizó el presente estudio recibió las puntuaciones de las diversas dimensiones de la prueba en distintas métricas sin una identificación precisa del tipo de puntajes y escalas

en el que se mostraban los resultados. Así, todos los factores, exceptuando las variables cognitivas, eran escalas de intervalo. En cambio, las variables cognitivas estaban en algún tipo de escala ordinal. Puesto que la ficha técnica no contiene baremos ni estadísticos descriptivos, el usuario carece de parámetros que le permitan indagar sobre la naturaleza de las puntuaciones recibidas. En este aspecto una de las mejores guías para evaluar y desarrollar una prueba de calidad, incluyendo su documentación, la proporciona el Cuestionario de evaluación de tests revisado (CET-R), actualizado por Hernández, et al. [38] que está en la página web del Consejo General de Psicología de España y que se puede descargar directamente de: <http://www.cop.es/uploads/pdf/CET-R.pdf>. El CET-R es un cuestionario que permite hacer una evaluación cuantitativa y cualitativa de cualquier prueba y que también se la puede usar como pauta para mejorar pruebas existentes [32].

Respecto a las limitaciones del estudio y como es frecuente en el análisis de validez concurrente [10,11] el tipo de muestreo fue no probabilístico por conveniencia. Si bien el tamaño de la muestra representaba la casi totalidad de los vendedores de la organización, es claro que por el muestreo y el tamaño no es posible generalizar los resultados quedando confinados a la empresa en la que se hizo el estudio.

El diseño concurrente tiene algunas limitaciones inherentes: las personas evaluadas pueden no ser representativas de los candidatos ya que han sido seleccionadas, entrenadas y tienen experiencia en el desempeño del cargo, lo que restringe especialmente el rango del criterio [39]. Debido a que la correlación del predictor se hace con un criterio actual y no futuro, la capacidad predictiva como tal se la considera relativa [10, 11]. Sin embargo, la restricción de rango también ocurre en un diseño predictivo y una ventaja del diseño concurrente es que los participantes por ya tener el trabajo están menos motivados para distorsionar las respuestas [10].

Si bien los criterios de desempeño provinieron de un análisis del trabajo del cargo de vendedor y fueron considerados relevantes y esenciales por la organización en la que se efectuó el estudio, es posible que el cumplimiento de presupuesto tenga algún grado de contaminación ya que las medidas objetivas de ventas están afectadas no solamente por el desempeño del vendedor, sino también por factores externos sobre los que el ocupante no tiene control [40].

También se puede mejorar la fiabilidad y precisión de la medición de las competencias técnicas con el método del *assessment center* que basa el juicio de evaluación en la calibración de varios observadores [41]. En la presente situación la mayoría de las evaluaciones de las competencias provenían del juicio de un solo evaluador lo que conlleva probablemente uno de los errores de evaluación más habituales: el efecto de halo [42].

Este estudio representa el primer esfuerzo empírico e independiente por examinar la validez de criterio de la prueba Kudert y el resultado evidencia varias oportunidades de mejora entre las que destacan: (a) la necesidad de usar la Teoría de respuesta al ítem como metodología psicométrica para el tratamiento de las escalas ipsativas de la prueba, (b) la realización de estudios de validez de constructo en el que la prueba se correlacione con otros tests sobre la base de hipótesis de convergencia y discriminancia, (c) la investigación de la validez criterial en base a diseños concurrentes y predictivos con variedad de cargos y criterios de desempeño, (d) la mejora del

contenido del material que se ofrece a los usuarios teniendo presente que aun cuando no son generalmente expertos en psicometría [43], las recomendaciones de diversos autores e instituciones especializadas como las citadas en este artículo, son inequívocas: los usuarios de las pruebas deben tener acceso principalmente a la información psicométrica de la prueba y (e) la consideración de la llamada validez consecuencial [28], esto es: los potenciales impactos sociales positivos o negativos que tienen las decisiones tomadas a partir de los resultados de un test en el que la calidad de las decisiones depende finalmente de la disponibilidad de evidencias de validez de la prueba [26].

FUENTES DE FINANCIAMIENTO

Este artículo no fue financiado por ninguna persona o entidad de ayuda económica.

DECLARACIÓN DE CONFLICTO DE INTERÉS

Los autores no tienen conflictos de intereses morales, económicos, laborales o de investigaciones relacionadas al presente estudio.

APORTE DEL ARTÍCULO EN LA LÍNEA DE INVESTIGACIÓN

Es la primera revisión independiente de la herramienta Kudert que se utiliza ampliamente en el contexto organizacional ecuatoriano. La información permite a los usuarios hacer una valoración más equilibrada de las características del instrumento. También contiene sugerencias específicas sobre cómo mejorar la prueba lo que puede ser útil para sus creadores.

DECLARACIÓN DE CONTRIBUCIÓN DE CADA AUTOR

Luis Puertas contribuyó con la recolección de datos en campo, el análisis del cargo, la identificación de los criterios de desempeño y el análisis estadístico inicial. Jaime Moreno integró la información, revisó y replicó los análisis estadísticos efectuados, indagó la literatura y redactó el manuscrito.

AGRADECIMIENTOS

Se agradece a la organización que permitió realizar el levantamiento de información de los predictores y criterios en su fuerza laboral. Sin su apoyo este estudio no hubiera sido posible.

REFERENCIAS

- [1] Ortiz, G., Soto, D., Rivera, C., & Recalde, J. (Enero de 2020). *Ficha Técnica modelo Kudert Ecuador*. Obtenido de Kudert.com.
- [2] Kudert (29 de noviembre de 2015). La ciencia detras de Kudert. UTPL Inscripciones. https://inscripciones.utpl.edu.ec/sites/default/files/la_ciencia_detras_de_kudert_compressed.pdf



- [3] Owen, J.E., Mahatmya, D., & Carter, R. (2020). Dominance, Influence, Steadiness, and Conscientiousness (DISC) Assessment Tool (pp. 1185 – 1189). In Zeigler-Hill, V., & Shackelford, T.K. (Eds.). *Encyclopedia of personality and individual differences*. Cham: Springer. https://doi.org/10.1007/978-3-319-24612-3_301995
- [4] Ortiz, G.J., Rivera, C.M., Recalde, R.J., & López, V.M. (2021). Propiedades psicométricas de un instrumento para medir estilos conductuales en población ecuatoriana. *Revista Electrónica de Psicología Iztacala*, 24(1), 164-181. <https://www.iztacala.unam.mx/carreras/psicologia/psiclin/vol24num1/Vol24No1Art8.pdf>
- [5] Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory* (second edition). Mason, OH: Cengage Learning.
- [6] Castillo, R., & Rosero, J. (2018). *Instituto Nacional de Estadística y Censos, Encuesta Nacional de Empleo y Desempleo Urbano, ENEMDU*. Quito: INEC.
- [7] Seisdedos, N., & Cordero, A. (1996). *Inventario de Personalidad para Vendedores*. Madrid: TEA.
- [8] Brannick, M. T., Cadle, A., & Levine, E. L. (2012). Job analysis for knowledge, skills, abilities, and other characteristics, predictor measures, and performance outcomes (pp. 119 – 146). In Schmitt, N. (Ed.). *The Oxford handbook of personnel assessment and selection*. New York: Oxford University Press.
- [9] Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human performance*, 15(1-2), 187-210. <https://doi.org/10.1080/08959285.2002.9668091>
- [10] Tippins, N., Sackett, P., & Oswald, F. (2018). Principles for the Validation and Use of Personnel Selection Procedures. *Industrial and Organizational Psychology*, 11(S1), 1-97. <https://doi.org/10.1017/iop.2018.195>
- [11] Shultz, K. S., Whitney, D. J., & Zickar, M. J. (2021). *Measurement theory in action: Case studies and exercises*. (3rd ed.). New York, NY: Routledge. <https://doi.org/10.4324/9781003127536>
- [12] Morgeson, F. P., Brannick, M. T., & Levine, E. L. (2019). *Job and work analysis: Methods, research, and applications for human resource management*. Thousand Oaks, CA: Sage Publications.
- [13] Highhouse, S., Doverspike, D., & Guion, R. M. (2015). *Essentials of personnel assessment and selection*. Routledge. <http://45.114.134.178:9000/digi/TF01/2015/TF01T0000030.pdf>
- [14] Arias, K. (2018). *Elaboración de una propuesta de instrumentos de selección basados en competencias para la Empresa Pública de Agua Potable y saneamiento de la ciudad de Quito*. Quito: Pontificia Universidad Católica del Ecuador. <http://repositorio.puce.edu.ec:80/xmlui/handle/22000/17814>
- [15] Sonnentag, S., & Frese, M. (2002). Performance concepts and performance theory. *Psychological management of individual performance*, 23(1), 3-25.
- [16] Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, 9(1), 6. <https://doi.org/10.7275/qf69-7k43>
- [17] Bobko, P. (2001). *Correlation and regression: Applications for industrial organizational psychology and management*. Thousand Oaks, CA: Sage Publications.
- [18] Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781420036268>
- [19] IBM Corp. Released 2015. *IBM SPSS Statistics for Windows, Version 23.0*. Armonk, NY: IBM Corp.

- [20] Vinchur, A. J., Schippmann, J. S., Switzer III, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of applied psychology*, 83(4), 586. <https://doi.org/10.1037/0021-9010.83.4.586>
- [21] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum. <https://doi.org/10.4324/9780203771587>
- [22] Benjamin, D.J., Berger, J.O., Johannesson, M. et al. Redefine statistical significance. *Nature human behaviour*, 2(1), 6-10. (2018). <https://doi.org/10.1038/s41562-017-0189-z>
- [23] Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, 73(sup1), 1-19. <https://doi.org/10.1080/00031305.2019.1583913>
- [24] Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 79(4), 518. <https://doi.org/10.1037/0021-9010.79.4.518>
- [25] Abad, F. J., Schames, R., Sorrel, M., Nájera, P., García-Garzón, E., Garrido, L. E., & Jiménez, M. (2022). Construyendo tests adaptativos de elección forzosa “on the fly” para la medición de la personalidad. *Papeles del Psicólogo*, 43(1), 29-35. <http://www.papelesdelpsicologo.es/pdf/2982.pdf>
- [26] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- [27] Chan, E. K. (2014). Standards and guidelines for validation practices: Development and evaluation of measurement instruments. In *Validity and validation in social, behavioral, and health sciences* (pp. 9-24). Springer, Cham. DOI: 10.1007/978-3-319-07794-9_2
- [28] Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan Publishing Company. <https://psycnet.apa.org/record/1989-97348-002>
- [29] Horstmann, K.T., Knaut, M. & Ziegler, M. (2020). Criterion Validity (pp. 933 – 935). In Zeigler-Hill, V., & Shackelford, T.K. (Eds.). *Encyclopedia of personality and individual differences*. Cham: Springer. https://doi.org/10.1007/978-3-319-24612-3_301995
- [30] Downing, S. M. (2006). Twelve steps for effective test development (pp. 3 – 25). In S.M. Downing & T.M. Haladyna (Eds.). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- [31] Barbero, L. (2014). La norma ISO 10667, herramienta para la evaluación de personas en los entornos laborales. *Capital Humano*, 284, 92-96. <https://dialnet.unirioja.es/servlet/articulo?codigo=4571717>
- [32] Muñoz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1). DOI: 10.7334/psicothema2018.291
- [33] Ximénez, C., & Calderón, C. (2012). El análisis factorial de datos ipsativos: un estudio de simulación. *Psicothema*, 24(2), 302-309. <http://hdl.handle.net/11162/4217>
- [34] Calderón, C., & Ximénez, C. (2014). Análisis factorial de ítems de respuesta forzada: una revisión y un ejemplo. *Revista Latinoamericana de psicología*, 46(1), 24-34. [https://doi.org/10.1016/S0120-0534\(14\)70003-2](https://doi.org/10.1016/S0120-0534(14)70003-2)
- [35] Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, 18(1), 36. <https://doi.org/10.1037/a0030641>
- [36] Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology Press. <https://doi.org/10.4324/9781410605269>
- [37] Brown, A. (2016). Item response models for forced-choice questionnaires: A common framework. *Psychometrika*, 81(1), 135-160. <https://doi.org/10.1007/s11336-014-9434-9>

- [38] Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G., & Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 37(3), 192-197. <https://www.redalyc.org/pdf/778/77847916006.pdf>
- [39] Guion, R.M. (2011). *Assessment, measurement, and prediction for personnel decisions*. New York: Routledge. <https://doi.org/10.4324/9780203836767>
- [40] Borman, W., & Smith, T. N. (2012). The use of objective measures as criteria in I/O psychology (pp. 532 – 542). In Schmitt, N. (Ed.). *The Oxford handbook of personnel assessment and selection*. New York: Oxford University Press.
- [41] Thornton III, G.C., Mueller-Hanson, R.A., & Rupp, D.E. (2017). *Developing organizational simulations: A guide for practitioners, students, and researchers* (second edition). New York: Routledge. <https://doi.org/10.1111/peps.12288>
- [42] Murphy, K.R., Cleveland, J.N., & Hanscom, M.E. (2019). *Performance appraisal & management*. Thousand Oaks, CA: Sage Publications.
- [43] Reynolds, C. R., Altmann, R.A., & Allen, D.N. (2021). *Mastering modern psychological testing: Theory and methods* (second edition). Austin: TX: Springer International Publishing.

NOTA BIOGRÁFICA



Jaime Moreno. **ORCID iD**  <https://orcid.org/0000-0001-7296-8596>
PhD en Psicología por la Pontificia Universidad Católica del Perú. Master en Ciencias en Educación por la Universidad de Kansas, Estados Unidos y licenciado en Psicología Industrial por la PUCE. Profesor Principal e investigador de la Pontificia Universidad Católica del Ecuador en la ciudad de Quito. Sus áreas de interés académico son los sistemas de evaluación y gestión del desempeño, los estilos negativos de liderazgo y el estudio de los riesgos psicosociales y las variables de bienestar psicológico en el trabajo.



Luis Puertas. **ORCID iD**  <https://orcid.org/0000-0003-1076-7021>
Obtuvo su grado de Psicólogo Organizacional en la Pontificia Universidad Católica del Ecuador, tiene una Maestría en Psicología mención Comportamiento humano y Desarrollo Organizacional. Su línea de investigación es Psicología del trabajo. Actualmente es investigador independiente en la ciudad de Quito Ecuador.



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



Moreno & Puertas. Estudio de validez de criterio de un instrumento para medir estilos conductuales en población ecuatoriana.

Enero – Junio 2022

<http://dx.doi.org/10.33210/ca.v11i1.385>

