

## **Comparación entre-sujetos e intra-sujetos: ¿incompatibles o complementarias?**

*(Between-subject versus within-subject comparison:  
Incompatible or complementary?)*

**Rafael Martínez y Rafael Moreno\***

Universidad de Sevilla

Contemplando la posibilidad de integración de procedimientos y conceptos de las teorías conductual y de tests psicométricos en Psicología, hemos asumido como posición de partida que es posible explorar ciertas compatibilidades entre ambas orientaciones. Es necesario reconocer que esta asunción se enfrenta a una habitual desconfianza e ignorancia mutua. Ante este debate una opción es adoptar la noción de inconmensurabilidad paradigmática que desarrollara Kuhn (1962/1977) en su teoría de las revoluciones científicas, y justificar que los investigadores de una u otra orientación no tomen en cuenta los conceptos y procedimientos de orientaciones distintas a la propia. Otra opción, que es la que adoptamos aquí, es asumir la visión evolutiva de la ciencia que ofrece Toulmin (1977), en la que se considera que entre distintas orientaciones conceptuales pueden encontrarse elementos contradictorios y de ruptura, pero también elementos compartidos.

Nuestro interés en explorar posibles elementos compartidos entre las orientaciones conductuales y psicométricas en psicología se sustenta en consideraciones de distinto tipo. En primer lugar porque históricamente ambas son las que más énfasis han puesto en la medición a partir de datos observables de comportamiento como criterio básico para el carácter científico de los estudios psicológicos. Ese énfasis común en los datos observables ha estado relacionado con la influencia que ambas orientaciones recibieron

\*Dirigir toda correspondencia a los autores a:Departamento de Psicología. Facultad de Psicología. Calle Camilo José Cela s/n. 41018 Sevilla (España)

Correo electrónico: rmcervan@us.es, rmoreno@us.es

del positivismo y el operacionalismo, ilustrado en los trabajos de Skinner, Boring y Stevens (veáse para más detalle Houts, 1994). En segundo lugar, y como consecuencia de sus logros en obtener datos objetivos, otro motivo para nuestro interés es el que ambas han servido de fundamento para algunas de las aplicaciones más usadas en la práctica profesional de la psicología.

Así lo demuestra por una parte el uso generalizado de tests por parte de los profesionales de la psicología, utilizados como medios (a menudo estandarizados) para obtener respuestas con las que pueden relacionarse comportamientos humanos en otros contextos; como por ejemplo cuando se pretende predecir el éxito en la escuela mediante tests de aptitudes. En contextos tales como servicios de orientación, clínicas de salud mental y hospitales psiquiátricos, los tests de habilidad y personalidad suelen ser útiles para el diagnóstico y evaluación de problemas de comportamiento. Asimismo empresas y gobiernos por igual utilizan de forma profusa tests para seleccionar a sus trabajadores. Por otra parte el análisis conductual aplicado, las terapias de conducta, las técnicas de modificación de conducta y la evaluación conductual, suponen una aplicación de principios derivados de la filosofía conductista, y de conocimientos derivados experimentalmente de la teoría del aprendizaje, al diagnóstico y tratamiento de problemas de comportamiento. Estas técnicas se aplican con éxito en múltiples patologías tales como enuresis, tics, fobias, tartamudeos, conductas obsesivo-compulsivas, adicciones, conductas neuróticas, así como a la instrucción en el campo del retraso en el desarrollo. En el campo educativo también se usan con profusión técnicas basadas en los estudios sobre el aprendizaje, como las economías de fichas o los tiempos-fuera.

Silva (1993) ha realizado una profunda revisión de las diversas posturas contrapuestas entre los planteamientos psicométricos y conductuales, así como de los supuestos y conceptos que subyacen a ambas. Según este autor la clave para entender una pretendida incompatibilidad entre ambos enfoques reside en las preconcepciones acerca de la naturaleza y epistemología de los constructos psicológicos. Lo anterior se concreta en tres aspectos que enfrentan a los modelos psicométrico y conductual, respectivamente: i) una concepción estable y consistente del comportamiento frente a una concepción variable y situacional; ii) un modelo de explicación estructural intra-sujeto frente a un modelo de explicación funcional entre sujeto y ambiente; y iii) un análisis del comportamiento desde un marco de comparación inter-individual frente a un análisis desde un marco de comparación intra-individual.

Este último enfrentamiento de carácter más metodológico, el de la comparación intra-sujeto frente a la comparación entre-sujetos, es el que nos interesa desarrollar en este trabajo. Respecto a esta cuestión suele asociarse a las orientaciones conductuales un interés exclusivo en las comparaciones de la ejecución de un sujeto consigo mismo, llevadas a cabo fundamentalmente mediante análisis visual. En cambio, a las

orientaciones psicométricas se les supone un interés exclusivo por la comparación de la ejecución de un sujeto con la de otros, realizada en este caso mediante el recurso a métodos estadísticos. Estas asociaciones alimentan prejuicios que, como ha señalado Viladrich (1991), llevan a que en cada determinado contexto de investigación sustantiva los investigadores se auto-limiten al uso de determinados procedimientos y técnicas de análisis de datos, criticando con ferocidad otras técnicas por deformadoras de la información original, al tiempo que son condescendientes con las deformaciones de las técnicas a las que están habituados. Este tipo de actitud es la que parece encontrarse en algunos analistas de orientación conductual cuando expresan un tajante rechazo a la psicometría y los tests estandarizados, ampliado a veces a la estadística. En el ámbito psicométrico por el contrario lo que suele encontrarse es ignorancia de las orientaciones conductuales más que un rechazo sistemático y expreso, al considerarlas producto de un paradigma históricamente superado en línea con las afirmaciones de la denominada revolución cognitiva de 1956 (Bruner, 1991). El problema con estas asociaciones es que simplifican la realidad y no toman en consideración que dentro de una y otra orientación, psicométrica y conductual, puede haber múltiples variaciones y matices que desmienten esas simplificaciones. En los próximos apartados tratamos de mostrar algunos matices que pueden ayudar a superar estas simplificaciones.

## **PSICOMETRÍA, TESTS NORMATIVOS Y TESTS REFERIDOS A UN CRITERIO**

Hablar de orientaciones psicométricas como un todo homogéneo es una denominación vaga e imprecisa que se justifica sólo por el contexto introductorio en el que la hemos usado. La psicometría es la disciplina que estudia la medición de lo psicológico (Muñiz, 1998), incluyendo habitualmente en su interior ramas tan diversas como la psicofísica, la teoría de la medida o la teoría de los tests. A su vez cada una de estas especialidades puede incluir varias orientaciones no siempre coincidentes. Así ocurre con las teorías axiomática y representacional de la medida, o con la teoría clásica de los tests, la teoría de la generalizabilidad o la teoría de respuesta a los ítems. Centrándonos en los tests, las teorías psicométricas se han especializado en los aspectos lógicos y en los modelos matemáticos asociados a la construcción y uso de pruebas psicológicas (Crocker y Algina, 1986). Estas teorías han permitido generar metodologías estadísticas para evaluar entre otras cosas los criterios de fiabilidad y validez de las pruebas psicológicas, entendidos como requisitos necesarios de toda medida. En este contexto los tests psicológicos se consideran muestras de conducta para realizar inferencias (Muñiz, 1998b), que pueden ir desde pruebas de aptitudes y de rendimiento, inventarios de intereses, escalas de actitudes con opciones cerradas

de respuesta, hasta escalas de valoración, exámenes abiertos, listas de observación o instrumentos proyectivos en los que un evaluador debe valorar las respuestas (Suen, 1990). En todo caso, estos instrumentos se denominan psicométricos en la medida que supongan procedimientos estandarizados desarrollados de acuerdo a criterios que aportan validez y fiabilidad, constituyendo un marco definido y estable para la interpretación de las respuestas individuales.

No todos los tests psicométricos llevan asociada una comparación entre sujetos. Esa característica se ajusta a los tests normativos, pero no se corresponde sin embargo con los tests referidos a un criterio. La clave que determina la diferencia entre uno y otro tipo de tests está en la manera de interpretar el comportamiento individual en una prueba. Para interpretar la puntuación de un individuo en una prueba es necesario compararla con algún referente, ya que en ausencia de datos adicionales que ayuden a la interpretación una puntuación directa carece de significado (Embretson y Hershberger, 1999). Señalar que un individuo ha resuelto correctamente 15 problemas de una prueba matemática, o identificado 34 palabras de una prueba de vocabulario, nos da poca información sobre sus habilidades o aptitudes en cualquiera de estos dominios. Tampoco resuelve completamente este problema el cálculo de los porcentajes de aciertos. Una puntuación del 60% de aciertos en una prueba matemática (p. ej. acertar 15 de un total de 25 problemas) puede suponer una habilidad equivalente a acertar el 30% en otra prueba o el 90% en una tercera. La selección de los items que componen cada prueba y las dificultades que presentan determinará el significado de cada puntuación. Tanto las sumas de aciertos (puntuaciones directas) como sus porcentajes sólo pueden ser interpretados en términos de un marco de referencia claramente definido y uniforme.

En el ejemplo anterior no sabemos si el individuo es mejor en matemáticas o en vocabulario. Si la prueba de matemáticas es mucho más fácil que la de vocabulario, acertar 15 de 25 (60%), ello no indica que el sujeto resuelve mejor en ese dominio que en el de vocabulario en el que acierta 34 de 100 palabras (34%). Si la prueba de vocabulario es mucho más difícil, los porcentajes no son directamente comparables. Las normas o baremos ayudan a interpretar las puntuaciones en una prueba en comparación a la ejecución en esa prueba de un grupo normativo. Los baremos se establecen determinando empíricamente lo que una muestra representativa de personas hace en una prueba. De este modo, la interpretación de la puntuación individual se hace por comparación con la ejecución del grupo representativo. Las denominadas normas de desarrollo permiten comparar la puntuación individual con la puntuación media por grupos de edad, mientras que las normas de curso permiten comparar la puntuación individual con la media obtenida por grupos representativos de cada nivel escolar. Por su parte, las normas intragrupo permiten que la ejecución individual sea interpretada en comparación con la ejecución del grupo más cercano, como por ejemplo respecto a

la puntuación con otros niños de la misma edad o del mismo curso escolar. Con tal fin en lugar de las puntuaciones directas se suelen utilizar puntuaciones derivadas como los percentiles y las puntuaciones típicas. El percentil indica la posición relativa del individuo respecto a la muestra estandarizada. La puntuación típica expresa la distancia individual respecto a la media del grupo en términos de la desviación típica de la distribución. Estas medidas facilitan además la comparación de la ejecución de un mismo sujeto en distintas pruebas, siempre que las distribuciones tengan una forma parecida. En la figura 1 mostramos una distribución normal, que es la que se suele ajustar a las distribuciones de las puntuaciones obtenidas en pruebas psicológicas y su correspondencia en percentiles.

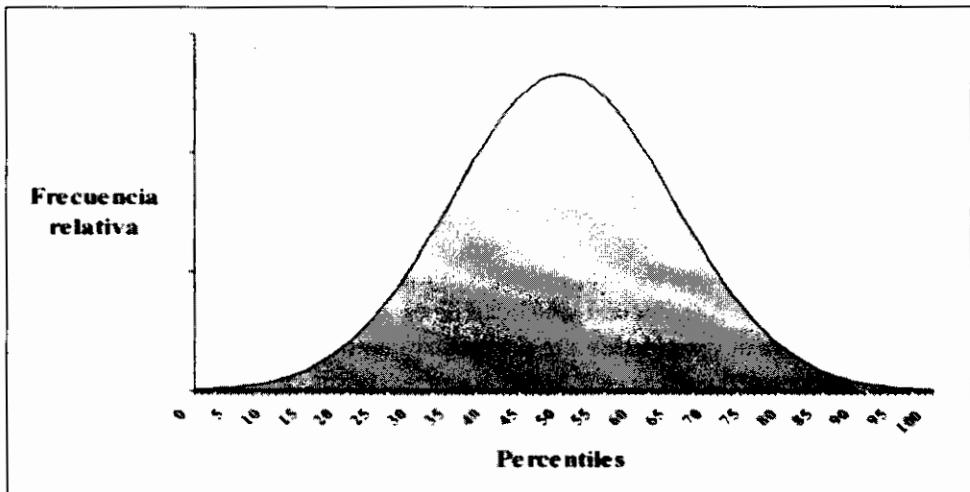


Figura 1.- Distribución normal de frecuencias relativas según los percentiles.

El principal problema de los baremos es que son relativos y dependen de las características del grupo con el que se compare (Anastasi y Urbina, 1997). Si un sujeto obtiene un percentil 30 en aptitud verbal en una muestra de sujetos muy hábiles, mientras que obtiene un percentil 70 en aptitud espacial respecto a una muestra con aptitudes bajas, no resulta posible considerar que el sujeto tenga mejor aptitud espacial que verbal, ya que un percentil 30 puede mostrar en realidad una mejor ejecución en aptitud verbal que un percentil 70 en aptitud espacial. Esta dependencia respecto a la composición de la muestra de sujetos en que se elabora hace especialmente importante

que las muestras sean representativas para que los baremos sean realmente útiles. Por ejemplo, comparar la altura y el peso de un niño con las distribuciones de alturas y pesos de niños de la misma edad aporta información útil para valorar el desarrollo físico del niño siempre y cuando las distribuciones procedan de grupos representativos de la población a la que pertenezca dicho niño. Esta comparación nos informa, por ejemplo, de cuando un niño tiene un peso por debajo de lo esperado o tiene una altura que sobrepasa de lo normal para su edad. En el mismo sentido de permitirnos anticipar posibles problemas de comportamiento puede ser útil comparar la ejecución de un niño en matemáticas o en lengua con la ejecución de una muestra representativa de niños de su misma edad y curso escolar. Una ejecución de un niño excesivamente baja en comparación con lo que son capaces de hacer los niños de su mismo grupo de referencia nos alerta de poca competencia en esas materias y de la necesidad de alguna ayuda adicional para alcanzar unos niveles adecuados.

Aunque la comparación de la puntuación individual con el baremo estadístico obtenido en un grupo normativo no nos informa de lo que un individuo realmente sabe hacer ni de las posibles causas de su diferencia comparativa, sin embargo sí nos informa sobre la frecuencia relativa de ese comportamiento en la población. Es posible hacer de ese modo una interpretación probabilística de una determinada ejecución en una prueba psicológica. Esto permite valorar si una determinada ejecución es esperable o no, por exceso o por defecto, en individuos con las mismas características que definen a su grupo de referencia. La utilidad de esta información es sobre todo de tipo práctico, orientada al diagnóstico del comportamiento individual y en su caso a una posterior intervención.

Pero además de los tests normativos, en el ámbito psicométrico se ha desarrollado la noción de tests referidos a un criterio (Glaser, 1963) desde una perspectiva completamente diferente. La interpretación en estos tests está referida a un contenido o a un dominio. Este tipo de pruebas utiliza como marco de referencia para la interpretación un dominio específico de contenido, en lugar de una población de personas. En los tests referidos a un criterio la ejecución individual es interpretada en términos de los tipos especiales de contenidos que domina (por ejemplo el tipo de operaciones aritméticas o el tamaño estimado del vocabulario que maneja). Este tipo de tests se utiliza fundamentalmente en el ámbito educativo y se ajustan sobre todo a la evaluación de habilidades básicas y del rendimiento. Se utilizan, entre otras cosas, para indicar si un individuo ha alcanzado o no un nivel preestablecido de ejecución competente. Por ejemplo, cuando se trata de habilidades básicas, suele exigirse una ejecución casi completamente correcta (desde un 80 u 85% de ítems acertados).

La ejecución se interpreta en términos del significado del contenido, poniendo el interés en lo que el examinando puede hacer y conoce, y no en cómo resulta en la

comparación con otros sujetos. Para la interpretación de la ejecución individual en estas pruebas es clave determinar qué proporción de ítems acertados va a utilizarse como criterio para que se pueda establecer de forma fiable el nivel de competencia de un sujeto, para lo que la psicometría ha desarrollado una tecnología estadística que permite establecer esos puntos de corte con garantías de fiabilidad (Muñiz, 1998b). Un requisito fundamental para construir una prueba de este tipo es definir de forma clara un dominio, a partir de la cual se seleccionan las muestras de ítems que van a componer las pruebas. Sin una especificación cuidadosa y un control del contenido, el resultado de un test referido a un criterio degenera en algo idiosincrásico y no interpretable.

En todo caso, la existencia de este tipo de tests psicométricos contradice la afirmación de que la psicometría sólo se ocupa de la comparación estadística entre sujetos. La preocupación por la ausencia de significado de la interpretación de los resultados de los tests en referencia a normas o baremos estadísticos es algo de lo que tradicionalmente se han ocupado los propios psicómetros. Como señalan Embretson y Hershberger (1999), los trabajos pioneros de Thurstone (1928) y Thorndike, Bregman, Cobb y Woodyard (1926) trataban precisamente de desarrollar métodos de interpretación de las puntuaciones referidos al contenido de los ítems. La emergencia de la teoría de respuesta a los ítems dentro de la psicometría parece haber renovado ese interés por superar la interpretación de las puntuaciones en referencia a normas estadísticas. Así lo ilustran trabajos como el de Wright (1999), quien aplica un modelo de Rasch a una escala de valoración comportamental de estilos de vida para evaluar auto-cuidados en ancianos, o el de Woodcock (1999) que, utilizando también el citado modelo psicométrico, desarrolla distintos índices referidos al contenido de los ítems para la interpretación de las puntuaciones obtenidas en su batería sobre dominio lingüístico, obteniéndose en ambos casos información sobre lo que la persona puede hacer.

## **ESTUDIOS CONDUCTUALES, ERROR DE MEDIDA Y PRUEBAS PSICOLÓGICAS**

Al igual que comentamos respecto a las orientaciones psicométricas, hablar de orientaciones conductuales es una forma vaga de referirse a múltiples variantes y especialidades básicas y aplicadas, no siempre acordes entre sí. En general todas estas variantes denominadas conductuales se basan en mayor o menor medida en el conductismo como filosofía de la psicología, ya sea en su versión metodológica o en su versión radical. Dentro de la teoría de la conducta se incluyen desde las teorías de los reflejos condicionales de Pavlov y la triple relación de contingencia de Skinner, con la distinción respondiente/operante, hasta modelos más recientes como los de Staats (1996; 2003), la teoría de los marcos relacionales de Hayes, Barnes-Holmes y

Roche (2001) o el modelo de campo y paramétrico de Ribes (Ribes y López, 1985; Ribes y Sánchez, 1990; 1992). El análisis experimental del comportamiento se agrupa en torno a la metodología y las técnicas de investigación. Por su parte el análisis conductual aplicado, las técnicas de modificación de conducta, las terapias de conducta y la evaluación conductual, representan ramas aplicadas al diagnóstico y tratamiento de problemas de comportamiento. En general, lo que comparten todas estas modalidades es el acuerdo en la conducta observable como el dato básico a medir.

Johnston y Pennypacker (1980) señalan que en general las medidas conductuales no necesitan más que un pequeño número de sujetos, que observados de manera intensiva mediante diseños experimentales potentes permiten eliminar el mayor número de fuentes de error. La clave para estos autores está en que las medidas conductuales son idemnóticas, lo que supone el uso de unidades estándares y absolutas que permiten la definición de respuestas y facilitan la observación directa. A diferencia, las medidas vaganóticas, propias de la psicometría, requieren gran número de sujetos, comparación entre grupos, uso de la estadística inferencial y asumen error inter-sujetos. La conclusión de Johnston y Pennypacker (op.cit) es que las diferencias entre ambas estrategias de medida son irresolubles, porque dependen de la concepción de la naturaleza del objeto de estudio. Según estos autores, los que utilizan medidas vaganóticas asumen que su objeto de estudio es intrínsecamente variable, lo que no sucede con los que utilizan medidas idemnóticas como las conductuales. En este último caso, el interés por los eventos comportamentales públicos motiva a describirlos como unidades de medidas estándares y absolutas, por lo que el establecimiento de criterios para la fiabilidad de la medida —o de su grado de error- es fácil de realizar.

El problema del error es una de las preocupaciones centrales de la teoría de la medida. Inicialmente se pensó que los errores de medida podían llegar a ser eliminados a través del refinamiento de principios científicos y del desarrollo de la tecnología. Esta creencia ya no es mantenida hoy en día por la mayoría de los científicos, y en la actualidad es habitual que las medidas sean acompañadas de alguna indicación de su precisión o del grado probable de error. Sin embargo Sidman (1960), ha destacado que sólo en teorías como las de la física, dado el grado de precisión que han logrado en sus medidas, el asumir un grado inherente de variabilidad de error no acarrea costes en su capacidad explicativa. En el ámbito de las ciencias de la conducta, con mucho menor poder explicativo, el asumir esa variabilidad como error inherente conllevaría el abandono de intentar reducirlo y desanima de la búsqueda de sus causas.

Este argumento se puede revertir, sin embargo, respecto a la consideración de las medidas conductuales como idemnóticas. Si asumimos que son unidades estándares y absolutas, propias de un modelo determinista, eso nos lleva a abandonar la búsqueda de posibles errores de medida y desestima cualquier tecnología para evaluar su



fiabilidad, que es lo que ocurre en muchos estudios conductuales. Así lo afirma Silva (1993), cuando comenta que la evaluación conductual, por sus orígenes en la psicología experimental, no se preocupó por asuntos como la validez o la fiabilidad de las medidas. Según este autor el modelo de base determinista y experimental, contrario a una alternativa probabilista y correlacional, es el que ha producido un tenaz rechazo de la metodología estadística que se ha extendido a los criterios métricos de calidad. Para él, los altos costes y las dificultades de aplicación de muchas técnicas de evaluación conductual invitan a una integración de procedimientos psicométricos.

La integración que reclama Silva (op. cit.) es algo más complejo y con más dimensiones de las que hemos tratado aquí. Sin embargo, quizás una ventana a través de la cual sea posible explorar cierta complementariedad pueda ser el error de medida. Dada su ubicuidad en toda ciencia resulta difícil pensar que la investigación conductual se haya librado de los errores de medida de tal forma que no necesite evaluarlos. Para este propósito la tecnología psicométrica podría ser sin duda de gran ayuda para los estudios conductuales. Para ello sería útil diferenciar los principales tipos de error que configuran el error de medida. Por un lado se encuentran los errores de observación, que incluyen errores de los instrumentos y personales, sistemáticos y aleatorios. Por otro lado se encuentran los errores de muestreo, en tanto que selección de las circunstancias en las que se recogen los datos y que incluyen la determinación al menos de a quién, dónde, cuándo y cómo se observa la conducta.

La noción de error de muestreo inmediatamente invita a pensar en el muestreo de sujetos como principal fuente de problemas, lo cual no es motivo de preocupación en los estudios conductuales, en tanto que desestiman las inferencias a una población de sujetos más amplia. Sin embargo el concepto de error de muestreo abarca más aspectos que el de los sujetos. Por ejemplo, en el caso del desarrollo de medidas conductuales a partir de la ejecución del sujeto en una prueba surge la misma cuestión que ante el desarrollo de un test psicométrico: La selección o muestreo de las conductas que se van a observar en una determinada prueba (Muñiz, 1998b). Este muestreo de conductas, se realiza indirectamente a través de los estímulos que se le presentan al sujeto en cada ensayo o ítem de la prueba –lo que supone a su vez un muestreo de estímulos-. El principal problema asociado al muestreo de estímulos proviene de que las medidas utilizadas sobre la ejecución de los sujetos, estén basadas en las respuestas obtenidas en pruebas o tareas específicamente diseñadas ad hoc de acuerdo a las hipótesis de una investigación, lo que se ha denominado como medidas a la medida (Meltzoff, 2000). Es un ‘pecado’ en todo caso no específico de los estudios conductuales, sino compartido por la mayor parte de la investigación básica en psicología, cualquiera que sea su orientación, como analizan Schmidt y Hunter (1996).

En cualquier caso, la medida básica proviene de un recuento de las respuestas que corresponden a un determinado criterio –por ejemplo que se consideren aciertos– y que sumadas por grupos de ítems o por sesiones experimentales proporcionan las frecuencias absolutas, a partir de las cuales se suelen calcular otros estadísticos descriptivos como las frecuencias relativas y los porcentajes de respuestas, componentes habituales de los gráficos utilizados para el análisis visual en los estudios conductuales. El uso de estos estadísticos y gráficos supone aceptar un modelo de medida que no suele hacerse explícito ni se evalúa. Dicho modelo implica asumir que es correcto y posible sumar las respuestas de los sujetos en los diferentes ensayos puesto que responden a una misma dimensión. Si realmente es correcta esta suposición es algo que sin embargo no se suele evaluar. En este sentido, la aportación de cada ítem o ensayo a la coherencia u homogeneidad del conjunto va a determinar la calidad de la medida final. La clave para este objetivo está de nuevo en el muestreo de estímulos; es decir, en la selección de los ítems que componen una prueba, una vez que su propósito ha sido establecido y decidido el formato de los ítems que se van a usar (Moreno, Martínez y Muñiz, 2004). Para esa selección de estímulos es importante contar con algún modelo explicativo que relacione las propiedades de los ítems con las de las respuestas. Este modelo debería ser probado empíricamente, además de contar con información empírica sobre las propiedades de los ítems. Con los datos de un estudio ya recogidos la tecnología psicométrica permite un análisis detallado del comportamiento de los ítems, y así evaluar su aportación a la calidad de la medida y su pertinencia como elemento de una prueba.

## MATRIZ DE DATOS COMPARTIDA

Si retomamos la inicial influencia positivista compartida por los tests psicométricos y la teoría de la conducta y nos fijamos en los datos observables de ambas orientaciones, llama la atención la similitud de algunos de sus procedimientos de recogida de datos. Eso ocurre por ejemplo entre los tests de igualdad (Anastasi y Urbina, 1997) y algunas de las tareas de discriminación condicional aplicadas en humanos (por ejemplo, Ibáñez, 1992). La aplicación de uno u otro tipo de pruebas a unos sujetos nos va a proporcionar unas respuestas que pueden representarse en similares matrices de datos de sujetos por estímulos (tabla 1), donde  $S_n$  representa al Sujeto  $n$ ;  $e_m$  corresponde al estímulo  $m$ ; y  $r_{nm}$  sería la respuesta dada por el sujeto  $n$  al estímulo  $m$ .

De las respuestas dadas por los sujetos ( $r_{nm}$ ), en cada celda de la matriz se puede registrar bien algún aspecto morfológico –como por ejemplo la opción seleccionada por el sujeto en cada ítem o ensayo, su latencia o su duración– o bien aspectos competenciales –como al registrar la precisión de la respuesta en términos de acierto o

error-. En este último caso, mientras que las sumas de los aciertos por filas de la matriz (sumas de  $r_{11}$  a  $r_{1m}$ ) nos puede informar de la aptitud o habilidad en la prueba del sujeto  $S_1$ , las sumas de los aciertos por columnas (sumas de  $r_{11}$  a  $r_{n1}$ ) nos puede informar de la dificultad de elementos que componían el ítem o estímulo  $e_1$  de la prueba. Si en los estudios conductuales se considera legítima la suma de los aciertos obtenidos por un mismo sujeto, igualmente será legítima una suma de los aciertos obtenidos por un mismo estímulo o ítem entre distintos sujetos, ya que se parte de la misma matriz. Este último dato dividido por el total de sujetos que intentan responder al ítem, no es otro que el índice psicométrico de dificultad de los ítems. Este índice nos permite escalar una propiedad básica de los ítems, su dificultad, que se relaciona directamente con las propiedades psicométricas globales de la prueba y nos aporta información muy útil para la tarea de muestreo de estímulos necesaria para configurar una prueba y reducir el error de medida.

**Tabla 1.-** Respuestas obtenidas en una matriz de sujetos por estímulos

	$e_1$	$e_2$	$e_3$	...	$e_m$
$S_1$	$r_{11}$	$r_{12}$	$r_{13}$	...	$r_{1m}$
$S_2$	$r_{21}$	$r_{22}$	$r_{23}$		$r_{2m}$
$S_3$	$r_{31}$	$r_{32}$	$r_{33}$		$r_{3m}$
.	.				.
.	.				.
.	.				.
$S_n$	$r_{n1}$	$r_{n2}$	$r_{n3}$	...	$r_{nm}$

Esta propiedad de los estímulos es legítima considerarla dentro de los estudios conductuales, como ilustra un trabajo de Ribes y Martínez (1990) el que plantean diferencias de dificultad entre ensayos -o ítems- de discriminación condicional con figuras de dos modalidades de forma y color. Estos autores diseñan así una prueba en la que consideran ensayos de tres niveles de dificultad, basándose en el número de formas y de colores de los estímulos así como en el número de estímulos de elección correctos, aunque este supuesto sobre la dificultad de los ítems no queda evaluado empíricamente.

La matriz de datos que hemos presentado en la tabla 1 se ajusta a lo que se denomina un marco de referencia de dos vías, en el que los conceptos a medir son complementarios y se definen unos en términos de los otros mediante una definición

constitutiva o conjunta. La noción de fenómenos que se definen conjuntamente fue adecuadamente expresada por Yela (1982) respecto a los sujetos psicológicos y su medio estimulante cuando señalaba que no hay por una parte un sujeto y por otra un medio estimulante y, luego, una relación entre ellos, sino que los dos se definen conjuntamente en necesaria y constante interacción. Un marco de referencia de dos vías es el de menor orden posible, que en el caso de las medidas obtenidas a través de pruebas psicológicas implica dos clases de entidades: una formada por las respuestas (las cuales se suman o agrupan de acuerdo a los sujetos que las emiten) y otra formada por los ítems (los cuales se suman o agrupan de acuerdo a la prueba psicológica de la que forman parte).

Mientras que para escalar respuestas es necesaria una comparación del mismo sujeto –intra- a través de diferentes estímulos, para escalar estímulos es necesario comparar el mismo estímulo a través de diferentes sujetos –entre-. De este modo se entiende que el escalamiento de ambos tipos o clases de entidades (respuestas e ítems) es complementario, a causa de lo cual se denomina medición simultánea conjunta o medición conjunta. Puede parecer que cuando se utiliza una prueba sólo el sujeto es medido, y sin embargo en este caso el escalamiento conjunto se ha realizado anteriormente como parte de la construcción del instrumento, realizándose la medición por tanto en términos de ese escalamiento conjunto. Las comparaciones dentro de una clase de un marco de referencia de dos vías (p. ej. respuestas de los sujetos o ítems de la prueba), son independientes de las comparaciones dentro de la otra clase, pero se complementan al ser constitutivas de una misma matriz.

## **A MODO DE CONCLUSIÓN**

El argumento que hemos tratado de desarrollar en este trabajo es que quizás sea posible compatibilizar algunos procedimientos de la teoría de tests y de la teoría de la conducta. Al margen de lo convincentes que hayamos sido, no nos parece una empresa del todo absurda al ser compartida por otros autores. Se pueden citar a modo ilustrativo algunos intentos de complementar estudios conductuales y psicométricos. Así por ejemplo, Fernández (1989) considera viable un test para medir la denominada personalidad Tipo A, el test GFP70, fundamentándose en el análisis experimental del comportamiento, como un ejemplo de lo que denomina psicometría operante. En el ámbito de la personalidad, trabajos como los de Doval (1995), Doval y Viladrich (1994), Doval, Viladrich, Ribes y Villanueva (1993), Doval, Viladrich y Riba (1998), Hernández, García, Rubio y Santacreu (2004) y Santacreu y García (2000) han explorado de manera sistemática algunas posibilidades de la metodología psicométrica para el estudio de la personalidad desde una perspectiva conductual. Viladrich y Doval (1998)

subrayan la coincidencia entre el estudio conductual y psicométrico de la personalidad en la búsqueda de consistencias en el comportamiento individual a través de diversas situaciones espacio-temporales, aunque se enfrentan en el tipo de medidas pertinentes o en las inferencias que permiten. En el ámbito de la evaluación de aptitudes intelectuales desde perspectivas interconductuales como las de Ribes (1990), se han llevado a cabo procedimientos para la medición de diferencias individuales, como ocurre con el trabajo de Ibáñez (1992), quien desarrolla un procedimiento informatizado para la medición de aptitudes intelectuales basado en la tarea de igualación de la muestra para discriminar entre sujetos de alta y baja competencia intelectual, y que propone para la selección de estudiantes universitarios.

Relacionado con este último campo, nosotros mismos hemos explorado con cierto éxito compatibilidades entre la metodología psicométrica y la utilización de tareas de discriminación condicional (Martínez, 2000; Martínez y Moreno, 2002). Entendemos que aspirar a complementar conceptos y procedimientos psicométricos desarrollados en el ámbito de la teoría de tests con el análisis funcional del comportamiento propio de la teoría de la conducta puede ser de utilidad para aspectos tales como la mejora de instrumentos de medición en psicología, y ello mediante la reducción de errores de medida gracias a una mejor comprensión del control experimental de la relación entre estímulos y respuestas. Pero del mismo modo puede servir para fundamentar adecuadamente la fiabilidad y validez de las medidas utilizadas en áreas de la psicología realizadas desde perspectivas conductuales.

## REFERENCIAS

- Anastasi, A. y Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice-Hall.
- Bruner, J. (1991). *Actos de significado. Más allá de la revolución cognitiva*. Madrid: Alianza.
- Crocker, L. y Algina, J. (1986). *Introduction to classical and modern test theory*. Nueva York: Holt, Rinehart and Winston.
- Doval, E. (1995). *Estudio del estilo interactivo de tendencia al riesgo: metodología para la evaluación de las diferencias individuales*. Tesis doctoral no publicada: Universidad Autónoma de Barcelona.
- Doval, E. y Viladrich, M. C. (1994). *Predicción del comportamiento en situaciones arriesgadas*. Comunicación presentada en el II Coloquio sobre Psicología Interconductual, Madrid.
- Doval, E., Viladrich, M. C. y Riba, M. D. (1998). Estudio de las diferencias individuales en el riesgo asumido en decisiones consecutivas. *Revista Mexicana de Psicología*, 15 (2), 83-93
- Doval, E., Viladrich, M. C., Ribes, E. y Villanueva, A. (1993). *Multitrait-multimethod analysis of risk-taking patterns*. Comunicación presentada en el 1993 European Meeting of the Psychometric Society. Universidad Pompeu Fabra, Barcelona.
- Embretson, S. E. y Hershberger, S. L. (1999). Summary and future of psychometric methods in testing. *The new rules of measurement. What every psychologist and educator should know* (Cap. 7. pp 243-254). Mahwah, NJ: Lawrence Erlbaum Associates.

- Fernández, P.G.(1989).Psicometría operante.*Revista Mexicana de Análisis de la Conducta*,15(1),115-1
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, 39, 93-104.
- Hayes, S. C., Barnes-Holmes, D. y Roche, B. (eds.) (2001). Relational frame theory: A post-Skinnerian account of human language and cognition. New York: Kluwer Academic/Plenum Publishers.
- Hernández, J.M., García, O., Rubio, V. J., Santacreu,J.(2004). La persistencia en el estudio conductual de la personalidad. *Psicothema*, 16 (1), 39-44.
- Houts, A. C. (1994). Operational Analysis, Behavior Analysis, and Epistemology in Science and Technology Studies. *Revista Mexicana de Análisis de la Conducta*, 20 (monográfico), 101-143.
- Ibáñez, C. (1992). Medición de aptitudes intelectuales mediante procedimientos interactivos. *Revista Mexicana de Análisis de la Conducta*, 18(1-2),103-140.
- Johnston, J. M y Pennypacker, H. S. (1980). Traditions of behavioral measurement. *Strategies and tactics of human behavioral research* (Cap. 4). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kuhn, T. S. (1977). *La estructura de las revoluciones científicas*. Madrid: Fondo de Cultura Económica. (Trabajo original publicado en 1962)
- Luzoro, J. (1998, noviembre). El conductismo radical: un muerto que goza de muy buena salud. Comunicación presentada en el *IV International Congress on Behaviorism and The Sciences of Behavior*, Sevilla.
- Martínez, R. (2000). *Diseño de ítems y niveles de interacción psicológicos*. Tesis Doctoral no publicada: Universidad de Sevilla.
- Martínez, R. y Moreno, R. (2002). Integración de teoría sustantiva, diseño de pruebas y modelos de análisis en la medición psicológica. En A. Bazán y A. J. Arce (eds.) *Estrategias de evaluación y medición del comportamiento en psicología* (87-119). Ciudad Obregón, Méx.: Instituto Tecnológico de Sonora.
- Meltzoff, J. (2000). *Crítica a la investigación. Psicología y campos afines*. Madrid: Alianza.
- Moreno, R., Martínez, R. y Muñoz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, 16 (3), 490-497.
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Lawrence Erlbaum.
- Muñoz, J. (1998). La medición de lo psicológico. *Psicothema*, 10 (1), 1-21.
- Muñoz, J. (1998b). *Teoría clásica de los tests*. Madrid: Pirámide.
- Ribes, E. (1990). *Psicología General*. México: Trillas.
- Ribes, E. y López, F. (1985). *Teoría de la conducta. Un análisis de campo y paramétrico*. México: Trillas.
- Ribes, E. y Martínez, H. (1990). Interaction of contingencies and rule instructions in the performance of human subjects in conditional discrimination. *The Psychological Record*, 40, 565-586.
- Ribes, E. y Sánchez, S. (1990). El problema de las diferencias individuales: Un análisis conceptual de la personalidad. En E. Ribes (Ed.) *Problemas conceptuales en el análisis del comportamiento*. México: Trillas.
- Ribes, E. y Sánchez, S. (1992). Individual behavior consistencies as interactive styles: their relation to personality. *The Psychological Record*, 42, 369-387.
- Santacreu, J. y García, O. (2000). La utilización de tests comportamentales informatizados en el estudio de la personalidad: La evaluación de la persistencia. *Psicothema*, 12 (1), 93-98

- Schmidt, F. L. y Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1(2), 199-223.
- Sidman, M. (1973). *Tácticas de investigación científica* (3ª ed.). Barcelona: Fontanella (Trabajo original publicado en 1960)
- Silva, F. (1993). *Evaluación conductual y criterios psicométricos*. Madrid: Pirámide.
- Staats, A. W. (1996). *Behavior and personality. Psychological behaviorism*. New York, NY: Springer Publishing
- Staats, A. W. (2003). A psychological behaviorism theory of personality. En M. J. Lerner y T. Millon (eds.). *Handbook of psychology: Personality and social psychology*, Vol. 5. (pp.135-158). New York: John Wiley & Sons
- Suen, H. K. (1990). *Principles of tests theories*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Toulmin, S. (1977). *La comprensión humana. I. El uso colectivo y la evolución de los conceptos*. Madrid: Alianza.
- Thorndike, E. L., Bregman, E. O., Cobb, M. V., y Woodyard, E. (1926). *The measurement of intelligence*. New York: Teachers College Bureau of Publications.
- Thurstone, L. L. (1928). The absolute zero in the measurement of intelligence. *Psychological Review*, 38, 406-427.
- Viladrich, M. C. (1991). Análisis de datos y psicología interconductual: alternativas al análisis visual. *Revista Mexicana de Análisis de la Conducta*, 17, 141-161.
- Viladrich, M. C. y Doval, E. (1998). ¿Estilos interactivos o la psicometría de sujeto único? *Acta Comportamental*, 6 (Mono), 113-125.
- Woodcock, R. W. (1999). What can Rasch-based scores convey about a person's test performance? En S. E. Embretson y S. L Hershberger (eds.) *The new rules of measurement. What every psychologist and educator should know* (Cap. 5. pp 105-127). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wright, B. D. (1999). Fundamental measurement for psychology. En S. E. Embretson y S. L Hershberger (eds.) *The new rules of measurement. What every psychologist and educator should know* (Cap. 4. pp 65-104). Mahwah, NJ: Lawrence Erlbaum Associates.
- Yela, M. (1982). El sujeto y la conducta: introducción general. En *El sujeto en la psicología científica actual. Actas del 7º Congreso Nacional de Psicología* (pp. 7-12). Santiago de Compostela.

## RESUMEN

Se analiza una de las divergencias usuales entre las teorías conductuales y las teorías psicométricas: la prioridad dada por las teorías conductuales a la comparación de diferentes estímulos de un mismo sujeto –intra- y la que suelen otorgar las teorías psicométricas a la comparación de unos mismos estímulos o preguntas en diferentes sujetos –entre-. A pesar de tal diferencia, llama la atención que ambos análisis parten de una única y misma matriz de datos, de respuestas de los sujetos por preguntas de una prueba, definidos ambos elementos mutua y conjuntamente. De ahí que las orientaciones conductuales y psicométricas puedan considerarse complementarias más que incompatibles y radicalmente diferentes, algo que desarrolla y discute esta ponencia.

Palabras clave: comparaciones intrasujetos e intersujetos, teorías psicométricas, teorías conductuales, personalidad

## **ABSTRACT**

This paper analyses one of the typical divergences between the behavioral and psychometrics theories of personality: the priority given by behavioural theories to the comparison of different stimuli in the same subject (within-subject approach) and that is usually given by psychometric theories to the comparison of the same stimuli or questions in different subjects (between-subject approach). Despite these differences, it is notable that both perspectives begin from a single shared data matrix, the responses given by a subject in a test, which mutually and conjointly defines both elements. This presentation argues that the behavioral and psychometric orientations may be considered complementary, rather than incompatible or radically different.

**Key words:** between and within-subject comparison, psychometric theories, behavioral theories, personality.