



Desarrollo de escalas psicométricas para la medición de competencias personales

Development of Psychometric Scales for Personal Competencies Assessment

Marco Vinicio Fournier Facio ^a y Tomás Vargas Halabí ^{b*}

^a Universidad de Costa Rica; ^b Instituto Costarricense de Electricidad (ICE)

(Recibido 11 de enero de 2008; aceptado 12 de marzo 2008)

Resumen

La implementación de los modelos de competencias plantea desafíos organizacionales importantes en el nivel de medición de variables psicológicas, especialmente en las áreas de selección y capacitación. En este artículo, se expone el proceso de diseño de 23 escalas psicométricas, para la evaluación del diccionario de competencias de una institución pública autónoma costarricense. Se describen la estrategia metodológica seguida, los índices de confiabilidad y validez obtenidos, así como los hallazgos con respecto a la relación entre el rendimiento en las pruebas validadas y el desempeño en labores de atención al cliente en una muestra de 60 funcionarios.

Palabras clave: psicometría, pruebas psicológicas, selección de personal, modelos de competencias, evaluación de competencias

Abstract

The implementation of the competency models involves important organizational challenges in regards to the measurement of psychological variables, mainly in the selection and training aspects. This article describes the building process of 23 psychometric scales designed to assess the competency dictionary of a public Costa Rican organization. It describes the methodological approach of the research, the reliability and validity coefficients obtained, as well as findings showing the relationship between scales scores and customer's service tasks performance. The study was conducted in a sample of 60 employees.

Keywords: psychometrics, psychological tests, personnel selection, competency models, and competency assessment

* Correspondencia. Correo electrónico: tvargas@ice.go.cr



Introducción

En el área de la administración, se ha venido enfatizando cada vez más, la importancia de los recursos y capacidades de las organizaciones como base para su éxito (Barney, 2001; Mahoney & Pandian, 1992; Peteraf, 1993; Wernerfelt, 1984); se establece con ello, un balance con los enfoques predominantes durante la década de los ochentas, que resaltaban principalmente la relación entre las condiciones ambientales y el desempeño organizacional (Boxall & Purcell, 2000; Grant, 1991). Esto ha permitido reposicionar la función de recursos humanos como un aspecto crítico para la supervivencia, crecimiento y éxito de cualquier organización (Boxall & Purcell, 2000; Wright, Dunford & Snell, 2001; Wright, MacMahan & MacWilliams, 1994).

Por otra parte, en la actualidad, los modelos de competencias gozan de una amplia difusión como plataforma para los sistemas de gestión de recursos humanos (Horton, 2000; Markus, Cooper-Thomas & Allpress, 2005; Shippmann, Ash, Battista, Carr, Eyde, Hesketh et al., 2000). Este es el caso de la Institución pública, en la cual se realizó el proceso de investigación y desarrollo que se describe en este artículo, cuya Dirección de Recursos Humanos (DRH) definió el enfoque de competencias como su filosofía orientadora.

Diversos autores han señalado los beneficios derivados de implementar la gestión por competencias, por ejemplo Rodríguez, Patel, Bright, Gregory & Gowing (2002) señalan que permiten reclutar, seleccionar y retener personas de alto desempeño, así como crear y mantener una dinámica cultural que fomente el logro. Otros autores han señalado que contribuye a eliminar la falta de flexibilidad característica de los enfoques basados en el análisis ocupacional (Capaldo, Iandoli & Zollo, 2006), orienta a la organización hacia el futuro, facilita la comunicación con los empleados (Hayton & McEvoy, 2006) y les apoya para enfrentar con éxito el cambio (Yeung, 1996).

No obstante, los beneficios proclamados por algunos autores y consultores, es necesario señalar que la implementación de dichos modelos no se encuentra exenta de dificultades y riesgos. Por una parte, plantea importantes desafíos técnicos en cuanto a la identificación, descripción, evaluación y desarrollo de competencias (Capaldo, Iandoli & Zollo, 2006). Por otra, se ha llegado a cuestionar el beneficio de su implementación, por ejemplo Laber y O'Connor (2000), Markus, Cooper-Thomas y Allpress (2005) y Sparrow (1995) señalan que se carece de investigaciones empíricas rigurosas que demuestren los beneficios que usualmente se le atribuyen a la gestión por competencias.

Independientemente de la valoración global que se haga de los modelos de competencias, en el campo de la selección de personal, facilitan la implementación de mejores herramientas de evaluación. Esto es relevante en tanto la selección de personal es, sin duda, una actividad fundamental para cualquier organización (Gatewood & Feild, 2001; Schmidt & Hunter, 1998; Schmidt, Hunter, McKenzie & Muldrow; 1979; Schmidt, Hunter, Outerbridge & Trattner, 1986). En el caso de la Institución pública, en la cual se realizó la investigación descrita en este artículo, se han venido utilizando herramientas como las entrevistas conductuales y las simulaciones en los procesos de selección. Estas metodologías son comúnmente aceptadas como herramientas válidas para la evaluación de competencias (Spencer & Spencer, 1993; Wood & Payne, 1998); no obstante, su aplicación masiva es costosa. Esto ha puesto de manifiesto la necesidad de contar con otro tipo de instrumentos cuyo formato requiera una menor inversión de tiempo para su aplicación.

Por otra parte, las competencias comúnmente se definen como características subyacentes de la persona causalmente relacionadas con un desempeño superior en el trabajo (Yeug, 1996). Dichas características, de acuerdo con Spencer y Spencer (1993), se asocian a variables psicológicas, tales



como motivos, rasgos, actitudes, valores, autopercepciones, conocimientos y habilidades. Desde este punto de vista, resulta consistente con el enfoque de competencias, la evaluación de dichos atributos mediante los métodos y teorías implicados en la medición de variables psicológicas. La disciplina que se encarga de ello, es la psicometría (Muñiz, 2002), razón por la cual se plantea la utilización de pruebas psicológicas desarrolladas a partir de un enfoque psicométrico, como una alternativa para la evaluación de competencias.

Como cualquier instrumento de medición científica, las pruebas psicológicas deben poseer adecuados niveles de confiabilidad y validez (Anastasi & Urbina, 1998; Matesanz, 1997; Muñiz, 2002). Adicionalmente, el contenido debe adaptarse a las necesidades y la cultura organizacional correspondiente, asimismo, sus resultados deben estandarizarse con el fin de contar con normas de interpretación ajustadas a la población específica de interés (Gatewood & Feild, 2001).

Si bien existen pruebas psicométricas en el mercado, estas tienden a medir rasgos de personalidad y habilidades de tipo general que no necesariamente corresponden a la especificidad de las necesidades de cada organización. Adicionalmente, en la mayoría de los casos, no se cuenta con procesos de estandarización en el nivel nacional, y menos aún, de la organización. Por otra parte, muchas de estas pruebas se utilizan en múltiples centros de trabajo, lo cual provoca procesos de aprendizaje en los oferentes, lo que a su vez reduce la exactitud de la medición.

Por todo lo anterior, resulta particularmente importante para la Institución, contar con pruebas diseñadas específicamente para medir, de forma confiable y válida, las competencias establecidas en su diccionario. Con el objetivo de dotar a dicha Institución de tales herramientas, se desarrolló, con el patrocinio y apoyo de la DRH, un proceso de más de tres años de investigación, orientado al diseño de pruebas psicométricas. El objetivo del presente artículo es describir dicho proceso y sus resultados principales.

Estrategia metodológica

El proyecto se desarrolló en varias etapas:

Definición de competencias y construcción de modelos teóricos

En el momento de iniciar el estudio, el diccionario institucional contenía 76 competencias, algunas de las cuales mostraban un importante nivel de traslape entre sí. Dadas las limitaciones de recursos, se procedió a identificar aquellas competencias de mayor uso, para lo cual se contó con la colaboración de dos funcionarios con amplia experiencia en el levantamiento de perfiles de competencias y en procesos de selección. En los casos en los cuales se observó un nivel importante de traslape, se optó por unir las competencias en una nueva dimensión, asimismo, se identificaron tres competencias o rasgos generales no incluidos en el diccionario, pero que se consideraron de relevancia.

Las competencias o rasgos generales identificados fueron: estabilidad emocional (TEE), inteligencia emocional (TIE), capacidad cognitiva (TCG). Las competencias específicas definidas fueron: administración de recursos humanos (TARH), asertividad (TAS), automotivación (TAUT), compromiso con la tarea (TCOTA), comunicación escrita (TCOES), creatividad (TCREA), disposición a compartir información (TCOINF), empatía (TEMP), estilos de dirección (TESDI), exactitud y precisión (TEXAC), flexibilidad (TFLEX), gestión de recursos (TGEREC), iniciativa (TINI), liderazgo (TLIDE), manejo de crisis (TCRISIS), organización (TORG), seguridad en sí mismo (TSEG) y trabajo en equipo (TTEQUIP).



Una vez seleccionadas las 18 competencias específicas y las tres generales, se procedió a elaborar una definición para cada una. En el caso de las primeras se tomó como referencia el contenido del diccionario institucional. Con base en estas definiciones, se realizó una revisión de la literatura, con el fin de elaborar un modelo teórico para cada competencia. Con esta información, se elaboró una tabla de especificaciones o modelo final, el cual describe de manera sintética, las características que poseen las personas con un alto o un bajo nivel en cada competencia, en tres diferentes dimensiones, a saber: cognitiva, emocional e interactiva. En algunos casos, por la naturaleza de la competencia, no fue necesario elaborar el modelo en las tres áreas, por ejemplo, capacidad cognitiva.

Elaboración de las versiones preliminares de las pruebas

El proceso de selección de la Institución requiere que se evalúe a cada candidato en un número importante de competencias, mediante pruebas psicológicas, con el fin de contar con una orientación general que luego se amplía con herramientas de tipo cualitativo. Por ello, el área de selección planteó como requerimiento que las pruebas diseñadas fuesen instrumentos cortos, por lo que se definió un número aproximado de 10 ítems para las competencias específicas y de 20 ítems para los rasgos generales.

Tomando como base el modelo teórico de cada competencia, se procedió a redactar el doble de los ítems que debía contener finalmente cada prueba. Se analizaron diferentes formatos de ítems para cada competencia y se seleccionó el que mejor se ajustó a los usos propios de selección de personal. La gran mayoría de los instrumentos se diseñaron como pruebas de poder, con reactivos de diversos niveles de dificultad. Cuando no fue posible utilizar este enfoque, los instrumentos se diseñaron como cuestionarios o inventarios de personalidad. También se elaboró un modelo matemático para el cálculo de la calificación final de cada una de las pruebas.

En el caso de la competencia capacidad cognitiva, se decidió elaborar tres versiones diferentes de la prueba, con el fin de que se ajustara mejor a las características de los principales grupos ocupacionales de la organización, a saber: profesionales en ingeniería, otros profesionales (por ejemplo en ciencias sociales), técnicos y administrativos.

Una vez redactados los ítems de prueba, se sometieron a la revisión independiente de dos profesionales en psicología de la DRH, conocedores del diccionario institucional. Para tal efecto, se les facilitaron las definiciones y los modelos teóricos diseñados en el proyecto. Adicionalmente, los ítems fueron sometidos a la evaluación de dos profesionales del Programa de Equidad de Género, con el fin de evitar la inclusión de sesgos de género en el vocabulario y en la redacción.

Análisis de las instrucciones, vocabulario y tiempos de las pruebas

Se procedió a seleccionar una muestra de 30 funcionarios de la DRH, bajo la condición de que no fuesen psicólogos, debido al conocimiento que este tipo de profesionales posee sobre los instrumentos psicométricos. El objetivo de esta aplicación fue analizar el nivel de comprensión de las instrucciones, el vocabulario utilizado en los ítems y establecer los tiempos requeridos para responder cada prueba.

Las competencias autonomía, compromiso con la tarea, estabilidad emocional, liderazgo, manejo de crisis y seguridad en sí mismo se evaluaron con pruebas tipo inventario de personalidad, por lo que para hacer menos evidente el constructo medido, se combinaron y aplicaron en un solo instrumento. Tratándose de una cantidad grande de pruebas, la aplicación se realizó en dos sesiones, en cada una se distribuyeron los test, controlando y balanceando los niveles de dificultad



y la duración.

El análisis de la claridad de las instrucciones y del vocabulario permitió identificar una cantidad importante de correcciones, las cuales se incorporaron en cada uno de los instrumentos correspondientes. Por otra parte, se definió el tiempo límite, como una desviación estándar sobre la media del tiempo consumido por los sujetos de la muestra en responder cada prueba. En el caso de las pruebas combinadas en un solo inventario, al no tratarse de pruebas de poder, no se estableció tiempo límite.

Aplicación de las versiones preliminares de las pruebas y recopilación del criterio de los supervisores

Se aplicaron todas las escalas a 558 funcionarios de una Unidad de Negocios de la Institución. Se evaluó prácticamente a la totalidad de los funcionarios con un nivel educativo mínimo de secundaria incompleta. Esta Unidad mantiene instalaciones en la mayoría de las regiones del país.

Para realizar la aplicación de las escalas, se conformó un equipo de colaboradores, a quienes se les informó acerca de los objetivos del estudio y se les entrenó en los procedimientos de aplicación. El equipo estuvo formado por profesionales en administración y en psicología de los departamentos de selección y gestión por competencias, así como por dos estudiantes avanzados de la carrera de psicología. Este equipo tuvo a su cargo la aplicación de las versiones preliminares de las pruebas en todo el país. En todos los casos, la aplicación fue supervisada directamente por un profesional en Psicología. De forma previa a la aplicación, se implementó una estrategia de comunicación, para informar al personal de los objetivos del estudio, aclarando el carácter confidencial de los resultados y enfatizando que estos no tendrían implicaciones administrativas. Posteriormente, se citó al personal con la debida anticipación.

Por diversas razones, algunos funcionarios no pudieron asistir a las dos sesiones, por lo cual el número final de pruebas aplicadas varía ligeramente entre 503 y 558 sujetos para cada una de las escalas. Caso aparte lo constituyen las pruebas de capacidad cognitiva, ya que el número final de aplicaciones dependió del número de funcionarios en cada grupo ocupacional, a saber profesionales en ingeniería con 73, otros profesionales con 50 y técnicos y administrativos con 434.

A continuación, se detallan las características de los sujetos incluidos en la muestra:

Tabla 1. Sexo de la muestra

Sexo	%
Masculino	86.8
Femenino	13.2

Tabla 2. Nivel educativo de la muestra

Nivel educativo	%
Secundaria incompleta	9.5
Secundaria completa	22
Educación técnica	17.7
Universitaria incompleta	26.8
Bachillerato universitario	8.9
Licenciatura	10.7
Maestría	4.3



Tabla 3. Edad de la muestra

Categoría de edad	%
30 años o menos	19.1
De 31 a 40 años	41.1
Más de 40 años	39.7

De forma posterior a la administración de las escalas, el equipo de aplicadores se dio a la tarea de recopilar la evaluación del supervisor en una sub-muestra de 227 funcionarios. Este tipo de criterio ha sido utilizado con frecuencia en la literatura de selección para validar los predictores (Gatewood & Feild, 2001).

Una vez aplicadas las versiones preliminares de las pruebas y recopilado el criterio externo, se procedió a realizar los análisis de confiabilidad y validez, con el fin de obtener la versión final de las escalas. Estos análisis se exponen con detalle en el apartado de análisis de datos y resultados.

Finalmente, para facilitar la interpretación de las pruebas, se transformaron las puntuaciones de acuerdo con el procedimiento sugerido por Fournier y Garnier (1990), de modo que todas las escalas varían de 0 a 100. Para el establecimiento de los baremos, se calcularon percentiles cada 5 puntos, tomando como base los puntajes totales transformados de las versiones finales de las pruebas.

Análisis de la relación entre las versiones finales de las pruebas y el rendimiento en labores de servicio al cliente

El área de servicio al cliente constituye un foco importante de interés para la Institución, por tal razón, se intentó determinar la relación entre el rendimiento en las pruebas una vez depuradas y el desempeño en labores de atención al cliente. Para tal efecto, la Unidad de Negocios de Servicio al Cliente realizó un proceso de evaluación basado en criterios de desempeño previamente definidos, lo cual permitió clasificar a 60 funcionarios en dos grandes categorías: alta y baja aptitud para ejecutar las funciones de servicio al cliente que brinda esta Unidad. En el primer grupo quedaron, seleccionados 39 funcionarios y en el segundo 21.

Valga señalar que la estrategia de grupos contrastados ha sido ampliamente utilizada en la validación de pruebas de personalidad (Anastasi & Urbina, 1998). Por otra parte, el objetivo primordial de los instrumentos de selección consiste en brindar parámetros objetivos para poder estimar el desempeño laboral de los candidatos. Dado lo anterior, resulta de relevancia evaluar la capacidad que poseen un conjunto de pruebas para discriminar entre funcionarios de alto y bajo desempeño. Esto, aunque no ofrece una validación específica de cada escala con respecto al constructo que pretende medir, permite una comprobación adicional de la validez y utilidad de las escalas en conjunto, para la selección de un tipo de ocupación en particular. Esto a su vez, valida uno de los perfiles por competencias que la Institución posee en el área de Servicio al Cliente. Este análisis se discute con detalle en el siguiente apartado.

Análisis de datos y resultados

En este apartado, se exponen los análisis realizados y los resultados obtenidos. Estos fueron calculados con el SPSS (*Statistical Package for the Social Sciences*) versión 10.0.



Análisis de confiabilidad

Para la mayoría de las pruebas, se calculó el *alfa de Cronbach*. Para ello, se fueron eliminando todos aquellos ítems que permitieran incrementar la confiabilidad y reducir el tamaño total del test hasta un mínimo aproximado de 10 o 20 ítems, dependiendo del tipo de escala (competencias específicas o rasgos generales). Para la escala Inteligencia Emocional, fue necesario respetar la representatividad de los diferentes contenidos, con el fin de garantizar que la versión final contuviera las dimensiones contempladas inicialmente, aun cuando esto redujera parcialmente la confiabilidad final. En este caso, se consideró conveniente estimar la confiabilidad por medio del método de división por mitades, puesto que no se trata de un constructo homogéneo.

Para cada escala, se ensayaron diversas opciones de calificación, con el fin de determinar la que garantizara una mayor confiabilidad. En general, se trabajó con diversas estrategias para la diferenciación de los puntajes, según las diferentes categorías de respuesta, así como entre éstas y la “no respuesta”. Tal y como puede observarse en la Tabla 4 en todos los casos, la confiabilidad es superior a .8, lo cual es satisfactorio dado el número de ítems en cada escala.

Tabla 4. Confiabilidad de cada escala

Escalas	Confiabilidad	Número final de ítems
TARH	.83	10
TAS	.84	10
TAUT	.81	13
TCOES	.87	10
TCOTA	.91	15
TCREA	.82	16
TCOINF	.87	13
TEMP	.88	10
TESDI	.84	14
TEXAC	.85	14
TFLEX	.87	10
TGEREC	.89	14
TINI	.94	10
TLIDE	.88	13
TMCRIS	.85	14
TORG	.85	10
TSEG	.87	13
TTEQUIP	.87	14
TEE	.92	20
TIE	.94	22
TCC ingenieros	.9	23
TCC otros profesionales	.86	13
TCC técnicos administrativos	.87	19



Análisis de validez

Relación entre el rendimiento en las pruebas y la evaluación de los supervisores

Para recopilar el criterio externo, se diseñó un formulario que se aplicó a los supervisores, el cual incluía tres ítems para cada competencia. Para el análisis, se sumó el puntaje de estos tres ítems, este resultado se utilizó como criterio externo para cada competencia y se le consideró como una variable continua que varía de 3 a 15. Se asignó en cada uno de los tres ítems, un 5 al sujeto que ocupa el primer lugar, y un 1 al último lugar. De este modo, se tienen 3 ítems por competencia, con un máximo de 5 puntos y un mínimo de 1, lo cual da un total máximo de 15 y un mínimo de 3.

El análisis de la distribución de las calificaciones de la evaluación, realizada por los supervisores, mostró dos tipos de problemas. Por una parte, muchos de los evaluadores tendieron a ser poco consistentes dentro de cada dimensión o competencia evaluada. Por otro lado, y en sentido opuesto, muchos evaluadores tendieron a presentar una respuesta estereotipada.

En el primer caso, para poder evaluar el grado en que dicho problema ocurría, se procedió a calcular la desviación estándar de los tres ítems correspondientes a cada competencia y para cada sujeto evaluado. Muchos de los sujetos mostraron desviaciones que implicaban variación, prácticamente de un extremo a otro, en la posición que le asignaba su jefe dentro de la misma competencia, es decir, dependiendo del ítem, algunos sujetos fueron calificados en una ocasión como los mejores y en otras, como los peores en la misma competencia. Algunas competencias mostraron tendencias más extremas que otras, pero en general, la proporción de sujetos con alta variabilidad es grande. Todo lo anterior evidencia falta de coherencia del supervisor dentro de cada dimensión. Esto podría deberse a dos posibles causas: el supervisor respectivo no respondió con la concentración adecuada, o dentro de la misma dimensión, pretendió medir aspectos bastante diversos.

El segundo problema, es decir, la tendencia a producir una respuesta estereotipada es precisamente la situación opuesta a la observada en el primer problema. En este caso, muchos supervisores tendieron a colocar al evaluado en la misma posición o tan solo a un lugar de diferencia, a través de todas las competencias evaluadas. Esta situación permite considerar dos posibles causas: el evaluador respondió de manera automática, sin considerar el contenido específico de cada ítem y, sobre todo, de cada competencia evaluada, o no logró discriminar adecuadamente entre las diferentes dimensiones evaluadas.

Todo lo anterior implica limitaciones importantes en el proceso de medición del criterio externo, lo cual no es casualidad, ya que han sido una preocupación permanente en Psicología Organizacional, las dificultades inherentes a la medición del rendimiento (Blum & Taylor, 1968; De Quijano, 1992; Dunnette & Hough, 1992). Siendo así, inevitablemente los índices de validez se verán afectados por esta situación.

Con el fin de reducir los efectos de los problemas encontrados, se procedió a establecer varias correcciones, antes de realizar los análisis necesarios para calcular la validez. En primer lugar, se realizó una depuración, con el objetivo de eliminar aquellos sujetos cuyas evaluaciones mostraban mucha variabilidad dentro de cada dimensión, pero también, aquellos sujetos que mostraban, por el contrario, poca variabilidad de una dimensión evaluada a otra. Esta limpieza se realizó de manera específica y diferencial para cada competencia evaluada. Seguidamente, se procedió a calcular la correlación simple entre el puntaje de la prueba y el criterio externo en cada competencia. Dicha correlación se corrigió en función de las confiabilidades de las escalas y del criterio externo, así



como por reducción de la variabilidad (Lord & Novick, 1968). Esto último, por cuanto el proceso de selección de la Institución impone un filtro importante. Los resultados se presentan en la Tabla 5.

Tabla 5. Índices de validez de cada escala según correlación corregida por baja confiabilidad y reducción de la varianza

Escalas	Validez
TARH	.26
TASER	.33
TAUTOM	.25
TCOES	.26
TTAREA	.28
TCREA	.26
TCOINF	.20
TEMP	.31
TESDI	.31
TEXAC	.24
TFLEX	.35
TGEREC	.33
TINI	.58
TLIDER	.30
TCRISIS	.21
TORG	.33
TSEGUR	.24
TTEQUIP	.37
TESTAB	.39
TIE	.43
TCC	.46

Tal y como puede observarse, a pesar de las limitaciones con el criterio externo, la mayoría de los índices de validez tienden a mostrar valores aceptables. Sin embargo, el proceso de limpieza del archivo tiende a reducir considerablemente el número de sujetos (aproximadamente 100 por competencia), por lo que los índices observados deben analizarse con mucho cuidado.

Se utilizó otra opción para calcular la validez sin reducir tanto el número de sujetos. Tal y como se ha anotado, una posible hipótesis para explicar la variabilidad a lo interno de cada dimensión evaluada consiste en la posibilidad de que el supervisor haya establecido una discriminación sutil a través de los tres ítems de cada dimensión, de modo que no esté midiendo un concepto homogéneo. Esta situación es contemplada por la Teoría de Respuesta al Ítem, la cual propone la existencia de diversos rasgos latentes en toda dimensión cuantificada, así como la necesidad de contemplar el conjunto de estos rasgos, para poder establecer un análisis preciso de la contribución de cada ítem y su respectivo poder discriminativo. Al mismo tiempo, esta teoría contempla la necesidad de conceptualizar cada ítem por separado (Osterlind, 2006). Siendo así, es posible establecer una medición de la validez, en donde la correlación no se realice entre el criterio externo y el total del test, sino entre el criterio externo y cada uno de los ítems de la escala.



Para tal fin, se calculó la variabilidad total para cada evaluado, seguidamente, se seleccionaron aquellos evaluados que presentaban una variabilidad media, comprendida entre los valores .65 y 1.3 desviaciones estándar. En este caso, el proceso de depuración fue menos exigente, de modo que al final se trabajó con 152 sujetos para las pruebas generales. En el caso de las pruebas de capacidad cognitiva, se trabajó con 25 ingenieros, 22 profesionales de otras carreras y 105 técnicos o administrativos.

El cálculo de la validez se estableció mediante dos métodos alternativos: regresión múltiple y análisis de varianza. La diferencia entre ambos radica en cómo se clasifiquen los ítems como escala de medición. Si se consideran continuos, la regresión es el análisis más adecuado. En este análisis, se incluyó cada uno de los ítems calificados de la prueba como una variable independiente o predictora del puntaje obtenido en el criterio externo. Si por el contrario, se clasifican como discretos, entonces se considera más adecuado el análisis de varianza. En este caso, cada ítem calificado se incluye como una variable categórica, de tal forma que se utilizan todos los ítems como variables independientes. En esa situación se utilizó el análisis de varianza múltiple univariado (varias variables independientes discretas – ítems - y una variable independiente continua – criterio externo-).

En el caso del análisis de regresión múltiple, se obtuvo el R cuadrado y en el caso del análisis de varianza el Eta cuadrado. Ambos estadísticos indican la cantidad de variabilidad del criterio externo que logran explicar todos los ítems juntos. Calculando la raíz cuadrada de cada uno, se obtiene una estimación de la correlación múltiple entre los ítems de la prueba y el criterio externo. En ambos casos, se estableció una corrección de la validez mediante la fórmula de atenuación por baja confiabilidad (Lord & Novick, 1968). Además, en el caso de la regresión múltiple, se estableció una corrección por disminución de variabilidad por selección (Lord & Novik, 1968). Para las pruebas de capacidad cognitiva, puesto que se tienen tres versiones diferentes, el número de sujetos para las dos primeras resulta demasiado pequeño, por lo tanto, se procedió a establecer un promedio de los índices de validez obtenidos para cada una de ellas. Estos resultados se muestran en la Tabla 6.



Tabla 6. Validez según las dos opciones de análisis

Escalas	Validez con regresión	Validez con <i>Anova</i>
TARH	.33	.48
TAS	.60	.58
TAUT	.40	.67
TCOES	.54	.54
TCOTA	.44	.60
TCREA	.43	.81
TCOINF	.31	.53
TEMP	.35	.51
TESD	.50	.47
TEXAC	.41	.49
TFLEX	.43	.68
TGEREC	.36	.49
TINI	.38	.59
TLIDE	.40	.70
TMCRIS	.33	.68
TORG	.33	.45
TSEG	.42	.69
TTEQUIP	.38	.34
TEE	.60	.87
TIE	.41	.72
TCC (PROMEDIO DE - EG-CCING-A, - EG-CCPROF-A, - EG-CCTEC-A)	.53	.62

Tal y como puede apreciarse, los índices de validez superan en todos los casos la correlación simple entre el criterio externo y el total del test. Además, estos ítems han sido calculados con una muestra muy superior en cuanto al número de sujetos, excepto en los casos de capacidad cognitiva para ingenieros y capacidad cognitiva para otros profesionales, razón por la cual los análisis deben considerarse como preliminares.

Relación entre el rendimiento en las pruebas y el desempeño en labores de atención al cliente

En vista de que la clasificación de los funcionarios realizada por la UEN de Servicio al Cliente es una variable dicotómica, mientras que los totales de los tests se distribuyen como variables continuas, se realizó un análisis de regresión logística, utilizando como variable dependiente, la clasificación de los funcionarios, y como independientes, los totales de las 23 escalas.

Ahora bien, la subrutina correspondiente del SPSS versión 10.0 no cuenta con el método paso a paso (*stepwise*), por lo que se procedió a calcular la regresión logística utilizando dos acercamientos complementarios.



En el primer caso, se realizaron pruebas *t* de *Student* para la diferencia entre promedios de muestras independientes, lo que permitió identificar nueve competencias en las que ambos grupos se diferencian de manera estadísticamente significativa (Tabla 7). En todos los casos, el grupo con alto desempeño mostró promedios más altos (Tabla 8). Lo anterior implica, de por sí, una relación importante entre la evaluación realizada y la aptitud de los funcionarios en las labores de atención al cliente. En otras palabras, los sujetos considerados con desempeño superior en su centro de trabajo obtienen calificaciones estadísticamente superiores a los de bajo desempeño. Interpretando el nivel de significancia de las diferencias obtenidas, fue posible determinar un orden específico para su entrada en la regresión logística calculada con el método *enter*.

Tabla 7. Pruebas de significancia para la diferencia de promedios entre el grupo de alto y el de bajo desempeño

Escalas	<i>t Student</i>	<i>p</i>	Diferencia entre medias
TGEREC	3.910	.000	15.3665
TTAREA	3.305	.002	9.3974
TLIDER	2.807	.007	8.5045
TSEGUR	2.272	.030	8.0188
TCOES	2.173	.034	9.5640
TEXAC	2.037	.046	10.6966
TCREA	1.935	.062	9.1142
TTEC	1.862	.068	8.4276
TINI	1.824	.073	6.1957



Tabla 8. Promedio y Desviación Estándar para las competencias seleccionadas en los grupos de alto y bajo rendimiento

Escalas	Grupos	N	Promedio	DE
TGEREC	Alto rendimiento	38	51.9737	15.8584
	Bajo rendimiento	21	36.6071	11.4027
TTAREA	Alto rendimiento	38	78.6932	10.1955
	Bajo rendimiento	21	69.2957	10.9242
TLIDER	Alto rendimiento	38	56.6568	10.3633
	Bajo rendimiento	21	48.1524	12.4527
TSEGUR	Alto rendimiento	38	82.9979	10.0061
	Bajo rendimiento	21	74.9790	14.3623
TCOES	Alto rendimiento	38	84.4592	14.7360
	Bajo rendimiento	21	74.8952	18.5791
TEXAC	Alto rendimiento	38	57.0013	24.2696
	Bajo rendimiento	21	46.3048	15.9269
TCREA	Alto rendimiento	38	57.7303	13.5928
	Bajo rendimiento	21	48.6161	19.0683
TTEC	Alto rendimiento	38	66.1776	16.8928
	Bajo rendimiento	21	57.7500	16.1816
TINI	Alto rendimiento	38	85.3347	11.1277
	Bajo rendimiento	21	79.1390	14.6854

Con el método *enter*, el modelo de nueve competencias resulta significativo ($X^2 = 29.603$, $p < .01$). Se obtuvo un coeficiente de determinación de .542 (R^2 de Nagelkerke) para una correlación múltiple de .74, lo cual es relativamente positivo, puesto que se logra explicar, mediante los tests involucrados, más de la mitad de la variabilidad en la clasificación establecida por el centro de trabajo. Si se compara la pertenencia actual de los funcionarios en los dos grupos establecidos de acuerdo con el desempeño, con la clasificación que se lograría utilizando las nueve pruebas, se observó que en el grupo de funcionarios de alto desempeño, que son los que más interesa seleccionar, se lograría acertar su clasificación en el 92.1% de los casos, mientras que en el grupo con baja aptitud, el porcentaje de acierto es de 66.7%. Esto conduce a una clasificación correcta general del 83.1% de los casos. Debe advertirse que estos valores resultan siempre inflados por el hecho de que la ecuación de clasificación se establece precisamente con los sujetos que luego van a ser clasificados (Hair, Anderson, Tatham & Black, 1999); sin embargo, los resultados obtenidos son realmente satisfactorios.

De acuerdo con las recomendaciones de la UEN de Servicio al Cliente, se recalculó el modelo mediante el método *enter*, incluyendo además de las nueve competencias identificadas inicialmente, otras que se consideraron relevantes, a saber: asertividad, empatía, automotivación, organización, manejo de crisis, inteligencia emocional, disposición a compartir información. El modelo resultó significativo ($X^2 = 39.319$, $p < .001$), se obtuvo un coeficiente de determinación de .668 (R^2 de Nagelkerke) y una correlación múltiple de .817. El porcentaje de clasificación correcta para el



grupo de alto desempeño fue de 94.7% y para el de bajo desempeño, 81%, para un total promedio de 89.8%.

El segundo acercamiento al análisis de regresión logística se realizó con el método hacia delante (*forward*), en el cual cada prueba entra en el modelo, si logra el nivel de significancia preestablecido (.05), una vez eliminada la influencia de los otros tests que ya han entrado en el análisis. Tómese en cuenta que las escalas poseen relación entre sí, por lo cual el peso que pueda tener sobre la predicción, podría estar contemplado en otros tests que ya han entrado en el análisis, de modo que su contribución termina sobrando, al estar ya contabilizada por las otras pruebas con las cuales tiene relación. Siendo así, aunque el test en cuestión establezca diferencias significativas entre ambos grupos de acuerdo con la prueba *t de Student*, su contribución, en combinación con otras escalas, puede perder importancia. Este tipo de análisis permite reducir y simplificar el modelo, aunque no necesariamente lo hace más preciso, y puede perderse información relevante con el nivel cualitativo. Por esta razón, se han incluido en el análisis, tanto este método como el *enter*.

Con el método hacia delante (*forward*), se observan las mismas tendencias, sin embargo, puesto que este toma en cuenta las intercorrelaciones entre los tests, y de estas se deriva un análisis para determinar el grado de importancia de cada uno, los resultados en este caso resumen el modelo a solo dos competencias ($X^2 = 23.176, p < .001$): gestión de recursos y compromiso con la tarea. Estos resultados no implican que los tests excluidos carezcan de relación con la calidad del rendimiento de los funcionarios, sino más bien, la información que estas escalas proporcionan, es compartida, en parte, por los que sí entraron en el análisis, por lo que su contribución no tiene un peso estadístico importante. En este caso, el coeficiente de determinación baja a .446 (R^2 de Nagelkerke), para una correlación múltiple de .67, que sigue siendo alta, pero menor a la obtenida con los dos modelos calculados con el método *enter*. Los porcentajes de clasificación correcta también tienden a disminuir. En este caso, se obtuvo un 84.2% de acierto para los sujetos con alta aptitud y un 66.7% para los de baja aptitud, con un 78% de clasificación general correcta.

Finalmente, se calcularon los costos y se estimó la utilidad derivada del uso de las pruebas durante el primer año, de acuerdo con el procedimiento establecido por Schmidt, Hunter, McKenzie y Muldrow (1979) y los parámetros sugeridos por Schmidt y Hunter (1983), con el fin de validar el proyecto frente a los tomadores de decisiones. Los datos de este análisis no se presentan por razones de confidencialidad, sin embargo, el beneficio estimado durante el primer año supera varias veces los costos de desarrollo de los instrumentos.

Conclusiones

Los modelos de competencias, con independencia de cómo se juzgen, plantean importantes desafíos en relación con la medición de variables psicológicas, especialmente en las áreas de empleo y formación. En el caso de la Institución, el proyecto desarrollado ha permitido dotar a la organización de herramientas de medición, con índices de confiabilidad y validez satisfactorios, de las principales competencias definidas en su diccionario. Asimismo, se ha sentado una base para el ajuste y actualización continua de los baremos de las pruebas. Si bien deben continuarse realizando estudios técnicos con el fin de poder utilizar estos instrumentos para la toma de decisiones, esta experiencia constituye un avance importante en el área de evaluación de competencias.

Por otra parte, en el nivel técnico, la utilización de la evaluación del supervisor como criterio externo no ha estado exenta de dificultades. En primer lugar, a pesar del plan de comunicación desarrollado, la discusión nacional en torno a la posible apertura de algunos servicios que brinda



la Institución, generó temores en relación con la evaluación. En segundo lugar, la gran cantidad de competencias evaluadas plantea exigencias cognitivas importantes para los supervisores. En este sentido, una alternativa para continuar con la validación de las herramientas es constituida por la utilización de la metodología de grupos contrastados. Esta reduce sensiblemente las demandas que se les plantean a los supervisores, permite validar las pruebas para diferentes ocupaciones con muestras más pequeñas, a la vez que se validan empíricamente los perfiles de puestos levantados.

Finalmente, la contribución estimada del proyecto en términos económicos supera varias veces los costos incurridos. Esto es crítico frente a los tomadores de decisiones, especialmente en el contexto que enfrentan actualmente las instituciones públicas. Un beneficio adicional de gran relevancia lo constituye el proceso de aprendizaje interno generado, el cual ha redundado en mejoras, en las prácticas de medición, en diferentes áreas de la gestión de recursos humanos.

Referencias

- Anastasi, A. & Urbina S. (1998). *Test psicológicos*. México: Prentice Hall.
- Barney, J. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of Management*, 27, 643-650.
- Blum, M. & Taylor, J. (1968). *Industrial psychology. Its theoretical and social foundations*. New York, USA: Harper & Row.
- Boxall, P. & Purcell, J. (2000). Strategic human resource management: Where have we come from and where should we be going?. *International Journal of Management Review*, 2(2), 183-2003.
- Capaldo, G., Iandoli, L. & Zollo, G. (2006). A situationalist perspective to competency management. *Human Resource Management*, 45(3), 429-448.
- De Quijano, S. (1992). *Sistemas efectivos de evaluación del rendimiento: Resultados y desempeños*. Barcelona: PPU.
- Dunnette, M. & Hough, L. (1992). *Handbook of industrial organizational psychology*. Palo Alto, California: Consulting Psychologists Press.
- Fournier, M. & Garnier, L. (1990). Fórmula abreviada para la transformación de puntajes de variables a una misma escala. *Revista Latinoamericana de Psicología*, 12(3), 421-426.
- Gatewood, R. & Feild, H. (2001). *Human resource selection*. Forth Worth: Harcourt Publishers.
- Grant, R. (1991). The Resource-Based theory of competitive advantage: implication for strategy formulation. *California Management Review*, 33(3), 114-135.
- Hair, J., Anderson, R., Tatham, R. & Black, W. (1999). *Análisis multivariante*. Madrid, España: Prentice Hall.
- Hayton, J. & McEvoy, G. (2006). Competencies in practice: an interview with Hanneke C. Frese. *Human Resource Management*, 45(3), 495-500.
- Horton, S. (2000). Introduction – competency movement: its origins and impact on the public sector. *The International Journal of Public Sector Management*, 13(4), 306-318.
- Laber, M. & O'Connor, J. (2000). Competency modeling: ready, set, research. *The Industrial Organizational Psychologist*, 37, 91-96.
- Lord, F. & Novick, M. (1968). *Statistical theories of mental test scores*. Massachusetts, USA: Addison-Wesley.



- Mahoney, J. & Pandian, R. (1992). The resource-based view within the conversation of strategic management. *Strategic Management Journal*, 13(5), 363-380.
- Mantesanz, A. (1997). *Evaluación estructurada de la personalidad*. Madrid, España: Pirámide.
- Markus, L., Cooper-Thomas, H. D. & Alpress, K. (2005). Confounded by competencies? Anevaluation of the evolution and use of competency models. *New Zealand Journal of Psychology*, 34(2), 117-126.
- Muñiz, J. (2002). *Teoría clásica de los test*. Madrid, España: Pirámide.
- Osterlind, S. (2006). *Modern measurement. Theory, principles, and applications of mental appraisal*. New Jersey, USA: Pearson.
- Peteraf, M. (1993). The cornerstone of competitive advantage: a resource - based view. *Strategic Management Journal*, 14, 179-191.
- Rodríguez, D., Patel, R., Bright, A. Gregory, D. & Gowing, M. (2002). Developing Competency models to promote integrated human resource practices. *Human Resource Management*, 41 (3), 309-324.
- Schmidt, F. & Hunter, J. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology*, 68(3), 407-414.
- Schmidt, F. & Hunter, J. (1998). The validity and utility of selections methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Schmidt, F., Hunter, J., McKenzie, R. & Muldrow, T. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology*, 64(6), 609-626.
- Schmidt, F., Hunter, J., Outerbridge, A. & Trattner, M. (1986). The economic impact of job selection methods on size, productivity and payroll costs of federal work force: an empirical based demonstration. *Personnel Psychology*, 36, 1-29.
- Shippmann, J., Ash, R., Battista, M., Carr, L., Eyde, L., Hesketh, B., Kehoe, J., Pearlman, K., Prien, E. & Sánchez, J. (2000). The practice of competency modeling. *Personnel Psychology*, 53, 703-740.
- Sparrow, P. (1995). Organizational Competencies: A valid approach for the future?. *International Journal of Selection and Assessment*, 3(3), 168-177.
- Spencer, L. & Spencer, S. (1993). *Competence at work. Models for superior performance*. New York, USA: John Wiley & Sons.
- Wernerfelt, W. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5(2), 171-180.
- Wood, R. & Payne, T. (1998). *Competency based recruitment and selection. A practical guide*. Chichester, England: John Wiley & Sons.
- Wright, P., Dunford, B. & Snell, S. (2001). Human resources and the resource-based view of the firm. *Journal of Management*, 27, 702-721.
- Wright, P., McMahan, G. & McWilliams, A. (1994). Human resources and sustained competitive advantage: a resource-based perspective. *International Journal of Human Resource Management*, 5(2), 302-326.
- Yeung, A. (1996). Competencies for HR professionals: an interview with Richard E. Boyatzis. *Human Resource Management*, 35(1), 119-131.



Reseña de los autores

Marco Vinicio Fournier Facio

Catedrático retirado de la Universidad de Costa Rica (UCR), realizó sus estudios universitarios en la misma institución donde obtuvo el grado de licenciatura y máster en Psicología, lugar donde laboró como docente por 30 años. Asimismo, fungió como miembro activo de dicha institución educativa, donde coordinó diversos proyectos de investigación, en el área de mercadeo, opinión pública, educación, violencia, política y juventud, autor de diversas publicaciones a nivel nacional e internacional.

Tomás Vargas Halabí

Psicólogo y administrador de negocios, realizó sus estudios universitarios en la Universidad de Costa Rica, donde obtuvo el grado de bachiller y licenciatura en Psicología. Posteriormente obtuvo su maestría en Administración de Empresas en el Instituto Tecnológico de Costa Rica (ITCR), actualmente es estudiante activo del Doctorado en Dirección de Empresas en la Universidad de Valencia, obteniendo el Diploma de Estudios Avanzados (DEA), encontrándose en periodo de realización de su tesis doctoral. Ha laborado para el Instituto Costarricense de Electricidad (ICE), Grupo de Análisis y Consultoría A.C., Programa de Evaluación del ITCR, Coopesa R.L., Banco de Costa Rica, ha sido docente de la UCR y del ITCR.