



PARTE SEGUNDA.  
ESTRUCTURA DEL CAMPO  
PSICOMETRICO.

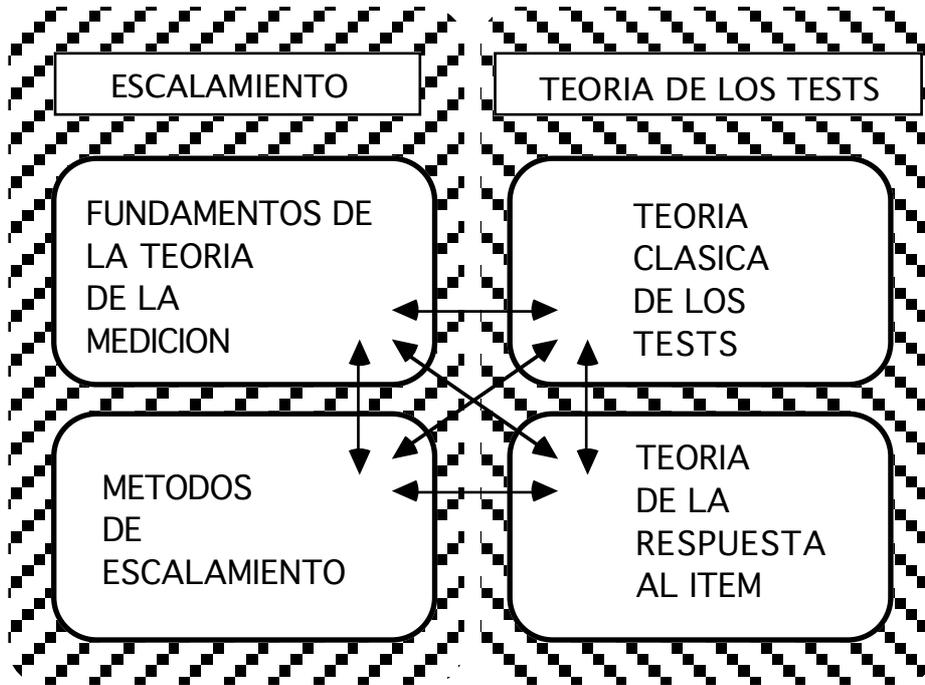
#### 4. CAMPOS DE LA PSICOMETRIA.

##### 1. Los cuatro grandes campos de la psicometría.

Cuando afirmábamos que la teoría psicométrica, como núcleo conceptual de la psicometría, se refiere a la teoría de la medición psicológica, estábamos utilizando la expresión 'teoría de la medición psicológica' con un sentido muy amplio. Si utilizamos un significado más restringido de la expresión 'teoría de la medición' entonces podemos distinguir en la teoría psicométrica tres grandes campos: la teoría de la medición, el escalamiento psicológico y la teoría de los tests (Santisteban, 1.982). A su vez, la teoría de los tests puede descomponerse en dos grandes cuerpos teóricos, el de la teoría clásica de los tests (TCT) y el de la teoría de la respuesta al ítem (TRI). De este modo pueden distinguirse, como se representa en la *figura 3*, cuatro grandes campos psicométricos: el de la fundamentación de la *teoría de la medición*, el del *escalamiento*, el de la *teoría clásica de los tests* y el de la *teoría de la respuesta al ítem*.

La distinción entre estos campos quizás depende más del estado de desarrollo de la psicometría, y de las tradiciones de investigación sobre las que se elaboran, que de una división fuerte del objeto de la psicometría. Los cuatro se refieren a la medición psicológica, los cuatro requieren una fundamentación matemática, los cuatro forman parte de la teoría psicométrica y los cuatro contribuyen de modo más o menos directo a la aplicación psicométrica.

Figura 3. Cuatro Grandes Campos de la psicometría.



Es posible ordenar los principales núcleos teóricos de la psicometría de otros modos. Por ejemplo, se puede pensar en los dos primeros campos como uno solo, referido a medición y escalamiento; y en los dos últimos como uno solo, referido a teoría de tests; definiendo de este modo solo dos grandes troncos psicométricos: el de la teoría de la medición y el escalamiento, y el de la teoría de los tests (Yela, 1.968). También es posible, en fin, destacar más algunas grandes partes de estos troncos principales desglosando la psicometría en otras combinaciones. Por ejemplo, se pueden distinguir la teoría de la medición, el escalamiento psicofísico, el escalamiento psicológico y la teoría de los tests. O bien, pueden desglosarse la teoría de la medición, el escalamiento, la teoría clásica de los tests, la teoría del espacio muestral de items, y las teorías del rasgo latente, además de otras combinaciones más o menos justificables. En realidad se trata tan solo de

diversos modos de agrupar o clasificar los mismos contenidos conceptuales, sin que estas distinciones aporten o discutan nada particularmente esencial.

Aquí optaremos por distinguir en la psicometría los cuatro grandes troncos conceptuales mencionados en primer lugar: la teoría de la medición, los métodos de escalamiento, la teoría clásica de los tests, y la teoría de la respuesta al ítem, dejando otras distinciones para las subdivisiones de estos campos.

## **2. Teoría de la medición.**

Distinguiremos primero, pues, la *teoría de la medición*, que es un campo que ha tomado cuerpo por sí mismo desde un punto de vista matemático, y cuyos contenidos se han desarrollado referidos tanto a las ciencias físicas como a las ciencias sociales.

*Concepto. La teoría de la medición se ocupa de la discusión acerca de las condiciones necesarias y suficientes de la medición y de la relación que deben mantener entre sí los numerales fruto de la medición y las propiedades de objetos, sujetos o eventos a las que se refieren.*

La aproximación más extendida a la teoría de la medición, que defienden autores como Suppes y Zinnes (1.963), Krantz et al. (1.971) y Roberts (1.979), considera la *medición como la construcción de un modelo*, lo que implica someter los modelos de medición al tipo de consideraciones comunes acerca de los modelos matemáticos referidos al mundo empírico. Puede considerarse que el campo de la teoría de la medición incluye los llamados "problemas de la medición" (Coombs et al., 1.981) y en consecuencia la discusión, de nuevo de actualidad, acerca de las escalas de medida, en el sentido en que Stevens (1.951) usara esta expresión. (Stevens propuso una bien conocida teoría de las escalas de medida según la cual estas se clasifican en escala nominal, ordinal, de intervalo y de razón; quizás el modo de pensar acerca del

nivel de medición de una variable que resulta todavía más popular en Psicología.)

En psicología, las relaciones entre teoría psicológica, teoría de la medición, instrumentos de medición, y aquello que se mide, no siempre resultan sencillas y claras. La teoría de la medición tiene la vista puesta justamente en la conexión entre la medición y la realidad medida, el difícil hilo que enlaza la teoría con los hechos. Esto significa obtener y elaborar la fundamentación matemática de la medición y el desarrollo de los axiomas bajo los que es posible la medición.

Se ha propuesto una aproximación axiomática a la teoría de la medición que distingue una teoría de la medición fundamental y una teoría de la medición derivada (Santisteban, 1.982). Roberts (1.979), por ejemplo, distingue dos tipos de medición, la medición fundamental y la medición derivada; dos problemas en la medición fundamental, el de la representación y el de la unicidad; y dos modos de ver los axiomas en el teorema de la representación, el prescriptivo o normativo, y el descriptivo. Coombs y otros (1.981), por ejemplo, han distinguido cuatro problemas básicos en la teoría de la medición: el problema de la representación, el problema de la unicidad, el problema de la significatividad y el problema de la construcción de escalas. Narens y Luce (1.986) ofrecen una clasificación en cinco puntos de los principales temas de la teoría de la medición muy semejante, y aportan información sobre el recorrido histórico de la teoría de la medición.

Jañez (1.989) presenta una síntesis histórica sobre las teorías de la medición, con tres grandes etapas: una primera época de idealismo platónico, una segunda época en la que aparece la teoría de las escalas de medición, básicamente representada por el trabajo de Stanley Smith Stevens, y una tercera época de teoría representacional, que podría considerarse inaugurada por el trabajo de Suppes y Zinnes de 1.963, y bien representada por los de Krantz, Luce, Suppes y Tversky de 1.971 y de Roberts de 1.979. Para Jañez (1.989) "la moderna teoría de la medición se

levanta sobre dos conceptos básicos: el de sistema relacional y el de homomorfismo entre sistemas relacionales. Sobre esa base aborda tres problemas: el de representación, el de unicidad, y el de la significación."

### 3. Escalamiento.

*Concepto de escalamiento.* Estrechamente ligado al campo de la teoría de la medición se encuentra, en segundo lugar, el campo del *escalamiento*, o si se prefiere, de los métodos de escalamiento. Es difícil condensar en pocas palabras cual es el campo del escalamiento, y buena prueba de ello es que los autores rehuyen ofrecer definiciones sobre qué es el escalamiento y frecuentemente prefieren explicar con ejemplos o métodos concretos en qué consiste. Para empezar, es posible que los términos "escalamiento" o "escalación", traducciones demasiado fonéticamente guiadas del inglés "scaling", no signifiquen nada más que "medición".

Realmente puede decirse en general que *todos los procedimientos de escalamiento permiten asignar números a objetos, eventos o sujetos en función de alguna o algunas propiedades de los mismos*. Esto es tanto como decir que los procedimientos de escalamiento permiten medir, en un sentido amplio, esos objetos, eventos o sujetos en función de sus propiedades. Frecuentemente el énfasis de los métodos de escalamiento está en "*crear o construir*" la escala de medición, la correspondencia entre los objetos, eventos o sujetos en función de sus propiedades, y un conjunto estructurado de números.

*Concepto de escala.* Un modo de comenzar a comprender el ámbito del escalamiento en psicología es analizar el significado del *concepto de escala* en algunos ejemplos bien conocidos de diversas ciencias. Un conjunto de minerales pueden ser ordenados en función de la comparación dos a dos "quién raya a quién" en una escala de dureza. Por ejemplo, la escala de Mohs consiste precisamente en diez minerales (talco, yeso, calcita,

fluorita, apatito, ortosa, cuarzo, topacio, corindón y diamante) ordenados y numerados de 1 a 10 por este procedimiento, que, a su vez, sirven para medir la dureza de otros minerales. Un conjunto de sonidos pueden ser escalados en función de su volumen. Por ejemplo, la escala de Beranek clasifica los sonidos en seis categorías (muy quedo, quedo, moderadamente quedo, ruidoso, muy ruidoso e intolerablemente ruidoso). Un conjunto de tarjetas de color pueden escalarse en función de su intensidad. Por ejemplo, la escala de Linke consta de ocho tarjetas ordenadas con grados de azul numeradas del 2 al 16 que se emplean para medir la intensidad del azul del cielo, reservándose los números impares para el caso en que el observador encuentre un azul situado entre dos tarjetas. Por ejemplo, la escala de Forel es una escala de amarillos, verdes y azules para registrar el color del mar. Los terremotos pueden ser escalados en función de una serie de efectos observables. Por ejemplo, la escala de Richter, de 1 a 9 puntos; por ejemplo, la escala de Mercalli, de 1 a 12 puntos. Los vientos pueden también ser ordenados en escalas; por ejemplo, la escala de vientos de Beaufort, la escala de vientos del Norte de las Montañas Rocosas, etc. Un ejemplo particularmente importante de un conjunto de escalas bien articuladas entre sí y con una gran aceptación internacional lo constituye el sistema métrico decimal. Muy conocido es el concepto de escala musical. También son muy conocidas las numerosas escalas de temperaturas; por ejemplo, Fahrenheit, Rankine, Dalton, Glaque, Kelvin, Celsius, etc.

Cada uno de esos ejemplos tiene sus particularidades, pero en todos se ha establecido *una relación entre propiedades de objetos o eventos y un conjunto numérico, según ciertas reglas que se hacen explícitas en alguna operación de medición definida*. Del mismo modo, un conjunto de items pueden ser ordenados en una escala en función del grado en que representan una propiedad. Así puede hablarse de la escala de inteligencia de Binet-Simon, de la escala F de Adorno, de la escala de inteligencia infantil de Cattell, de la escala B de inteligencia del test 16PF.

Obsérvese a través de todos los ejemplos que hay *dos procesos* distintos. Primero, *cómo establecer la escala*, es decir, como asociar un conjunto de objetos, eventos o sujetos en función de sus propiedades a un conjunto de números. Es el proceso de generar o crear la escala de medida. Segundo, *cómo usar esa escala para medir o 'escalar' otros objetos*, eventos o sujetos en función de la misma propiedad en la que se construyó y opera la escala. Los "métodos de escalamiento" y el verbo "*escalar*" en psicología se refieren más bien a la primera operación, mientras que la palabra "*medir*" suele reservarse para la segunda. Obsérvese también que todas las escalas, como todos los lenguajes, tienen un componente de *arbitrariedad* y otro de *convención* entre sus usuarios.

Sin embargo, este modo de presentar las cosas es solo una primera aproximación. Si el campo del escalamiento en psicometría hubiera de contener todo lo relativo al proceso de construcción o elaboración de la escala de medida, entonces una parte muy importante de la teoría clásica de tests y de la teoría de la respuesta al ítem deberían considerarse también escalamiento. Tradicionalmente esto no es así, y aunque las fronteras entre estas aproximaciones es en muchos casos difusa, el campo del escalamiento y las obras características del mismo, pueden distinguirse de los de la teoría clásica de tests y de los de la teoría de la respuesta al ítem, aun tratándose de una distinción convencional.

*Escalamiento psicofísico y escalamiento psicológico.* El campo del escalamiento psicológico tiene su origen en la psicofísica. *La psicofísica se preocupa, esencialmente, por establecer la relación entre una magnitud física* (por ejemplo, peso) *y la sensación que el sujeto tiene de esa magnitud* (por ejemplo, la sensación de peso). Este es un modo de abordar un aspecto de un problema central en psicología, el de la relación entre el mundo físico y el mundo mental. Para ello la psicofísica necesitaba disponer de mediciones tanto de la magnitud física, aspecto usualmente resuelto por otras ciencias, como de la sensación, y de este modo tuvo que ocuparse de "escalar" las sensaciones de los

sujetos. *El escalamiento psicofísico se refiere precisamente al escalamiento efectuado sobre sensaciones que tienen un referente en una magnitud física.* Por contraposición, *el escalamiento psicológico se refiere a aquel escalamiento referido a dimensiones psicológicas que no tienen un referente físico inmediato.*

El escalamiento psicológico surgió a partir del escalamiento psicofísico, básicamente gracias a Thurstone, y se desarrolló después con las aportaciones de Likert, de Guttman, de Coombs y de otros muchos. Muy frecuentemente la distinción entre escalamiento psicofísico y psicológico, y la distinción entre los modelos de escalamiento de estos autores, constituyen un modo de clasificación de los métodos de escalamiento.

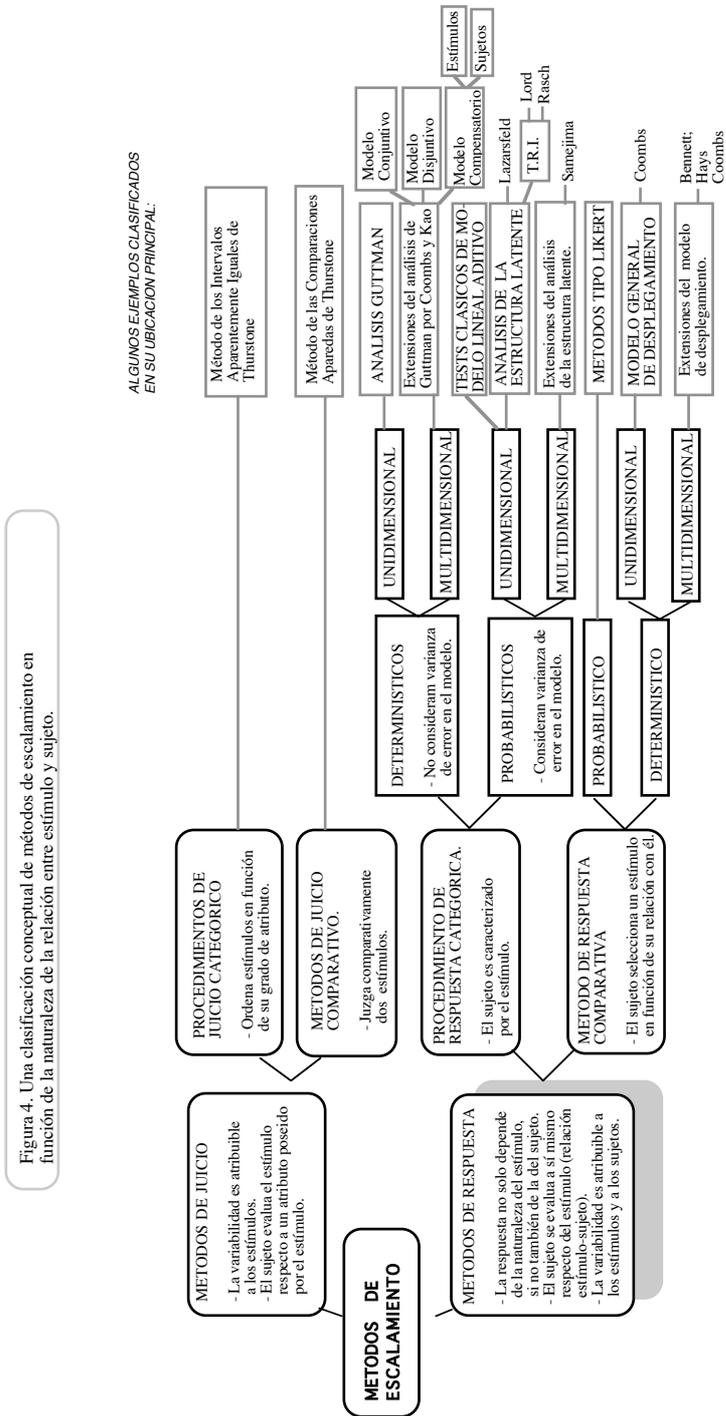
A pesar de que existen evidentemente otros muchos trabajos posteriores importantes sobre escalamiento psicológico, la obra tradicional más emblemática de este campo, que sigue constituyendo una referencia obligada, es quizás la de Torgerson de 1.958, que permite una visión estructurada sobre la que articular los desarrollos posteriores.

*Clasificación en función del objeto a escalar.* Los métodos de escalamiento pueden clasificarse *en función de qué se está escalando.* Con este criterio se distingue entre el escalamiento de personas, de estímulos y la a veces llamada 'aproximación de respuesta' que implica escalar simultáneamente personas y estímulos. Es decir, se pueden escalar personas, estímulos o ambos a la vez, simultáneamente. La mayoría de los métodos de escalamiento psicológico presentan dos fases con dos cometidos: primero se escalan un conjunto de estímulos - frecuentemente los ítems de un cuestionario-, y una vez escalados estos se usan para medir o "escalar" sujetos. Escalar o "calibrar" los ítems significa decir que posición ocupa cada ítem en una escala de medida de algún atributo. Escalar o "medir" los sujetos significa decir que posición ocupa cada persona en una escala de medida de algún atributo.

*Clasificación en métodos de juicio y métodos de respuesta.* Los métodos de escalamiento pueden ser clasificados en función de la naturaleza de la relación entre el estímulo y el sujeto que da la respuesta, tal como se muestra en la *figura 4* .

Tal como puede verse, los métodos de escalamiento se dividen de este modo, siguiendo la nomenclatura de Torgerson (1.958), en métodos de juicio y métodos de respuesta, orientados los primeros al escalamiento del estímulo y los segundos, básicamente, al escalamiento de la persona.

En la figura se han incluido también algunas subdivisiones más precisas de esta clasificación y la ubicación de los modelos de escalamiento de algunos autores en su punto más representativo, si bien estas ubicaciones pueden depender de que trabajos de un autor se estén considerando principalmente.



*Métodos de juicio.* Los *métodos de juicio* se caracterizan porque el estímulo a escalar es evaluado respecto de alguna característica del estímulo mismo, sin referencia, teóricamente, al sujeto que efectúa el juicio. Se trata de escalar el estímulo mismo, no de escalar al sujeto

ni de poner en relación estímulo y sujeto. Hay dos grandes tipos de métodos de escalamiento de juicio: el método de juicio comparativo y el procedimiento de juicio categórico.

En el *método de juicio comparativo* al sujeto se le presentan los estímulos en pares y se le pide que diga cuál de los miembros del par posee más (o menos) de una característica común a ambos estímulos. Por ejemplo, se le pueden dar dos objetos a un sujeto y solicitarle que diga cual es más pesado, o más agradable, o más útil. Por ejemplo, se le pueden dar dos enunciados a un sujeto y solicitarle que diga cual expresa una actitud más belicista, o más conservadora, o más ecologista. El *método de las comparaciones apareadas de Thurstone*, basado en su ley del juicio comparativo, es la expresión más conocida de este tipo de métodos de escalamiento. El *procedimiento de juicio categórico* implica que el sujeto dispone al mismo tiempo de un conjunto de más de dos estímulos, y se le solicita que los ponga en orden en función de alguna propiedad o característica que poseen esos estímulos. Obsérvese que desde un punto de vista lógico puede considerarse una extensión realizada en un mismo tiempo del método de juicio comparativo. Por ejemplo, el sujeto dispone de una lista de cinco ciudades y se le solicita que las ordene en función de su tamaño, o de su grado de polución, o en función de su calidad turística. Por ejemplo, el sujeto dispone de una lista de diez conductas y se le pide que las clasifique en función del grado en que representan agresividad, o inteligencia, o habilidad social. Quizás el *método de los intervalos aparentemente iguales de Thurstone* sea el ejemplo más clásico de un método de escalamiento de juicio categórico.

*Métodos de respuesta.* Los *métodos de respuesta* implican que al sujeto se le solicita que exprese su relación con el estímulo que se le presenta. Ello implica que la variabilidad de las respuestas que se obtenga a través de una muestra de sujetos habrá de atribuirse, en principio, tanto a los estímulos como a los sujetos. Los métodos de respuesta se dividen en procedimientos de respuesta categórica y métodos de respuesta comparativa.

En el *método de respuesta categórica* el sujeto dice ante un estímulo determinado si éste representa su posición en la dimensión objeto de estudio. Un ejemplo de este modo de operar se produce cuando a un sujeto se le presenta un ítem y se le pide que conteste si éste representa o no su modo de pensar o de proceder.

En el *método de respuesta comparativa* el sujeto dispone simultáneamente de varios estímulos y debe expresar de algún modo cual es su relación con los mismos, por ejemplo, seleccionando aquel que representa su posición en la magnitud objeto de medición.

Obsérvese que esta distinción entre dos tipos de métodos de respuesta es paralela a la distinción entre dos tipos de métodos de juicio.

*Modelos deterministas y probabilísticos.* Los modelos de escalamiento que utilizan métodos de respuesta pueden ser de naturaleza *determinista*, cuando el modelo no está diseñado para soportar desajustes con la realidad bajo la forma de error, o de naturaleza *probabilista*, cuando admiten varianza de error en el modelo.

*Modelos unidimensionales y multidimensionales.* Los modelos de escalamiento que utilizan métodos de respuesta pueden operar sobre una única dimensión, -un único atributo, rasgo, aspecto o característica- en cuyo caso se denominan *unidimensionales*, o sobre dos o más dimensiones a la vez, en cuyo caso se denominan *multidimensionales*.

La ubicación de algunos ejemplos de métodos de escalamiento bien conocidos puede contribuir a aclarar estas distinciones. Por ejemplo, el *escalograma de Guttman*, que solicita a los sujetos que digan si o no ante un conjunto de estímulos, que trabaja típicamente sobre una única dimensión, y que no admite error en el modelo, es un método de respuesta, que utiliza procedimiento de respuesta categórica, determinístico y unidimensional. Por ejemplo, el método de *desplegamiento unidimensional de Coombs* es un método de respuesta comparativa, y en su

formulación original de naturaleza determinista y unidimensional.

*Escalamiento unidimensional y escalamiento multidimensional.* En las clasificaciones sobre métodos de escalamiento es inevitable hacer mención a la distinción entre escalamiento unidimensional (UDS) y escalamiento multidimensional (MDS). Esta es una distinción muy importante por cuanto que diferencia dos inmensas áreas de trabajo e investigación.

El escalamiento unidimensional se ocupa de representar el conjunto de objetos, (estímulos o sujetos,) bajo consideración, en una única dimensión. Para decirlo gráficamente, los escala sobre una recta.

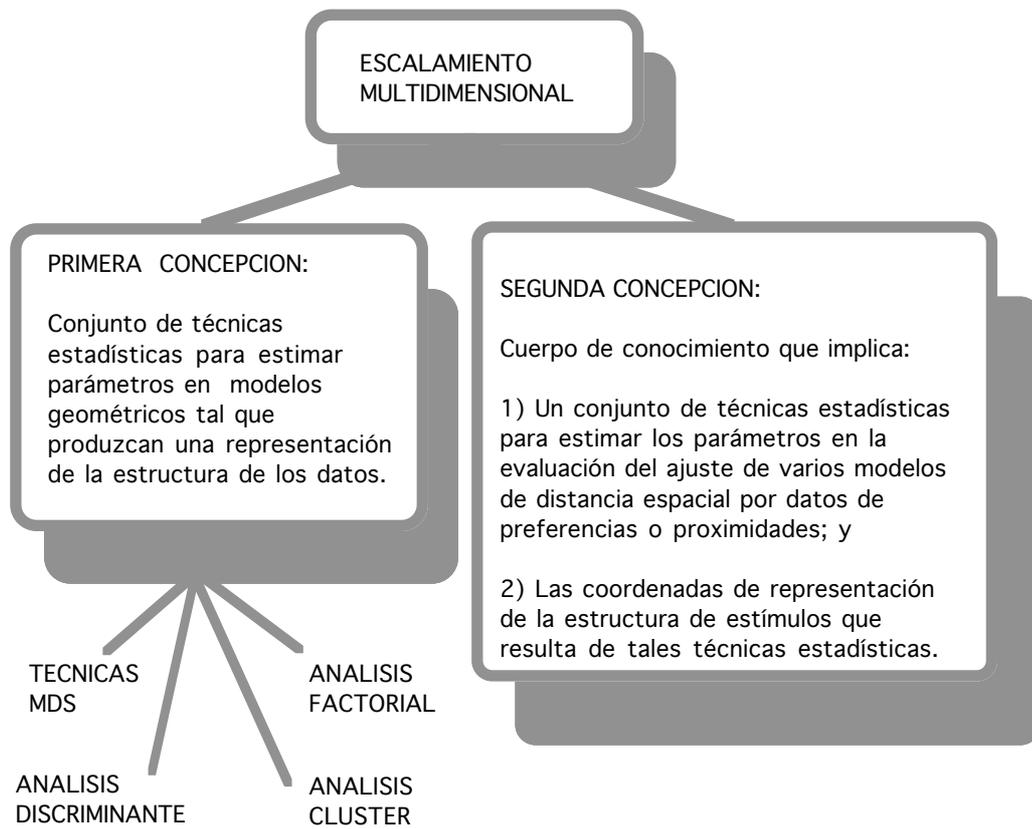
El escalamiento multidimensional se ocupa de representar al conjunto de objetos bajo consideración en dos o más dimensiones simultáneamente. Para decirlo gráficamente, los ubica en un plano (bidimensional) o en un espacio (tridimensional) o en algún espacio de más de tres dimensiones.

Aunque la idea de 'unidimensionalidad' es muy intuitiva el análisis de la dimensionalidad es una de las cuestiones más complejas e interesantes de los modelos de escalamiento.

*Concepciones generales del escalamiento multidimensional.* Se han presentado en la literatura dos concepciones generales del escalamiento multidimensional, como se muestra en la *figura 5*.

Cuando el escalamiento multidimensional se mira desde el punto de vista de la estadística, se trata de un *conjunto de técnicas de análisis que permiten representar la estructura de un conjunto de datos*. En ese sentido, las técnicas de escalamiento multidimensional son técnicas de análisis multivariado muy próximas a otras como el análisis factorial, el análisis de componentes principales, el análisis factorial de correspondencias o el análisis cluster.

Figura 5. Dos concepciones generales del escalamiento multidimensional.



Quando se mira desde el punto de vista de la psicometría, desde el punto de vista de la medición, las técnicas del escalamiento multidimensional se clasifican en la extensión de las técnicas de escalamiento unidimensionales aportando la posibilidad de escalar simultáneamente en más de una dimensión.

Por supuesto, no se trata tanto de dos concepciones contrarias cuanto que complementarias. En general puede decirse que la primera concepción es considerada demasiado amplia por la mayoría de los autores, de modo que se suele restringir la definición del MDS ("MultiDimensional Scaling" o Escalamiento Multidimensional) a la segunda concepción que es la defendida por Carroll y

Arabie (1.980). Lo usual, en este sentido más restringido, es *no* considerar técnicas multivariadas como el análisis discriminante, el análisis cluster o el análisis factorial como parte del escalamiento multidimensional, aunque algunas de ellas, como el análisis factorial, pueden utilizarse también con propósitos de escalamiento.

#### **4. Teoría de los tests.**

La teoría de los tests es el bloque de contenido tradicionalmente considerado como principal dentro de la Psicometría, y probablemente el que mayor atención académica ha recibido. Para algunos autores, más por omisión o selección que por afirmación explícita, diríase que la teoría de los tests, y en particular la teoría clásica de los tests, es la Psicometría. Evidentemente la teoría de los tests no es toda la psicometría, aunque histórica y conceptualmente ocupa, desde luego, un lugar principal en la misma.

Figura 6. Clasificación de la Teoría de los Tests en dos grandes ramas y algunos de los principales trabajos en cada línea.



*Clasificación de la teoría de los tests.* La teoría de los tests puede subdividirse en la actualidad, tal y como se muestra en la *figura 6*, en dos grandes ramas: la teoría clásica de los tests y la teoría de la respuesta al ítem.

La *teoría clásica de los tests* (TCT) se inició con los trabajos de Spearman a principios de siglo. Las obras principales de esta corriente son las de Guilford de 1.936 y 1.954, y sobre todo la de Harold Gulliksen de 1.950 y la de Lord y Novick de 1.968. En particular la obra de Lord y

Novick ha sido considerada algunas veces como la obra más importante de la psicometría.

Por su parte la *teoría de la respuesta al ítem* (TRI) tiene en las obras de Lazarsfeld (1.950), Rasch (1.968), Birnbaum (1.968), Lord (1.980), Wainer y Messick (1.983) y De Gruijter y Van der Kamp (1.984) algunos de sus principales puntos de referencia. El conjunto de contenidos de la teoría de la respuesta al ítem, a su vez, admite diversas clasificaciones.

Hambleton y Van der Linden (1.982) han expresado con toda claridad la distinción de dos grandes ramas dentro de la teoría de los tests (Teórica Clásica *versus* Teoría de la Respuesta al Ítem) que antes señalábamos: Teoría de la Respuesta al Ítem, o teoría del rasgo latente, como se la ha llamado hasta hace poco. Actualmente la Teoría de la Respuesta al Ítem (TRI) esta teniendo un mayor impacto en el campo de los tests. Los modelos derivados de la TRI se están usando para desarrollar tests, para igualar puntuaciones de tests no paralelos, para investigar sesgos en los items, para obtener puntuaciones, así como para manejar otros acuciantes problemas de medición (ver, p.e., Hambleton, 1.983; Lord, 1.980).» (Hambleton y Van der Linden, 1.982; pag. 373.)

*Dificultades de clasificación en teoría de los tests.* Por supuesto es posible considerar otras clasificaciones de la teoría de los tests. Ni siquiera las denominaciones de 'teoría clásica' y 'teoría de la respuesta al ítem' son unívocas. A veces diferentes cosas se engloban bajo esas denominaciones, en función de que se tomen las etiquetas con más o menos generalidad. A veces a esas mismas teorías se las denomina de modo diferente. Sin embargo, estas etiquetas parecen haber hecho mayor fortuna a través de revisiones y presentaciones sumarias de la teoría de los tests, y la tendencia actual parece ser que se incrementará la generalidad de su uso. Tal y como aquí se usan estas denominaciones generales, pretenden abarcar la teoría de los tests, de modo que algunas teorías, variantes,

y desarrollos, tendrán que ser incluidos como subcampos de estas dos grandes ramas.

Este modo de clasificación tiene una serie de dificultades que deben considerarse. Por ejemplo, muchas veces se usa la denominación teoría clásica de los tests excluyendo la *teoría de la generalizabilidad* y el *modelo del espacio muestral de items*. Sin embargo, sólo con una pretensión ordenadora, aquí consideraremos estos cuerpos teóricos dentro de la aproximación clásica. Más dificultades se encuentran en la clasificación de algunos desarrollos particulares, como los *tests referidos al criterio*, que de suyo entran en contraposición con los tests de norma grupal, lo que implica otros criterios de clasificación distintos, y que han sido estudiados bajo modelos 'clásicos' y también de la teoría de la respuesta al ítem.

En realidad, la teoría clásica de tests, entendida como la teoría que postuló Spearman, es mucho más reducida de lo que aquí se pretende. La cuestión está en que esa teoría fue desarrollada y extendida una y otra vez, y aunque ya no es un blanco favorito de la investigación psicométrica, todavía puede considerarse que su cuerpo teórico y sobre todo sus variantes siguen creciendo. Lo mismo puede decirse, dentro de la TRI, del análisis de la estructura latente de Lazarsfeld o del modelo que Rasch postuló.

*Principales temas de la Teoría Clásica de los Tests.* La teoría clásica presenta cuatro grandes temas: La fiabilidad, la validez, la tipificación y el análisis de items. Algunos añadirían como un quinto tema el proceso de construcción de los tests, pero en realidad se trata de la integración ordenada de los contenidos de los cuatro temas señalados.

**Fiabilidad.** La *fiabilidad* es el tema más clásico y más conocido de la Psicometría.

*Concepto de fiabilidad.* Se dice que un test es fiable cuando "mide bien aquello que esta midiendo." La fiabilidad se refiere a la constancia de la medida, al grado en que un

instrumento de medida psicológica no deformará el resultado de una medición debido a cambios, fluctuaciones o variaciones del instrumento mismo.

*Componentes de la fiabilidad.* La fiabilidad tiene dos grandes componentes: la consistencia interna y la estabilidad temporal.

La *consistencia interna* se refiere al grado en que los distintos items, partes o piezas de un test miden la misma cosa. Significa la constancia de los items para operar sobre un mismo constructo psicológico de un modo análogo.

La *estabilidad temporal* se refiere al grado en que un instrumento de medida arrojará el mismo resultado en diversas mediciones concretas midiendo un objeto o sujeto que ha permanecido invariable. Si un test no tiene "fiabilidad" entonces no se puede "confiar" en la medición que este ofrece.

**Validez.** La *validez* es la cualidad más importante de un instrumento de medida.

*Concepto de validez.* Se dice que un test es válido si "mide lo que dice medir". Esta cuestión de medir lo que se pretende medir y no otra cosa, no es en absoluto trivial en gran parte de la Psicología.

*Componentes de la validez.* La validez puede decirse que tiene tres grandes componentes: la validez de contenido, la validez de criterio o criterial y la validez de constructo.

La *validez de contenido* se refiere al grado en que el test presenta una muestra adecuada de los contenidos a los que se refiere, sin omisiones y sin desequilibrios de contenido.

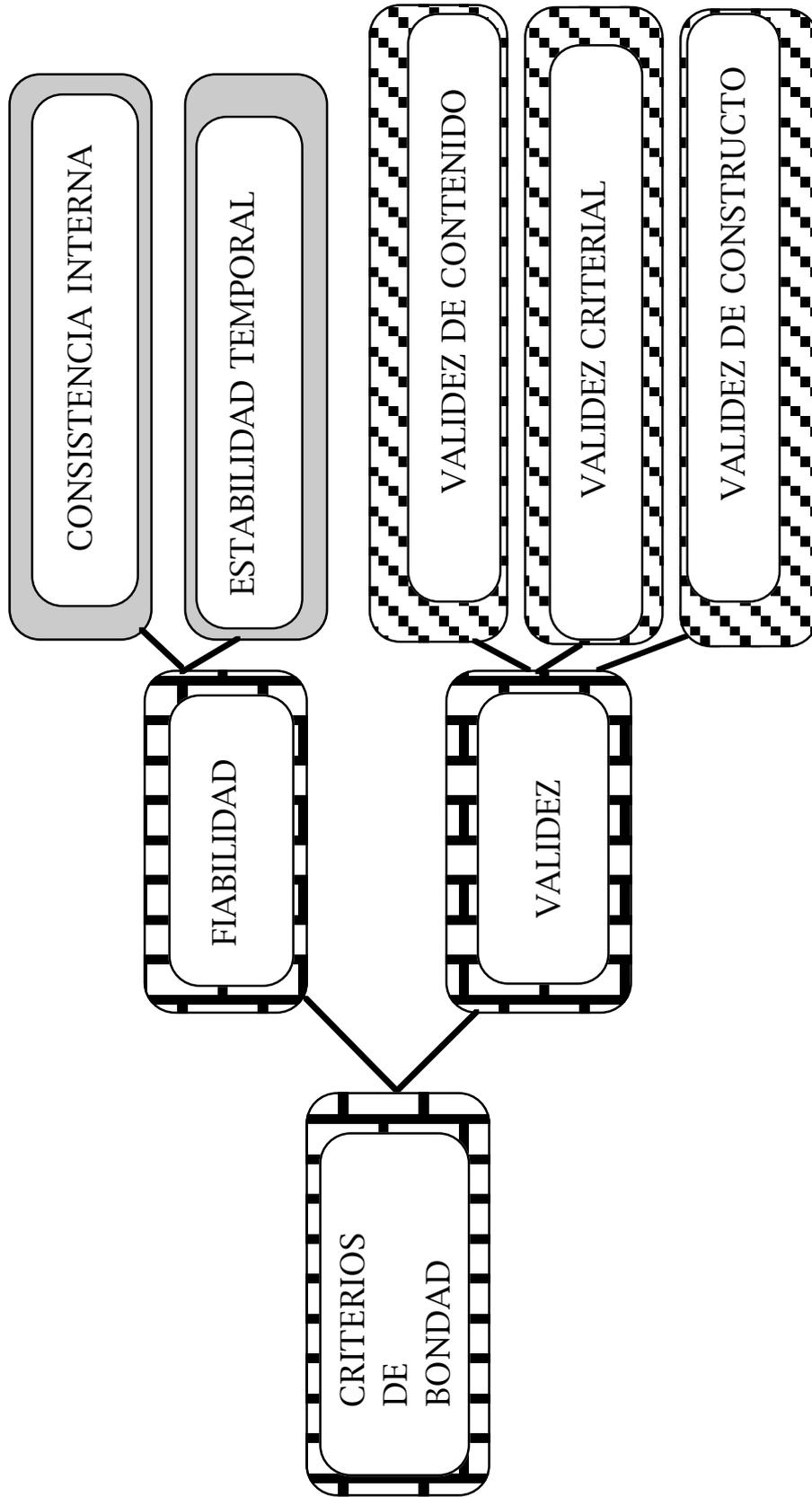
La *validez criterial* se refiere al grado en que el test correlaciona con variables ajenas al test, -criterios,- con los que se espera por hipótesis que debe correlacionar de

determinado modo. Un *criterio* es una variable distinta del test que se toma como referencia, que se sabe que es un indicador de aquello que el test pretende medir o que se sabe que debe presentar una relación determinada con lo que el test pretende medir. Se denomina *coeficiente de validez* a la correlación del test con un criterio externo.

La *validez de constructo* es un concepto más complejo y todavía más difícil de sintetizar en una frase. En general, se refiere al grado en que el instrumento de medida cumple con las hipótesis que cabría esperar para un instrumento de medida diseñado para medir precisamente aquello que se deseaba medir. La validez de constructo puede considerarse que es un concepto general que abarca los otros tipos de validez.

*Criterios de bondad de un test.* Dado que tanto la fiabilidad como la validez son cualidades o criterios de calidad que un test debe presentar para considerarse adecuado a veces se denominan con la expresión "criterios de bondad". La *figura 7* presenta una síntesis de los principales criterios de bondad que debe cumplir un test.

FIGURA 7. PRINCIPALES CRITERIOS DE BONDAD DE UN TEST PSICOMETRICO.



*Tipificación.* La *tipificación* se refiere al conocimiento disponible acerca de cómo responden de hecho los sujetos al test y cual es la distribución de puntuaciones que la población de referencia presenta. La tipificación se refiere a la elaboración de las normas muestrales del test que permiten la interpretación de las puntuaciones de los sujetos. La teoría clásica no permite hacer mediciones absolutas, abstractas. Cada medición efectuada con un test desarrollado bajo teoría clásica no es más que una comparación de la puntuación del sujeto en el test con las puntuaciones que han obtenido los miembros de una muestra adecuada en ese test. De una muestra concreta y determinada de sujetos. Básicamente, lo que permite la tipificación es afirmar qué posición ocupa un sujeto en una muestra. Por ejemplo, que tanto por cien de sujetos son peores que él en el test. Hay muchas formas de tipificar las puntuaciones pero el procedimiento más popular consiste en calcular los *percentiles* con una muestra adecuada.

*Análisis de items.* El *análisis de items* se refiere al proceso de selección a que se somete a los items aspirantes a formar parte de un test. Como es sabido, un ítem es cada uno de los enunciados o cuestiones de un test que demandan una respuesta del sujeto. Antes de que un conjunto de items puedan considerarse un test han de ser sometidos a una serie de análisis que garanticen que *cada ítem* es por sí lo suficientemente válido y fiable.

Las técnicas y procesos por los que se efectúa esa comprobación es lo que se conoce como análisis de items. (Además, en el proceso de elaboración del test, no bastará con que los items uno a uno sean válidos y fiables, después, el test como conjunto habrá de demostrar también que es válido y fiable, antes de proceder a su tipificación.) Así, las dos cualidades que básicamente se piden a un ítem son las mismas que básicamente se piden al test: fiabilidad - homogeneidad o correlación con el total del test- y validez - correlación con uno o más criterios externos-.

*Teoría de la Respuesta al Item.* La TRI o Teoría de la Respuesta al Item puede considerarse un intento de superar

esa concepción relativa de la medición que ofrece la Teoría Clásica en la que a fin de cuentas lo que se obtiene después de medir a un sujeto es tan solo la posición del sujeto en un grupo. La Teoría de la Respuesta al Ítem puede considerarse un intento de obtener mediciones que sean independientes de la muestra o población a la que pertenece el sujeto. Básicamente la TRI trata con los modos para obtener este tipo de mediciones, por ello todo el tema de la fiabilidad, tan importante en la Teoría Clásica, está aquí ausente.

*Curva característica del ítem (c.c.i.).* El concepto central de la TRI es el de *función característica del ítem* y el asociado de *curva característica del ítem (c.c.i)* que es la representación gráfica de dicha función. La función y la curva característica son la representación -matemática o gráfica- de la relación que existe entre la probabilidad de que el sujeto dé determinado tipo de respuesta a un ítem y el rasgo o dimensión subyacente que ese ítem pretende medir. Normalmente, *la curva característica de un ítem representa la relación entre una capacidad, que se suele representar con la letra griega  $\theta$ , y la probabilidad de que el sujeto acierte el ítem.* En general, a mayor nivel  $\theta$ , mayor probabilidad de acertar el ítem, según determinada función matemática. Esas funciones pueden ser de diversos tipos, pero usualmente son de tipo logístico. Un modo de diferenciar diversos modelos o variantes de TRI consiste en considerar qué tipo de función se utiliza y cuantos parámetros considera esa función.

La TRI ha presentado condiciones óptimas para algunas *aplicaciones* en las que la Teoría Clásica tenía sus dificultades. Por ejemplo, la TRI ha permitido un gran desarrollo de tests computarizados contruidos "a la medida" del sujeto que se está evaluando. Este es el campo que denominan de los "tailored" tests, test computarizados, tests recortados, etc. Otra aplicación interesante está en los llamados test "de maestría" ("Mastery tests"), test diseñados para evaluar si un sujeto supera (master) o no un determinado punto de capacitación. Los temas de la igualdad de tests y el estudio de los sesgos de los items y

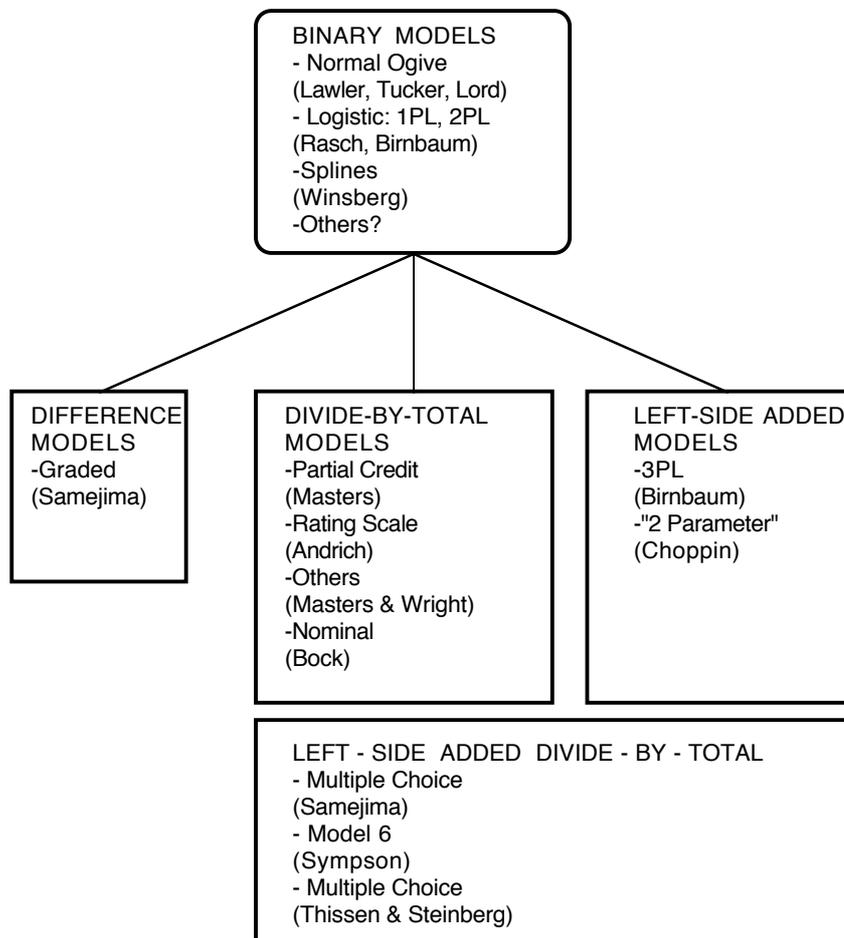
de los tests se cuentan también entre las más importantes aplicaciones.

*Dificultades de clasificación de los modelos de TRI.* Los modelos de Teoría de la Respuesta al Item pueden ser clasificados bajo distintos criterios: supuestos, tipo de curva característica de ítem, consideración del error de medida, número de parámetros que consideran en su función, procedimientos de estimación de parámetros, procedimientos de evaluación del ajuste a los datos, etc.

Existe una cierta confusión en el modo en que los modelos son etiquetados. A veces se etiquetan por el autor que los ha elaborado, pero ello complica las cosas cuando un autor ha elaborado varios modelos con diversas variantes, como es el caso de Samejima, o de F.M. Lord. Otras veces se clasifican no por su forma misma y sus precisas características, sino por analogía con otros modelos principales semejantes, siendo ésta una de las prácticas que llevan a denominaciones más amplias pero también menos precisas.

Además, existen dificultades que no son solamente de denominación. Las relaciones entre modelos, que están en continuo desarrollo, vuelven todavía más imperfecta cualquier clasificación. Por ejemplo, los tests referidos al criterio están en estrecha relación con los '*mastery tests*' (*tests de maestría*), cuyo análisis más idóneo parece corresponder a los modelos de análisis de clase latente. Por ejemplo, el análisis de clase latente puede distinguirse en sentido estricto de la Teoría de la Respuesta al Item, sin embargo, al usar la etiqueta 'Teoría de la Respuesta al Item' con un propósito abarcador, incluiremos el conjunto de modelos referidos al análisis de la estructura latente bajo esta denominación.

Figura 8. Taxonomía de los modelos de Respuesta al Ítem elaborada por Thissen y Steinberg (1.986). (Debido a que se trata de una taxonomía, hemos preferido respetar en la tabla la denominación original en inglés).



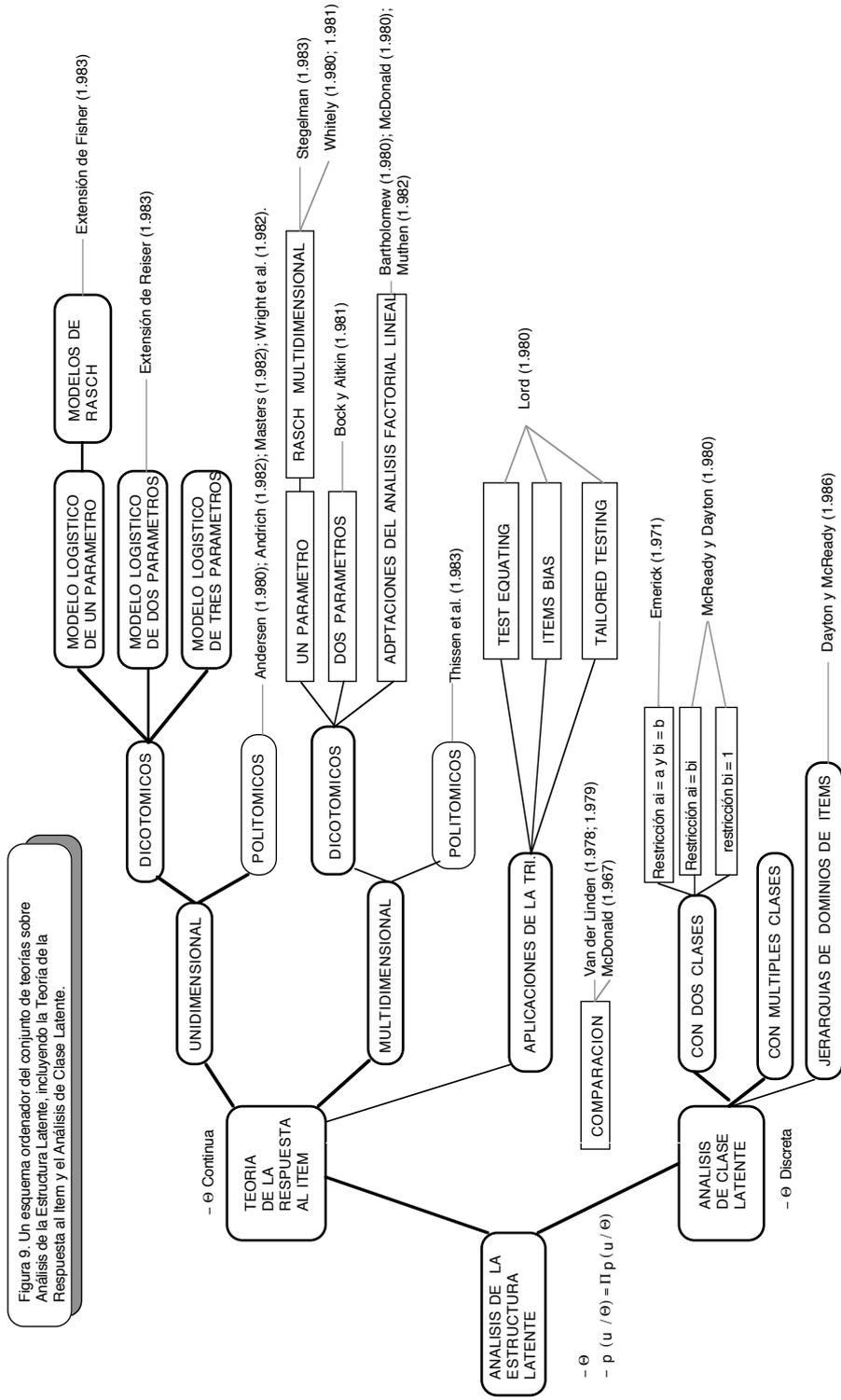
*Taxonomía de modelos de TRI de Thissen y Steinberg (1.986).* Bajo esas condiciones resulta difícil, como Thissen y Steinberg (1.986) han puesto de manifiesto, decir cuando un desarrollo es lo suficientemente diferente para considerarlo un modelo nuevo o distinto que merece otra etiqueta. No obstante, a pesar de estas dificultades, estos autores han hecho un esfuerzo por elaborar una taxonomía que presentamos con las denominaciones originales -para no crear más confusión,- en la *figura 8*. Lo que todavía es pronto para juzgar es si esta taxonomía original de los modelos de la teoría de la respuesta al ítem hará suficiente fortuna para poder considerarla admitida por convención

dentro del campo. A priori, a pesar del esfuerzo formal de estos autores, hay que admitir que sus 'etiquetas' no son precisamente las más populares en la literatura especializada.

*Clasificación de los modelos de análisis de la estructura latente.* La taxonomía de Thissen y Steinberg (1.986) se ha mostrado como un ejemplo de intento de ordenación diferente al esquema ordenador más amplio que proponemos en la *figura 9*. El esquema de la *figura 9* ha sido elaborado teniendo presente la revisión de la literatura de Traub y Lam (1.985), pero sin ceñirse única ni estrictamente al planteamiento de estos autores.

Puede claramente apreciarse en el esquema como el análisis de la estructura latente, que se refiere al conjunto de modelos que implican un vector de rasgo latente  $\theta$  y el supuesto de independencia local, puede desglosarse en dos, la teoría de la respuesta al ítem, que supone  $\theta$  continua, y el análisis de clase latente, que supone  $\theta$  discreta.

La teoría de la respuesta al ítem se divide a su vez en dos ramas, la de los modelos unidimensionales y la de los multidimensionales, en función de que  $\theta$  se considere único o múltiple. La mayoría del trabajo se ha efectuado en modelos TRI unidimensionales para ítems dicotómicamente valorados, categoría que a su vez puede clasificarse en función del número de parámetros que se consideran en la función que describe la c.c.i. (curva característica del ítem) del modelo.



Se habla de funciones logísticas, pero, como es sabido, también es posible utilizar otras funciones de uno, dos y tres parámetros, especialmente la función de ojiva normal. Sin embargo, se menciona la logística porque es la que más fortuna ha hecho en la investigación.

Los modelos de Rasch, que actualmente pueden considerarse una familia de modelos más que 'un' modelo, se clasifican dentro de los logísticos de un parámetro. El esquema también recoge algunas de las aplicaciones de la TRI.

Otras muchas aplicaciones podrían añadirse al esquema, y también podría considerarse una subclasificación de los modelos de análisis de clase latente. Por supuesto, sólo una parte de la literatura representativa de algunos puntos del esquema ha sido incluida en el mismo.

## **5. Esquema de los principales núcleos de la Psicometría.**

En síntesis, hemos destacado tres grandes núcleos conceptuales en la psicometría: la teoría de la medición, los métodos de escalamiento y la teoría de los tests, subdividida en teoría clásica y teoría de la respuesta al ítem. A su vez, hemos efectuado una serie de distinciones dentro del apartado del escalamiento, tales como escalamiento psicofísico versus escalamiento psicológico, y escalamiento unidimensional versus multidimensional. Todavía dentro de estos hemos presentado algunas distinciones. También hemos efectuado algunas distinciones dentro del apartado de teoría de los tests. En primer lugar, hemos distinguido entre los dos grandes troncos, la teoría clásica de los tests y la teoría de la respuesta al ítem, considerando estas denominaciones con amplitud.



la *figura 10*. La figura presenta en su parte superior los grandes apartados considerados. Primero los fundamentos de la medición psicológica. Segundo, el escalamiento. Y, tercero, la teoría de los tests ya desglosada en teoría clásica y teoría de la respuesta al ítem. Bajo cada uno de estos encabezamientos se presenta un esquema de algunos de sus contenidos principales.

En los fundamentos de la medición psicológica se mencionan dos campos, uno denominado introducción a la medición, y otro referido a las escalas de medida en la acepción original de Stevens (1.951).

El escalamiento se presenta desdoblado en tres grandes núcleos temáticos: el escalamiento psicofísico, el escalamiento unidimensional y el escalamiento multidimensional. El esquema insiste sobre todo en el escalamiento unidimensional para destacar los modelos de escalamiento de Thurstone, de Likert, de Guttman y de Coombs. Por supuesto, no sólo existen estos modelos y métodos de escalamiento. Se han destacado por su valor histórico y por representar líneas clásicas principales.

El esquema presta su mayor atención a la teoría de los tests, y, particularmente, a la teoría clásica de los tests. De la celdilla de la teoría clásica de los tests se han desglosado: el proceso de construcción de tests psicológicos, la fiabilidad y la tipificación. La teoría de la fiabilidad ha constituido el núcleo de desarrollo de la teoría clásica. Bajo la teoría de la fiabilidad se ubica el tema del error de medida, la aproximación de medidas paralelas, los factores que inciden sobre la fiabilidad de un test, incluyendo homogeneidad y longitud, la aproximación de consistencia interna y se introduce también el tema de la teoría de la generalizabilidad, próximo al análisis de la fiabilidad mediante análisis de varianza, debido a Hoyt, y también a la cuestión de la validez. Los grandes apartados de validez y análisis de ítems han sido considerados en un espacio intermedio entre la teoría clásica (en la que suelen incluirse) y la teoría de la respuesta al ítem, dado que contienen elementos utilizables desde ambas. Bajo la celdilla

de validez se consideran los subcampos clásicos de validez de contenido, de criterio y de constructo. Bajo la validez de criterio se incluye también el análisis de modulación y bajo la validez de constructo se considera también la extensa cuestión de la estructura factorial y los modelos de ecuaciones estructurales que aunan modelos path y modelos de análisis factorial. Dentro del análisis de ítems se destaca el tema del banco de ítems.

La teoría de la respuesta al ítem requiere una introducción dedicada a considerar sus supuestos, y el concepto de curva característica de ítem, entre otros contenidos. La presentación de los principales modelos, con atención a las cuestiones de estimación de parámetros y ajuste, y el tratamiento de las principales aplicaciones de la misma completan el esquema.

La figura 10 contiene una selección de algunos de los principales núcleos conceptuales de la Psicometría. Necesariamente, omite y simplifica muchos temas importantes, con un propósito interpretativo y ordenador. La complejidad subyacente quedará mejor reflejada en la aproximación histórica. El lector podrá ir adquiriendo una comprensión más precisa del mismo -y también de sus limitaciones- a medida que vaya avanzando en su conocimiento del campo.