

Artículo



La tecnología de evaluación ipsativa y el sistema APP-APT de Thomas International

Ipsative Assessment Technology and the Personal Profile Assessment by Thomas International

JORGE RODRÍGUEZ GONZÁLEZ

Thomas International Spain

RESUMEN

El objetivo fundamental del presente artículo es aportar más información para diferenciar entre las distintas pruebas a fin de que los profesionales de RRHH podamos elegir para evaluar al personal. Para ello es necesario ahondar en la distinción entre pruebas normativas e ipsativas, que son las pruebas más frecuentes en nuestro ámbito de actuación. Por último, y como objetivo final, este artículo pretende presentar los informes de validez del Análisis del Perfil Personal© de Thomas International®.

Las pruebas de elección múltiple y las pruebas de elección forzosa son distintas a varios niveles. El objetivo de ambas es común, evaluar a la persona, la diferencia es cómo cada una responde al criterio de rentabilidad (para los que es crítica la validez).

Las pruebas normativas funcionan de manera muy eficiente cuando cumplen los supuestos; pero veremos que en determinadas situaciones, y especialmente en contextos de selección, estos supuestos pueden llegar a distorsionar las puntuaciones de una persona en una desviación típica. Esta diferencia es más que suficiente para incluir en la siguiente fase a un mal candidato o excluir a uno bueno.

Por otro lado, las pruebas ipsativas no requieren supuestos tan rígidas, por lo que la información intraindividual que obtenemos corre menos riesgo de sufrir variaciones graves. Sin embargo, las pruebas ipsativas son más débiles en la comparación interindividual cuando analizamos dimensiones aisladas. En el caso de la selección de personal, implica que la comparación entre personas incluye siempre todo el perfil.

El uso de pruebas ipsativas nos asegura disponer como mínimo de la misma validez que las pruebas normativas, e incluso superior, a la vista de cómo resuelve la comparación entre individuos.

Por último, los datos del Análisis del Perfil Personal (APP) de Thomas International demuestran que los datos de validez, tanto para población mundial como para la población en España, son altos y estadísticamente significativos. Esto era esperable por la ausencia de fuentes de error tan acusadas en las pruebas ipsativas en general y el APP en particular.

ABSTRACT

In order to help human resources professionals differentiate among the wide range of selection tests, this paper provide them with useful information. A first differentiation is between normative and ipsative tests. This paper also reports about validity of Personal Profile Assessment (PPA) by Thomas International®.

Multiple choice tests are different from forced choice tests in a number of ways. Both types of tests share the goal of individual assessment, but they differ in how they meet cost-benefit standards, where validity is critical. Normative tests are very efficient when they meet assumptions, although these may distort an individual outcome by a standard deviation. This difference might lead to a wrong decision, i.e. accept a bad candidate or reject a good one. On the other hand, ipsative tests do not require such rigid assumptions and hence intra-individual information is less likely to vary abruptly. Nevertheless, for separate dimensions ipsative tests are weaker in inter-individual comparisons. In personnel selection, this means that between-person comparisons always include the whole profile. The use of ipsative tests ensures at least the same validity as normative tests or even higher in view of the way they manage between-individual comparison.

Lastly, data from Thomas International PPA show high and significant validity indices both for the world population and for the Spanish population. This was something to be expected thanks to the lack of marked error sources in ipsative tests, particularly in Personal Profile Analysis.

PALABRAS CLAVE

Selección de tests, Evaluación normativa, Evaluación ipsativa, APP, Validación.

KEY WORDS

Selection Tests, Normative Tests, Ipsative Tests, PPA, Validity.

INTRODUCCIÓN

El objetivo fundamental del presente artículo es aportar más información entre las diferentes pruebas que cualquier profesional de RRHH puede elegir para evaluar a su personal. Para ello es necesario ahondar en la distinción entre pruebas normativas e ipsativas, que son las pruebas más frecuentes en nuestro ámbito.

Seguramente todo el mundo esta de acuerdo en que el criterio central en el que nos debemos de fijar es la validez. Pero para llegar a ella es necesario recorrer un camino previo que nos permita poner en contexto cada prueba. La validez, como cualquier índice matemático, tiene una serie de características y asunciones que es necesario tener en cuenta. Dicho de otro modo, estos índices se calculan en unas condiciones determinadas y con una población determinada. En la medida en la que las condiciones en que nosotros utilizamos esa pruebas sean idénticas a las descritas en los estudios, podemos confiar más en el valor matemático. A modo de ejemplo, imaginemos un proceso de selección en España en el año 2004, de poco o nada nos sirve saber que una determinada prueba A, tiene una validez de .99 si el estudio se hizo en el año 1967 con población japonesa que vive en Perú.

Este ejemplo es tan extremo que resulta evidente que la validez que nosotros obtenemos al realizar un proceso de selección en nuestra empresa cuando menos será distinta.

No obstante, la concepción de este documento es el desarrollo y análisis de las tecnologías de evaluación está aplicada a realidad empresarial española, lo cual implica

que nosotros consideramos que la validez no es un fin en sí mismo sino un medio. El verdadero objetivo es optimizar el proceso de evaluación de las personas. Dicho de otro modo, el objetivo es llegar una solución que brinde la más alta calidad en los resultados, con los menores recursos posibles.

El mejor modo que tenemos ahora para evaluar en qué medida las características técnicas aportan valor a esa optimización es la validez. Sin embargo, no es el único criterio que consideramos que es relevante. Volvamos a poner un ejemplo extremo. Supongamos que tenemos que elegir entre dos pruebas para un proceso de selección. Podríamos optar por la prueba B, que tiene una validez real para nuestro grupo objetivo de 0.70 pero que requiere un alto coste de licencia, tener a los candidatos y a los evaluadores 5 horas trabajando (y por tanto espacios físicos) y que además requiere tener formadas a las personas que realizan las pruebas (recordemos que cuando extrapolamos los índices de validez suponemos que todo el mundo esta igual de formado para aplicar, corregir e interpretar las pruebas). Por otro lado, podemos optar por la prueba C, que tiene una validez de 0.60, implica gastos en función de cada persona que evaluamos, se puede aplicar en 1 hora (incluso vía Internet) y tanto la aplicación como la interpretación las realiza un programa de ordenador (por lo que no dependemos que personal especializado en el uso de esa prueba). La pregunta es: ¿Cuál es la prueba que maximiza la eficacia de nuestro proceso de selección?

El problema es que no hay una respuesta única a este dilema. En general podemos decir que hay dos posibles afrontamientos. Las personas más "academicistas" seguramente estarán convencidas de la bondad y

superior calidad de la primera; mientras que las personas más “pragmáticas”, cuando menos considerarán si una diferencia de 0.10 justifica la diferencia en costes.

Estamos tratando de evaluar el coste que supone equivocarse en un proceso de selección ya que el problema se basa en cómo nosotros podemos relacionar indicadores de RRHH con la cuenta de resultados o con la creación de valor.

Para intentar responder a ambos grupos de personas, y a diferencias de las revisiones de la literatura, nuestra revisión de las características de cada tipo de prueba, incluirá ambos tipos de elementos. De este modo, ambos grupos de profesionales se podrán hacer una imagen completa de todas las diferencias.

Hemos considerado que el mejor modo de responder a este objetivo es revisar en detalle los dos tipos de pruebas más usuales y aplicadas en el mundo del trabajo. Estas pruebas son las normativas y las ipsativas (Véase Michell, 1990; para una lista completa todos los tipos de pruebas).

De igual modo, debemos limitar nuestra revisión a parte del espectro de características personales relevantes a la hora de evaluar a una persona. En concreto, este trabajo se centra exclusivamente en la revisión de las pruebas de personalidad y/o comportamiento. La mayoría de las conclusiones son aplicables al resto de áreas, pero presentan algunas peculiaridades (Michell, 1990).

Para cada tipo de prueba, revisaremos: las definiciones que están más extendidas, el nivel de medida que emplean, qué supuestos hacen y, por último, las venta-

jas e inconvenientes que presentan. Además, para completar el análisis describiremos las relaciones que hay entre ambos tipos de medidas. Dicho de otro modo, si el carácter normativo o ipsativo depende sólo del instrumento o también si hay cuestiones estadísticas y de tratamiento de las puntuaciones. Lo que no vamos a hacer, es una revisión de los índices de validez encontrados para cada prueba concreta, el autor interesado puede encontrarlo en (Baron, 1995; Irvine, Mettam y Syrad, 1994).

Tras la necesaria revisión teórica, planteamos los datos obtenidos por el Análisis del Perfil Personal© (a partir de ahora APP) de Thomas International®. Dentro de este apartado abordaremos tanto los aspectos más técnicos, es decir índices de fiabilidad y validez, como los aspectos más pragmáticos de su utilización.

Si al final de este documento, el lector considera que tiene una idea más clara de la diferencia entre los dos tipos de pruebas y un mayor conocimiento estadístico del APP podremos considerar que hemos cumplido nuestro propósito.

PRUEBAS NORMATIVAS

El primer tipo de pruebas que vamos a analizar son las pruebas normativas. Este es el tipo de pruebas más comunes en la evaluación de RRHH.

Siendo fieles a nuestro esquema vamos a comenzar por definir de manera más precisa que es una prueba normativa y generar nuestra propia prueba con datos ficticios, que después nos servirán para hacer un análisis de las posibles fuentes de error.

Definición

Las pruebas normativas se caracterizan por emplear una tarea de evaluación, normalmente una cuestión o una frase descriptiva, ante la que la persona tiene que elegir en que grado es aplicable a sí mismo o en que grado esta de acuerdo. Para ello dispone de distintas alternativas que se gradúan progresivamente entre estos dos extremos.

Una característica fundamental de este tipo de pruebas es que, desde el punto de vista del candidato, cada ítem o pregunta es independientemente entre si. Yo puedo contestar lo que considero oportuno en cada pregunta. Será después cuando combinemos el conjunto de preguntas en constructos o con la escala de sinceridad.

A modo de ejemplo, vamos a realizar

un cuestionario de evaluación de la personalidad. Nuestro cuestionario tendrá sólo ocho cuestiones, porque para nosotros el tiempo es importante, y el candidato tendrá cinco opciones de respuesta. Tras pensar razonablemente cada una de las preguntas nuestro cuestionario es el que se muestra en la tabla 1.

Como puede fácilmente apreciar el lector, todas preguntas han sido inventadas. Sin embargo, otras se podrían parecer a algún test. Se han elegido esas preguntas porque en nuestra práctica profesional son las que algunos profesionales de RRHH han considerado más “sorprendentes”. No vamos a entrar a analizar su validez aparente en este documento pero sin duda es otro elemento importante a analizar en la eficacia de un proceso de evaluación y en la venta de la imagen de la empresa.

Tabla 1						
EJEMPLO FICTICIO DE PRUEBA DE ELECCIÓN MÚLTIPLE						
		Muy bajo	Bajo	Medio	Alto	Muy alto
1.	En qué medida se considera usted respetuoso con la gente					
2.	Me relaciono con facilidad con la gente					
3.	Siempre ayudo a la gente que me encuentro					
4.	Prefiero estar en el salón que en el dormitorio					
5.	Conozco mucha gente					
6.	Hago amigos con facilidad					
7.	Siempre contesto de manera simpática a la gente					
8.	Trabajo bien en equipo					

Como se puede imaginar el lector la realidad nunca es tan sencilla. Por el contrario, lo habitual es tener un modelo teórico que sirva de soporte a la evaluación, aunque en algunos casos, este modelo surge de los datos recogidos. La repercusión en RRHH de esta diferencia, sería otra interesante discusión que no podemos tratar en este momento.

En cualquiera de los dos casos, el modelo dicta los criterios para el análisis de los ítems, y en concreto para su agrupación. Véase Kline (1998) para una referencia completa de la diferencia entre comenzar desde un modelo teórico o hacerlo de manera empírica.

Podemos dar un paso más para facilitar la comprensión de las pruebas normativas.

Supongamos que hemos aplicado esta prueba a cuatro candidatos. Tras las fases iniciales de análisis y de desarrollo teórico, hemos llegado a la conclusión de que el modo adecuado de agrupar los ítems es el que se muestra en la tabla 2. Aquí tenemos ahora cuatro constructos, factores o rasgos representados por las letras A, B, C y D. Para la corrección sustituimos el valor muy bajo por 1, el bajo por 2; y así sucesivamente hasta muy alto que corresponde con 5.

Esta tabla muestra para cada uno de las personas (en las columnas) las puntuaciones que han marcado en cada uno de los ítems (filas) y la posterior agrupación en factores (filas en negrita).

Para completar el análisis de estas res-

Tabla 2.
RESULTADOS DE CUATRO CANDIDATOS A NUESTRA PRUEBA DE ELECCIÓN MÚLTIPLE

		Persona A		Persona B		Persona C		Persona D		Suma por rasgo
Ítem 1	Rasgo A	1	2	3	6	5	6	5	10	14
Ítem 5		1		3		1		5		
Ítem 2	Rasgo B	1	2	3	6	5	6	5	10	14
Ítem 6		1		3		1		5		
Ítem 3	Rasgo C	1	2	3	6	5	6	5	10	14
Ítem 7		1		3		1		5		
Ítem 4	Rasgo D	1	2	3	6	5	6	5	10	14
Ítem 8		1		3		1		5		
Suma por persona		8		24		24		40		

puestas, hemos calculado la suma total de las puntuaciones para cada columna y para cada fila. La primera nos indica la suma de todas las puntuaciones de cada uno de los candidatos, mientras que las filas nos dicen la suma para cada factor de todas las personas.

Un análisis típico de una prueba normativa incluiría una comparación con los datos normativos; es decir la población de referencia. La puntuación de 2 de la persona A en el rasgo 1, no significa nada en sí mismo; depende de la puntuación del grupo de referencia. De este modo contestamos a la siguiente pregunta: ¿En qué parte del continuo de puntuaciones de todo el grupo situamos esa puntuación?

Nivel de medida

El nivel de medida de una prueba depende del formato de respuesta y de las variables medidas. Como se puede apreciar en nuestro ejemplo, el formato de respuesta usado aquí es mediante elección múltiple. Para cada uno de los ítems la persona dispone de varias opciones de respuesta (en nuestro caso 5). Lo más habitual es que esta elección múltiple sea del tipo Likert, de 5 ó de 7 opciones (Véase Michell, 1990, para un referencia completa de sobre el concepto de nivel de medida).

Este modo de recoger la información esta asociado con un nivel de medida de intervalo. Las características psicológicas, medidas de esta forma, tienen validez ecológica ya que nos permite medir o bien la cantidad o fuerza de un sentimiento, idea, predisposición hacia una persona, cosa o acción o la calidad de esa fuerza. La primera se evalúa mediante el uso de adver-

bios; mientras que la segunda mediante adjetivos.

En principio no hay límite para la cantidad de dimensiones o constructos que se pueden evaluar con una prueba de tipo normativo, ya que cada una de las evaluaciones se hace de manera independiente. Sin embargo, hay numerosos estudios (Lord y Novick, 1968) que demuestran que no manejamos más allá de 9 dimensiones, siendo lo normal 5. Lo cual muestra una evidente relación teórica y empírica con la amplitud de la memoria de trabajo.

A este problema debemos agregar que en realidad sabemos que las evaluaciones en cada una de las dimensiones no son independientes, sino que varían en el grado de relación entre ellas. La solución que ha dado la teoría de la evaluación psicológica es la utilización del análisis factorial. El propósito de esta técnica es agrupar dimensiones muy relacionadas para facilitar su comprensión y poder dotarlas de significado psicológico.

Supuestos

El nivel de medida es lo que más fácil se puede apreciar en cualquier prueba, y es accesible incluso para el candidato que hace nuestras pruebas. Pero elegir un tipo de nivel de medida condiciona una serie de supuestos. Estas suposiciones previas de partida determinan en que grado nuestras conclusiones son validas.

La primera suposición de las pruebas normativas es que las categorías son continuas. Dicho de otro modo que a cualquier candidato cuando haga nuestra prueba le resultará igual de fácil pasar del muy bajo

al bajo, que del bajo al medio. Hay una abundante investigación sobre la dificultad que supone asumir que esto es cierto (Baron, 1996), por lo que se estudia el efecto real que produce.

La segunda suposición que hacemos al emplear una prueba de tipo normativa es que la escala de medida es de intervalo. Este supuesto es la operativización matemática del primero, pero tiene grandes implicaciones respecto al uso de las herramientas estadísticas, por ejemplo el análisis factorial paramétrico.

Tan importante como el anterior es la suposición de la existencia de independencia entre las varianzas error. Lo que nos equivocamos en cada una de las dimensiones no está relacionado entre sí. Dicho de otro modo, no habrá factores que afecten de manera consistente y relacionada a todas las dimensiones.

Por último, algunas pruebas asumen un tipo determinado de distribución. Las personas, cuando consideramos la población de manera global, mostrarán puntuaciones agregadas ajustadas a una distribución normal (pocos en los extremos y muchos en el medio).

Más tarde analizaremos en detalle las repercusiones de estos supuestos en el uso de las pruebas normativas en un contexto laboral, pero antes de avanzar es necesario hacer al menos dos consideraciones sobre la última de las suposiciones. ¿Cree alguien que la distribución en una dimensión de personalidad concreta de los candidatos que se presentan a un proceso de selección es normal? Por ejemplo, seleccionamos personas para un puesto de ingeniero (por usar un tópico que cada vez es

menos frecuente). Cualquier persona habituada a trabajar con este colectivo sabe que en términos de personalidad es poco frecuente encontrar tantas personas caóticas, desorganizadas como personas ordenadas, precisas, sistemáticas.

La segunda consideración a realizar sobre este punto se refiere al propósito del proceso de selección. En términos globales el objetivo es predecir conducta, pero de manera pragmática es diferenciar entre candidatos. No nos basta saber si dos personas tienen el potencial, sino entre ellas cuál se adaptará más rápido o cuál aguantará la presión o cuál se ajustará al equipo y a la cultura. ¿Es la distribución normal lo más adecuado para dar respuesta a ese objetivo?

Como se puede suponer estos criterios tienen un increíble peso en la validez. Los índices de validez solo se pueden extrapolar cuando se cumplen todos estos requisitos. Un índice de 0.99 puede no ser aplicable a nuestra situación si no tenemos la certeza de que cumplimos esas características.

Inconvenientes

Los principales inconvenientes de este tipo de pruebas se refieren a su aplicación práctica. De manera resumida la literatura recoge tres inconvenientes:

- Son muy sensibles al engaño.
- Más largos en tiempo.
- En selección no cumplen los supuestos.

Cuando decimos que son más sensibles al engaño, significa que la persona cons-

ciente o inconscientemente puede engañar o falsear los datos. La diferencia entre ambos tipos de engaños son la intencionalidad de la persona y no el grado de autoconocimiento o consciencia.

Dentro del engaño no intencional, los principales problemas con los que se enfrentan este tipo de pruebas son la tendencia central, la aquiescencia, la deseabilidad social y la tendencia a infravalorarse.

Cuando hablamos de problemas denominados de tendencia central, nos referimos a la inclinación que tienen ciertas personas de evitar usar puntuaciones extremas. Es el ejemplo de la persona B de la Tabla 1. Con personas de este tipo podemos encontrarnos con dimensiones diferenciadas e incluso extremas en función de la población de referencia. Supongamos que la persona B obtiene un percentil de 70 en la dimensión A, ¿cómo debemos interpretar ese dato teniendo en cuenta que la persona contestó a todas las preguntas como medio?

La segunda fuente de error es la aquiescencia, que se define como la tendencia a mostrarse de acuerdo con la dirección de la frase. En nuestro caso podría ser la persona D. Sus puntuaciones demuestran que está siempre de acuerdo en grado máximo con la dirección de las preguntas. Normalmente el diseño de una buena prueba debe equilibrar los enunciados formulados positiva y negativamente. Nuestra prueba no lo tiene.

En tercer lugar, según su relevancia en la literatura, se han encontrado problemas de deseabilidad social. Hay personas que al contestar a la prueba se perciben a sí

mismos de manera más positiva, acorde con la imagen socialmente deseable. En nuestro pequeño ejemplo, podría ser la persona D. La misma persona puede estar en ambos tipos de engaño porque nosotros no hemos realizado un equilibrio de las respuestas.

Por último, y en oposición al anterior hay personas que tienen tendencia a infravalorarse. Estas personas se puntúan muy bajo. En algunos casos está asociado a una baja autoestima. Nuestro ejemplo se corresponde con el ejemplo A.

Es difícil estimar el efecto que tiene estas fuentes de error. Lo que si sabemos (Baron, 1996) es que pueden modificar la puntuación de las personas en un factor, entre media y una desviación típica (Ones, Viswesvaran y Korbin, 1995; Viswesvaran y Ones, 1999).

Es de esperar que en un contexto de trabajo a estas fuentes de engaño inconsciente haya que sumar la predisposición de la persona de dar una imagen más positiva de sí misma. Lo cual seguramente tiene un aspecto muy positivo, asociado con la motivación por adquirir el puesto, pero es de esperar que incremente el riesgo de que aparezcan estas fuentes de engaño, e incluso, la presencia de engaño intencionado. A nadie se le escapa que hay academias y centros de formación que entrenan a la gente para contestar a las pruebas de forma intencionada.

El segundo inconveniente es que este tipo de pruebas requieren más tiempo, en parte para paliar estas fuentes de errores y, en parte, porque emplean muchas escalas. Sea cual sea la causa, implica que en un contexto laboral se incrementan los costes,

tanto de la organización como del candidato evaluado.

Finalmente, hay dudas (Baron, 1995) de que este tipo de pruebas cumplan todos los requisitos matemáticos y metodológicos en un contexto de selección de personal. El ejemplo más claro es el nivel de medida por intervalos. No parece cierto que la distancia entre el valor 1 y el 2 de la escala sea igual a la distancia entre el 4 y el 5, por poner un ejemplo.

Ventajas

Después de leer lo expuesto anteriormente, uno puede preguntarse: ¿Por qué razón se emplean todavía pruebas de tipo normativo? En primer lugar debemos decir, que estas dificultades no se aplican en igual grado a todas las pruebas normativas. Son características asociadas al paradigma de medida. La diferencia entre pruebas estará en el grado en que solventan estas dificultades a priori y el coste que esto implica (por ejemplo: en tiempo, restricciones de aplicación, etc.).

No obstante, las pruebas normativas también tienen características positivas. Las más destacadas son:

- Son fáciles de interpretar.
- Especialmente cuando queremos comparar sujetos entre sí.
- Hay gran cantidad de literatura a nivel internacional.
- Algunos modelos acumulan abundante investigación (por ejemplo, el Big Five).

La cuestión no es si es adecuada o no, que es obvio que lo sea, sino en que con-

textos sus debilidades se hacen más patentes.

Es evidente que este tipo de pruebas tiene aspectos positivos y negativos; y que el equilibrio entre ellos varía según los contextos de utilización. La siguiente pregunta es: si no uso pruebas normativas, ¿qué uso?

PRUEBAS IPSATIVAS

Una de las posibles respuestas a esta pregunta es usar pruebas de tipo ipsativo. Pero antes de saber para qué contexto y de qué modos es rentable este tipo de pruebas es necesario hacer una revisión de sus características.

Definición

Por ipsativa entendemos aquella prueba que recurre a puntuaciones forzadas, es decir, que todas las escalas de medida en esa prueba, son constantes para todas las personas. Dicho de otro modo, la persona tiene una serie de elecciones o de puntos que repartir entre varias opciones, pero asignar una puntuación a una de las opciones implica no poder hacerlo con las otras. De manera más formal, podemos decir que las pruebas ipsativas son medidas multiescala que restringen la varianza intraindividual (Cattell, 1988).

Hay dos modos de llegar a una restricción de este tipo: mediante un cuestionario específico o a través del tratamiento matemático de los datos. En este apartado consideramos puntuaciones ipsativas sólo a aquellas que se derivan directamente de las respuestas del candidato, y

dejaremos el análisis de las puntuaciones *ipsativizadas* para el siguiente apartado. El motivo de esta distinción es que los datos parecen indicar que no tienen las mismas propiedades ambos tipos de datos ipsativos.

El modo directo de obtener puntuaciones ipsativas es mediante cuestionarios de elección forzosa. El grado de elaboración de estos cuestionarios varía notablemente, pero el más fácil que nos podemos imaginar sería forzar al sujeto a que repartiera 10 puntos entre una lista restringida de vinos, según su preferencia. Independientemente del resultado y de la elección individual, la suma total de los puntos asignados por cada uno de nosotros será la misma, ya que era una restricción de la tarea.

Si lo complicamos en el sentido de introducir varias escalas en lugar de una

bas ipsativas, porque este adjetivo es una característica de los datos y de la medida. Como después veremos se pueden obtener datos ipsativos de pruebas de elección múltiple.

A título de ejemplo, si dejáramos que el candidato eligiera tantos adjetivos sin límite como quisiera no sería ipsativo. Cattell (1988) diferencia un tipo de medidas como parcialmente ipsativo si se establece restricciones de otro tipo. Las pruebas que no lo hace no son ni ipsativas ni normativas, sino de tipo idemórfico.

Un ejemplo de prueba de elección forzosa se muestra en la Tabla 3. En este caso hemos ideado un ejemplo relativamente simple de prueba con sólo cuatro cuestiones. El candidato debe marcar en cada una de las casillas el adjetivo que considera que es más aplicable para describirse en un ambiente de trabajo.

<p align="center">Tabla 3 CUESTIONARIO DE ELECCIÓN FORZOSA</p>					
cortés		influenciable		Jovial	minucioso
persuasivo		Audaz		preciso	obediente
modesto		Leal		valiente	inconquistable
original		encantador		ecuánime	travieso

sola, nos vamos progresivamente aproximando a cuestionarios multiescala de elección forzada. En este caso, se establece una restricción para cada escala.

A modo de precisión purista se habla de pruebas de elección forzosa y no de prue-

Este tipo de cuestionario es de elección forzosa porque de los cuatro adjetivos, sólo se puede elegir uno de ellos. Dicho de otro modo, la elección de cada una de las preguntas no es libre sino que ordena las opciones según la calidad de la preferencia. Podría ser que las cuatro

lo fueran o que ninguna lo fuera, pero el candidato esta forzado a elegir en cada columna.

Hay cuestionarios más elaborados que le obligan al candidato a elegir la característica que más se aplica y la que menos.

más elaborados y mayor profundidad de los análisis.

Supongamos que hemos aplicado nuestra mini prueba a cuatro candidatos, la matriz de datos resultante podría ser la que se muestra en la tabla 4.

Tabla 4 RESULTADOS DE CUATRO CANDIDATOS A NUESTRA PRUEBA DE ELECCIÓN FORZOSA						
		Persona 1	Persona 2	Persona 3	Persona 4	Suma por Dimensión
Audaz	Dimensión 1	+			+	8
Cortés		+	+			
Ecuánime		+		+		
inconquistable		+		+		
Encantador	Dimensión 2		+			3
Valiente					+	
Minucioso						
Modesto						
Obediente	Dimensión 3		+			2
Influenciable						
Jovial						
persuasivo					+	
Original	Dimensión 4			+		4
preciso			+			
travieso					+	
Leal				+		
Suma por persona		4	4	4	4	

En términos conceptuales no hay diferencia, las puntuaciones siguen siendo ipsativas. Sin embargo, disponemos de dos puntuaciones por dimensión en lugar de una; lo cual permite un tratamiento de los datos

La suma de las filas define la puntuación de cada persona en cada una de nuestras cuatro dimensiones. Esta información varía para cada dimensión. Sin embargo, la suma por columnas nos informa de las pre-

guntas contestadas por cada persona; pero es igual para todas las personas evaluadas.

Como se puede apreciar en estos resultados, si analizamos a las personas tenemos que el primer candidato ha concentrado todas sus respuestas en el dimensión 1, mientras que en el resto no tiene ninguna respuesta. Por el contrario, el resto de los candidatos han repartido sus respuestas de manera diferente. Así los candidatos 2 y 4 han repartido por igual todas sus respuestas, pero eligiendo diferente adjetivos. Por último el candidato 3 se ha centrado únicamente en dos dimensiones.

Nivel de medida

Cuando empleamos una prueba de elección forzosa debemos tener en cuenta que las dimensiones están relacionadas entre sí. Nuestra medida obliga al candidato a elegir entre las distintas opciones. Si marca una no podrá marcar las otras. Esto implica que los factores de una prueba de este tipo siempre muestran una correlación media.

El grado de relación entre las escalas depende tanto del modelo teórico en el que se basa como de la propiedad psicométrica de la propia prueba. Pero en cualquier caso, serán unas características a tener en cuenta en los análisis y en la interpretación posterior.

Las puntuaciones que obtenemos en cada una de las dimensiones son relativas. Dicho de otro modo, no decimos que esta persona tiene un 4 en la dimensión 1, sino que esta persona participa más de este factor que de los demás. El motivo de esta diferencia con las pruebas de elección múltiple es que en las pruebas ipsativas no se

fija un cero absoluto para la interpretación de cada una de las escalas, sino que referencia el cero relativo en cada factor al resto.

Una de las primeras conclusiones de esta diferencia es la validez ecológica. Las pruebas de elección forzada reproducen de manera más ecológica la necesidad de elegir entre los distintos tipos de comportamientos disponibles en el repertorio conductual de una persona. De este modo, nos acercamos más a cuáles pueden ser los comportamientos más característicos de cada persona

Algunos estudios afirman que este tipo de prueba está más cercana a la realidad, es decir explican mejor el comportamiento que queremos predecir, en la medida en que las opciones están relacionadas con una situación real (Closs, 1996).

La idea que subyace es que no se puede medir todo simultáneamente y por tanto hay que elegir. Es más, algunos de los factores o sentimientos que en las pruebas normativas se miden de manera independiente podrían ser contradictorios, y por tanto excluyentes.

Asunciones

El primer supuesto implícito que hacemos al emplear una prueba de elección forzosa es que estamos en un nivel de medida ordinal. Esto es, podemos ordenar puntuaciones, pero no hacer cálculos sobre la distancia entre ellas.

Si a esta primera asunción le sumamos que consideramos que las dimensiones están relacionadas entre sí, seguramente

veremos claro porque no necesitamos hacer supuestos sobre la distribución o la independencia de la varianza de error (Véase la Karpatschof y Elkjaer, 1999, para una referencia completa de la métrica de la medida ipsativa frente la normativa).

Excepto el nivel de medida, las pruebas ipsativas no necesitan hacer supuestos para poder evaluar personas. Esto es muy útil en la medida en que es aplicable a gran variedad de contextos y resistentes a los requerimientos más exigentes, como puede ser la situación de evaluación en un proceso de selección.

Aunque no sean propiamente supuestos, la medida con pruebas de elección forzosa encuentra que los factores que se emplean son casi siempre de carácter bipolar. Es decir, que podemos claramente identificar dos polos cualitativamente diferentes en la interpretación del comportamiento de una persona en la misma dimensión.

Por otro lado, prácticamente la totalidad de las pruebas de elección forzosa muestran altas correlaciones entre los factores medidos. Estas características no condicionan la utilidad de la prueba ni la calidad de las conclusiones; pero sí impone condiciones sobre utilización e interpretación, tal y como veremos en el análisis de los inconvenientes y de las ventajas.

Inconvenientes

El mayor inconveniente del uso de las pruebas de elección forzosa es que se interpretan y se emplean de manera diferente. En concreto, una prueba de elección forzosa que genera puntuaciones ipsativas

es incompatible con la comparación de dos personas en factores aislados. Dicho de otro modo, nosotros no podemos decir que el candidato 1 tiene más orientación a comportarse según el factor 1 de nuestro ejemplo, que el resto de los candidatos. En palabras de Horst (1965):

“...las medidas ipsativas, ya sea que se deriven de procedimientos experimentales o de computación, no pueden ser normativas en el sentido de que una persona puede ser comparada con otra con relación a una variable única. Por supuesto, esto es así porque alguna constante arbitraria ha sido añadida o sustraída de todas las puntuaciones de cada individuo medido. Puesto que no es la misma constante para todos los individuos, las medidas resultantes no son comparables de un individuo a otro. A menudo se pasa por alto esta limitación. Con frecuencia, se compara erróneamente a las personas por lo que respecta a las medidas ipsativas”.

Las implicaciones de esta diferencia son muy importantes, y es especialmente seria en el caso de la evaluación de personas en el contexto laboral. Trabajos clásicos demuestran que cuando se emplean indiscriminadamente en decisiones sobre personal tienen una utilidad limitada (Johnson, Wood y Blinkhorn, 1988).

Dada la ausencia de supuestos matemáticos sobre los datos, los análisis estadísticos son más complicados. No podemos emplear los mismos estadísticos para analizar ambos tipos de pruebas. Como veremos más adelante, esta conclusión se debe de matizar por el número de factores y por la correlación entre los factores.

Ventajas

A pesar de las dos limitaciones anteriores, sin duda, hace pensar seriamente el hecho de que, las pruebas ipsativas deben tener grandes ventajas objetivas para que haya una gran cantidad de personas y organizaciones que las sigan empleando.

El propósito de este apartado es mostrar esas ventajas y la relación que hay con las limitaciones previamente señalados.

Pero en lugar de hacerlo de manera general y sin fundamentos estadísticos, lo referiremos a un ejemplo concreto de prueba ipsativa que es el APP © de Thomas International. Todas las descripciones de las ventajas y características se mostrarán en el último apartado.

RELACIONES ENTRE AMBOS TIPOS DE MEDIDAS

Las relaciones entre ambos tipos de medidas tiene un importante componente estadístico, lo que implica que se pueden hacer transformaciones entre uno y otro tipo de puntuaciones.

Nosotros haremos una rápida revisión de estas transformaciones, pero pondremos el énfasis en los aspectos de la interpretación psicológica. Que las puntuaciones sean transformables no significa que sean igualmente interpretables.

Tal y como hemos visto, hacer ipsativa una puntuación originalmente normativa es muy fácil. Sólo tenemos que restarle la medida de ese ítem. De este modo, tenemos puntuaciones que cumplen los requisi-

tos de la definición de ipsativo, forzar a que la suma sea igual.

Las datos (Cross, 1996) parecen indicar que estas variables no tienen significado psicológico. Para las personas más críticas con la estadística y sobre todo con el significado psicológico de la media, esta conclusión no será en grado alguno sorprendente.

Una de las mayores diferencias que hemos visto entre ambos tipos de pruebas es el punto de referencia. En las puntuaciones normativas esta forzado a cada variable y en la ipsativas es relativa a las otras variables. Lo que no se puede interpretar es la mezcla de ambos.

Por este motivo, nosotros hemos insistido en la distinción entre pruebas de elección forzosa, que dan puntuaciones ipsativas “naturales” y puntuaciones ipsativas, que pueden ser forzadas estadísticamente.

Lo que no podemos hacer es convertir una puntuación ipsativa en normativa, porque hemos forzado una serie de grados de libertad y nos faltaría información para hacerlo correctamente.

En resumen, aunque el empleo de las medidas del Análisis del Perfil Personal (APP) exige precaución, la elección forzosa no es una vía inválida para valorar a individuos aislados, en tanto lo que se haya medido sea fiable y válido para ese individuo. Por supuesto, este es el aspecto crítico con el que se enfrentan todos los diseñadores de escalas, ya sean éstas ipsativas o referidas a la norma. Teniendo presentes estas precauciones y limitaciones ahora podemos dedicarnos a lo que el Análisis del Perfil Personal se propone medir, al modo

en que lo hace y al éxito demostrado durante más de 25 años de aplicación.

EL SISTEMA APP/APT DE THOMAS INTERNATIONAL

En esta sección, se describe brevemente el Análisis del Perfil Personal de Thomas International, pero hay que recordar que no es un instrumento de evaluación excluyente. En realidad, debe incluirse dentro de un proceso conjunto que hace que tanto el candidato como el entrevistador traten el problema de qué características laborales deben ser satisfechas por un trabajador eficiente en un trabajo particular dentro de la organización, con un jefe particular (por extensión el equipo) y con una cultura determinada.

En las organizaciones, normalmente el superior inmediato y el responsable de RR.HH. contribuyen al proceso rellenando consensuadamente Formulario de Análisis del Puesto de Trabajo (APT). Por otra parte, el candidato rellena el Análisis del Perfil Personal (APP). Una vez que el entrevistador ha compilado un perfil de las cualidades ideales para el trabajo del APT, éste se contrasta con el perfil aportado por el candidato con el APP.

Thomas International trabaja con la congruencia del perfil completo en lugar de puntuaciones aisladas para predecir, mediante la evaluación de las personas, la probabilidad de éxito y satisfacción en el trabajo de los empleados o candidatos evaluados.

El objetivo de este documento es analizar la diferencia entre pruebas normativas e ipsativas, utilizando el APP de Tho-

mas como una prueba ipsativa de referencia.

El APP de Thomas International, fundamento teórico

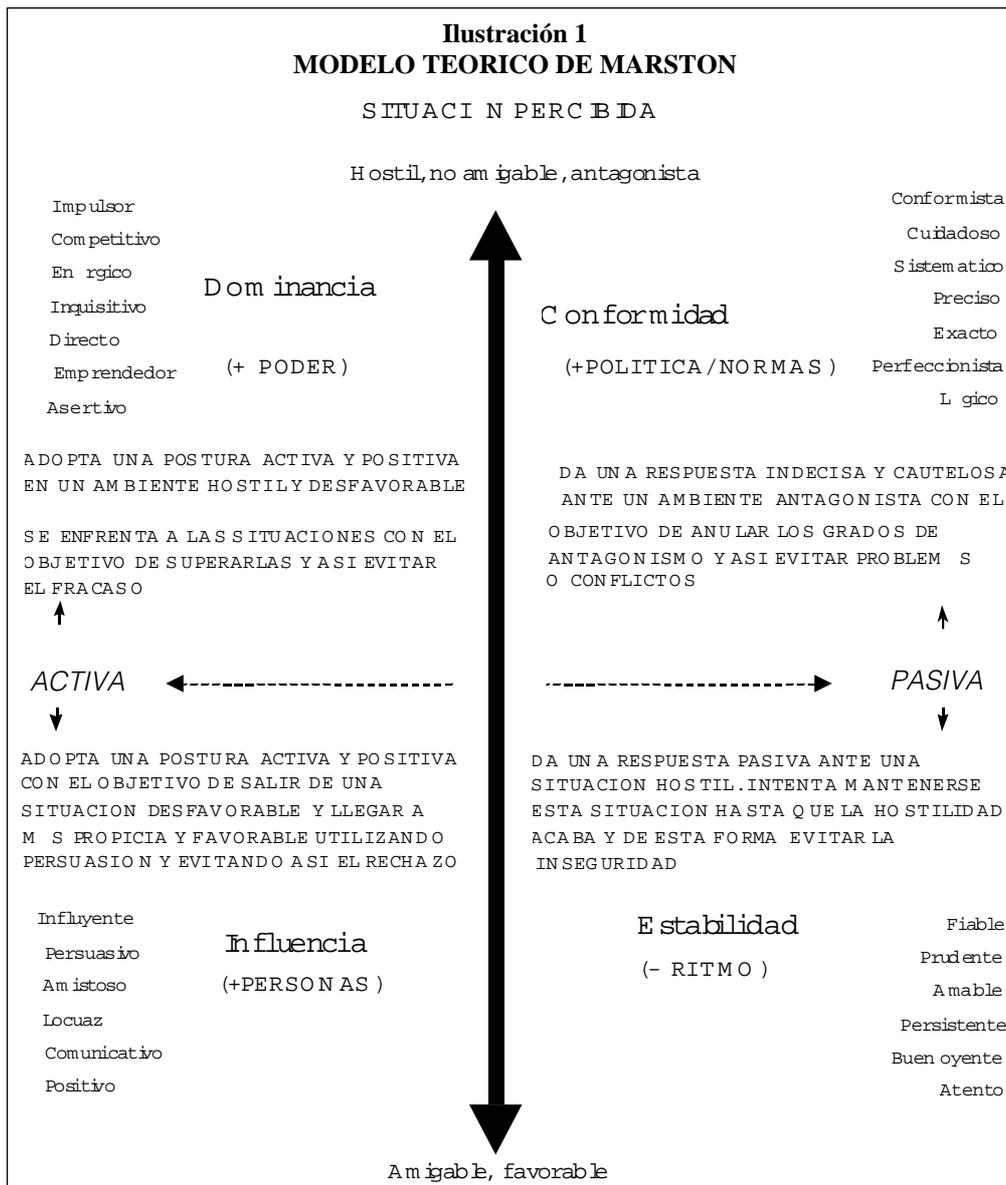
El modelo teórico que guía todo el desarrollo del APP es la teoría biopsicosocial de William Marston. Esta teoría holística poco conocida, percibe la conducta como la respuesta biosocial necesaria de cualquier individuo a los medios creados por las circunstancias externas y por otros individuos. A su vez, estas respuestas modifican lo que se percibe a continuación en las interacciones humanas.

Marston consideró que la conducta surge de la naturaleza intrínseca del medio, que puede ser percibida por los seres humanos orientados hacia una meta como un continuo, antagónico en un extremo y favorable en el otro. Estas percepciones producen lo que Marston denomina “fuerzas vitales” o emociones. En medios antagónicos, ya sean ecológicos o sociales, se necesitan formas de adaptación particulares para sobrevivir. En medios favorables, predominan otras conductas de adaptación. Estas formas de adaptación son ajustadas más o menos por el organismo a través de una dimensión conductual actividad – pasividad preferida o habitual, mediada por emociones.

A partir de la interacción de medios favorables – antagónicos con las reacciones de disposición actividad – pasividad de los individuos, Marston postuló cuatro estrategias de supervivencia de adaptación básica, que resultan de los conjuntos cartesianos formados por el continuo ambiental antagonista – favorable y la dimensión

conductual actividad – pasividad. Por definición, estas son adaptaciones activas y pasivas a medios rígidos y limitadores, por una parte, y a circunstancias favorables y gratificantes, por otra. Hemos

representado esta formulación teórica en la figura 1, donde se muestran los dos ejes, las motivaciones y algunos de los adjetivos asociados con cada uno de los cuatro cuadrantes.



La psicología intercultural ha aportado, un sólido núcleo de investigación que define las estructuras ecológicas para la adquisición de las habilidades sociales y cognitivas, Berry y la mayoría de los psicólogos interculturales (para una visión panorámica adecuada, véase Berry e Irvine, 1986) han apoyado y preparado durante muchos años esta postura ecológica general.

El elemento crítico diferenciador de esta teoría es la distinción en la elaboración de perfiles personales es la distinción entre el “parecer” y el “ser”. En cuestionarios, declaraciones o informes propios, hechos en respuesta de elección forzosa al estímulo, las palabras muestran más de lo que la persona pueda imaginarse.

En síntesis, los creadores de los formularios trataron de combinar dos ideas básicas. En primer lugar, surgirían diferentes modelos de acuerdo con las percepciones de los climas. Los ambientes hostiles evocan una conducta diferente de la que evocan los ambientes amistosos. En segundo lugar, lo que los entrevistados dicen acerca de sus percepciones de sí mismos y del ambiente no siempre puede revelar la profundidad de su propio conocimiento de sí mismo. Las respuestas a las preguntas pueden estar influidas tanto por el deseo social como por el autoengaño y otros mecanismos para ocuparse de las emociones.

Las ideas de Allport son importantes para la comprensión de la razón fundamental que subyace al APP. Para Allport, existen cualidades que son comunes a todos, y conjuntos de cualidades que son específicas de los individuos. De acuerdo con este acento sobre la plasticidad del individuo, se opuso a los que postulaban que la con-

ducta presente tiene su explicación en la inconsciencia pasada. Consideró que era mejor buscar explicaciones de los motivos en función de las situaciones y sentimientos contemporáneos.

De esto se deduce que cualquier respuesta inmediata a un conjunto de estímulos estructurados, como los adjetivos que describen a una persona en función de características o cualidades que son comunes a muchas personas, aportará una comprensión de las respuestas actuales del individuo sobre el rol percibido en este contexto. El trabajo que Allport hizo sobre los nombres de las cualidades fue investigado extensamente por los creadores del APP para las descripciones de cualidades que conforman las cuatro dimensiones de Marston. Esto no quiere decir que todas las cualidades que puedan haber sido medidas fueron incluidas en las cuatro interacciones emocionales – ambientales básicas.

Puede consultarse (Thomas International, 1981a) para una revisión completa del modelo y los trabajos posteriores para el análisis de los datos históricos de validez y fiabilidad (Thomas International, 1983b, c, d; 1986a, b; 1987a, b, c).

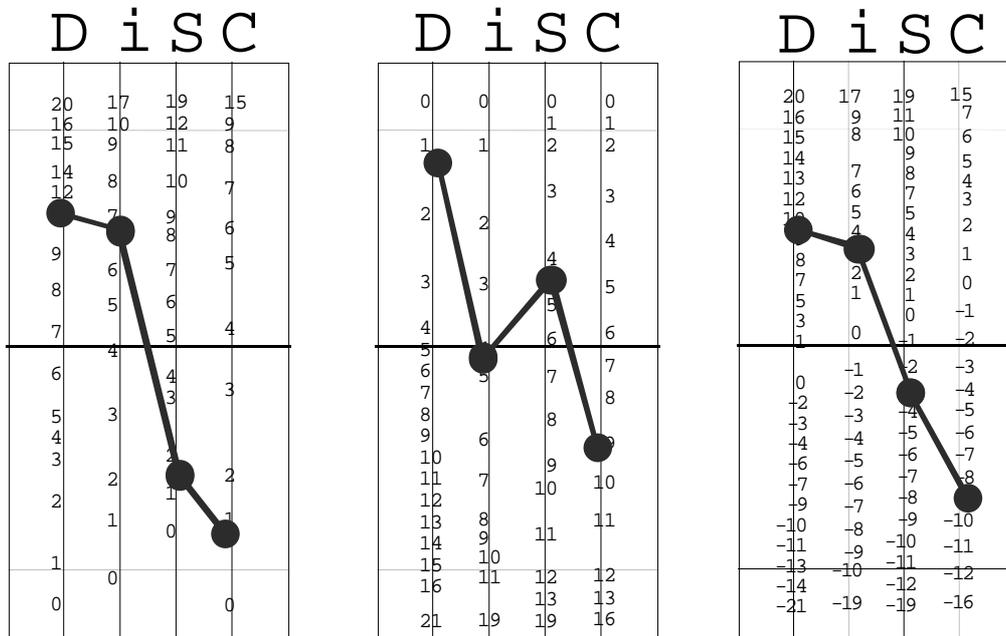
Mucho más importante que el marco teórico es la interpretación práctica que nos puede ofrecer el sistema. El sistema APP/APT ofrece tres gráficos, cada uno nos da información.

El sistema nos da información sobre la persona en tres escenarios. Por un lado nos informa de cómo la persona está en la situación actual, en una situación de presión y la autoimagen de la persona. Esta triple información es de gran utilidad para poder diferenciar qué elementos del com-

portamiento de la persona son estables (en el tiempo) y consistentes (recurrentes en distintas situaciones). La Ilustración 2 muestra un ejemplo real de una persona evaluada en estos tres escenarios, mediante las cuatro dimensiones de nuestro sistema.

Como hemos visto la utilidad potencial del sistema es grande en la medida que nos aporta gran cantidad de información con variedad de perspectiva temporal. Decimos utilidad potencial porque antes será necesario verificar que la validez del sistema es la adecuada.

Ilustración 2
EJEMPLO DE GRAFICO



La existencia de estos tres gráficos concede una gran potencia predictiva a esta prueba ipsativa porque como hemos visto uno de sus puntos fuertes es la comparación intraindividual. Esta comparación no se realiza únicamente entre las escalas, como veíamos en nuestro ejemplo inicial, sino que el caso exclusivo del APP además se puede hacer entre momentos temporales y entre situaciones.

Datos psicométricos

Disponer de datos Psicométricos relevantes del Sistema APP/APT es una dimensión importante de este apartado. Se trata de poder confirmar estadísticamente, lo que los clientes están viviendo día a día, ya que cada uno de los países en los que está operativa esta herramienta, puede avalar, para tranquilidad de los usuarios, que

se han realizado los pertinentes análisis para adaptarla a su población laboral. A veces los análisis estadísticos se repiten en el tiempo una vez que he transcurrido un tiempo considerado como prudencial, para repetir tales estudios matemáticos.

En la medida en que el objetivo de este apartado es informar de los datos psicométricos de la prueba, consideramos que es necesario hacer un breve recorrido por los datos de otros países más relevantes. En segundo lugar, presentaremos de manera más extensa los datos propios para la población española.

Otros países

Los primeros datos que vamos a mostrar deben necesariamente estar referidos a la estructura interna. Hemos descrito el modelo teórico, pero ahora es necesario demostrar que los datos confirman estos supuestos.

En la siguiente tabla se muestra las correlaciones entre las cuatro variables del modo, para el gráfico 3, para 6 muestras de países distintos.

Tal y como se presentan en la tabla estos datos muestran que la estructura interna de la prueba es constante para todos los países. Esto es de gran importancia a la hora de conceder credibilidad a la prueba, especialmente en aplicaciones internacionales o para uso su internacional en el caso de un tratamiento intercultural de la información psicométrica recogida.

Otro elemento a tener en cuenta es que las variables están relacionadas entre si.

Esta característica, como hemos visto, es propia de los modelos ipsativos ya que la respuesta en un factor condiciona el resultado en los factores restantes.

Por último, es relevante que se demuestre también el carácter bipolar de las dimensiones del APP. Los valores de la D y la I están correlacionados negativamente con la S y la C. Tal y como hemos visto en el modelo, la D y la I son activos en su polo positivo, en tanto que la S y la C son, pasivos, es decir negativos en el continuo bipolar.

Un análisis especialmente importante en cualquier prueba se refiere a la fiabilidad de la misma. En este caso, y para facilitar la lectura sólo vamos a presentar los datos referidos al Reino Unido. Los datos completos de todos los países se pueden consultar en (REF).

Los análisis de fiabilidad que vamos a presentar son los habituales, con cada factor analizado independientemente. Este tipo de análisis no suele beneficiar a los valores que obtenemos de una prueba ipsativa, pero no obstante nos permite comparar los datos con el resto de las pruebas. Un análisis más fino de la fiabilidad de una prueba ipsativa debería de considerar el perfil conjunto en lugar de un análisis diferenciado.

En la Tabla 6 se ofrecen los resultados de la fiabilidad para población inglesa. Estos datos demuestran que cada una de las escalas cumple los requisitos impuestos para las pruebas normativas. Hay que recordar que analizar la fiabilidad de esta manera implica hacer una serie de asunciones que distorsionan e infravaloran la fiabilidad real de las escalas ipsativas.

Tabla 5 TABLA DE CORRELACIONES ENTRE LOS FACTORES DEL DISC EN SEIS PAÍSES				
	D	I	S	C
Dominancia		.03	-.65	-.51
Influencia			-.50	-.58
Estabilidad				.47
Conformidad				
RUSIA: 600 sujetos				
	D	I	S	C
Dominancia		.11	-.70	-.60
Influencia			-.31	-.57
Estabilidad				.38
Conformidad				
HOLANDA: 127 sujetos				
	D	I	S	C
Dominancia		.04	-.72	-.47
Influencia			-.32	-.52
Estabilidad				.28
Conformidad				
TURQUIA: 214				
	D	I	S	C
Dominancia		-.09	-.73	-.50
Influencia			-.31	-.44
Estabilidad				.32
Conformidad				
DINAMARCA: 539				
	D	I	S	C
Dominancia		.05	-.78	-.54
Influencia			-.45	-.61
Estabilidad				.46
Conformidad				
USA: 1512				
	D	I	S	C
Dominancia		-.15	-.75	-.49
Influencia			-.21	-.40
Estabilidad				.24
Conformidad				
REINO UNIDO: 4083				

Tabla 6				
RESULTADOS DE FIABILIDAD PARA LA POBLACIÓN INGLESA				
Escala compuesta	D	I	S	C
Media APP	-.2	1.9	.1	0
Desviación Típica APP	4.2	4.1	4.5	3.4
Media APP	.8	1.2	.3	-.5
Desviación Típica APP	4.4	4.2	4.7	3.6
Test – Retest	.78	.64	.64	.52

Nota: N= 1323.

Los datos se obtuvieron de población británica, de un total de 123 empresas.

Para nuestro propósito de predicción en el contexto de RRHH es más la validez, ya que requiere una fiabilidad previa. En nuestro contexto de trabajo, la validez predictiva referida al rendimiento es el paso previo inicial para determinar si una prueba es útil para su uso en un contexto laboral determinado (tabla 7).

Además de estos datos, Thomas dispone a nivel internacional de datos sobre la validez de constructo y la validez de criterio. Son especialmente relevantes los estudios que correlacionan el APP con el 16PF y el OPQ que por razones obvias no reproducimos en este documento ya que exceden el propósito de este trabajo en particular.

Tabla 7				
VALIDEZ PREDICTIVA				
Escala	D	I	S	C
El total de 1323 individuos	.39	.25	.44	.14
798 con APT en la entrevista	.48	.29	.56	.24

N=1323; 0.05, P<0.17

N=798; 0.05, P<0.22

Población española

Los datos españoles sobre la estructura interna se muestran en la siguiente tabla. Los resultados muestran que mantiene la misma estructura que los datos internacionales (tabla 8).

les eran mujeres), pertenecientes a 12 empresas (tabla 9).

Las pruebas se aplicaron a personas que llevaban en el mismo puesto un tiempo no inferior a un año y la diferencia entre el primer APP y el segundo, se realizó al

Tabla 8				
TABLA DE CORRELACIONES ENTRE FACTORES PARA POBLACIÓN ESPAÑOLA				
	D	I	S	C
Dominancia		-.05	-.69	-.53
Influencia			-.37	-.51
Estabilidad				.35
Conformidad				

N= 3723 sujetos

Fiabilidad:

Los datos de fiabilidad están divididos en dos estudios. El primero fue llevado a cabo entre los años 1992-1993. El número total de personas es de 84 (23% de los cua-

lido de 3 meses. Las muestras pertenecían a empresas de Madrid y Barcelona preferentemente.

En el segundo estudio los datos fueron obtenidos entre 1999 y el año 2000. El

Tabla 9					
FIABILIDAD EN POBLACIÓN ESPAÑOLA					
		D	I	S	C
Momento 1	Media	.4	1.1	.2	.1
	Desviación típica	4.1	4.2	4.3	3.6
Momento 2	Media	.1	1.3	.4	0
	Desviación típica	4.3	4.3	4.7	3.8
Test – Retest		.81	.67	.62	.54

N= 84

número total de personas es de 375, de los cuales el 25% son mujeres; pertenecientes a 25 empresas (tabla 10).

sustantivos de cierta complejidad semántica por frases descriptivas.

Tabla 10					
FIABILIDAD EN POBLACIÓN ESPAÑOLA					
		D	I	S	C
Momento 1	Media	-.73	.21	-.85	-.81
	Desviación típica	4.59	3.20	4	3.29
Momento 2	Media	.07	.14	-.04	-.58
	Desviación típica	4.84	3.4	3.79	3.41
Test – Retest		.89	.74	.76	.77

N= 375

Pruebas paralelas

Un segundo tipo de estudios se refiere a las pruebas paralelas de las que dispone Thomas Internacional. La primera de estas pruebas es el APP PLUS. Este cuestionario se usa cuando a la persona le cuesta entender los adjetivos empleados habitualmente, sustituyendo algunos adjetivos o

Los resultados obtenidos se muestran en la tabla siguiente. Estos datos fueron obtenidos en el estudio de validación llevado a cabo en el año 2000 (tabla 11).

Las pruebas se aplicaron a personas que llevaban en el mismo puesto un tiempo no inferior a un año y la diferencia entre el primer APP y el segundo, se realizó al

Tabla 11					
CORRELACIÓN APP Y APP PLUS					
		D	I	S	C
APP	Media	-.08	.57	.13	.23
	Desviación típica	5,11	3,53	4,07	3,27
APP Plus	Media	-.43	-.34	.43	.05
	Desviación típica	5.03	3.58	3.40	3.22
Test – Retest		.78	.78	.74	.73

Tabla 12					
CORRELACIÓN APP CASTELLANO Y APP EUSKERA					
		D	I	S	C
Castellano	Media	-0.16	-2.16	.49	1.86
	Desviación típica	3.44	2.88	3.09	3.14
Euskera	Media	-.83	-1.4	-.38	1.02
	Desviación típica	3.89	3.27	3.9	3.28

N=113

cabo de 3 meses. Las muestras pertenecían a empresas de Madrid y Barcelona preferentemente.

La segunda prueba desarrolla es el APP en euskera. Este estudio se llevo a cabo durante el año 2004. Los resultados se muestran en la tabla 2.

Validez

La primera estimación de la validez es la Observación Conductual. Los resultados obtenidos se muestran en la Tabla 13.

Respecto a los 84 personas entrevistados (23% mujeres, 77% hombres) pertenecientes a 12 empresas, los jefes a

partir del APT hicieron una escala likert, previamente reajustada y después de realizar la entrevista se procedió a su evaluación, a la vez que se aplicaba el 2º APP.

Todos los sujetos, a la vez que se hacía la 2ª aplicación re-test, fueron evaluados por observación en su trabajo usando las descripciones del APT. Estos resultados, en principio, superan a los valores de la Tabla 7.

Conclusiones

De acuerdo con los resultados de los puntos anteriores, se puede concluir:

Tabla 13				
VALIDEZ APP-ENTREVISTA EN POBLACION ESPAÑOLA				
ESCALAS	D	I	S	C
Sobre toda la muestra 84 sujetos	.38	.31	.42	.23

N= 84; 0.05 p < 0.19

Tabla 14
VALIDEZ APP-OBSERVACIÓN EN POBLACIÓN ESPAÑOLA

ESCALAS	D	I	S	C
Sobre 46 sujetos con APT y entrevista	.49	.32	.60	.28

N= 46

a) Que la validez y la fiabilidad son estadísticamente significativas para la muestra española, e incluso superan los valores de otros países.

b) Que el concepto DISC de Thomas International a través de los análisis APP/APT, son perfectamente transferibles a la población española.

c) En consecuencia se puede considerar que la comercialización del sistema APP/APT cumple con los estándares de calidad psicométricos requeridos para pruebas de este tipo

Conclusiones

Llegados a este punto es bueno hacer una revisión crítica del significado que tiene para nosotros la información presentada y comentada en este artículo. Especialmente porque seguro que queda en el aire la siguiente reflexión: "Estas teorías están muy bien, pero ¿cómo contribuye a solucionar mis problemas de evaluación de las personas?". Vamos a intentar resumir nuestra posición.

Las pruebas de elección múltiple y las pruebas de elección forzosa son claramente distintas. Ambas tienen áreas de aplicación, formas de utilización y por tanto modos de utilización distintas. El objetivo

de ambas es común, evaluar a la persona, la diferencia radica en como cada una responde al criterio de rentabilidad (dentro de la cual es crítica la validez).

Las pruebas normativas funcionan de manera muy eficiente cuando cumplen los supuestos que hemos analizado (las categorías son continuas, nivel de medida de intervalo, independencia de las varianzas error y nivel de distribución), pero hemos visto que en determinadas situaciones, y especialmente en contextos de selección, estos supuestos pueden llegar a distorsionar las puntuaciones de una persona en una desviación típica. Esta diferencia es más que suficiente para incluir en la siguiente fase a un mal candidato o excluir a uno bueno.

Esto no quiere decir, que siempre que empleamos pruebas normativas cometemos sistemáticamente errores, sino que debemos de ser conscientes de que en ciertas ocasiones el riesgo es alto, y puede inducir a pérdidas de precisión de una desviación típica.

En cambio, las pruebas ipsativas, como el APP, no hacen supuestos tan rígidos, por lo que la información intraindividual que obtenemos corre menos riesgo de sufrir variaciones graves; lo que evita que cometamos errores estadísticamente significativos aunque no se cumplan los supues-

tos de las pruebas normativas. No obstante, las pruebas ipsativas son algo más frágiles en la comparación interindividual cuando analizamos dimensiones aisladas. Afortunadamente en el caso de la selección de personal, la toma de decisiones siempre implica que la comparación de todo el perfil de varias personas, por lo que no se muestran pérdidas de potencia en las pruebas ipsativas.

Si consideramos que la situación en la cual vamos a evaluar puede tener el riesgo de cometer errores o poner en peligro los supuestos de las pruebas normativas, entonces es recomendable emplear pruebas de tipo ipsativo, pero que hayan probado poseer las siguientes características.

En primer lugar, la prueba ipsativa debe tener al menos la misma validez que las pruebas normativas, sino más. De nada nos sirve una prueba que no incumpla ningún supuesto si luego no dispone de validez suficiente, porque jamás alcanzaremos nuestro objetivo de rentabilidad.

El segundo criterio crítico es el número de escalas. Es deseable que la prueba tenga un número bajo de escalas. Cuando el número es alto, se pierden las ventajas estadísticas propias de las pruebas de elección forzosa. Tener más escalas no mejora la predicción, como hemos visto la puede incluso perjudicar en ambos tipos de pruebas. Además es probable que la presencia de un mayor número de escalas reduzca la ventaja de tiempo que supone hacer una prueba de elección forzosa.

El análisis del número de escalas es pertinente ligarlo a la correlación que hay entre las distintas escalas. No es lo mismo contar con cuatro escalas que no tiene nada

que ver entre sí, que disponer del mismo número de escalas y además relacionadas. Para el análisis de las puntuaciones lo ideal es mejor que las escalas sean pocas, muy relacionadas y bipolares.

Para maximizar los beneficios de una prueba ipsativa debemos de analizar en detalle el modo en que ha resuelto la mayor dificultad de este paradigma de evaluación, que es la comparación entre individuos. Para poder confiar plenamente en una prueba de elección forzosa requiere que tenga resuelta las herramientas de comparación intraindividual.

Por último, hemos mostrado como el sistema APP/APT cumple los requisitos exigidos a las pruebas ipsativas y puede resolver esos problemas en términos de validez, rentabilidad de la inversión y características psicométricas de las escalas.

Con respecto a los datos de fiabilidad y validez, tanto para población española como para la población internacional, demuestra que los valores son altos y significativos. Esto era esperable por la ausencia de fuentes de error, pero era necesario probarlo. Este dato todavía cobra más valor teniendo en cuenta que las muestras utilizadas se han obtenido en situaciones cotidianas de procesos de selección, evaluación y desarrollo análogas a las que hace todos los días cualquier profesional de RRHH. Todos sabemos, aunque sólo se por simple observación, que esos candidatos poco o nada tienen que ver con los estudiantes de cualquier universidad o escuela de negocios donde son utilizados casi como un grupo experimental.

En definitiva, creemos que el sistema APP/APT de Thomas International es una

prueba ipsativa de elección forzosa, sólida, rápida con una excelente capacidad predictiva y que ha sido desarrollada para servir de apoyo eficaz a la gestión racional de los recursos humanos en las organizaciones.

REFERENCIAS BIBLIOGRÁFICAS

- Baron, H. (1995). Strengths and limitations of Ipsative Measurement. *Journal of Occupational and Organisational Psychology*.
- Barrett, P. (1999). Test Design, Measurement and Assessment. En <http://www.liv.ac.uk/pbarrett/paulhome.htm>
- Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organisational Psychology*, 69, 25-39.
- Berry, J.W. y Irvine, S.H. (1986). Bricolage: savages do it daily. In R.J. Sternberg y R.K. Wagner, (Eds.). *Practical Intelligence*. New York: Cambridge.
- Catell, R. B. (1988) The data box: Its ordering of total resources in terms of possible relational systems. In J.R. Nesselroade y R. B. Catell (Eds.) *Handbook of multivariate experimental psychology*. NY: Plenum Press.
- Closs, S. J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organisational Psychology*, 69, 41-47.
- Hendrickson, T.M. (Undated). *Personal profile analysis: a technical manual*: Marlow: Thomás International Management Systems (Europe) Ltd.
- Irvine, S.H.; Mettam, D. y Syrad, T. (1994). Valid or more valid? Keys to understanding personal appraisal practices at work. *Current Psychology*, Spring 1994, Vol 13, nº1, 27-59.
- Johnson, C.E., Wood, R. y Blikhorn, S.F. (1988). Spuriousness and spuriousness: the use of Occupational Psychology. 57: 67-76.
- Karpatschof, B. y Elkjaer, H. (1999). The reliability of Ipsative scores- Examined by espiral data and simulation study. Research report number 1. Department of Psychology. University of Copenhagen.
- Kline, P. (1983). *Personality: measurement and theory*. London. Hutchinson.
- Kline, P. (1998). *The new Psychometrics: Science, Psychology and Measurement*. Routledge.
- Lord, F. M. y Novick, M. R. (1968) *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley.
- Marston, W.M. (1928). *Emotions of normal people*. New York: Harcourt, Brace.
- Michell, J. (1990). *An introduction to the logic of Psychological Measurement*. Lawrence Erlbaum.
- Ones, D. S., Viswesvaran, C. y Korbín, W. P. (1995) Meta-analyses of fakability estimates: Between subjects versus within subjects designs. Paper presented at the 10th annual meeting of the Society of Industrial and Organisational Psychology, Orlando, FL.
- Thomas International (1981a). *Personal*

Profile Analysis (Revised) Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1981b). Human Job Analysis (Revised) Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1983a). Interviewing and screening guide. Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1983b). Advanced analysis course manual. Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1983c). Selecting effective people. Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1983d). PPA advanced profile guide. Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1986a). TIRR-1. Concurrent validity among instruments used by industry for personnel purposes, 16 PF, OPQ and PPA Part.1 (B. Cripps y E. Geelhoed). Marlow: Thomas Interna-

tional Management Systems (Europe) Ltd.

Thomas International (1986b). TIRR-3. Concurrent validity among instruments used by industry for personnel purposes, 16 PF, OPQ and PPA Part.2 (B. Cripps y E. Geelhoed). Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1987a). TIRR-2. Description of 4083 PPA results: correlations and distributions (B. Cripps y E. Geelhoed). Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1987b). TIRR-4. Description of 5242 PPA results: correlations and distributions (B. Cripps y E. Geelhoed). Marlow: Thomas International Management Systems (Europe) Ltd.

Thomas International (1987c). TIRR-6. Predictive validity, illusion or reality (B. Cripps y E. Geelhoed). Marlow: Thomas International Management Systems (Europe) Ltd.

Viswesvaran, C. y Ones D. S. Meta-Analyses of Fakability Estimates: Implications for Personality Measurement Educational and Psychological Measurement. 1999; 59: 197-210.