



Artículo

Tests de Juicio Situacional: introducción y revisión de la investigación

Situational Judgment Tests: introduction and research review

FILIP LIEVENS¹

Fecha de Recepción: 10-10-2006

Fecha de Aceptación: 19-02-2007

RESUMEN

Este artículo tiene dos objetivos principales. El primer objetivo es exponer los detalles prácticos de los tests de juicio situacional (TJS), describiendo qué son los TJS, cuál es su historia, cuáles son las características principales y cómo se desarrollan. El segundo objetivo es debatir la creciente evidencia empírica que apoya el uso de los TJS, señalando cuáles son las fortalezas y debilidades empíricas de los TJS. El artículo termina presentando algunos de los caminos claves para la investigación futura.

ABSTRACT

This paper has a twofold aim. Firstly, it will provide an outline of situational judgement tests (SJT), including SJT development and main features. Secondly, empirical evidence supporting the use of the JST will be discussed, showing what their empirical strengths and weaknesses are. Finally, some suggestions are given for future research.

PALABRAS CLAVE

Test de juicio situacional, Test basado en video, Validez, Impacto adverso, Validez de criterio.

KEY WORDS

Situational judgment test, Video-based test, Validity, Adverse impact, Validity of criteriy.

¹ Ghent University, Bélgica.

TESTS DE JUICIO SITUACIONAL: INTRODUCCIÓN Y REVISIÓN DE LA INVESTIGACIÓN

En los últimos años, varios de los clientes del Departamento de Reclutamiento solicitaron una información más detallada sobre las competencias interpersonales de los candidatos. En concreto, estas compañías querían disponer de tal información ya en la primera etapa de selección. No obstante, no querían invertir en un costo centro de evaluación (assessment center). Es más, eran muy escépticas con respecto al uso de los inventarios de personalidad del tipo auto-informe. Por tanto, la oficina de reclutamiento estaba buscando desesperadamente una forma eficiente de medir las competencias interpersonales en las etapas preliminares de selección.

Las Fuerzas Armadas de Estados Unidos se enfrentaban con el problema de un alto índice de renuncias entre los oficiales recién empleados. Aparentemente, los oficiales contratados tenían una visión muy idealizada del Ejército como el empleador. Por esa razón, las Fuerzas Armadas querían incluir en la página web dedicada al reclutamiento una presentación realista del trabajo. Así, en la página web se colocó un test donde a los candidatos potenciales a oficiales se les presentaban unas situaciones cortas relacionadas con el liderazgo militar (incluyendo fotos) y se les preguntaba que harían si se enfrentasen a ellas. Les fue proporcionado feedback sobre cada respuesta correcta.

En muchos países, los predictores predictores cognitivos se usan habitualmente en los exámenes de admisión a las universidades y a los cursos de postgrado. Aunque estos tests cognitivos son predictores válidos del rendimiento académico, un

inconveniente es que también presentan un gran impacto adverso. Por esta razón, se están buscando pruebas estandarizadas que amplíen las competencias evaluadas y que puedan ser aplicadas a grandes grupos de personas al mismo tiempo.

Estas tres situaciones son solamente unos ejemplos que podrían ilustrar por qué los tests de juicio situacional (TJS) están ganando cada vez más popularidad en los procesos de selección de personal en los Estados Unidos (Chan y Schmitt, 2002; Weekle y Jones, 1999). Asimismo, las investigaciones empíricas sobre los TJS han estado floreciendo en los últimos años. Esto también ha sido demostrado con la publicación del primer libro sobre los TJS (Weekly y Ployhart, 2006).

A pesar de este desarrollo práctico y de la investigación, los TJS no han tenido gran impacto en la práctica de selección de personal en Europa. Una de las razones podría ser que los profesionales no están muy familiarizados con los TJS como procedimiento selectivo (Ryan, McFarland, Baron, y Page, 1999). Otra razón podría ser que los profesionales están familiarizados con los TJS, pero son escépticos con respecto a su efectividad.

Sea cual sea la explicación correcta, hay una clara necesidad de dar a conocer los TJS a los profesionales en Europa. Por eso, el objetivo de este artículo es doble. Primero, exponer los detalles prácticos de los TJS. ¿Qué son los TJS? ¿Cuál es su historia? ¿Cuáles son sus características principales? ¿Cómo se les desarrolla? Segundo, debatir la creciente evidencia empírica que apoya el uso de los TJS. En otras palabras, ¿cuáles son las fortalezas y debilidades empíricas de los TJS? se terminará el artículo presentando algunos de

los caminos claves para la investigación futura.

TESTS DE JUICIO SITUACIONAL. INTRODUCCIÓN

Definición y Ejemplos de los TJS

Los TJS son instrumentos que plantean a los candidatos situaciones laborales hipotéticas y muestran las posibles soluciones a éstas. Los candidatos tienen que decidir cuál de las respuestas alternativas elegirían. Los TJS se usan habitualmente para seleccionar a los individuos, aunque también pueden utilizarse para objetivos tales como la formación o el desarrollo (Fritzsche, Stagl, Salas, y Burke, 2006).

A continuación se presenta un ejemplo de un ítem de un TJS, a saber, uno que fue incluido en un vídeo-TJS para los candidatos al puesto de enfermería (Weekley and Jones, 1997, p.38-39): *Una auxiliar de enfermería entra en el comedor y observa al hijo de uno de los pacientes recién ingresados, que parece estar muy triste. [La escena empieza cuando la auxiliar de enfermería entra en el comedor y se acerca al hijo del residente.]*

Sandy: ¿Señor Andrews? ¿Está usted bien?

Arnie: [Evidentemente ha estado llorando, pero intenta serenarse] Oh, no sé... Esta es una de las cosas más difíciles que he tenido que hacer. Me siento muy culpable de que mi padre esté aquí.

Sandy: Parece que él está bien aquí. Se está adaptando bien.

Arnie: Pero no está tan cómodo como lo estaba en casa.

Sandy: Me doy cuenta de que es un momento difícil para usted, pero estoy segura que esto saldrá bien cuando su padre se acostumbre a estar viviendo aquí.

Arnie: No sabía que otra cosa hacer, sé que el no podía quedarse más tiempo sólo, pero me siento avergonzado y triste por su situación... (El guión se acaba en este momento).

Si tú fueras la auxiliar de enfermería, ¿qué harías con respecto al hijo del residente?

a) Decirle que personalmente te ocuparás de su padre.

b) Presentarle a otros familiares de los residentes que se enfrentan la misma situación.

c) Sugerirle que hable con un trabajador social.

d) Darle un poco de privacidad.

Breve historia de los TJS

Los TJS tienen una larga historia. De acuerdo con Weekle, Ployhart, y Holtz (2006), el primer TJS fue utilizado durante los exámenes del Servicio Civil de EE.UU. en 1873. Un ejemplo de ítem era: «Un banco le pide asegurar protección para un dispositivo, como una marca registrada, que usarán en sus documentos. ¿Cómo respondería a esa solicitud?» Hay también rastros del uso de los TJS durante la Segunda Guerra Mundial. Por ejemplo Ansbacher (1941) mencionó que se utilizaron ítems tales como el mencionado a continuación: “Tu equipo deportivo está

planeando un viaje a Berlín para asistir al partido de fútbol Alemania-Inglaterra, que tendrá lugar dentro de dos semanas. Te han confiado todos los preparativos y la gestión del viaje. ¿Qué harías?” (p. 381). En la misma línea, los psicólogos del ejército estadounidense intentaron evaluar el juicio de los soldados durante la Segunda Guerra Mundial. Estas pruebas de juicio consistieron en guiones con varias respuestas alternativas. Las soluciones estaban basadas más en la habilidad de cada persona de recurrir a su sentido común, experiencia y conocimiento general, que en el razonamiento lógico.

En los años 1940, cierto número de TJS fueron desarrollados para medir el potencial de supervisión, para medir la habilidad de los supervisores para funcionar efectivamente en las situaciones que requerían decisiones que implicaban personas, o para medir el conocimiento de los supervisores y el entendimiento de las relaciones humanas en la industria. A fines de los 1950 y a principios de los 1960, los TJS eran utilizados por las grandes organizaciones como parte de la batería de tests selectivos para predecir el éxito directivo.

No obstante, el uso de estos prototipos no fue común hasta que la versión moderna fue “reinventada” por Motowidlo, Dunette, and Carter (1990). Estos autores examinaron el uso de los TJS para seleccionar a los directivos de primer nivel. Esta medida planteaba a los candidatos una serie de situaciones laborales descritas de manera verbal y varias alternativas de cómo actuar en cada situación. Los candidatos tenían que elegir la respuesta con la que estarían más y menos de acuerdo en cada situación. (Para una revisión más detallada de la historia de los TJS, véase

McDaniel, Morgeson, Finnegan, Campion, y Braverman, 2001).

Características principales de los TJS

Un análisis de los ejemplos de ítems muestra las *características comunes* de los TJS (véase también McDaniel y Nguyen, 2001; Weekley, Ployhart, y Holtz, 2006). Primero, normalmente los TJS no se desarrollan para medir a priori algunos rasgos específicos subyacentes. Esto los diferencia de las herramientas de selección más tradicionales, tales como los tests de habilidad cognitiva o los inventarios de personalidad. En cambio, los TJS dan una muestra de la conducta representativa de un área de interés, la cual puede ser un puesto específico (p. ej. de vendedor). Esto implica también que la consistencia conductual es la principal base teórica de los TJS. Esto es, la suposición que el rendimiento del candidato durante la prueba de selección (TJS) será coherente y por tanto será predictivo con respecto a su desempeño en el puesto.

Como segunda característica común a todos, los TJS plantean a los candidatos las situaciones relacionadas con el trabajo. La noción de que las situaciones presentadas están relacionadas con el puesto para el que el candidato está siendo considerado, incrementa la relevancia y el realismo de los ítems de los TJS. No obstante, los TJS pueden variar en la fidelidad o el realismo con el que presentan los estímulos (p. ej. las situaciones). La “fidelidad de la tarea estímulo” se refiere al grado hasta que la forma de la tarea estímulo es coherente con la manera en que se encuentra dicha situación en el lugar de trabajo. En este punto debe hacerse una distinción por un lado entre

los TJS de papel y lápiz o escritos y los de tipo vídeo o multimedia por otro lado. Teniendo en cuenta lo anterior, un TJS se califica como un *test escrito* cuando el guión está presentado en formato escrito y a los candidatos se les pide elegir e indicar una alternativa de respuesta apropiada en el cuadernillo adjunto al test. Por lo tanto, los TJS de papel y lápiz o escritos tienen una baja fidelidad del estímulo. En cambio, un típico test en vídeo consiste en un número de escenas de vídeo, cada una de ellas seguida por una pregunta cerrada. Cada escena representa una manera individual de solucionar una situación típica para un puesto concreto. En el crítico “momento de la verdad”, la escena se acaba y al candidato se le pide que elija entre varias maneras de actuar. Las preguntas y las opciones de respuestas se presentan de manera visual y se acompañan de una narración. Los TJS *multimedia* son básicamente iguales que los TJS basados en vídeo. La única excepción es que el VCR ya no sirve más como medio de comunicación. En su lugar, mediante los CDs o DVDs, se utilizan fragmentos informatizados de películas como video-clips cortos. Claramente, los TJS basados en vídeo y multimedia permiten que el contexto del ítem sea descrito detalladamente, aumentando de este modo la fidelidad del estímulo.

Tercero, los TJS tienen frecuentemente un formato de ítems de respuesta múltiple. Esto significa que los candidatos tienen que elegir una alternativa entre un listado de diferentes opciones de respuesta. De nuevo, estas alternativas de respuesta pueden ser presentadas bien de forma escrita (baja fidelidad de respuesta) o en vídeo (fiabilidad de respuesta media). En cualquiera de los casos, no se pide a los candidatos que demuestren su comportamiento

real. Esta característica diferencia los TJS con respecto a los estímulos de alta fidelidad, tales como los centros de evaluación (assessment centers) o las muestras de trabajo, que dan a los candidatos la posibilidad de responder de la manera que reflejaría su comportamiento laboral verdadero, observado y evaluado por unos asesores entrenados.

Otras modalidades de respuesta a los TJS también merecen atención. En algunos casos, la respuesta del candidato a una situación dada determina la siguiente situación a presentar. Así, los candidatos se enfrentan con las consecuencias de sus elecciones. Esto significa que todos los candidatos no responden a los mismos ítems. Estos TJS se llaman “ramificados” o “anidados” o “interactivos”. Una posibilidad tecnológica de desarrollar un TJS interactivo es inherente a los TJS multimedia que presentan a un candidato varios fragmentos de vídeo en función de las respuestas del mismo a un fragmento del vídeo anterior. Esto permite a los TJS estimular la dinámica de interacción, manteniendo a la vez un cierto grado de estandarización. Además de la ramificación, otra modalidad de respuesta es que las instrucciones sobre las contestaciones alternativas pueden tener formato con tendencia conductual (“¿Qué es lo que probablemente harías?”) o formato de conocimiento (“¿Cuál es la mejor respuesta?”).

Una característica final de los TJS es que la matriz de soluciones (respuestas correctas) se desarrolla a priori. Esto significa que no hay asesores o jueces para evaluar a los candidatos. Por el contrario, los TJS se puntúan de acuerdo con la opinión de los especialistas o sobre una base empírica. Se discutirá esta cuestión a lo largo del apartado siguiente.

En conclusión, estas características principales de los TJS clarifican los paralelismos y diferencias entre los TJS y los mejor conocidos tests basados en muestras de trabajo, tales como los centros de evaluación (assessment centers). Las similitudes entre los TJS y los centros de evaluación (assessment centers) incluyen que ambos se basan en el principio de la consistencia conductual y los principios de fidelidad psicológica. Además, ambos son métodos basados en las muestras de trabajo o de conducta que pueden captar una multitud de constructos, incluso cuando no fueron pensados originalmente para medir un constructo específico. Las diferencias clave incluyen que los TJS pueden ser aplicados a grandes grupos y también por Internet, mientras que los centros de evaluación (assessment centers) se usan habitualmente para muestras más pequeñas y en lugar concreto. Como se mencionó anteriormente, los centros de evaluación (assessment centers) son unas simulaciones de alta fidelidad debido a que los asesores observan y evalúan el comportamiento real del candidato mientras dura dicho comportamiento. En contraste, los TJS son simulaciones de baja fidelidad debido a que los candidatos deben elegir una respuesta correcta entre un conjunto limitado de respuestas predeterminadas. Por consiguiente, la estandarización está contenida en las características de los TJS porque todos deben responder a las mismas situaciones (con la excepción de los TJS ramificados) y la matriz de soluciones se determina a priori.

Desarrollo de los TJS

Tal y como describen Motiwidlo y colaboradores (1990), el desarrollo de los TJS habitualmente sigue tres etapas. La primera etapa del desarrollo de un TJS es

un análisis del puesto, durante el cual se recogen incidentes críticos del puesto proporcionados por expertos en la materia en cuestión (los EMC) (p. ej. ocupantes del puesto, supervisores, clientes) o, en algunos casos, de las fuentes de archivo. Los incidentes críticos son relatos que destacan los comportamientos excepcionalmente buenos o malos en las situaciones que se enfrentan realmente en el puesto. Una vez recogidos los incidentes críticos, el creador del test los agrupa en las áreas de contenido similar, elige guiones representativos para cada una de estas áreas y los edita para que tengan una longitud y un formato similar.

En el segundo paso, se pide al mismo grupo o a uno diferente de EMC o a un grupo de los empleados que generen una o más *respuestas* a cada situación. Generalmente, estas personas tienen que indicar que harían ellas con más probabilidad y que actuación creen que sería la mejor. Los EMC son útiles porque deberían ser capaces de identificar las mejores respuestas y también de generar algunas alternativas menos óptimas. Los empleados inexpertos son útiles porque ofrecen respuestas que representan una amplia gama de efectividad (laboral). Después de haber recogido las respuestas alternativas, el creador edita todas las alternativas de respuestas de tal modo que cada ítem disponga de un amplio rango de efectividad (laboral). Habitualmente se mantienen sólo cuatro alternativas de respuesta, aunque esta es sólo una regla general.

En el último paso, se desarrolla la *matriz de soluciones*. Aunque en la literatura que concierne los TJS se discuten varios métodos de puntuación, aquí comentaremos solamente los enfoques más utilizados: el basado en las opiniones de

los expertos y el empírico. Los TJS son a menudo desarrollados racionalmente (p. ej. derivados de las opiniones de los expertos) utilizando el grupo de los EMC o de los empleados excelentes. Estos expertos juzgan la efectividad de varias respuestas o identifican las mejores y las peores opciones. Las opciones identificadas como “las mejores” se puntúan como correctas y las opciones identificadas como “los peores” se puntúan como incorrectas.

Una posibilidad alternativa de puntuar es emplear una matriz de soluciones desarrollada empíricamente. Con este fin, se aplica el TJS a una gran muestra piloto. A continuación se seleccionan los ítems (o las respuestas alternativas) y/o se ponderan según el grado de evidencia de que estos ítems (o las respuestas alternativas) diferencien entre las personas que puntuaron a diferentes niveles en la variable criterio (pe. desempeño en el puesto de trabajo). Las opciones de respuestas que fueron seleccionadas con relativa frecuencia por los individuos que demuestran un alto rendimiento en el criterio serán puntuadas como correctas. Las opciones seleccionadas con relativa frecuencia por los individuos de bajo rendimiento serán puntuadas como incorrectas.

El uso de los EMC en todas las etapas del proceso de desarrollo de un TJS tiene dos ventajas principales. Primero, el realismo de la parte central de cada ítem y de las alternativas de respuesta es más alto cuando en el desarrollo participan los EMC además de los creadores del test. Y segundo, un grupo numeroso de los EMC puede generar un amplio campo de respuestas para cada situación dada.

Para ilustrar el desarrollo de los TJS, se describirá el proceso de desarrollo de un

TJS en vídeo, que forma parte del Examen de Admisión Flamenco para “Estudios Médicos y Dentales” en Bélgica en 2000 (Lievens y Coetsier, 2002; Lievens, Buyse, y Sackett, 2005). El objetivo del TJS original en vídeo era medir otras aptitudes diferentes de la capacidad cognitiva, como las habilidades interpersonales y de comunicación. Por eso, el TJS consistía en viñetas cortas grabadas en cinta de vídeo sobre las situaciones interpersonales claves con las que se encuentran los médicos tratando a los pacientes. En el primer paso del proceso de desarrollo, se recogió una muestra representativa de los incidentes críticos de médicos expertos y de profesores de medicina general. Luego, se examinó la literatura relevante y a otro grupo de expertos (médicos experimentados y profesores de medicina general) se le solicitó aportar ejemplos de comportamientos laborales eficaces e ineficaces en las situaciones correspondientes. A continuación, se escribieron los guiones, que representaban palabra por palabra el diálogo de las partes involucradas. Se examinó el realismo de las escenas y, cuando hizo falta, se reescribieron los guiones. Se contrataron actores semiprofesionales y se grabó en vídeo su interpretación de los guiones en un estudio de grabación. Para garantizar el realismo, un médico experto ayudó en el plató. Después, se extrajeron las preguntas (los ítems situacionales) y las respuestas a partir de las actuaciones grabadas en vídeo y de los comportamientos críticos. Y finalmente, se pidió a otro panel de expertos que desarrollase las normas de puntuación. Su tarea era leer los guiones, observar las actuaciones grabadas en vídeo e indicar de manera independiente la mejor respuesta para cada una de las preguntas. Los expertos llegaron a un acuerdo satisfactorio y se resolvieron todas las discrepancias mediante una discusión (p. ej.

cambiando la pregunta o una de las alternativas de respuesta) para conseguir la regla de puntuación.

PUNTOS FUERTES Y DÉBILES DE LOS TJS. RESUMEN BASADO EN LA INVESTIGACIÓN

Desde que Motowidlo y colaboradores (1990) “reinventaron” los TJS, un número considerable de estudios ha sido dedicado al estudio de la efectividad de los mismos. Por consiguiente, tenemos una imagen bastante buena de las fortalezas y debilidades de los TJS, aunque la evidencia empírica disponible todavía es menos extensa que con respecto a otros predictores tales como los tests de capacidad cognitiva y los inventarios de personalidad. En este apartado, se presenta una visión general sobre los beneficios y problemas relacionados con los TJS, basándose en un extenso resumen de la investigación empírica sobre los TJS. La efectividad de los TJS se revisó según los siguientes criterios psicométricos “duros” y “blandos”, ampliamente conocidos: validez relacionada con el criterio, validez añadida, validez relacionada con el constructo, utilidad, impacto adverso, percepción de los candidatos, falseamiento y susceptibilidad a los efectos de la práctica y del entrenamiento.

VALIDEZ RELACIONADA CON EL CRITERIO

La pregunta clave en cualquier práctica selectiva es si el procedimiento selectivo es capaz de predecir los criterios laborales. En este sentido, varios estudios examinaron si los TJS son buenos predictores del desempeño en el trabajo. McDaniel y

colaboradores (2001) llevaron a cabo el primer meta-análisis en el ámbito laboral (basándose en 95 estudios) de la validez de los TJS relacionada con el criterio. Concluyeron que los TJS son buenos predictores del desempeño laboral ($\rho=.34$ con 102 correlaciones). Además, una considerable variabilidad en los coeficientes relacionados con el criterio encontrada en los estudios sugirió la presencia de los moderadores. El moderador clave de la validez de los TJS fue la aplicación o no del análisis de puesto para desarrollar el TJS. Los TJS basados en el análisis de puesto mostraron valores de validez más altos con respecto a los que no estaban basados en el análisis de puesto (.38 frente a .29). Además de la buena validez de los TJS en el ámbito laboral, las investigaciones recientes también han demostrado que los TJS pueden ser buenos predictores en los contextos educacionales, como parte de los exámenes de admisión (Lievens y colaboradores, 2005; Oswald, Schmitt, Gillespie, Ramsay, y Kim, 2004).

Hay tres supuestos generales que explican por qué los TJS predicen el desempeño laboral (Motowidlo y colaboradores, 1990). La primera explicación indica que el mejor predictor del comportamiento futuro es el comportamiento pasado (es decir, el principio de consistencia conductual). El segundo supuesto, muy relacionado, es que los TJS miden las intenciones y metas de los candidatos (“la teoría del establecimiento de metas”). La explicación final es que los TJS miden constructos tales como conocimiento tácito, inteligencia práctica, capacidad cognitiva general o rasgos de personalidad (véase más adelante) que han mostrado ser predictores del desempeño laboral importantes, robustos o útiles. En la actualidad, existe todavía una considerable disputa sobre los

mecanismos exactos que vinculan a los TJS con el desempeño laboral.

La cuestión de si los TJS escritos son mejores predictores del desempeño laboral que los basados en vídeo (multimedia) es muy interesante. De un lado, los TJS basados en vídeo o multimedia tienen mayor fidelidad debido a que la información que presentan es más rica y más detallada, lo que lleva al mejor encaje con el comportamiento-criterio tal y como éste ocurre en el trabajo. Esto debería resultar en una validez de criterio más alta. No obstante, por otra parte, como la capacidad cognitiva es un importante predictor del desempeño laboral, los TJS basados en vídeo y multimedia pueden tener menos validez ya que no requieren de comprensión de lectura, teniendo así menos carga cognitiva. Y además, estos pueden también introducir información contextual irrelevante y aumentar el error dentro de los TJS, resultando en una validez más baja. Dos estudios intentaron examinar estas afirmaciones. Lievens y Sackett (2006) demostraron que el cambio de un TJS basado en vídeo por uno escrito (manteniendo el contenido invariable) redujo substancialmente la validez de criterio del test. Además, la versión escrita tuvo una relación significativamente más alta con la capacidad cognitiva. En otro estudio, Chan y Schmitt (1997) encontraron que los TJS basados en vídeo tienen un impacto adverso significativamente más bajo con respecto a las personas de raza negra que los TJS escritos. Además, las personas reaccionaron más favorablemente a los TJS basados en vídeo.

VALIDEZ AÑADIDA

Aparte de la validez de cada uno de los predictores, es crucial tanto teórica como

prácticamente examinar la validez predictiva de los TJS sobre otros predictores establecidos. Desde un punto de vista teórico, existe una necesidad de entender si los TJS miden en realidad constructos parecidos o diferentes. Desde un punto de vista práctico, el saber si añadir un TJS explicará una varianza adicional en el criterio puede ayudar a afinar el proceso de selección. Llamada de manera habitual validez añadida, el uso de predictores adicionales tiene valor, desde el punto de vista de la utilidad, solamente cuando añade una varianza explicada adicional al criterio, más allá de aquella que aportaron otros predictores más baratos (Schmidt y Hunter, 1998).

Los resultados demuestran que los TJS explican una varianza adicional en la predicción del desempeño laboral por encima de la capacidad cognitiva, los conocimientos de puesto, la experiencia laboral, y los Cinco Grandes factores de personalidad (Chan y Schmitt, 2002; Clevenger, Pereira, Weichmann, Schmitt, y Schmidt-Harvey, 2001; McDaniel et al., 2001; Weekly y Jones, 1997, 1999). Esto significa que el rendimiento futuro del candidato será predicho con más exactitud si una batería de tests compuesta por un test de capacidad cognitiva y un inventario de personalidad se complementa con un TJS. Estos resultados fueron reproducidos en los entornos educativos para predecir el rendimiento académico (Lievens et al., 2005; Oswald et al., 2004).

VALIDEZ RELACIONADA CON EL CONSTRUCTO

A lo largo de los años, varios constructos fueron relacionados con los TJS. Según Wagner y Sternberg (1985), el pro-

pósito de un TJS es medir algo diferente de la inteligencia académica (capacidad cognitiva). Sugirieron que los TJS miden “conocimiento tácito” o “inteligencia práctica” (es decir, el conocimiento práctico que habitualmente no está expresado o declarado abiertamente y que tiene que ser adquirido en ausencia de la instrucción directa). Otras investigaciones no apoyan esta posición y revelan que los TJS están relacionados con la capacidad cognitiva. En el meta-análisis de McDaniel y colaboradores (2001) se encontró que los TJS muestran una correlación significativa moderada ($r=.46$) con la capacidad cognitiva, aunque detrás de esta estimación existe una variabilidad considerable. Por ejemplo los TJS basados en vídeo tuvieron la correlaciones con la capacidad cognitiva más bajas que los TJS escritos (Weekley y Jones, 1997). Otro ejemplo es que los TJS basados en el análisis de puesto están habitualmente más relacionados con la capacidad cognitiva general que los que no están basados en el análisis de puesto (.50 frente a .38). Aun así, otros investigadores proponen que los TJS son medidas alternativas de *conocimiento de puesto*, de la *experiencia laboral* o de las *variables interpersonales*. Por ejemplo, McDaniel y Ngyuen (2001) encontraron correlaciones de pequeñas a moderadas entre los TJS y la experiencia laboral y algunos de los Cinco Grados Factores de personalidad: Concienciación, Estabilidad Emocional y Amigabilidad (Weekley y Jones, 1999).

Resumiendo, el grado con el que los TJS cubren diferentes constructos parece variar mucho. Esto no es una sorpresa, ya que los ítems de los TJS pueden referirse a un amplio abanico de situaciones e incluir diferentes tipos de contenido a los que los candidatos tienen que estar atentos cuando toman las decisiones. Además, las res-

puestas a los TJS con opciones múltiples son resultado de una combinación de capacidad, experiencia y personalidad. Por lo tanto, un TJS se puede conceputar mejor como un método de medición con el que se puede evaluar una variedad de conocimientos, habilidades y capacidades laborales (CHCO) (McDaniel et al., 2001; Weekley y Jones, 1999), que como un método que evalúa cualquier constructo de diferencias particulares individuales. La naturaleza multidimensional de los TJS también se evidencia por el análisis factorial de los ítems que éstos contienen. Tal investigación de los TJS basada en el análisis factorial revela habitualmente una plétora de factores difíciles de interpretar (Schmitt y Chan, 2006). Por consiguiente, la mayoría de los TJS tienen bajos coeficientes de consistencia interna.

UTILIDAD

Una de las críticas a los TJS está relacionada con sus altos costes. Evidentemente, los costes del desarrollo de los TJS son considerables. Una fuente de costes de desarrollo de los TJS proviene de la relación de los mismos con un puesto concreto; lo que requiere la repetición del procedimiento de desarrollo para cada uno de los TJS específicos, lo que consume mucho tiempo y es costoso. Además, los TJS basados en vídeo y multimedia implican la creación de los guiones, la contratación de los actores (semi-)profesionales, el rodaje y la edición de estos cortometrajes.

Al contrario que los costes de desarrollo, los costos de administración de los TJS escritos son bajos. Para los TJS basados en vídeo y multimedia los costes administrativos son más altos, ya que se

tienen que hacer las inversiones tecnológicas (VCR, los PC) para administrarlos.

A pesar de estos costes de desarrollo y administración, se debería tener en cuenta que los TJS (escritos o basados en vídeo) pueden utilizarse para evaluar grandes grupos de candidatos al mismo tiempo. Esta es una ventaja muy importante respecto a las entrevistas y los centros de evaluación (assessment centers). No obstante, no se tiene consciencia de ninguna investigación empírica que haya examinado la utilidad del uso de los TJS.

IMPACTO ADVERSO

Este problema trata la cuestión de si los TJS son instrumentos imparciales o si grupos concretos (p. ej. las personas de raza blanca, hombres, y los candidatos jóvenes) obtienen de manera sistemática resultados más altos. Algunos estudios examinaron los efectos de la raza y del sexo en el rendimiento durante la aplicación de los TJS.

Con respecto a la raza, las diferencias en las puntuaciones promedio de los TJS entre los subgrupos raciales son normalmente menores que las encontradas para los tests de capacidad, incluyendo los de capacidad cognitiva. Aunque las personas de raza blanca obtienen consistentemente puntuaciones más altas en los TJS, las diferencias entre las personas de raza blanca y negra son sustancialmente menores que la diferencia habitual de una desviación típica completa, encontrada en el caso de los tests de capacidad cognitiva. En realidad, en el meta-análisis de Nguyen, McDaniel, y Whetzel (2005) se encontró una diferencia de puntuación media entre las personas de raza blanca y negra de .38

de desviación estándar a favor de los candidatos blancos. La correlación de los TJS con la capacidad cognitiva explicó casi toda la varianza de las diferencias entre las medias de los grupos raciales a través de los estudios. De este modo, las diferencias de los subgrupos, entre las personas de raza blanca y negra, parecen reducirse considerablemente por la evaluación a través de los TJS de un amplio rango tanto de capacidades cognitivas como no cognitivas, tal como las habilidades interpersonales. Además, los TJS basados en vídeo parecen tener un menor impacto adverso que los TJS escritos, ya que los TJS basados en vídeo tienen una menor carga cognitiva (Chan y Schmitt, 1997).

Con respecto al sexo, las mujeres parecen puntuar levemente mejor en los TJS que los hombres. En su meta-análisis, Nguyen y los colaboradores (2005) encontraron una diferencia en los resultados promedios entre mujeres y hombres de .10 de la desviación típica a favor de las mujeres. Esta sutil tendencia de género puede deberse a las diferencias en los rasgos de personalidad entre los sexos, desencadenadas por las situaciones presentadas en los TJS, que a menudo tienen una naturaleza interpersonal. En general, las mujeres tienden a puntuar más alto en los rasgos tales como Conciencia y Amigabilidad.

PERCEPCIONES DE LOS CANDIDATOS

En general, los candidatos prefieren las herramientas de selección que perciben como relacionadas con el trabajo. Esta es una de razones por las que las muestras de trabajo normalmente reciben valoraciones favorables (Hausknecht, Day, y Thomas, 2004). En este sentido, no es sorprendente

que la investigación sobre las reacciones de los candidatos a los TJS mostrara que fueron percibidas favorablemente y que los formatos en vídeo resultaron ser incluso más apreciados que los formatos escritos (p. ej. Chan y Schmitt, 1997). Además, Richman-Hirsch, Olson-Buchanan, y Drasgow (2000) mostraron que un TJS multimedia tiene más validez aparente y es considerado más ameno y moderno que las formas computerizadas y escritas del mismo.

Las reacciones positivas de los candidatos son substanciales porque juegan un papel importante en los procesos selectivos y de evaluación en cuanto a la motivación hacia la prueba y en su desempeño (Chan y Schmitt, 1997). Además, las reacciones positivas de los candidatos a las pruebas de selección contribuyen a la reputación de la organización percibida por los mismos (Richman-Hirsch et al., 2000).

FALSEAMIENTO

Debido a que los TJS son estímulos de baja fidelidad y utilizan el formato de auto-informe, es relevante examinar hasta que punto son susceptibles a la distorsión de la respuesta (p. ej. fingimiento del candidato de ser "bueno"). El problema de falseamiento de los TJS, ha empezado recientemente a ganar atención en la literatura. Hooper, Cullen, y Sackett (2006) resumieron la evidencia empírica disponible y descubrieron que las diferencias en las puntuaciones medias entre los candidatos a quienes se les pidió de responder honestamente y los a quienes se pidió "fingir" varía entre .08 y .89 desviaciones estándar. También concluyeron que los efectos de falseamiento de los TJS son menores que en el caso de las medidas de personalidad.

Curiosamente, Hooper y los colaboradores (2006) también identificaron varios moderadores que podían facilitar el falseamiento de los TJS y que podrían explicar grandes diferencias a través de los estudios. En primer lugar, cuanto mayor era la carga cognitiva de los TJS, menos fueron susceptibles al falseamiento (Peeters y Lievens, 2005). En segundo lugar, los ítems más transparentes fueron más susceptibles al falseamiento. En tercer lugar, el tipo de las instrucciones sobre las respuestas fue un factor clave, debido a que afectaba a la carga cognitiva y el tamaño de la distorsión de la respuesta en los TJS (Nguyen, Biderman, y McDaniel, 2005). Instrucciones con tendencia conductual (p. ej. "¿Qué es lo que probablemente harías?") mostraron las correlaciones más bajas con la capacidad cognitiva y menor impacto adverso, pero fueron más susceptibles al falseamiento que las instrucciones basadas en el conocimiento (p.ej. "¿Cuál es la mejor respuesta?"). Finalmente, el tipo de diseño del estudio que se utilizó desempeñó un papel importante. Los resultados de los experimentos del laboratorio fueron lo peor que podría pasar en comparación con lo que suele ser el falseamiento en los procesos de selección en la vida real. Los diseños tales como el experimento de laboratorio manipulan el falseamiento e investigan si los candidatos *pueden* fingir durante la prueba. Esto no es la misma cuestión que si los candidatos realmente fingen durante la prueba en los procesos de selección verdaderos.

SUSCEPTIBILIDAD A LOS EFECTOS DE LA PRÁCTICA Y DEL ENTRENAMIENTO

Cuando un procedimiento de selección de personal empieza a ser popular, se

puede asumir que los candidatos participarán en cursos de entretenimiento comerciales y adoptarán estrategias con el fin de mejorar sus resultados, incrementando de este modo sus oportunidades de ser seleccionado. Este último asunto suscita la cuestión clave: ¿Los resultados de los TJS pueden mejorarse a través del entrenamiento?

Hasta ahora solamente un estudio ha abordado este problema. Cullen, Sackett, y Lievens (2006) examinaron la susceptibilidad al entrenamiento de los TJS desarrollados para su consideración como instrumentos selectivos durante los exámenes de alto impacto (en este caso, los del proceso de admisión a la Universidad). Se crearon las estrategias para aumentar la puntuación en cada test y, mediante un programa de entrenamiento basado en vídeo, los estudiantes universitarios fueron entrenados en el uso de dichas estrategias. Los resultados indicaron que algunos de los TJS fueron susceptibles al entrenamiento. Estos resultados demuestran que se debería ser cauteloso con respecto al uso de los TJS en el contexto de exámenes de alto impacto.

Una cuestión semejante es si los candidatos pueden mejorar sus puntuaciones cuando vuelvan a realizar los TJS. Otra vez, la investigación al respecto es escasa. Lievens, Buyse, y Sackett (2005b) demostraron que los efectos de volver a realizar los TJS no fueron mayores que los de las pruebas tradicionales, tales como los tests de capacidad cognitiva. Junto a ello, un moderador importante es la opción de desarrollar formas alternativas de los TJS. Lievens y Sackett (en prensa) compararon varias aproximaciones al desarrollo de formas alternativas de los TJS que se diferenciaban en términos de similitud de los

ítems que incluían. La aproximación basada en una pequeña similitud entre formas alternativas de los TJS (p. ej. asignación al azar de los ítems entre las varias formas alternativas de los TJS), tuvo menor efecto en la repetición de la prueba.

SUGERENCIAS PARA FUTURAS INVESTIGACIONES

Aunque los TJS han demostrado ser predictores válidos en los dominios laborales y educativos, estamos comenzando a entenderlos mejor. Por eso, un hueco importante en la investigación existente es cómo las diferentes características de los TJS influyen en su eficacia. Ya se han emprendido los primeros pasos en este camino. Tal y como ha sido descrito con anterioridad, las investigaciones previas (Chan y Schmitt, 1997; Lievens y Sackett, 2006; Nguyen et al., 2005) ya han identificado el nivel de fidelidad del estímulo (el formato escrito frente al basado en vídeo) y el tipo de las instrucciones sobre las respuestas (basadas en el conocimiento frente a las basadas en la tendencia conductual) como factores clave a la hora de determinar la carga cognitiva de los TJS. No obstante, esta investigación solo roza la superficie. Por lo tanto, se necesitan más estudios que investiguen la influencia de otras características de los TJS sobre variables relevantes, tales como el impacto adverso y la validez. Algunos ejemplos de tales características serían: el uso de los ítems ramificados en contraposición a no ramificados, el tipo de EMC (compañeros, supervisores, clientes) utilizados para generar los incidentes críticos, el nivel de especificidad de los ítems, el nivel de interdependencia entre las situaciones presentadas por las preguntas, o la extensión de las últimas.

Segundo, se necesita aumentar nuestro conocimiento de por qué los TJS predicen la conducta laboral. Tal y como ha sido descrito anteriormente, las investigaciones anteriores ante todo examinaron la capacidad cognitiva, la experiencia y la personalidad como antecedentes del rendimiento durante la realización de los TJS. De nuevo, esto es sólo el principio. Recientemente, el conocimiento procedimental y las políticas de rasgos implícitos han sido propugnados como dos plausibles explicaciones alternativas de por qué los TJS predicen el comportamiento laboral (Motowidlo, Hooper, y Jackson, 2006). Esto puede abrir un abanico de posibilidades para una mayor investigación teórica con respecto a los TJS.

Si se quiere que los TJS realmente hagan avances en las prácticas de selección internacionales, una tercera y crucial área de investigación es su capacidad de ser transportable entre las diferentes culturas (Lievens, 2006). Es decir, ¿pueden los TJS desarrollados en una cultura ser transportados y utilizados como un predictor válido en otra? Aunque esta pregunta puede ser respondida afirmativamente en el caso de los tests de capacidad cognitiva (Salgado, Anderson, Moscoso, Bertua, y De Fruyt, 2003; Salgado, Anderson, Bertua, De Fruyt, y Rolland, 2003), se duda de que este sea el caso de los TJS. Por ejemplo, Such y Schmidt (2004) examinaron la validez del mismo TJS en varios países. Dicho TJS fue válido en la mitad de los países, en concreto en el Reino Unido y Australia. En cambio, no fue predictivo en México. La posibilidad de generalizar los TJS a otros contextos puede resultar peligrosa si los TJS fueron utilizados en contextos diferentes (p. ej. laborales, organizacionales, culturales) y para un criterio diferente para el

que originalmente fue pensado. Esto conduce a dos implicaciones. Primero, la interpretación de la respuesta conductual correcta o apropiada a una situación específica presentada en un TJS podría variar en función de los valores culturales. En otras palabras, la matriz de soluciones podría variar de una cultura a otra. Segundo, los TJS podrían tener una validez diferencial entre las distintas culturas si los resultados de los TJS no se ajustan a los datos del criterio recogidos en otra cultura. En las aplicaciones interculturales de los TJS, la adaptación de la matriz de soluciones a una cultura ajena podría ser una manera de ajustar los predictores y los criterios. Es necesaria más investigación para verificar este supuesto. Hasta ahora, ningún estudio ha explorado las diferencias culturales en términos de los enlaces entre las situaciones, opciones de las respuestas, o el tipo de constructo de las opciones de respuesta de los TJS.

CONCLUSIONES

El objetivo de este artículo era presentar los TJS, incluyendo sus características principales, junto con sus puntos fuertes y sus puntos débiles. Puesto que se pretende fomentar el uso de los TJS en la práctica de selección, se finaliza este trabajo resumiendo la evidencia empírica de la efectividad de los TJS con respecto a otras herramientas de selección de reconocido prestigio. También se recapitula en la Tabla 1 la evidencia empírica existente.

Probablemente el punto más fuerte de los TJS es que tienen unos niveles de validez predictiva significativos, relativamente comparables a los de las medidas de selección ampliamente usadas, tales como los centros de evaluación (assessment cen-

Tabla 1.
Revisión de la Investigación sobre los TJS

Tema	Resultados de la Investigación	Fuente
Validez Relacionada con el Criterio	<ul style="list-style-type: none"> a. Los TJS tienen una correlación media con el rendimiento laboral de .34 b. La validez de los TJS fue establecida en varios entornos. c. Los TJS basados en video son más válidos que los de tipo escrito. 	<p>McDaniel et al. (2000)</p> <p>Lievens y Sackett (2006)</p>
Validez Añadida	<ul style="list-style-type: none"> d. Los TJS explican varianza adicional sobre de la explicada por la capacidad cognitiva, la personalidad, los conocimientos del puesto y la experiencia. 	<p>Chan y Schmitt (2002); Clevenger et al. (2001); Weekley y Jones (1997, 1999)</p>
Validez Relacionada con el Constructo	<ul style="list-style-type: none"> e. Los TJS son métodos de medición para evaluar una variedad de constructos. f. Análisis factorial de los TJS muestra que estos son multidimensionales. g. Los TJS tienen baja consistencia interna. 	<p>Schmitt y Chan (2006)</p>
Utilidad	<ul style="list-style-type: none"> h. No existe investigación al respecto. 	
Impacto Adverso	<ul style="list-style-type: none"> i. Las personas de raza blanca puntúan .38 <i>SD</i> mejor que las personas de raza negra. j. Las mujeres puntúan .10 <i>SD</i> mejor que los hombres. k. Los TJS basados en video tienen un menor impacto adverso que los TJS escritos. 	<p>Nguyen, McDaniel, y Whetzel (2005)</p> <p>Chan y Schmitt (1997)</p>
Percepciones de los Candidatos	<ul style="list-style-type: none"> ▪ Los TJS basados en video (multimedia) están mejor percibidos que los TJS escritos. 	<p>Chan y Schmitt, (1997); Richman-Hirsch et al. (2000)</p>
Falseamiento	<ul style="list-style-type: none"> ▪ El falseamiento puede mejorar el resultado de un TJS desde .08 hasta .89 <i>SD</i>. ▪ El falseamiento tiene un menor impacto en los TJS que en los inventarios de personalidad. ▪ La transparencia de los items, la carga cognitiva del TJS, el tipo de las instrucciones sobre las respuestas y el tipo del diseño de estudio moderan el falseamiento de los TJS. 	<p>Hooper, Cullen, y Sackett (2006)</p>
Susceptibilidad a la Práctica/al Entrenamiento	<ul style="list-style-type: none"> ▪ Los TJS son susceptibles a los efectos de la práctica y del entrenamiento. 	<p>Cullen, Sackett, y Lievens (2006)</p>

ters) y las entrevistas conductuales estructuradas. Además, los TJS muestran validez añadida sobre los tradicionales tests de capacidad cognitiva y de personalidad. Esto significa que somos más capaces de predecir el desempeño laboral del candidato, si a una batería de selección tradicional añadimos un TJS. Además, las reacciones de los candidatos son positivas debido a que perciben la relación de los TJS con el puesto de trabajo, y los TJS tienen un menor impacto adverso hacia las minorías que los tests de capacidad cognitiva (especialmente si la carga cognitiva de los TJS es baja). Finalmente, los TJS pueden ser empleados varias veces para un puesto, y lo que es más importante, para evaluar grandes grupos de candidatos al mismo tiempo. En cuanto a sus puntos débiles, debería destacarse que los TJS podrían ser susceptibles al falseamiento, a la práctica y a los efectos del entrenamiento (aunque a un menor grado que los inventarios de personalidad). Además, los TJS son instrumentos aplicables a un contexto específico, lo que hace necesario que se tengan que desarrollar para puestos y culturas específicos.

En suma, los TJS parecen combinar las ventajas de los tests de capacidad cognitiva y de personalidad, (p. ej. administración a grupos y resultado automático), y los de las entrevistas conductuales estructuradas y de los centros de evaluación (p. ej. correspondencia exacta con el puesto, un menor impacto adverso, y reacciones positivas de los candidatos). No obstante, cabe destacar que no se propone un reemplazamiento de las otras herramientas de selección más tradicionales por los TJS. En vez de ello, se expone que los TJS deberían ser considerados como unos útiles complementos a las herramientas de selección de reconocido prestigio.

REFERENCIAS BIBLIOGRÁFICAS

Ansbacher, H.L. (1941). German military psychology. *Psychological Bulletin*, 38, 370-392.

Chan, D., y Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.

Chan, D., y Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233-254.

Clevenger, J., Pereira, G.M., Wiechmann, D., Schmitt, N. y Schmidt-Harvey, V.S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410-417.

Cullen, M.J., Sackett, P.R. y Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, 14, 142-155.

Fritzsche, B.A., Stagl, K.C., Salas, E. y Burke, C.S. (2006). Enhancing the design, delivery and evaluation of scenario-based training: Can situational judgment tests contribute? In J. A. Weekley y R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 301-318). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Hausknecht, J.P., Day, D.V. y Thomas, S.C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639-683.

Hooper, A.C., Cullen, M.J. y Sackett, P.R. (2006). Operational threats to the use of SJTs: Faking, coaching, and retesting issues. In J.A. Weekley y R.E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 205-232). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lievens, F. (2006). International situational judgment tests. In J.A. Weekley y R.E. Ployhart (Eds.) *Situational Judgment Tests: Theory, measurement and application* (pp. 279-300). SIOP Frontier Series. Mahwah, NJ: Lawrence Erlbaum Associates.

Lievens, F. y Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, 10, 245-257.

Lievens, F. y Sackett, P.R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91, 1181-1188.

Lievens, F. y Sackett, P.R. (2007). Situational judgment tests in high stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology*, en prensa

Lievens, F., Buyse, T. y Sackett, P.R. (2005a). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90, 442-452.

Lievens, F., Buyse, T. y Sackett, P.R.

(2005b). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, 58, 981-1007.

McDaniel, M.A. y Nguyen, N.T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9, 103-113.

McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A. y Braverman, E.P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology*, 86, 730-740.

Motowidlo, S., Dunnette, M.D. y Carter, G.W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 640-647.

Nguyen, N.T., Biderman, M.D. y McDaniel, M.A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, 13, 250-260.

Nguyen, N.T., McDaniel, M.A. y Whetzel, D.L. (2005, April). *Subgroup differences in situational judgment test performance: A meta-analysis*. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Oswald, F.L., Schmitt, N., Kim, B.H., Ramsay, L.J. y Gillespie, M.A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, 89, 187-208.

- Peeters, H. y Lievens, F. (2005). Situational judgment tests and their predictive-ness of college students' success: The influence of faking. *Educational and Psychological Measurement*, 65, 70-89.
- Richman-Hirsch, W.L., Olson-Buchanan, J.B. y Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, 85, 880-887.
- Ryan, A.M., McFarland, L., Baron, H. y Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 52, 359-391.
- Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C. y De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, 56, 573-605.
- Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. y Rolland, J.P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology*, 88, 1068-1081.
- Schmidt, F.L. y Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmitt, N. y Chan, D. (2006). Situational judgment tests: Method or construct? In J. Weekley y R.E. Ployhart (Eds.), *Situational judgment tests* (pp.135-156). Mahwah, NJ: Lawrence Erlbaum.
- Such, M.J. y Schmidt, D.B. (2004, April). *Examining the effectiveness of empirical keying: A cross-cultural perspective*. Paper presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.
- Wagner, R.K. y Sternberg, R.J. (1985). Practical intelligence in real world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology*, 49, 436-458.
- Weekley, J.A. y Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, 50, 25-49.
- Weekley, J.A. y Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, 52, 679-700.
- Weekley, J.A. y Ployhart, R.E. (2006). *Situational judgment tests: Theory, measurement and application*. San Francisco, Jossey Bass.
- Weekley, J.A. Ployhart, R.E., y Holtz, B.C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J.A. Weekley y R.E. Ployhart (Eds.), *Situational Judgment Tests* (pp 157-182). Mahwah, NJ: Lawrence Erlbaum.