



**INSTITUTO UNIVERSITARIO
DE LA EMPRESA**

**INTRODUCCIÓN A LA MEDICIÓN OBJETIVA EN
ECONOMÍA, ADMINISTRACIÓN Y DIRECCIÓN DE
EMPRESAS: EL MODELO DE RASCH.**

OREJA RODRÍGUEZ, JUAN RAMÓN

SERIE ESTUDIOS 2005/ 47

SANTA CRUZ DE TENERIFE, ABRIL DE 2005
(Reedición ampliada diciembre 2005)



**UNIVERSIDAD DE
LA LAGUNA**

La investigación en epistemología de la medida en Economía, Administración y Dirección de Empresas está en un periodo de tensión creativa, a partir de la evidencia de las limitaciones de las teorías del test implícitamente utilizadas en este campo científico. Hay una gran necesidad de instrumentos de medida que permitan diagnosticar las distintas situaciones económicas y empresariales, permitiendo la adopción de las apropiadas acciones estratégicas. En un intento de aproximarnos a los planteamientos objetivos de las mediciones actitudinales presentamos un análisis previo de los requisitos de las mediciones, las limitaciones de la Teoría Clásica del Test y la discusión de los Modelos del Rasgo Latente, para finalmente destacar la importancia del Modelo de Rasch. Este modelo se utiliza actualmente en psicología, educación y medicina donde ha evidenciado su utilidad, a la vez que se está desarrollando en amplios campos científicos de las ciencias sociales. En este trabajo destacamos las principales características del modelo, su actual nivel de aplicación en la investigación de la Economía y Administración y Dirección de Empresas en la universidad española, junto con una experiencia concreta de un análisis simple utilizando el programa de computación Ministep / Winsteps.

The investigation in epistemology of the measurement in Economy and Business Administration is in a period of creative tension, to divide of the evidence of the limitations of the implicitly used test theories in this scientific field. There is a great necessity of measurement instruments that allow diagnosing the different economic and Business situations, allowing the adoption of the appropriate strategic actions. In an attempt to approximate us to the objective expositions of the attitudinal measurements we presented a previous analysis of the requirements of the measurements, the limitations of the Classical Tests Theory and the discussion of the Models of the Latent Trait, finally to emphasize the importance of the Model of Rasch. This model is used at the moment in psychology, education and health care where it has demonstrated his utility, simultaneously that is being developed in wide scientific fields of social sciences. In this work we emphasized the main characteristics of the model, its present level of application in the investigation of the Economy and Business Administration in the Spanish university, along with a concrete experience of a simple analysis using the program of computation Ministep/Winsteps.

INTRODUCCIÓN A LA MEDICIÓN OBJETIVA EN ECONOMÍA, ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS: EL MODELO DE RASCH.

OREJA RODRÍGUEZ, JUAN RAMÓN

SERIE ESTUDIOS 2005/ 47

LA LAGUNA, ABRIL DE 2005
(Reedición ampliada diciembre 2005)

Juan Ramón Oreja Rodríguez. Departamento de Economía y Dirección de Empresas. Facultad de Ciencias Económicas y Empresariales. Campus de Guajara. Universidad de La Laguna. Camino de La Hornera s/n 38071 La Laguna Tenerife Islas Canarias (España)
Teléfono: 922 31 70 56 Fax: 922 31 70 77 Correo electrónico: joreja@ull.es

ÍNDICE

PRESENTACIÓN

1.- INTRODUCCIÓN A LA MEDICIÓN OBJETIVA

- 1.1.- Concepto de medida
- 1.2.- Antecedentes epistemológicos
 - 1.2.1.- Planteamientos previos
 - 1.2.2.- La Teoría Clásica del Test: TCT
 - 1.2.3.- La Teoría del Rasgo Latente
 - 1.2.4.- La Teoría de la Respuesta al Ítem: TRI

2.- MODELO DE RASCH

- 2.1.- Elementos del modelo de Rasch
- 2.2.- Familia de modelos de Rasch
- 2.3.- El modelo de Rasch en ADE

3.- LA PRÁCTICA DE LA MEDICIÓN OBJETIVA: UNA APLICACIÓN DEL MODELO DE RASCH EN ADE

- 3.1.- Programas informáticos: Winsteps / Ministep
- 3.2.- Entradas
 - 3.1.1.- Fichero de control: Órdenes sintácticas básicas
 - 3.1.2.- Fichero de datos: del Excel al Wordpad
- 3.3.- Resultados
 - 3.3.1.- Fichero de resultados: creación y gestión
 - 3.3.2.- Ejecución del programa MINISTEP
 - 3.3.3.- Tablas de medición conjunta
 - 3.3.4.- Estadísticos de fiabilidad
 - 3.3.5.- Calibración de los ítems
 - 3.3.6.- Desajustes de los ítems
 - 3.3.7.- Medida de los sujetos
 - 3.3.8.- Desajuste de los sujetos
 - 3.3.9.- Escalograma de Guttman

BIBLIOGRAFÍA

PRESENTACIÓN

En línea con el proceso de mejora metodológica en que se encuentra inmerso nuestro campo científico hemos realizado una investigación sobre la medición y su calidad. Un primer resultado se presenta con este documento de trabajo, en donde tratamos de destacar la problemática de la medición y los requisitos de aceptabilidad de la misma a los niveles de calidad adecuados. Por ello, nos hemos centrado en la nueva línea metodológica de la medición objetiva derivada de los modelos probabilísticos de Rasgos Latentes, en particular el Modelo de Rasch, que suponen una superación de las limitaciones de la Teoría Clásica del Test.

En las revisiones bibliográficas a nivel nacional e internacional, se aprecia una incipiente uso en las Ciencias Económicas y Empresariales del Modelo de Rasch, por lo que nos parece oportuno acercarnos al conocimiento del mismo con un mayor detalle, de tal forma que nos permita evaluar su viabilidad como modelo apto para la interpretación de la información derivada de la utilización de los instrumentos de medida que aplicamos en la delimitación de nuestros constructos.

El trabajo que se presenta tiene un carácter introductorio, siguiendo la sistemática planteada por Leplège en su nota técnica de 2003 sobre la obra de George Rasch, por cuanto intenta contestar las preguntas relacionadas con el diseño de un sistema de medición objetivo. En ese sentido, es necesario enfrentar la crítica de Campbell a la posibilidad de medición de los “eventos sensoriales” con las respuestas que ofrece Stevens, para las ciencias sociales.

El camino moderno de la medición objetiva ha supuesto la incorporación a su acervo de parte de las propuestas de Thurstone, Guttman y Lazarsfeld, criticadas y, posiblemente, mejoradas constantemente por la investigación sobre epistemología de la medida.

Por otro lado, los instrumentos de medida han sido utilizados partiendo de unos modelos matemáticos determinados, por lo que la calidad de las mediciones depende del análisis de sus errores. Por ello, se presenta a nivel introductorio la Teoría Clásica del Test, junto a la Teoría de la Respuesta al Ítem, con el modelo probabilístico general que la soporta.

Dentro de la Teoría de la Respuesta al Ítem, nos encontramos con una fuerte discusión sobre si los planteamientos de Rasch, con su modelo de “un parámetro” son concluyentes o las propuestas generalizadoras de los modelos de “dos y tres parámetros” son más aceptables. La presentación de los argumentos de ambas partes nos ha llevado a centrarnos en el Modelo de Rasch.

La información que presentamos del Modelo de Rasch se inicia con una explicación sobre sus orígenes y las asunciones que posibilita que se cumplan los requisitos exigibles para una medición objetiva. La utilización del modelo logístico como modelo de ajuste de los datos que permite determinar la probabilidad de respuesta al ítem, la medición conjunta de los parámetros de los sujetos y de los ítems y la determinación de la unidad de medida

“*logit*” como expresión de la distancia entre los parámetros del modelo, que nos admite comparaciones en términos de la aritmética lineal

Presentamos el modelo matemático de Rasch y las implicaciones derivadas de los procesos de estimación de los parámetros mediante algoritmos de máxima verosimilitud, mediante casos que han utilizado los programas Ministep / Winsteps.

La utilización del Modelo de Rasch no puede ser inmediata si no se analizan las distintas aplicaciones ya existentes en la literatura de nuestro país, por ello presentamos unas fichas que describen las estructuras de los trabajos que aplican el Modelo de Rasch en Economía, Administración y Dirección de Empresas. Estos trabajos analizados es el resultado de cruzar los buscadores bibliográficos con la información disponible sobre grupos de investigación que trabaja con el Modelo de Rasch.

El conocimiento de los programas de computación permite el tratamiento de la información y genera un conjunto de estadísticos. Por ello, presentamos el programa Ministep / Winsteps disponible como recurso de investigación educativa en la Web del Ministerio de Educación y Ciencia, planteando un programa de control sintáctico básico, pero útil para las necesidades derivadas de los diseños de las investigaciones a realizar. En ese sentido diseñamos el programa con las distintas órdenes que lo componen, así como los ficheros de datos y resultados necesarios.

El análisis de la calidad de la medición se denota por la posibilidad de calibrar los ítems y considerar el desajuste de la información procedente de algún sujeto encuestado. Esta calidad se aprecia a partir de los coeficientes de consistencia interna, así como los de ajustes al modelo. El estudio de los ajustes y desajustes, junto con el análisis de los estadísticos y curvas disponibles ocupa parte de este trabajo.

Finalmente se presenta un repertorio bibliográfico que puede permitir a cualquier investigador interesado su introducción en este campo de la medición.

Este trabajo no se habría llevado a cabo sin hubiésemos contado con la ayuda de muchos compañeros y amigos. Quisiera destacar el papel jugado en la introducción en el campo del diseño de constructos, medidas y control de calidad de las mismas de los catedráticos Cesar Camisón (Universidad Jaime I) y Francisco J. Sarabia (Universidad Miguel Hernández). Al catedrático Pedro Álvarez (Universidad de Extremadura) le debo mi aproximación al Modelo de Rasch. Por otra parte, los compañeros del Grupo de Investigación del Análisis Estratégico del Departamento de Economía y Dirección de Empresas y del IUDE, que están aplicando en sus investigaciones esta nueva metodología.

A todos: Gracias.

Quedo a disposición de los profesores y alumnos, agradeciendo sus opiniones sobre el material que se presenta.

Juan R. Oreja
La Laguna
Abril de 2005

Agradecemos las revisiones que se nos han enviado desde distintas universidades sobre temas específicos que hemos procedido a desarrollar. Mantenemos abierto el documento de trabajo con objeto de incorporar las nuevas sugerencias que surjan.

Juan R. Oreja
La Laguna
Diciembre de 2005

1.- INTRODUCCIÓN A LA MEDICIÓN OBJETIVA

1.1.- CONCEPTO DE MEDIDA

En la metodología de la Economía, Administración y Dirección de Empresas se encuentra fuertemente enraizada la búsqueda de información a partir del análisis comparativo de entes o sujetos ante conceptos económicos. No obstante, como indica Everitt (1984), "...algunos conceptos [en Ciencias Sociales] no están bien definidos y hay muchas discusiones sobre el significado real de los términos...". Tales conceptos se vienen a definir como variables latentes dado que no se pueden observar directamente. En todo caso nos acercamos a conocer algunas de sus propiedades, rasgos, atributos o características. Estos conceptos son "...constructos hipotéticos inventados por los científicos con el propósito de entender algún área de interés...", desde esta perspectiva podemos considerarlos como una construcción social (Berger y Luckman, 2003). Los constructos no son directamente observables, pero algunas de sus manifestaciones sí lo son.

Al comparar las mediciones de las manifestaciones observables de los constructos nos encontramos con la problemática de la medición objetiva, dado nuestro interés de utilizar adecuadamente la información proveniente de la comparación de los sujetos ante los conceptos económicos y, en su caso, generalizar los resultados obtenidos.

En este contexto podemos indicar que toda medida es el resultado de una comparación entre sujetos económicos. Medir es asignar números a observaciones de acuerdo con cierto conjunto de reglas (Summers, 1976), por lo que siguiendo a Leplège (2003), desde una perspectiva empírica, se puede establecer que toda medida es considerada como la expresión de una graduación: un escalar. El aspecto básico de la medición, a efectos comparativos, es que el constructo pueda ubicarse en una línea, denominada continuo lineal. Esta línea se puede dividir en unidades iguales, que pueden ser mayores o menores, desde un origen (Andrich y Marais, 2005). Un número real que provenga de este proceso, debe haber estado vinculado a tres elementos:

- Un concepto (constructo) a medir
- Un instrumento de medida (que trata de cuantificar los conceptos abstractos no directamente observables).
- Un protocolo experimental que tiene por finalidad estandarizar las circunstancias de la comparación en el constructo a medir mediante el instrumento de medida.

La discusión sobre si se pueden medir conceptos abstractos no directamente observables (constructos) tuvo un momento álgido en 1932 cuando el físico británico Campbell negó la posibilidad de estimación cuantitativa de los eventos sensoriales (Campbell, 1932).

Stevens (1959) replantea el problema de medición mediante el cambio conceptual del problema pasando de unas operaciones inventadas (visión física) a escalas clasificadas (visión matemática). A tales efectos Torgerson (1958) indica que "la aproximación de

Stevens u otras similares se refieren a diferentes métodos para una sistemática clasificación de conjuntos limitados de objetos concretos, más que a métodos de medición de propiedades abstractas”

De acuerdo con Stevens medir es la asignación de un número a eventos u objetos de acuerdo con una regla. En su trabajo trata de precisar qué tipo de aritmética era válida para los diferentes números que resultaban de las mediciones (Wright, 1997 a), para lo cual estableció la siguiente tabla (Stevens, 1959):

ESCALA	OPERACIÓN	EJEMPLO	ESTADISTICO CENTRAL	ESTADISTICO DISPERSIÓN	ASOCIACIÓN	TEST
NOMINAL	Igualdad	Contar ciudades	Moda	-	-	χ^2
ORDINAL	Desigualdad > <	Dureza mineral Número calles Puntuaciones	Mediana	Percentiles	Correlación del rango de ordenación	Test del signo y ejecución
INTERVALO	Distancia	Temperatura Posición Tiempo estándar Puntuaciones	Media aritmética	Desviación estándar	Correlación Producto momento	Test t Test F
RATIO	Ratio	Contar Longitud Densidad Posición	Media geométrica Media armónica	Asociación porcentual		

La representación gráfica de la tipología de mediciones de acuerdo con la escala de Stevens, podría ser la siguiente (Wright, 1997):

Nominal	3	875	23	12		
Ordinal	1	2 3 4	5	6	7	
Intervalo	-3	-2	-1	0	1	2 3
Ratio	10^{-3}	10^{-2}	10^{-1}	10^0	10^1	10^2 10^3

Como destaca Wright (1997a) contar pertenece a la clase de escala de ratio. Para aplicar la aritmética lineal usual en la escala de intervalo se exige la aplicación de logaritmos a las

puntuaciones obtenidas. Si estamos contando objetos reales, como preguntas correctas o manzanas, las puntuaciones obtenidas sólo tienen valor ordinal debido a la variación del tamaño del objeto. La solución propuesta por Stevens (1959) no genera mediciones lineales, sino clasifica los números ya en uso, destacando la importancia de las escalas de ratios que pueden linealizarse mediante logaritmos.

Ha habido una fuerte discusión respecto a las escalas de medida, Siegel (1956) y Senders (1958) asumen los conceptos de Stevens, a la vez que destacan las técnicas estadísticas apropiadas de acuerdo a dicha escala. No obstante, Lord (1953) critica que una escala de medida deba dictar las estadísticas que hay que emplear. Kaiser (1960) discute las propuestas de Stevens y también Cliff (1992) destaca la debilidad del trabajo de Stevens por centrarse en la aparente implicación del investigador dado que, de algún modo, la naturaleza de una escala se define por éste, mientras que Duncan (1984) destaca la confusión que este sistema de ha generado en la ciencias sociales cuantitativas.

Actualmente, se vienen utilizando estas escalas de forma corriente (Sarabia, 1999) en la investigación social. Pero como indican Glass y Stanley (1970:13) es importante considerar las relaciones entre las medidas con lo que se está midiendo. La naturaleza de esta relación interesa más al especialista en medida, ya que los métodos estadísticos son medios para analizar los números en tanto que tales y no en tanto que atributos reales cuantificados. Cualquier conjunto de números es susceptible de procesarse estadísticamente mediante cualquier técnica (con las correspondientes limitaciones), pero no se conoce ninguna técnica que no pueda emplearse en razón de que los números que se le adscribieron no son los “indicados”. Los métodos estadísticos ni quitan ni ponen a la significación de los números con base en los cuales se efectúan. A tales efectos, destacan el pensamiento de Kaplan (1964: 205/206): “Las matemáticas nos ahorran el arduo trabajo de pensar por cuenta propia, pero esta ventaja se paga pensando antes y después de hacer entrar en juego a las matemáticas”

Se completa el proceso de medición con la consideración de los conceptos de validez y fiabilidad. La validez es un juicio evaluatorio integrado del grado en el cual la evidencia empírica y la racionalidad teoría apoya la adecuación y la corrección de las inferencias basadas en las puntuaciones (Messick, 1988). Una medida es válida si mide lo que se supone está midiendo.

Fiabilidad es la propiedad de un conjunto de puntuaciones de un test que indica la cantidad de error de medición asociado a las puntuaciones (Frisbie, 1988). De esta forma se está considerando la consistencia relativa de las puntuaciones del test (Traub y Rowley, 1991). Es la medición de la exactitud de las puntuaciones de un test. A la fiabilidad le afecta tanto la longitud del test, como la dificultad y discriminación de los ítemes, las características del grupo encuestado y diversos aspectos relacionados con la administración del test (Frisbie, 1988; Traub y Rowley, 1991, Andrich, 2005).

La fiabilidad no implica validez. Una medición puede ser exacta, lo cual no supone que se esté midiendo lo que se desea medir. Una medida válida debe ser fiable, ya que debe referirse a los resultados que con precisión reflejan los conceptos que se miden. La fiabilidad es la condición necesaria pero no suficiente de la validez.

1.2.-ANTECEDENTES EPISTEMOLOGICOS

Los actuales modelos de análisis de las medidas generadas por los instrumentos de medición de los constructos tienen como referencias seminales los trabajos de Thurstone, Guttman y Lazarsfeld. En paralelo, la teoría tradicional del test (Spearman) ha servido como base instrumental, con sus planteamientos lineales, de la interpretación de las puntuaciones observadas, de acuerdo con el paradigma de la puntuación verdadera. Del análisis de las limitaciones de la teoría tradicional, en el contexto de los desarrollos de los planteamientos previos, se presenta el paradigma del Rasgo Latente, dentro del cual se ha desarrollado con peculiaridades específicas el modelo de Rasch.

1.2.1.- PLANTEAMIENTOS PREVIOS

THURSTONE

Fue un pionero de la aplicación de los métodos estadísticos al estudio de los procesos psíquicos (Thurstone, 1928; Tourstone y Chave, 1929). Propuso cuatro criterios que deben cumplir todas las medidas en las Ciencias Sociales (Andrich, 1989, Leplège, 2003):

- Unidimensionalidad. La medición implica orden y magnitud en una única dimensión. Toda medida permite que se pueda hacer una comparación del tipo: “es mayor que”, lo que supone:
 - La existencia de un continuo,
 - La existencia de una relación de orden, en términos de graduación, de intensidad o de magnitud: la longitud del continuo. (La idea de medida implica la existencia de una suerte de continuo lineal).

Thurstone admite que numerosos fenómenos relevantes en las Ciencias Sociales son multidimensionales pero que para compararse es necesario un escalamiento lineal (Thurstone, 1959: 218-219). Ello nos lleva al establecimiento de una hipótesis operativa de unidimensionalidad, que se tratará de validar a lo largo del proceso de medición.

Para medir hace falta construir un instrumento de medida, que permite que las magnitudes de las propiedades de un constructo se pueda ubicar en un continuo lineal. Para Thurstone, la localización de los ítems, formalizan y definen ese continuo (Andrich, 1989). “Deberíamos, por lo tanto, localizar esos ítems de los test en la escala como marcas de los diferentes niveles de crecimiento lineal” (Thurstone, 1925:267).

- La construcción de un instrumento de medida implica que se puede evaluar la coherencia y validez de las medidas obtenidas por la aplicación de un instrumento dado mediante modelos matemáticos explícito.

“Debemos superar los meros procedimientos de correlación desde que es posible hacerlo mejor que simplemente correlacionando atributos.

Cuando un problema es tan complejo que ninguna formulación racional esta disponible, entonces alguna cuantificación es aún posible por medio de los coeficientes de correlación de las contingencias o similares. Pero tales procedimientos estadísticos constituyen un reconocimiento de fallos para racionalizar el problema y establecer las funciones que subyacen en los datos. Queremos medir la separación entre dos opiniones en un continuo de actitud y queremos comprobar la validez de lo asumido en el continuo por medio de su consistencia interna” (Thurstone, 1959:267).

Como indica Andrich (1989) el modelo básico procede de los métodos físicos en los cuales el diseño para la recolección de datos requiere que cada persona de una clase definida compare objetos físicos unos con otros con respecto a alguna propiedad tal como el peso. Estos valores subjetivos calculados a partir de las respuestas pueden localizarse en relación a conocidos valores físicos.

El mayor paso que dio Thurstone fue liberar las escalas de valores subjetivos y psicológicos de cualquier contrapartida física a través de su ley de los juicios comparativos (Thurstone, 1927), que lleva a la representación matemática de un proceso discriminante, en donde se lleva a cabo comparaciones entre pares de conjuntos de estímulos (atributos, rasgos, actitud). Para ello introduce el concepto de continuo psicológico.

Esta ley es en realidad un modelo de medición que aplicada en un determinado contexto constituye una hipótesis científica relacionada con los resultados de comparaciones entre conjuntos de objetos.

Los elementos de la ley de los juicios comparativos son:

Cuando una persona n se enfrenta a una cuestión i , se percibe que existe el valor A_{ni} definido como variable aleatoria continua de acuerdo a la expresión:

$$A_{ni} = \alpha_i + \epsilon_{ni}$$

En donde:

α_i : es el valor afectivo o localización del ítem i , que es constante para todas las personas

ϵ_{ni} : es el valor aleatorio asociado con la persona n (al contestar al ítem i)

En las Ciencias Sociales, los modelos matemáticos son inevitablemente de naturaleza estadística, probabilísticos. El uso de estos modelos proporciona una oportunidad para verificar la consistencia de las estimaciones.

- Un requisito consistente es que la localización del ítem debe ser aditiva. Thurstone planteó el criterio de coherencia: aditividad

$$\text{Si } A > B \quad A = B + x$$

$$B > C \quad B = C + y$$

$$A > B > C \quad A = C + z \Rightarrow A = C + x + y \quad ; z = x + y$$

Si se puede obtener esta relación, se puede admitir la existencia de un continuo. Pero si esta relación no está presente, entonces la validez del continuo asumido queda rechazada. No se ha logrado construir una variable.

La vía de construcción de la variable latente ha fallado, pero puede sugerir posibles modificaciones de los ítems o de los conceptos que representan en el constructo.

- Invarianza de la medida.
Un instrumento de medida de un constructo debe ser independiente de las circunstancias coyunturales de la medición.
Para que un instrumento sea considerado como válido, los valores atribuidos a los ítems no deben depender de las opiniones de las personas de la muestra que han servido para diseñar la escala.

“Si se ve una escala como válida, sus valores no pueden verse afectados por las opiniones de las personas que han ayudado a construirla. Esto puede llevar a tests exigentes, pero los métodos de creación de escalas deben establecer tales tests antes de que sea aceptado como una simple descripción de las personas que construyeron las escalas” (Thurstone, 1928: 547).

Este criterio de invarianza de la medida se configura como una segunda hipótesis operativa en el proceso de medición, al plantearse como un requisito de la localización de los ítems en el continuo lineal. Implica dos exigencias simétricas (Andrich, 1989):

1. Se pueden omitir ítems del test a diferentes niveles de la escala sin afectar las puntuaciones individuales.
2. Si una escala se considera que es válida, no debe verse afectada por las opiniones de las personas que han contribuido a construirla.

Si la localización de los ítems no evidencia invarianza a lo largo de la escala, entonces hay una diferencia en el tipo de escalamiento de los ítems y la localización relativa de los ítems no proporciona nada más que la descripción de las personas encuestadas. De esta forma no se puede realizar comparaciones de grado entre las personas. Cuando la construcción de una variable latente falla en este criterio es que no se verifica la hipótesis de invarianza.

GUTTMAN

La propuesta básica de la escala de Guttman (1944, 1947, 1950) es que entre los ítems que la forman hay una relación interna tal que una persona que concuerda con un ítem que ocupa una determinada posición en la escala, ratificará todos los ítems que estén por debajo de ella en la escala (Dotson y Summers, 1976). Guttman propone un tipo de escala de respuestas en donde los ítems de un constructo unidimensional se ordenan de acuerdo a su dificultad, lo que permite una ordenación conjunta de sujetos e ítems. Para ello Guttman impone un orden estricto a las respuestas de los ítems, bajo planteamientos deterministas. Así, si se responde SI a una cuestión es porque se ha contestado SI a cuestiones de nivel inferior (escalograma de Guttman).

Esta estructura permite identificar un continuo y clasificar a los individuos de una población en las clases correspondientes a grandes jerarquías. En este continuo, se plantea el cumplimiento de dos criterios:

1. La jerarquía de los sujetos no depende de los ítems de los cuestionarios.
2. La jerarquía de los ítems no depende de las personas que han servido de base para el diseño de la encuesta.

La información necesaria para clasificar a los sujetos es la puntuación obtenida en los ítems. De esta forma, los sujetos se ordenan de acuerdo a su puntuación total, la cual es un índice de su habilidad. La puntuación total de los sujetos es la suma de las puntuaciones de los encuestados en cada ítem. Para cualquier puntuación total se puede inferir la pauta de respuestas y supone que todos los sujetos con el mismo total tienen la misma pauta (escala perfecta). No obstante, en la realidad esta situación puede no cumplirse debido a determinadas circunstancias del diseño, la actitud de sujeto encuestado y administración de los test (Andrich, 1988; 2005). En estos casos la información que se obtienen se deriva de que la estructura esperada de acuerdo con Guttman no se cumple y la evidencia de esta situación permite determinar las causas que genera la excepción.

Los ítems también pueden ordenarse de acuerdo a su puntuación total, que es la suma de las puntuaciones de los sujetos en cada ítem. (Véase el ejemplo del escalograma de Guttman para un test con escalamiento de 1 a 5)

La estructura planteada por Guttman proporciona un excelente marco de referencia para el análisis de la información proporcionada por un cuestionario, destacando la necesidad de modelos matemáticos (Lazarsfeld, 1977). Por ello se trataría de diseñar cuestionarios que incluya ítems con dificultad relativa que permitan verificar la progresión de los sujetos a lo largo del constructo.

Ejemplo de un escalograma de Guttman para un test con escalamiento de 1 a 5

Encuestados	Ítemes										Total
	5	10	4	1	8	6	2	3	9	7	
10	5	5	5	5	5	4	4	3	5	3	44
5	5	5	4	5	4	4	4	4	4	4	43
14	5	5	5	5	5	4	3	3	4	3	42
1	5	5	5	5	5	4	3	3	3	3	41
4	5	5	5	3	5	3	4	3	4	4	41
13	5	5	5	4	4	4	3	3	5	3	41
7	5	5	5	4	5	4	3	3	3	3	40
2	5	4	5	4	4	4	3	3	4	3	39
6	5	5	4	4	3	4	3	3	3	3	37
8	4	4	4	4	3	4	4	4	3	3	37
3	5	5	4	3	4	3	3	4	2	3	36
11	3	4	4	4	3	3	4	4	3	4	36
9	5	5	5	5	3	3	3	3	1	2	35
12	5	5	4	3	3	2	3	3	2	3	33
Total	67	67	64	58	56	50	47	46	46	44	

Fuente: Elaboración propia (programa de computación Ministep)

La principal limitación de este planteamiento es que no se tiene en cuenta la incertidumbre derivada de las opiniones, valores y juicios de los encuestados.

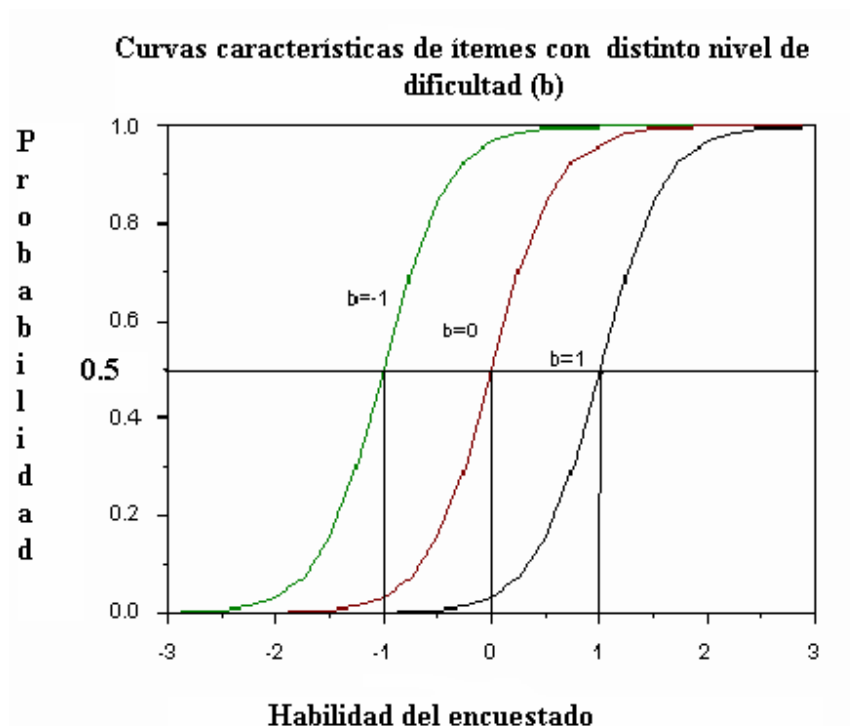
LAZARSFELD

Será el matemático y sociólogo Lazarsfeld quien en el planteamiento de sus análisis de las estructuras latentes desarrolló el modelo de Guttman, a partir de la consideración de las probabilidades que relacionan las observaciones obtenidas de los ítemes con los constructos (Lazarsfeld, 1977). Una serie de ecuaciones enlazan los parámetros latentes y las frecuencias de respuestas observadas empíricamente, de acuerdo al principio de la independencia local.

Parte de sus estudios de variables latentes, estableciendo la posibilidad de medir las características observables de las respuestas a los ítemes de los encuestados. Una actitud, habilidad, atributo puede representarse como una variable latente (Lazarsfeld, 1961).

Según la presencia en cada sujeto de una cierta intensidad de esta actitud la probabilidad de respuesta de forma específica a una cuestión dada será diferente (Boudon, 1999). De ahí que se pueda calcular una curva característica de respuesta al ítem, como representación de

la distribución de probabilidad de la respuesta a ese ítem, en el caso de funciones latentes continuas. El gráfico de la curva característica del ítem recoge una función continua que enlaza la probabilidad de respuesta positiva a un ítem con una posición sobre el continuo latente (Lazarsfeld, 1977). Hay una tipología de curvas características de los ítems, de acuerdo con las suposiciones que se realicen sobre la función de distribución. En respuestas dicotómicas, la expresión de estas curvas es del tipo logístico. El punto de inflexión se corresponde con la probabilidad de respuesta del 50%, expresando el nivel de dificultad de la cuestión. El análisis de las CCI permite identificar el poder discriminatorio correspondiente a la pendiente de la parte media de la curva.



Fuente: Brannic (2005) y elaboración propia.

En respuestas politómicas, las curvas características puede ser lineales, normales y logísticas (Martin, 1997).

1.2.2.- TEORÍA CLÁSICA DEL TEST

La mayoría de los trabajos de investigación en Economía, Administración y Dirección de Empresas que utilizan instrumentos de medida que pretenden medir conceptos percibidos por los sujetos económicos y empresariales utilizan de forma implícita la Teoría Clásica del Test (TCT). Ésta se desarrolló debido a la necesidad sentida por los psicólogos de evaluar las sensaciones percibidas y en general los fenómenos subjetivos, lo que les llevó a diseñar medidas que utilizaban métodos empíricos sin llegar a ser modelos matemáticos sofisticados, introduciéndose en los diferentes campos de las Ciencias Sociales.

El objetivo de la TCT es el establecimiento de la relación existente entre los fenómenos subjetivos que se quieren medir con los valores observados. Ello se realiza a partir de la aplicación del instrumento de medida, mediante modelos matemáticos simples.

Los tests serán los instrumentos de medida de los fenómenos subjetivos. Las puntuaciones obtenidas permiten inferir en qué grado se manifiestan dichos fenómenos.

Spearman (1904; 1910) planteó su Teoría del Resultado Verdadero, mediante un modelo de regresión lineal con dos variables cuyo supuesto fundamental es que si existe una serie de ítems en un test que sean indicadores del mismo constructo, entonces se puede cumplir que:

1. Cada persona v tiene una puntuación verdadera en el constructo que se denomina T_v .
2. El mejor indicador de la puntuación verdadera de la persona es la suma de las puntuaciones de los ítems que se denomina X_v (puntuación observada).
3. Estas puntuaciones observadas presenta un error en la medición de cada persona E_v .
4. Esos errores no están correlacionados con la puntuación verdadera.
5. En la población, la suma de errores es cero y están normalmente distribuidos.

Las principales ecuaciones que se pueden desarrollar la TCT, a partir de esos criterios son (Andrich y Marais, 2005: 26/27):

Ecuación general de la TCT:

$$X_v = T_v + E_v \quad [1],$$

En donde:

$E[X_v] = T_v$ La puntuación verdadera es la esperanza matemática de la puntuación observada.

$\rho[T_v, E_v] = 0$ La puntuación verdadera y el error son independientes (no están correlacionados).

$E[E_v] = 0$ La esperanza matemática de los errores de la medida es nula

Varianza del valor observado $S_x^2 = S_t^2 + s_e^2$ [2]

Fiabilidad del test $r_{xx} = S_t^2 / S_x^2$ [3]

En la estimación de un coeficiente de fiabilidad del test se parte de la idea que existen dos tests (tests paralelos) para medir el mismo constructo. Cada test tiene su error de medida, pero la puntuación verdadera es la misma. A partir de la ecuación [1] se puede determinar:

$$\begin{aligned} X_{v1} &= T_v + E_{v1} \\ X_{v2} &= T_v + E_{v2} \end{aligned}$$

Al estar interesados en la consistencia de los valores observados en los dos tests se calcula la covarianza entre X_{v1} y X_{v2} : C_{12} , cuyo valor es S_t^2 (varianza de la puntuación verdadera). La correlación entre estos dos tests sería:

$$r_{12} = C_{12} / S_1 S_2 ,$$

o lo que es lo mismo:

$$S_t^2 / S_x^2 = r_{xx}. \quad [3]$$

Esta ecuación expresa la proporción de la varianza de la puntuación total que se explica por la varianza de la puntuación verdadera. El rango de variación de este coeficiente de correlación varía entre 0 y 1.

Predicción de la puntuación verdadera $\hat{T}_v = \bar{X} + r_{xx} (X_v - \bar{X})$ [4]

Error estándar de la medición $S_e = S_x \sqrt{(1 - r_{xx})}$ [5]

LA VALIDEZ EN LA TCT

Una medida válida es la que está midiendo lo que se supone debe medir. La validez puede apreciarse en un variado número de formas alternativas. En las áreas de Economía y Empresa se está desarrollando un cuerpo científico construido a partir de la consideración de constructos. Éstos no pueden ser utilizados para desarrollo teóricos y empíricos ulteriores si no se logra evidenciar una alta correspondencia entre estas construcciones abstractas y los procedimientos usados para hacerlos operativos. Tal objetivo se logra a través del análisis de la validez del constructo que relaciona el grado de correspondencia entre los constructos y sus medidas (Peter, 1981), lo que en términos reales llevaría a establecer que una medida tiene validez de constructo en función:

- 1) Del grado que la medida valore la magnitud y la dirección de una muestra representativa de las características del constructo,
- 2) Del grado en que la medida no esté contaminada con elementos que pertenezcan al dominio de otros dominios o por errores.

Messick (1989) integra como dimensiones de la validez del constructo distintos aspectos tradicionales (Cronbach, 1971; Nunnally, 1978), tales como:

Validez de contenido, que puede ser establecida por medios de expertos que determinan que los ítems del instrumento de medida representan razonablemente el constructo a medir.

Validez concurrente, que se evidencia en el caso de los resultados obtenidos por la aplicación de un determinado instrumento de medida están relacionados dentro de lo esperado con los resultados de otros relevantes instrumentos de medida del constructo.

Validez predictiva, apreciada en el caso de que los resultados obtenidos en un determinado instrumento de medida serán consistentes con los que se obtendrán en el futuro en relación con el constructo analizado.

No obstante, el aspecto relevante del análisis de la validez es el significado, relevancia y utilidad de las medidas obtenidas en relación al contexto teórico (Peter, 1981) y a las consecuencias sociales de su uso (Messick, 1989).

La validez del constructo no se logra mediante un único estudio tal como indica Cronbach (1971), que destaca la necesidad de un proceso de agregación de múltiples resultados, que incluiría tanto razonamiento lógico deductivo como series de estudios de fiabilidad y validez (Peter, 1981).

En el análisis de la validez del constructo se consideran tanto los análisis de fiabilidad, como la validez convergente y discriminante de las medidas obtenidas. Centrándonos en los dos últimos, la forma tradicional de obtener información es mediante la utilización de análisis de correlación entre las medidas obtenidas en el instrumento de medida en relación con otros instrumentos con los que se encuentra relacionado (convergente) y con otros instrumentos con los que no se encuentren relacionados (divergentes) con el constructo considerado.

El modelo integrador desarrollado para investigar la validez del constructo es el análisis multirasgo-multimétodo de Campbell y Fiske (1959). Este análisis ha originado un amplio cuerpo de investigación que trata de superar las limitaciones y arbitrariedades de la aproximación original, lo que ha llevado a la inclusión de métodos complementarios relativos a análisis de varianza, análisis path y diversos análisis factoriales (Peter, 1981)

LA FIABILIDAD EN LA TCT

La hipótesis de partida en este planteamiento teórico es que los test son instrumentos falibles e imprecisos, por lo que mediante el modelo propuesto se trata de estimar los errores de medida para obtener el valor verdadero. La fiabilidad se puede definir como el grado en el que las medidas están libres de error, por lo que permite obtener resultados consistentes (Peter, 1979)

La fiabilidad de la medida se determina a partir de una amplia variedad de métodos que tratan de determinar la proporción de varianza sistemática de una escala de medición. Ello se debe a que la fiabilidad mide el nivel de error aleatorio existente en la medición observada y no el error sistemático, objeto de la valoración del criterio de validez.

Los métodos principales de estimación de la fiabilidad son:

Método del test-retest: la valoración de la fiabilidad se lleva a cabo con la misma escala de medida aplicada dos veces a los mismos sujetos bajo condiciones similares. Se estima mediante un coeficiente de correlación de Pearson entre las puntuaciones observadas de las dos administraciones. Esta correlación se puede interpretar como un índice de la estabilidad de las medidas a lo largo del tiempo.

Peter (1979) indica que el uso de este método puede implicar tres problemas básicos derivados del tiempo transcurrido entre las dos administraciones consecutivas del test, la posibilidad de que entre dichas administraciones surja un acontecimiento que influya en las puntuaciones obtenidas y de las correlaciones entre los ítems de la escala. Se sugiere que no sea utilizado como único método de medición de la fiabilidad, sino más bien como complementario con la estimación de la consistencia interna de cada administración, debido a la información que proporciona sobre la estabilidad de las medidas.

El método de la **consistencia interna** se sustenta en la división en dos mitades de las puntuaciones obtenidas en la aplicación del test, las cuales se utilizan para determinar un coeficiente de correlación. Trata de medir la homogeneidad de un conjunto de ítems. Peter (1979) destaca que el principal problema de este método se centra en la partición en mitades de las puntuaciones, ello se trata de superar calculando una fiabilidad media de todas las posibles particiones.

Para respuestas dicotómicas a los ítems Kuder y Richardson (1937) desarrollaron la denominada fórmula KR-20, cuya expresión es:

$$KR-20 = (k/k-1) [1 - (\sum p_i q_i / S_x^2)],$$

En donde:

p_i : es la proporción de respuestas a las que se le ha asignado 1 al ítem i

y

$q_i = 1 - p_i$. (proporción de respuestas a las que se le ha asignado 0)

S_x^2 = varianza de la puntuación observada en la encuesta

Hay un conjunto de variaciones de estas fórmulas denominadas KR21 y KR21' (Frisbie (1988).

La generalización de esta fórmula para los escalamientos politómicos se estableció posteriormente por Cronbach (1946; 1951), cuya alpha se puede determinar mediante la expresión:

$$\alpha = (k / k-1) [1 - (\sum S_i^2 / S_x^2)], \text{ siendo la variación de la } i = \{1, \dots, k\}$$

En donde,

S_i^2 = varianza de los ítems del test.

S_x^2 = varianza de la puntuación observada en la encuesta

El alpha de Cronbach es la fórmula actualmente más aplicada para la determinación de la fiabilidad de las puntuaciones de los instrumentos de medidas integrados de escalas con múltiples ítems en los estudios de Economía y Empresa. El rango de variación del coeficiente α es $\{0, \dots, 1\}$.

Las recomendaciones de Nunnally (1967; 1978), ampliamente difundidos en Economía y Empresa por los trabajos de Churchill (1979) sobre los mínimos de fiabilidad aceptables (modificas entre las dos ediciones indicadas) para los estudios preliminares, básicos e investigaciones aplicadas (0.70/0.80/90) no ha sido justificados teóricamente y han dado lugar múltiples estudios (véase por ej. Churchill y Peter, 1984; Peterson, 1994). Estos estudios han tratado de determinar los condicionantes de los distintos niveles de fiabilidad, tales como la muestra, tipo de muestra, tipo de escala, número de categorías en las escalas de ítems, número de ítems en las escalas, número de ítems suprimidos, modo de administración de la encuesta, la orientación de la escala, la naturaleza del constructo, tipo de investigación, ... , obteniendo diversos resultados que dan respuestas a las cuestiones relacionadas con los condicionantes específicos de la investigaciones económicas y empresariales, delimitando un marco de referencia para la interpretación de un determinado nivel de fiabilidad obtenida.

Método de los tests alternativos. La fiabilidad se determina, a partir de la consideración de los constructos se miden con dos test alternativos (conformados por escalas algo

diferentes) en dos momentos distintos del tiempo, mediante un coeficiente de correlación de las puntuaciones obtenidas. Peter (1979) indica que el principal problema de este método recae en el desarrollo de los tests alternativos.

CRÍTICA DE LA TCT

La principal ventaja del la TCT es su simplicidad, lo que hace que sea el modelo más utilizado en Ciencias Sociales. Sin embargo, el modelo TCT ha sido puesto en duda por distintos autores (Hambleton et al, 1991) debido a que:

- 1) Puntuaciones de los ítems depende de la muestra utilizada en el instrumento de medida.
- 2) Las estimaciones del concepto analizado depende de la elección del conjunto específico de ítems incluidos en el instrumento de medida,

Ello lleva a estos autores a concluir que las características de los encuestados y del test no pueden separarse. Cada uno de estos conjuntos sólo puede ser interpretado en el contexto del otro. (p. 2)

Gran parte de las críticas se centran en la calidad de las medidas, al no poderse asegurar la validez estadística de las mismas. Lord (1984) estableció que el problema está en el error estándar de la medición que es una función de la fiabilidad y de la varianza [5]. Se asume que es la misma para todos los sujetos encuestados, pero las puntuaciones de cualquier test son medidas no igualmente precisas para encuestados con distinto nivel en el atributo/constructo medido. Por lo que la asunción de igual error de medida para todos los encuestados es inverosímil.

Engelhard (2001) también centra sus críticas en la calidad de las medidas obtenidas al indicar que:

- La fiabilidad y validez del tests están correlacionados de forma inversa : Paradoja de la atenuación (el aumento de la consistencia interna puede tener efecto sobre la reducción de la validez del constructo) Loevinger (1957).
- No es posible asegurar directamente la validez de un test.

A la vez que corrobora que el principal problema, desde la perspectiva de una teoría de la medida es que el TCT conduce a una estimación de parámetros que no son independientes de las bases de datos particulares (muestras) que han sido utilizados para el diseño del test.

En conclusión, la fiabilidad que se obtiene en la TCT es sensible al método utilizado, ya que en cada caso son diferentes los factores que les pueden afectar. Por lo que no se puede esperar que la aplicación de los diferentes criterios determine el mismo nivel de fiabilidad de la medida.

La fiabilidad es una propiedad de las puntuaciones de una medición más que de la medición en sí misma, por lo que es dependiente de la muestra utilizada. La fiabilidad de un

instrumento aplicado a dos muestras diferentes será distinta, ya que la puntuación verdadera de la segunda población es diferente de la primera.

El análisis de las propiedades psicométricas, cuya invarianza de los valores o índices numéricos llamados medidas en el marco de la TCT reposa fundamentalmente en la interpretación de los coeficientes de correlación. Los coeficientes dependen de la variabilidad de la muestra en relación con el error de medida. Se debe contar siempre con las características de los sujetos que hayan sido consultados.

Se puede concluir que las medidas solo tienen sentido si se interpretan en el contexto de la muestra y del cuestionario utilizado.

Wright y Stone (1979), Lord (1980), Hambleton (1983), Hambleton y Swaminathan (1985) han planteado las características deseables de una alternativa a la TCT que serían (Hambleton et al, 1991):

- 1.- Características de los ítems que no sean dependiente de la muestra utilizada.
- 2.- Puntuaciones de las percepciones de los encuestados respecto al constructo analizado que no sea dependiente del instrumento de medida utilizado.
- 3.- Modelo de medición que se exprese a nivel de ítem y no de instrumento de medida
- 4.- Modelo que no exija un estricto test paralelo para valorar la fiabilidad
5. Un modelo que proporcione una medida de precisión por cada puntuación del nivel existente del constructo analizado

TEORÍA DE LA GENERABILIDAD

A partir de los planteamientos de la TCT, Cronbach et al (1963; 1972) profundizan en el análisis de las fuentes de los errores que pueden cambiar el significado del coeficiente de fiabilidad (Gleser et al, 1965) desarrollando la Teoría de la Generabilizabilidad. En esta teoría se usa el análisis de la varianza para destacar las distancias entre las distintas fuentes de error que se presentan en las puntuaciones obtenidas.

La Teoría de la Generabilizabilidad estudia las condiciones de la medición dentro de un universo de posibles condiciones de medición. Peter (1979) especifica que una condición de medición es un aspecto específico del procedimiento de medición. El término utilizado para designarlo es “una faceta de la medición”, tales como tiempo, instrumento u observadores.

Cronbach et al (1972) destacan la influencia que múltiples facetas tienen en el concepto de error. De esta forma, se refieren a cada una de las situaciones características de la medición que al variar condicionan las puntuaciones obtenidas.

Las fuentes de las facetas son (los orígenes se discriminan mediante ANOVA):

PRIMERA FUENTE: Diferencia de la presencia de los atributos a medir en los encuestados (siendo el objeto de la medición la habilidad de los encuestados).

SEGUNDA FUENTE: Dificultades de los ítemes que representan los atributos a medir.

TERCERA FUENTE: Interacción de las diferencias de los sujetos encuestados con los ítemes.

CUARTA FUENTE: Errores aleatorios (fuentes no identificadas)

Al formular los problemas de la fiabilidad en términos de generalizabilidad se reconoce que los procedimientos de medición pueden generar errores en la varianza. Aún el caso de que la fiabilidad se obtenga por los medios tradicionales, se está considerando una faceta de medición y un tipo de error de medición de forma explícita (Peter, 1979), por lo que un alto nivel de fiabilidad de un tipo no puede interpretarse como que el procedimiento de medición genera resultados consistente entre todas las posible facetas.

1.2.3.- LAS TEORÍAS DEL RASGO LATENTE

El proceso de revisión teórica de la TCT y los desarrollos de diferentes supuestos relativos a la medición de conceptos que no son directamente observables, a partir de los análisis de Lazarsfeld (1977) han posibilitado los análisis de Rasgos Latentes.

Los Análisis de Rasgos Latentes son procedimientos diseñados para medir rasgos, atributos, constructos, variables no observables directamente que puedan explicar los rendimientos observados en la aplicación de los instrumentos de medida. El supuesto central de estos análisis es la unidimensionalidad.

Con la unidimensionalidad se asume que en los procesos de medición, los atributos de los sujetos pueden aislarse y por ello medirse, sin que ningún otro aspecto le afecte. De la misma forma se asume que los instrumentos de medida que se construyen para medir esos atributos aislados solo son sensibles al mismo, sin que puedan verse influidos por ninguna otra circunstancia. Como indica Ryan (1983) estas características delimitan la asunción operativa de la unidimensionalidad, no siendo privativa de la medición psicológica, sino que está presente en las mediciones físicas (peso, estatura).

La unidimensionalidad es un objetivo que no siempre se logra. Su violación está generada por los sujetos, los ítemes del test o la propia administración del test. Por tanto, si un constructo es unidimensional dependerá del propósito de la medición. Los tipos de ítemes indican el nivel de precisión al que se aspira (Andrich y Marais, 2005). Como indican estos autores el uso de las puntuaciones totales de las encuestas implica una asunción de un único rasgo subyacente que el test está midiendo. Los índices de consistencia interna proporcionan una indicación directa de si hay una única dimensión. Si son bajos, puede haber una única dimensión pero se percibe medidas con un alto nivel de error o la presencia de otras dimensiones a los que los ítemes reflejan en diferente grado. La validación del constructo se centra en la identificación de la dimensión en un sentido sustantivo. En el

desarrollo del instrumento de medida tiene que estar claro a que constructo se está representando.

Los análisis de Rasgo Latente se centran en la interacción de los sujetos con los ítemes y los modelos estadísticos que recogen esta relación que se expresan gráficamente por las curvas características del ítem.

Hay dos grandes aproximaciones en la Teoría del Rasgo Latente. La diferencia entre ambas se refleja en el tipo de relación que se establece entre los datos y el modelo estadístico utilizado.

La primera de las aproximaciones se agrupa alrededor de las Teorías de la Respuesta al Ítem (TRI). En la TRI se trata de buscar un modelo que mejor se ajuste a los datos disponibles mediante la parametrización del rasgo latente y de las propiedades de los ítemes. Se les podrían considerar modelos de contraste. Como indican Thissen y Orlando (2001) los modelos de esta teoría miden lo que son los ítemes no lo que debería ser.

La aproximación alternativa, los modelos de Rasch (que tradicionalmente se viene considerando como un caso particular de los TRI, el que corresponde al modelo de un parámetro), se parte de un modelo determinado que dispone de unas propiedades específicas a las que los datos obtenidos se ajustan o no. Estos modelos se denominan de ajuste. Del análisis de los desajustes se obtiene la información necesaria para determinar las calibraciones de ítemes y las medidas convenientes en cada caso.

1.2.4.- LA TEORÍA DE LA RESPUESTA AL ÍTEM

La Teoría de la Respuesta al Ítem (Lord, 1980) es una de las modernas teorías de test más populares, dado el alto índice de penetración en el sector de los tests educativos y en algunos sectores de Ciencias de la Salud, ya que proporciona justificaciones teóricas a aspectos deficientemente cubiertos por la Teoría Clásica del Test. No obstante, en el resto de las ciencias sociales se emplea aún poco, como destaca Cortada (2004).

Esta Teoría propone un modelo probabilístico que permite conocer la información proporcionada por cada ítem. De acuerdo con Hambleton et al (1991) el comportamiento de un encuestado ante un ítem de un test puede explicarse por el rasgo latente. Las relaciones entre la respuesta al ítem y el rasgo latente se puede describir mediante una función denominada: Curva Característica del Ítem (CCI).

Supuestos básicos.

Se consideran hipótesis operativas de trabajo que serán validadas al relacionar los datos con los modelos TRI utilizados, determinando si son apropiados o no para explicar los datos Hambleton et al (1991):

1.- **Unidimensionalidad** del rasgo latente. En un modelo TRI sólo se mide un rasgo latente por un conjunto de ítemes del test.

El cumplimiento de este supuesto puede no ser exacto debido a factores tanto cognitivos, como de personalidad, así como relacionados a la propia administración del test. Si existe un componente dominante que influya en las respuestas estaremos frente a un rasgo latente, en ese caso se cumple el supuesto de unidimensionalidad del modelo TRI. En caso contrario estaremos en presencia de más de un rasgo latente (multidimensionalidad).

2.- **Independencia local.** Las respuestas a cualquier par de ítems son independientes entre sí. Ello implica que será sólo el nivel del rasgo latente del encuestado el que influye en sus respuestas a los ítems del test.

3.- La relación entre la variable latente y las respuestas observadas al ítem, denominada **Curva Característica del Ítem (CCI)**, se describe como una función monótona creciente (Si el nivel de rasgo latente disponible por el encuestado aumenta la probabilidad de respuesta correcta—en el caso de ítems dicotómicos— al ítem se incrementa).

Cuando el modelo es el apropiado para el conjunto de datos disponibles, verificando las hipótesis planteadas, se obtienen una serie de características deseadas (Hambleton et al, 1991):

- Las estimaciones del rasgo latente del encuestado no depende del test.
- Las estimaciones de los parámetros de los ítems no depende de la muestra de encuestados utilizada.
- Los rasgos latentes estimados obtenidos de diferentes conjuntos de ítems serán las mismas (si no se consideran los errores de medición).
- Los parámetros de los ítems estimados obtenidos de diferentes muestras serán los mismos (si no se consideran los errores de medición).

La gran diferencia con respecto a la TCT es que la TRI proporciona estimaciones del error estándar para los rasgos individuales estimados, no una estimación del error para todos los encuestados.

Propiedad de **Invarianza** de los parámetros del modelo.

Se supone que los parámetros que caracterizan un ítem no dependen de la distribución del rasgo latente de los encuestados. El parámetro que caracteriza a un encuestado no depende del conjunto de ítems del test. Se obtiene cuando el ajuste de modelo y datos es exacto en la población utilizada (Hambleton et al. 1991).

Modelos TRI

Partiendo del rasgo latente θ , los tipos de modelos de la TRI se clasifican en función de los parámetros que recogen, para respuestas dicotómicas a los ítems de los instrumentos de medida utilizados, así:

Modelo de un parámetro. Dificultad del ítem (b_i). Tradicionalmente se considerando al modelo de Rasch como un modelo TRI de un parámetro. No obstante, se pueden destacar

algunas diferencias entre ambos (RMT 2005) respecto a la motivación, sujetos, ítems, probabilidad, formulación, origen local de la escala, discriminación del ítem, y evaluación del ajuste, entre otros aspectos.

Modelo de dos parámetros: Tienen en cuenta la discriminación o pendiente de la CCI (a).

Modelo de tres parámetros: Tienen en cuenta el factor azar en la respuesta a ítems de alternativas múltiples (c) .

En el caso del modelo generalizado de TRI (tres parámetros), la expresión de la probabilidad de respuesta afirmativa a un ítem (dicotómico) sería:

$$P_i(\theta) = c_i + (1 - c_i) \left\{ \frac{\exp[1.7 a_i (\theta - b_i)]}{1 + \exp[1.7 a_i (\theta - b_i)]} \right\}$$

En donde:

θ : Rasgo latente: capacidad, habilidad, atributo, ..., del sujeto encuestado.

$P_i(\theta)$: Probabilidad de que un encuestado elegido al azar con habilidad θ conteste correctamente el ítem i

b_i : Parámetro de dificultad del ítem i

a_i : Parámetro de discriminación del ítem i

1.7: Valor arbitrario introducido para que la función logística se ajuste a la normal acumulada con exactitud 0.01

exp: Base del logaritmo neperiano: 2.178

c_i : Parámetro de pseudo azar: probabilidad de contestar correctamente un ítem de alternativas múltiple por azar (adivinar).

Para todo $i \in \{1, 2, \dots, n\}$ (número de ítems)

Modelo de un parámetro (1-PL TRI) (Birnbbaum, 1968)

$$P_i(\theta) = \frac{\exp[1.7 (\theta - b_i)]}{1 + \exp[1.7 (\theta - b_i)]}$$

El significado de los parámetros son los indicados en el caso general.

Tradicionalmente, se ha venido considerando a los modelos de Rasch incluidos en esta categoría, aunque presenten alguna diferencia matemática en sus formulaciones.

$P_i(\theta)$ es una curva logística cuyos valores varían entre 0 y 1 en la escala del rasgo latente. Las CCI han sido ajustadas para ser paralelas con una pendiente de 1.7, aproximadamente la pendiente de la ojiva normal acumulativa (RMT, 2005).

El parámetro b del ítem i es el punto de la escala del rasgo latente donde la probabilidad de respuesta correcta al ítem es 0.5 (la ubicación del ítem en el continuo lineal del rasgo latente).

Los ítems más difíciles se sitúan a la derecha del continuo lineal, mientras que los fáciles se sitúan a la izquierda. La asíntota inferior de la CCI es 0, que indica que los encuestados con muy bajo nivel de rasgo latente tienen una probabilidad 0 de contestar correctamente dicho ítem.

Modelo de dos parámetro (2-PL TRI) (Birnbaum, 1968)

$$P_i(\theta) = [\exp 1.7 a_i (\theta - b_i)] / [1 + \{ \exp 1.7 a_i (\theta - b_i) \}]$$

El significado de los parámetros son los indicados en el caso general.

El parámetro de discriminación del ítem a_i es proporcional a la pendiente de la CCI en el punto b_i de la escala del rasgo latente. Los ítems con pendientes más abruptas son más útiles para separar a los encuestados en diferentes niveles del rasgo latente que los ítems con pendientes más aplanadas (Hambleton et al. 1991).

Las observaciones que destaca Ryan (1983) respecto a este modelo se refiere a que el parámetro de discriminación incluido en el mismo no se usa para destacar la característica latente subyacente de los ítems del test, sino para proporcionar una más detallada descripción de las respuestas observadas a los ítems. Su inclusión hace la interpretación de las dificultades de los ítems ambigua.

Modelo de tres parámetro (3-PL TRI) (Birnbaum, 1968)

$$P_i(\theta) = c_i + (1 - c_i) \left\{ \frac{[\exp 1.7 a_i (\theta - b_i)]}{[1 + \{ \exp 1.7 a_i (\theta - b_i) \}]} \right\}$$

El significado de los parámetros son los indicados en el caso general.

El parámetro c_i proporciona una posible asíntota no nula para la CCI y representa la probabilidad de respuesta correcta al ítem de los encuestados de bajo nivel del rasgo latente. Se incorpora para representar el comportamiento al más bajo nivel del continuo de la variable latente, en donde la “adivinación” o pseudo azar puede ser un factor significativo ante los tests de respuestas múltiples (Hambleton et al. 1991).

Siguiendo a Ryan (1983) se podría destacar que este modelo asume que el encuestado al responder al ítem usa dos estrategias alternativas, su propio rasgo latente y la “adivinación”. Es difícil ver cómo un modelo que se acomoda a esas dos estrategias para contestar un ítem puede usarse para medir sujetos con un único rasgo latente unidimensional. También parece impropio modular la “adivinación” como una característica de los ítems del test, cuando “adivinar” parece más propio de los encuestados. Sin embargo, “adivinar” no parece que sea una característica ni de los ítems ni de los sujetos, ya que serán pocos los ítems que puedan ser adivinados por todos los encuestados y pocos los sujetos que adivinen todos los ítems.

Junto a los modelos dicotómicos planteados se puede destacar que también se han desarrollado diversos modelos politómicos como son el Modelos de Respuestas Ordenadas de Samejina (1969; 1997) que viene a recoger las posibilidades tradicionales de las alternativas de la escala de Likert y el Modelo Nominal de Bock (1972), que es una alternativa al anterior para ítems politómicos en donde no se exige previamente ninguna especificación del orden de las categorías mutuamente excluyentes del rasgo latente.

La estimación de los parámetros del modelo TRI, se puede realizar mediante métodos de máxima verosimilitud. Estos modelos permiten construir cuadros de valores estimados:

P_{ij}	δ_1	δ_2	δ_j	δ_{11}	$R_i = \sum P .$
β_1	P_{11}	P_{12}	P_{1j}	P_{111}	R1
β_2	P_{21}	P_{22}	P_{2j}	P_{211}	R2
β_i	P_{i1}	P_{i2}	P_{ij}	P_{i11}	Ri
β_{15}	P_{151}	P_{152}	P_{15j}	P_{1511}	R15
$S_{.j} = \sum P_{.j}$	S_1	S_2	S_j	S_{11}	

Así sería posible verificar la compatibilidad entre la matriz de los resultados observados y las estimados por el modelo y cuantificar la compatibilidad por un test X^2 para obtener un índice de ajuste (Leplège, 2003).

Siguiendo a Hampleton et al. (1991), la valoración del ajuste de los datos y el modelo se debe basar entre tipos de evidencias:

- 1) Validez de los supuestos del modelo para los datos del test.
- 2) Alcance sobre qué propiedades del modelo (por ej. invarianza de los parámetros de los ítems de los encuestados) se obtendrán.
- 3) Precisión de las predicciones del modelo cuando se usan datos reales (o simulados).

Será el estudio de los residuales estandarizados los que proporcionan información sobre el ajuste, mediante la aplicación de la Chi cuadrado.

Otras características destacables de los modelos TRI serían la posibilidad de obtener la Curva Característica del Test CCT y la función de información. La CCT recoge las sumas de las probabilidades de respuestas correctas a los ítems en el test, esta curva representa gráficamente el número esperado de ítems asumidos como una función del rasgo latente. Mediante una simple transformación de este valor esperado (dividiendo la suma de probabilidades por el número de ítems) se puede determinar la proporción correcta esperada como una función del rasgo latente (Weiss, 1995).

La función de información es un índice indicativo del nivel del rasgo latente en el cual un ítem es más útil para distinguir sujetos encuestados. Caracteriza la precisión de la medición de los sujetos encuestados a los diferentes niveles del rasgo latente. La curva de información del test se determina a partir de la suma de toda la información proveniente de los ítems del test (Lord, 1980)

Dos aspectos deben ser destacados entre la amplia variedad de posibilidades de los modelos TRI en relación con los TCT. Por una parte los procesos de equiparación de puntuaciones de los tests que miden el mismo rasgo latente y, por otro los Test adaptativos informatizados que proporcionan la medición más precisa de un encuestado cuando la dificultad del test se ajusta a la habilidad del encuestado.

Por otra parte, aunque gran parte del acervo científico común procedente del uso de variables latentes está presente en los modelos de TRI, hay una pronunciada crítica a estos modelos (véase por ej. Wright, 1992a y 1992b; Ryan, 1983; RMT, 2005). Se destaca que son esencialmente diferentes al utilizar procedimientos distintos. Mientras que los modelos de rasgo latente se refieren a la medición de las características subyacentes o latentes de los sujetos, los de respuesta al ítem describen y resumen estadísticamente las puntuaciones observadas (Ryan, 1983).

2.- EL MODELO DE RASCH

2.1.- ELEMENTOS DEL MODELO DE RASCH

INTRODUCCIÓN

George Rasch (1901-1980) fue un matemático danés que trabajó en el campo de las Ciencias de la Educación. Desarrolló un método estadístico que permitió la medida de los resultados educativos de los estudiantes, mediante el diseño racional de exámenes. Identificó una familia de modelos de medida probabilística que completó las perspectivas TCT. Los investigadores vienen considerando al Modelo de Rasch como un modelo que destaca por sus características matemáticas (Andersen, 1973; Fischer, 1973 y Wright, 1977). La característica estadística que distingue al modelo de Rasch es que los parámetros personas e ítems están algebraicamente separados y dan lugar a estadísticos suficientes (Andersen, 1973; Masters y Wright, 1984, Rasch, 1972).

Rasch inició sus investigaciones con el estudio de la distribución de los errores (de lectura de los encuestados) en observaciones en Dinamarca tras la segunda Guerra Mundial. Si el test había sido bien elegido, surgirían pocos errores, por lo que era razonable representar la distribución de los errores de lectura por una función de probabilidad de Poisson. (multiplicativa). La elección de este modelo de Poisson multiplicativo, se debía a las propiedades que le cualifican como un modelo de medida (Lord, 1953). El análisis de las propiedades que debían cumplir las medidas, le llevan a Rasch al uso de modelos aditivos exponenciales (modelos de medida): pasando de los modelos de Poisson multiplicativo al modelo logístico.

Rasch planteó su modelo dentro de la denominada Teoría de la medición conjunta, que parte de la consideración de que las medidas (observables) fundamentales tienen estructura aditiva.

Perline et al. (1979) han demostrado que los modelos pertenecientes a la familia identificada por Rasch son los únicos modelos compatibles con los principios de la teoría de la medida conjunta en el caso probabilístico.

De esta forma se determina el carácter crucial de los modelos de medida de Rasch para la metodología de la medida en las Ciencias Sociales (Cliff, 1992 y Mitchell, 1999). Los modelos de Rasch han sido utilizados en psicología, educación, economía aplicada, salud pública,..., para el desarrollo de instrumentos de medida o de identificación de continuos de medida en el conjunto de datos ya disponibles.

CARACTERÍSTICAS

Unidimensionalidad

Se asume que los ítems del instrumento de medida reflejan sólo un rasgo latente subyacente. Planteada la unidimensionalidad como una hipótesis operativa en el proceso de ajuste de los datos al modelo, la obtención del correspondiente índice de consistencia interna nos indicará si se valida esta hipótesis. La validez del constructo está vinculada a la identificación de la unidimensionalidad en un sentido sustantivo (Andrich y Marais, 2005).

En el modelo de Rasch, como modelo de rasgo latente, se considera que un constructo unidimensional permitirá la localización de los ítems en función del parámetro de los ítems (δ) y los sujetos encuestados en función del parámetro de su rasgo latente (β).

Suficiencia de la puntuación total

Rasch (1980) sigue el planteamiento de Fisher (1920) sobre estadísticos suficiente y establece que la puntuación total tiene toda la información estadística necesaria para determinar el parámetro del sujeto. Andrich y Marais (2005) indican que dada la puntuación total de los sujetos, la pauta de las respuestas no contiene ninguna ulterior información sobre el parámetro de las personas (su nivel del rasgo latente analizado).

Objetividad específica

Rasch (1980) utiliza este concepto para referirse a un aspecto esencial en la medición, como es que las posibles comparaciones entre sujetos sean independientes del instrumento de medida utilizado. De igual forma, debe ser posible comparar los instrumentos de medida independientemente de a qué sujetos se les aplica.

Esta idea ya había sido anticipada por Thurstone y Loevinger, pero fue Rasch quien lo estableció como requisito central de su modelo (Stenner, 1990).

Stenner (1994) destaca la diferencia entre la objetividad absoluta, más conocida por su presencia en las ciencias físicas y químicas y la objetividad específica desarrollada por Rasch. Indicando que en este caso la objetividad específica es una consecuencia de un conjunto de datos que se ajustan al modelo de Rasch. En ese caso la localización relativa de los parámetros de los sujetos y de los ítems en el continuo lineal del constructo son independientes de sus muestras. Sin embargo, las medidas absolutas se pueden obtener indirectamente introduciendo ítems o personas de referencia en el análisis.

La ventaja del modelo de Rasch es que esta propiedad de invarianza está en su propia estructura (Andrich y Marais, 2005), conformando una hipótesis verificable. Este aspecto ha dado lugar a tres interesantes desarrollos, como son el análisis del funcionamiento diferencial del ítem, la conformación de bancos de ítems a partir de la equiparación de medidas y la posibilidad de establecer test ajustados a los sujetos mediante procesos computerizados.

Ventaja respecto al uso de muestras

Las medidas del modelo de Rasch son estables aún cuando se usen muestras pequeñas (Linacre, 1994 y 2005; Fisher, 1997 y 2005). Fisher (1997) destaca cómo los teóricos de TRI Lord (1983), Hambleton y Cook (1977) también admiten este aspecto. Barnes y Wise (1991) han analizado un modelo modificado de Rasch (en su vertiente de 1-PL TRI) confirmando la preferencia de elección cuando el número de datos es limitado. Mientras que Wright y Stone (1979) sugieren tamaños mínimos para la longitud del test y de la muestra.

2.2.- FAMILIAS DE MODELOS DE RASCH

El modelo determinado por Rasch (1960/1980) recoge la probabilidad de un sujeto de asumir una categoría determinada dentro de un ítem como función del nivel del rasgo latente del sujeto (habilidad/capacidad,...) y la dificultad de los ítems.

Los principios de su modelo son:

- Los resultados (manifestación de la habilidad/capacidad del sujeto encuestado) se consideran una construcción abstracta (teoría del rasgo latente) situada a lo largo de un eje horizontal (continuo lineal)
- Los resultados se miden por medio de ítems, cuestiones o pruebas.
- Los ítems, cuestiones y pruebas pueden ser de respuesta simple, del tipo dicotómico (sí/no), o de respuestas múltiples.
- En cada cuestión o prueba se aplica una ecuación exponencial cuya variable es la diferencia entre el resultado del test y la dificultad de la cuestión o prueba.
- La curva representativa de la ecuación permite ubicar adecuadamente las cuestiones o pruebas a lo largo del eje, de acuerdo a su nivel de dificultad (de más fácil a más difícil).

Se puede obtener una serie de curvas paralelas que permiten cuantificar los resultados de los diferentes sujetos, de más al menos hábil, con las posibilidades predictivas que ello comporta.

EL MODELO DICOTÓMICO DE RASCH: ELEMENTOS BÁSICOS

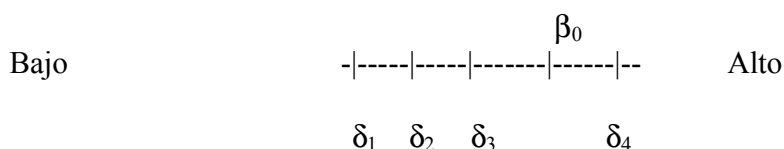
En el modelo de Rasch (1980) se considera una variable latente en la cual dos diferentes entidades interactúan. Por una parte las unidades de investigación por medio de los sujetos económicos encuestados y por otra los ítems del instrumento de medida diseñado para llevar a cabo la investigación: sujetos encuestados e ítems.

El objetivo del modelo es ubicar a encuestados e ítems en una escala simple representativa de la variable latente. En contraste con la TCT los sujetos encuestados y los ítems se ubican en una misma dimensión, el continuo lineal que representa la dificultad de los

ítemes en el contexto del constructo diseñado. En este continuo lineal hay una única dirección que implica una mayor presencia “más” de la variable latente (desde niveles bajos a altos).

Se definen dos parámetros β : encuestados y δ : ítemes que pueden situarse de forma conjunta en la misma escala lineal representativa de la variable latente. El término “más” significa que el parámetro en cuestión se encuentra ubicado a una mayor distancia a lo largo del continuo lineal.

Sea, por ejemplo, el caso de la ubicación del encuestado β_0 y los ítemes $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ en la escala lineal representativa de la variable latente. Los extremos de esta escala se establecen como: bajo y alto.



Los ítemes $\delta_1, \delta_2, \delta_3$ están más próximos al extremo definido como bajo que β_0 y el ítem δ_4 que están más próximos al extremo determinado como alto.

DETERMINACIÓN DE LA EXPRESIÓN MATEMÁTICA

Se define X_{ni} como la variable latente que representa la respuesta de un encuestado a un ítem del instrumento de medida. La puntuación obtenida de acuerdo, por ejemplo, a la utilización de una escala dicotómica 0 / 1, se puede expresar como $X_{ni} = \{0,1\}$ en términos tanto del parámetro β_n (encuestado n) y δ_i (ítem i del instrumento de medida utilizado).

Si se interpreta la posición de los sujetos en términos de probabilidad se podrían considerar las siguientes posibilidades: Si un encuestado se compara con un ítem que se encuentra en un nivel inferior en la escala, entonces podemos indicar que el encuestado β_n en relación con el ítem δ_i se sitúa en la relación $\beta_n > \delta_i$, es decir que $(\beta_n - \delta_i) > 0$, entonces se espera que el encuestado tenga una alta probabilidad de asumir el concepto representado por el ítem. Que puede expresarse como que la probabilidad de contestar afirmativamente al ítem es superior a 0.5

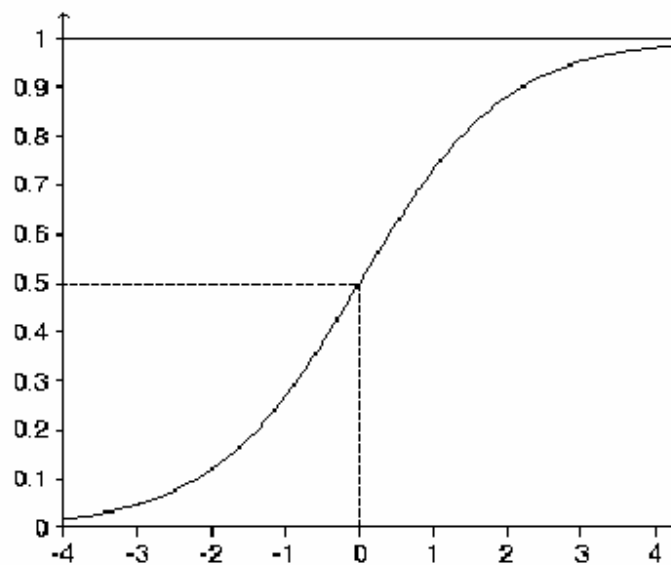
De igual forma podríamos indicar que si un encuestado se encuentra situado en la escala lineal representativa de la variable latente por debajo de un ítem su probabilidad de respuesta afirmativa es inferior del 0.5

Siendo $P [X_{ni} = 1]$ la probabilidad de contestar afirmativamente por el encuestado β_n al ítem δ_i , se puede expresar que:

- Si $\beta_n > \delta_i$, $(\beta_n - \delta_i) > 0$, entonces $P [X_{ni} = 1] > 0.5$
 Si $\beta_n < \delta_i$, $(\beta_n - \delta_i) < 0$, entonces $P [X_{ni} = 1] < 0.5$
 Si $\beta_n = \delta_i$, $(\beta_n - \delta_i) = 0$, entonces $P [X_{ni} = 1] = 0.5$

GRÁFICA DE LA PROBABILIDAD DE CONTESTAR AFIRMATIVAMENTE UN ENCUESTADO $P [X_{ni}]$ EN FUNCIÓN DE LA DIFERENCIA ENTRE PARÁMETROS $\beta_n - \delta_i$

$P [X_{ni}]$



$\beta_n - \delta_i$

Fuente: Elaboración propia

Representando por P ($P [X_{ni} = 1]$) la probabilidad de contestar afirmativamente un ítem, entonces $1-P$ representa la probabilidad contraria.

Una formulación que recoge la expresión de la probabilidad de los datos es la razón $P / 1-P$. Es conveniente formular las probabilidades como funciones de la diferencia de los encuestados y los ítems. El logaritmo neperiano de las probabilidades se puede establecer como la diferencia entre las posiciones del encuestado y el ítem como se expresa en la ecuación siguiente:

$$L [P / (1-P)] = \beta - \delta \quad (1)$$

Estas diferencias entre sujetos e ítems están expresadas en unidades de medida denominadas logits, dado que su origen es el logaritmo de la razón de desacuerdo (acierto/error) de la probabilidad de acertar correctamente un ítem por el sujeto encuestado. El logit, como unidad de medida, se obtiene por el uso del Modelo de Rasch que transforma las puntuaciones brutas procedentes de los datos ordinales en una razón logarítmica que posteriormente se puede pasar a una escala de intervalo.

La diferencia $(\beta_n - \delta_i)$ varía entre $-\infty$ y $+\infty$, mientras que la probabilidad $P [X_{ni}=1]$ entre 0 y 1, es decir que:

$$-\infty \leq (\beta_n - \delta_i) \leq \infty$$

$$0 \leq P [X_{ni}=1] \leq 1$$

La ecuación (1) se transforma en:

$$P / (1-P) = \exp (\beta - \delta) \quad (2)$$

Cuyo campo de variación es:

$$0 \leq \exp (\beta - \delta) \leq \infty$$

Despejando P, en la ecuación (2) se obtiene:

$$P[X_{ni}=1 | \beta_n, \delta_i] = \exp (\beta - \delta) / [1 + \exp(\beta - \delta)] \quad (3)$$

Que es el modelo obtenido por Rasch (Rasch, 1960), cuyo campo de variación es:

$$0 \leq \{ \exp (\beta - \delta) / [1 + \exp(\beta - \delta)] \} \leq 1$$

Los parámetros β_n y δ_i se estiman mediante el método de máxima verosimilitud en el caso del programa de computación Winsteps / Ministep (Linacre, 2005), usado en los estudios que se presentan en este documento de trabajo.

MODELO DE CRÉDITO PARCIAL (MASTERS)

Master (1982) y Wright y Master (1982) desarrollaron un modelo que comparte las características de la familia de modelos de Rasch: La puntuación total de los encuestados como estadístico suficiente para la medición del nivel de rasgo latente en el continuo lineal y la objetividad específica.

El modelo se establece a partir de la determinación de la probabilidad condicional de estar de acuerdo con una de las categorías de un ítem politómico en lugar de la anterior, que es monótona creciente a lo largo de la variable latente.

La función de respuesta

$$\Pr\{X_{vi} = x\} = \frac{1}{\gamma} \exp \sum_{k=0}^x (\beta_v - \delta_{ik})$$

Siendo:

v: sujeto encuestado

i: ítem contestado

k: categoría en el ítem i asumida por el encuestado.

γ = Suma de todos los posibles numeradores, que surgen de acuerdo con el número de categorías de los ítems,

$$\gamma = \sum_{h=0}^{mv} \exp \sum_{k=0}^x (\beta_v - \delta_{ik})$$

$x \in \{0, 1, \dots, m\}$

MODELO DE RASCH POLITÓMICO

El desarrollo de este modelo de Rasch para ítems politómicos se realizó a partir de los trabajos de Rasch (1961), Andersen (1977) y Andrich (1978 a y b). Andrich partió del modelo de Rasch para ítems dicotómicos y procedió a su generalización.

La probabilidad de asumir una determinada categoría del ítem i por parte del sujeto v sería:

$$\Pr\{X_{vi} = x\} = \frac{1}{\gamma} \exp [-\sum_{k=1}^x \tau_{ki} + x(\beta_v - \delta_i)]$$

Siendo:

v: sujeto encuestado

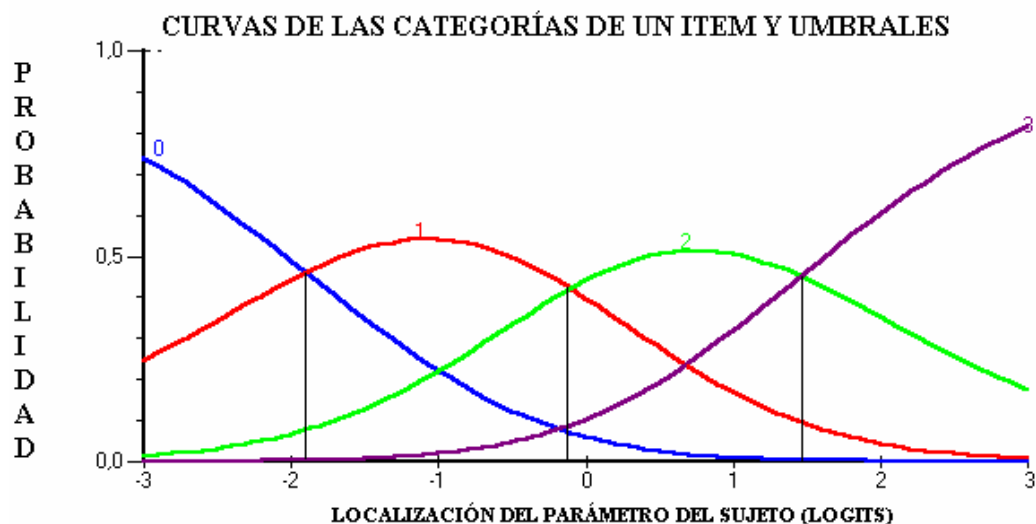
i: ítem contestado

γ = Suma de todos los posibles numeradores, que surgen de acuerdo con el número de categorías de los ítems.

En este proceso surgen los umbrales (τ) o puntos de cortes de las curvas características correspondientes a las distintas categorías de respuestas en los ítems (Andrich y Marais, 2005).

Los umbrales son unos parámetros que recogen los puntos en donde la probabilidad de respuesta de una categoría con la siguiente es equiprobable. En el caso de los ítems

dicotómicos sólo había un umbral, que era la dificultad del ítem, en donde coincidían la probabilidad de 0 ó 1. El número de umbrales es igual al número de categorías menos uno.



Fuente: Elaboración propia con el programa RUMM2020

El análisis del ajuste de los datos al modelo se realiza comparando los valores esperados y las proporciones observadas del número de respuestas en cada categorías de los ítems. En el programa de computación RUMM desarrollado por el proceso de ajuste se analiza a partir de test de Chi-cuadrado a los residuales que surgen en el ajuste.

MODELO DE LAS FACETAS MÚLTIPLES

Como indica Linacre (2002b), en el modelo de Rasch se enfrentan el nivel del rasgo latente que luce el sujeto (habilidad, competencia, capacidad, motivación,...) a la dificultad, facilidad, desafío,..., del ítem para producir una determinada puntuación. Los ítems y los sujetos son los elementos de una faceta: el enfrentamiento entre el sujeto y el ítem.

Si se da el caso de que en un par de sujetos, uno de ellos dispone de una ventaja, nos encontramos con una faceta más, siendo su expresión:

$$\log \left(\frac{P_{nm}}{P_{mn}} \right) = B_n - B_m + A_w$$

en donde, el sujeto n de habilidad B_n actúa en relación al sujeto m con la habilidad B_m y A_w es la ventaja que puede tener B_n .

Un caso típico de tres facetas nos la encontramos cuando los sujetos que se enfrentan a un ítem pasan a ser valorados por un tercero o juez. Éste juzga, con cierto grado de severidad o permisividad, la habilidad de un sujeto al enfrentarse a la dificultad de un ítem (como el

caso de una selección de personal). La puntuación que obtenga se puede deber a su habilidad, la dificultad del ítem y a la actitud del juez.

Estaríamos ante un modelo de cuatro facetas cuando un sujeto ejecuta una acción que es puntuada en los ítemes por un juez. Sería en el caso de Dirección estratégica, cuando el Consejo de Administración (juez) valora una estrategia formulada (acción) por el departamento de Marketing (sujeto), de acuerdo a los recursos y capacidades que sustenta una posible ventaja competitiva (ítemes).

$$\log\left(\frac{P_{nmijk}}{P_{nmij(k-1)}}\right) = B_n - A_m - D_i - C_j - F_{ik}$$

en donde, D_i es la dificultad del ítem i y,

F_{ik} es la estructura del ítem i , como por ejemplo un modelo de crédito parcial.

Las posibilidades de este modelo es amplia en Economía y Empresa, siendo desarrollado por Linacre (1989) y contando con un programa de computación propio como es el Minifac /Facets de de la empresa Winsteps.

OTROS DESARROLLOS

Los desarrollos de la familia de modelos de Rasch se puede encontrar en el trabajo de Fischer y Molenaar (1995).

CONTROL DE CALIDAD DEL AJUSTE DEL MODELO PARA ÍTEMES Y SUJETOS

FIABILIDAD

Andrich y Marais (2005) indican que el cálculo de índices de fiabilidad no es muy común en la Teoría del Rasgo Latente, aunque es posible calcular dichos índices. En el Modelo de Rasch utilizan índices de separación referidos a sujetos e ítemes.

El índice de separación de los sujetos es una estimación de la repetición de la ubicación de las personas a lo largo del continuo de la variable medida mediante otros ítemes que midan este mismo constructo (Wright y Masters, 1982). Andrich (1982) determinó su índice de separación de sujetos con una estructura similar al índice KR.20, pero mientras que éste se expresa en función tanto de la varianza de sujetos e ítemes, el índice de separación de Andrich se expresa únicamente en función de los parámetros asociados a los sujetos.

Como indican Bond y Fox (2001), la estimación se basa en el mismo concepto del índice de fiabilidad de TCT: el porcentaje de la varianza de respuestas observadas que se pueden repetir.

$$R_p = \frac{SA_p^2}{SD_p^2}$$

En donde,

SA_p^2 : representa la parte que se puede repetir de esta variabilidad en una nueva aplicación del cuestionario con distintos ítems que midan el mismo constructo.

SD_p^2 : representa la variabilidad total de las personas.

La cantidad de varianza que es repetible mediante el Modelo de Rasch se denomina variabilidad ajustada de las personas: SA_p^2 . Esta variabilidad se obtiene descontando la varianza del error (SE_p^2) al total de la varianza (SD_p^2), tal como se expresa a continuación:

$$SA_p^2 = SD_p^2 - SE_p^2$$

El índice de separación de ítems indica la repetición de la ordenación de los ítems colocados a lo largo del continuo si esos ítems fueran cumplimentados por otra muestra equiparable en sus características de habilidad a la anterior. Para un alto nivel de fiabilidad de ítems, se puede inferir que se ha desarrollado una ordenación jerárquica de los mismos en un continuo en donde algunos ítems tienen más dificultad y otros son más asequibles, por lo que se tiene confianza en estas inferencias (Bond y Fox, 2001). Su índice se determina de la misma manera que en el caso de los sujetos encuestados.

El índice de separación, tanto expresado en términos de las medidas de los parámetros de los sujetos y de los ítems, tiene un rango situado entre 0.00 y 1.00 (Wright y Masters, 1982). Cuanto más alto sea el valor obtenido, mejor es la separación existente y más precisa es la medición.

VALIDEZ DEL MODELO

Como indican Bond y Fox (2001), debido a que el Modelo de Rasch es una expresión matemática estricta de la relación teórica que podría existir entre todos los ítems y personas a lo largo del continuo del constructo, ningún ítem ni persona estará nunca perfectamente ajustado.

En el análisis de los ajustes se trata de determinar qué ítems y personas tienen pautas de respuestas que se desvían más de lo esperado por el Modelo de Rasch. El proceso se inicia con el cálculo de los residuales de las respuestas de cada persona e ítem (Wright y Masters, 1982).

$$Y_{ni} = X_{ni} - E_{ni}$$

En donde,

Y_{ni} : valor residual de la respuesta

X_{ni} : puntuación observada en la respuesta de la persona n al ítem i

E_{ni} : valor de la respuesta esperada por el Modelo de Rasch

Dado que habrá desviaciones en todas las personas e ítems (Y_{ni}), el diagnóstico de ajuste se resume en estadísticos de ajuste, que se expresa como un estadístico de los residuos medios estandarizados (elevados al cuadrado). A su vez, estos estadísticos se determinarán de acuerdo a dos enfoques diferenciados, representados por los índices OUTFIT e INFIT [Programa de computación Winsteps (Linacre, 2005)].

Siguiendo a Tristán (2001) se puede indicar que el estadístico OUTFIT correspondería al ajuste externo al ser sensible al comportamiento inesperado (de acuerdo con los planteamientos de Guttman) que afectan a ítems cuya dificultad está lejos del nivel de habilidad del sujeto. El estadístico OUTFIT es una estimación no ponderada del grado de ajuste de las respuestas a las estimaciones del Modelo de Rasch. Estos valores no ponderados tienden a sufrir la influencia de respuestas a ítems con gran dificultad por parte de personas y viceversa (Linacre, 2002)

El estadístico INFIT correspondería al ajuste interno (Tristan, 2001). Se trata de un valor sensible al comportamiento inesperado que afecta a ítems cuya dificultad está cerca del nivel de habilidad del sujeto encuestado. En este caso se ha ponderado el residuo estandarizado cuadrático para dar más valor a las observaciones objetivas. Este estadístico es más sensible a pautas de respuestas a ítems irregulares implícitas en personas y viceversa (Linacre, 2002a).

En ambos caso se expresa en dos formas (Linacre, 2002a):

- No estandarizado como media cuadrática: MNSQ
Es una transformación de los residuales generada por la diferencia entre la observación y la estimación del Modelo de Rasch, para una mejor interpretación. Se estiman mediante estadísticos Chi-cuadrados divididos por sus grados de libertad. Los valores son siempre positivos.

El valor esperado es 1.00, pero se viene admitiendo como aceptables valores entre 0.50 y 1.50.

Valores de esta media superiores a 1.50, pero inferiores a 2.0 indican desajuste que son improductivos para la construcción de la medida, pero utilizables.

Valores superiores a 2.0 indican desajustes que generan distorsiones para el sistema de medición. Puede estar causadas por una o dos observaciones.

Valores de esta media inferiores a 0.50 evidencian desajustes menos productivos para la medición, pero no generan distorsiones en la misma. Pueden llevar a conclusiones erróneas con coeficientes altos de fiabilidad y separabilidad.

- Estandarizado como test de hipótesis: ZSTD
Se presentan como valores normalizados z , con una distribución con media 0 y varianza 1. Usualmente se obtienen mediante la conversión del estadístico media cuadrática en una z -normalizada normalmente distribuida por medio de la transformación de Wilson-Hiferty (Linacre, 2002).

Los valores incluidos en el rango ± 1.9 están considerados como aceptables ($p < 0.05$)

Los valores superiores a $+2$ ($>+2$) están desajustados. Indican mayor variabilidad en los datos que los que predice el Modelo de Rasch, reflejando generalmente un mayor error de incertidumbre en los resultados.

Datos iguales o superiores a 3 están muy desajustados.

Los valores inferiores a -2 (<-2) están sobre ajustados. Indican menos variabilidad en los datos que los que predice el Modelo de Rasch, evidenciando generalmente dependencia de datos.

Un ejemplo de utilización de los criterios de ajuste para la calibración de los ítems de un cuestionario, antes de la determinación de los valores de las medidas de los sujetos, se puede encontrar en Montero y Oreja (2005 a), en donde analizan los ítems que delimitan el constructo “recursos culturales tangibles”, para determinar un cuestionario aplicable en el campo del marketing de turismo cultural. En el caso de la publicidad es interesante el trabajo de Salzberger (2004) en la determinación de una escala de medida de la actitud hacia la publicidad.

OTRAS APLICACIONES DE LOS MODELOS DE RASCH

Gran parte de las utilidades que presentamos se derivan de la validación o no, tanto total o parcial, de la hipótesis de invarianza del modelo, desarrollando el concepto de objetividad específica planteado por Rasch.

FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM (FDI / DIF)

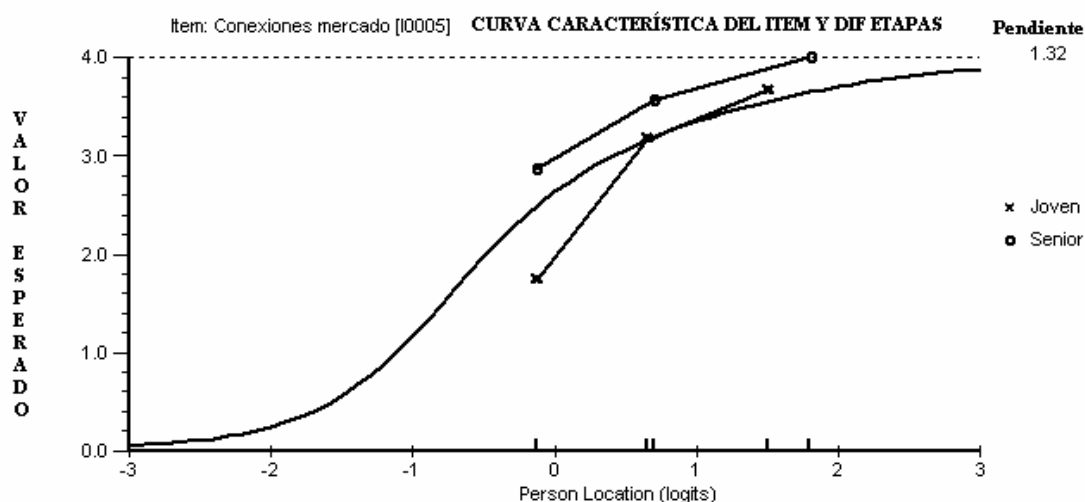
Es posible que a nivel del conjunto de datos se aprecie un ajuste global al modelos de los mismos permitiendo la calibración de los ítems del cuestionario y la medida de los sujetos encuestados. No obstante, puede que ese ajuste no sea el adecuado para determinados subconjuntos de la muestra de encuestados ante alguno de los ítems del instrumento de medida.

Esta clara reducción de la validez derivada del comportamiento diferencial de un grupo de encuestados ante unos determinados ítems permite al investigador profundizar tanto en las características determinantes de los grupos de encuestados con medidas diferentes como en los ítems que presentan ese desajuste, dado que sus puntuaciones nos están indicando una variedad de atributos diferentes de aquellos que la escala en cuestión trata de medir.

Andrich y Hagquist (2004) y Hagquist y Andrich (2004) han desarrollado un sistema, mediante el uso del programa de computación RUMM2020 (Andrich et al., 2004), para

determinar el alcance y contenido del posible funcionamiento diferencial de determinados ítems.

Plantean estos autores que hay dos aproximaciones para identificar el FDI/DIF. La primera se procedería a la estimación de un único conjunto de parámetros por cada ítem y posteriormente al estudio de los residuales identificados por los diferentes subconjuntos de la muestra de encuestados. Una segunda alternativa se centraría en estimar los parámetros de los diferentes grupos y luego compararlos entre sí. Nos centraremos en la primera alternativa que se ha desarrollado por Montero y Oreja (2006). En este caso se trataba de determinar si dos colectivos diferentes de artistas (jóvenes y senior) presentaban la misma actitud ante la acción comercial de los intermediarios (galerías de arte) en el mercado del arte. De esta forma se trataba de validar la hipótesis de invarianza del modelo de Rasch polítómico utilizado.



Fuente: Montero y Oreja (2005). Programa de computación utilizado RUMM2020.

El uso de las diferentes curvas características de los ítems para cada uno de los grupos en relación con el conjunto de la población y la utilización del análisis de la varianza entre los grupos determinó la existencia de un comportamiento diferencial de los grupos frente algunos ítems del cuestionario, (véase el caso del ítem “Conexiones con el mercado” en el gráfico: Curva característica del ítem y DIF etapas), generando la información necesaria para el desarrollo de acciones globales de marketing, diferenciadas para determinados ítems y determinados colectivos.

EQUIPARACIÓN DE ÍTEMES, ANCLAJES Y BANCOS DE ÍTEMES

La necesidad sentida de comparar medidas derivadas de instrumentos que recogen un determinado rasgo latente y entre diferentes poblaciones de encuestados a los cuales se les aplica estos instrumentos está constantemente presente en nuestros campos científicos. Sería una clara aplicación del principio de la objetividad específica y su evolución hacia la objetividad general (Stenner, 1994) a partir de la constatación de la hipótesis de invarianza del modelo de Rasch.

El objetivo del proceso de equiparación sería desarrollar un núcleo de medidas vinculadas a un banco de ítems que delimiten un determinado constructo de acuerdo a una métrica común. A partir de la calibración de los ítems dentro de esta métrica común se pueden crear y usar diferentes cuestionarios que reflejan el mismo constructo. Estos cuestionarios estarán conformados por distintos grupos de ítems que se seleccionan de acuerdo al colectivo a encuestar. De esta forma, encuestados que cumplieren diferentes cuestionarios pueden tener puntuaciones comparables de acuerdo a la misma métrica. Lo que nos permite procesos de bechmarking entre los sujetos económicos analizados.

El conjunto de ítems pueden compararse a partir de un proceso de anclaje. Este proceso implicaría la utilización de un conjunto de ítems que son comunes en los cuestionarios utilizados, siendo la base para el establecimiento de la métrica a partir de la cual los ítems no comunes de los diferentes instrumentos se pueden fijar en el continuo lineal del constructo analizado. Dada la complejidad del proceso se utilizan los programas de computación diseñados para tal efecto.

TEST ADAPTATIVOS INFORMATIZADOS (TAI /CAT)

Las posibilidades derivada de los modelos de Rasch, fundamentalmente del principio de objetividad específica, permite la selección por medio computarizado de los ítems más adecuados de cada banco de ítems del constructo analizado para determinar el nivel del rasgo latente existente en los encuestados, sin necesidad de responder a todas las cuestiones del cuestionario, pero contando con la precisión del Modelo de Rasch seleccionado para el ajuste de los datos. Este proceso de adaptación es frecuente en el campo de la educación, psicología educativa, medicina y la propia psicología contando con una bibliografía significativa en español. Una aplicación del TAI/CAT se puede encontrar en el trabajo de Styles y Andrich (1993). En la universidad española se puede consultar IIC (2003) en cuyas Jornada de Difusión Científico Técnica sobre Innovación en Capita Humano, realizadas bajo el lema "Las personas: el activo intangible de mayor valor para las organizaciones" se presento una experiencia TAI y en el trabajo de Pardo (2004) también sobre su utilización en los RRHH.

Los modelos de medidas identificados por Rasch y sus discípulos los han sido a causa de sus propiedades matemáticas. Su cualificación ha sido justificada con respecto a un cierto número de criterios exigentes entre los cuales se encuentra la unidimensionalidad y la invarianza. Si los datos observados se ajustan a estos modelos de medidas es posible afirmar que las puntuaciones transformadas son medidas en sentido fuerte (intervalos de medida), por oposición a las medidas ordinales tradicionalmente utilizados en los trabajos de economía y empresa, en base a datos actitudinales.

Una diferencia importante con la concepción clásica de la medida en las Ciencias Sociales y la aproximación de Rasch se refiere a la importancia relativa de los datos y el modelo de medida (Andrich, 1978 a y b). En otros términos, no se trata de explicar los datos obtenidos, sino de identificar, en el seno de estos datos, continuos compatibles con las operaciones de medida. Si los datos no se ajustan al modelo (cuidadosamente elegido) perteneciente a la familia de Modelos de Rasch, entonces es preciso buscar dónde dan

problemas los datos y tratar de comprender porque no se han producido verdaderas medidas en el marco de cuadro de referencia estudiado.

En este sentido, Dickes et al. (1994), establecen que: “La realización de una medida se inscribe entonces de forma natural en un proceso de verificación del objeto latente....Se trata de verificar si las relaciones observadas en el sistema empírico son compatibles con la representación que se ha hecho. Un modelo de medida no es más que la explicitación (y la formalización) de hipótesis a medir de acuerdo a unos hechos....Este reconocimiento del carácter teórico de la medida se acompaña de una explicitación mayor. Antes el modelo utilizado quedaba más implícito”

En conclusión, el modelo de Rasch es un test de la pertinencia de la demarcación entre instrumento de medida y simples inventarios de expresiones cuantitativas. Esta aproximación se opone a la comúnmente seguida por los analistas de datos de las ciencias sociales. La función de la medida será la descrita por Kuhn (1977): poner en evidencia las anomalías en los datos y verificar las teorías.

Estos modelos disponen de las siguientes ventajas:

1. Organización rigurosa de la concepción general del test en general: elección de las cuestiones y de las pruebas, formato de respuestas. En el caso de desviaciones se destacan los sujetos e ítems generadores y es posible considerar las causas.
2. Validación automática de los test construidos.
3. Valor predictivo del modelo
4. Universalidad del modelo

2.3.- EL MODELO DE RASCH EN ECONOMÍA, ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

En general, aunque se aprecia un cierto desarrollo de la investigación económica y empresarial mediante el uso de las Teorías del Rasgo Latente, sea mediante la TRI o los Modelos de Rasch, hay factores que están limitando su utilización. Por una parte, estas Teorías se han desarrollado en el campo de la Psicología y Educación, utilizando su terminología y cultura científica. Al igual que otros campos científicos en donde se aprecia un mayor impacto, como es en las Ciencias de la Salud, esperamos que se puedan superar las limitaciones culturales existentes, aplicando los isomorfismos pertinentes.

Se puede destacar otro factor de mayor envergadura, que se refiere al proceso de acorazamiento de la ciencia económica y empresarial actual y el proceso de construcción de teoría. La realidad es que la mayor parte de los desarrollos científicos que utilizan constructos se están realizando sobre la base de la Teoría Clásica del Test, aplicando ulteriores técnicas estadísticas que están basadas en el supuesto de que las medidas utilizadas y obtenidas son fiables y válidas (Churchill, 1979). Ya hemos indicado las limitaciones y dependencia de muestra de sujetos e ítems del cuestionario de estos

criterios de calidad de la medida en la TCT y las posibilidades que las Teorías de Rasgo Latente suponen al superar las mismas.

Un tercer factor limitativo del desarrollo de estas Teorías es la complejidad estadística de las mismas, que puede ser fácilmente superado por la amplia variedad de programas de computación que facilitan las diferentes tareas a realizar, antes de su interpretación científica.

Contando con estos factores limitativos, la utilización del Modelo de Rasch en España en las áreas de Economía, Administración y Dirección de Empresas está en la fase previa. Son escasos los trabajos que se han presentado en congresos o publicado en revistas nacionales e internacionales producidos por investigadores de las universidades españolas. Lo cual, por otra parte, no deja de ser un reflejo a nivel nacional de la utilización de los Modelos de Rasch a nivel internacional por investigadores de Economía y Empresa.

A partir de una búsqueda sistemática y puesto en contacto con los grupos de investigación que han realizado trabajos utilizando este sistema de medición objetiva, se puede presentar una representación de la producción científica realizada a nivel nacional.

TRABAJO EN ECONOMÍA

“Medida del desarrollo humano para los países de América Latina” **El Trimestre Económico** (Méjico) Vol. LXVIII (2) abril – junio 2001. N° 270, pp. 195 / 208

Juan Carlos Morán Álvarez* y Pedro Álvarez Martínez**
(*Universidad de Sevilla; **Universidad de Extremadura)

Los autores plantean como objetivo de la investigación la medición del desarrollo humano por medio de una variable unidimensional.

El constructo o variable latente definida es “desarrollo humano” de los países de América Latina, para lo cual se utiliza como variables las tres definidas para el caso por la PNUD, en su índice de desarrollo humano IDH.

Los datos utilizados son secundarios provenientes de las estadísticas publicadas para las variables a utilizar para medir el constructo por 19 países del área. Estos datos se han transformado mediante la aplicación del valor 1 para los datos mínimos de cada país y 10 para los correspondientes datos máximos. El resto de los valores categóricos del 2 al 9 se estiman por interpolación.

La metodología empírica se apoya en el Modelo de Rasch, a partir de la definición del modelo y establecimiento de la probabilidad de Rasch (1980). La estimación de los parámetros β_n y δ_i del Modelo se determina mediante algoritmos de máxima verosimilitud por el programa de computación Bigsteps (Wright y Linacre, 1997).

Los resultados obtenidos de la ejecución del programa Bigsteps, permite presentar un cluster de países de acuerdo con la medida alcanzada en su desarrollo humano (MDH). A

tales efectos se utiliza los mapas de distribución y las estadísticas de países (personas) clasificados por orden de medida.

A partir de los cluster de países se puede determinar los perfiles de los mismos, teniendo en cuenta los niveles que lucen de las distintas variables consideradas en el índice IDH (PNUD).

Como complemento de la información obtenida compara el posicionamiento de países logrados con el índice IDH y la medida MDH, logrando similares resultados. Lo cual lleva a los autores a destacar la fiabilidad del Modelo de Rasch utilizado.

Se baja al detalle del análisis realizado con el estudio de los desajustes evidenciados por las estadísticas de países (personas), a partir de la consideración de los residuales. El ajuste de los datos observados y los estimados a partir del análisis los residuales generados permite determinar la validez del Modelo.

Como conclusión presentan la utilidad del sistema de medición MDH (Modelo de Rasch) como alternativo del IDH, a la vez que permite un análisis detallado de los países que desajustan.

TRABAJOS EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

TRABAJOS EN RECURSOS HUMANOS

“Correlación entre los perfiles profesional y académico de los ejecutivos españoles” en la **Revista de Organización y Dirección**. Madrid. Nº 15 julio – septiembre de 1995.

Pedro Álvarez* y Julio García del Junco**

*Universidad de Extremadura; **Universidad de Sevilla)

El objetivo del estudio es determinar el ajuste entre los perfiles profesional y académico de los ejecutivos españoles.

Los constructos o variables latentes definidas han sido: “perfil académico” y “perfil profesional”, en ambos casos el instrumento de medida diseñado ha contado con los mismos 25 ítems.

Los datos son primarios obtenidos a partir de la remisión de los cuestionarios a directivos de las 500 mayores empresas del país. La muestra resultante, de acuerdo a las respuestas válidas recibidas, ha sido de 141 ejecutivos. El escalonamiento del cuestionario es de 10 categorías.

El trabajo empírico utiliza el Modelo de Rasch, presentando los autores el concepto de la medición objetiva, la formulación matemática del Modelo de Rasch, así como la formulación del sistema de estimación de los parámetros por medio de algoritmos de máxima verosimilitud. El programa utilizado es Bigsteps.

Los resultados de la utilización del Modelo de Rasch permiten presentar una ordenación de los ítems de las variables latentes definidas, a partir de las respectivas medidas obtenidas. A estas ordenaciones se le asigna un valor escalar del 1 al 25. Posteriormente se representan gráficamente, ubicando en el eje de abscisas el perfil profesional y en ordenadas el perfil académico.

La correlación entre la ordenación de los ítems de ambos constructos es negativa, por lo que destacan en conclusión la discrepancia de los perfiles. Pasando a analizar los ítems que se encuentran en posición $x > y$, $x = y$ y $x < y$.

Los autores concluyen sobre la importancia del método aplicado para la selección de aspirantes a los puestos de trabajo, así como la posible idoneidad de los solicitantes.

“Ética y Dirección de Empresas en Europa” **ESIC-MARKET** N° 95. 1997, pp. 189/198

Julio García del Junco* y María del Mar Ros Padilla*

* Universidad de Sevilla

El objetivo de este trabajo es la determinación de la importancia relativa de la ética en el perfil del directivo europeo.

El constructo o variable latente definido para esta investigación es “perfil del directivo europeo”, el instrumento de medida utilizado cuenta con 25 ítems, entre los cuales se incluye la ética.

El trabajo empírico utiliza el Modelo de Rasch, pero no se presenta explícitamente ningún desarrollo conceptual y matemático del mismo.

Los resultados obtenidos provienen de la utilización del mapa de medición conjunta de sujetos e ítems.

Se presenta la curva característica de los ítems (CCI).

Se presenta la clasificación ordenada de los ítems a partir del cuadro de estadísticas de ítems ordenadas por la medida obtenida por los distintos ítems. Se analizan los desajustes de los ítems, para terminar estudiando la importancia relativa de la ética en estas jerarquías.

La conclusión que presenta el artículo se centra en destacar la importancia relativa de la ética en el perfil del directivo europeo.

ESTRATEGIAS Y ENTORNO

“Computing the Perceived Environmental Uncertainty Functions by Rasch Model”
WSEAS Transactions on Business and Economics. Issue 4, vol. 1 October 2004, pp. 281–285

Vanessa Yanes – Estevez*; Juan Ramón Oreja – Rodríguez* y Pedro Álvarez**

*Universidad de La Laguna; **Universidad de Extremadura

El objetivo de este trabajo es proponer una nueva metodología matemática para el análisis de la incertidumbre percibida del entorno (PEU) por parte de la dirección de empresas, así como el de constatar la validez de planteamientos teóricos sobre PEU.

Los constructos planteados son tres:

- Complejidad del entorno (C)
- Dinámica del entorno (D)

Estos constructos serán medidos a partir de un instrumento que contiene 25 ítems (atributos del entorno, generadores de complejidad / dinámica) agrupados en cuatro subescalas, cuyos niveles de fiabilidad y validez ha sido previamente contrastados en estudios anteriores. Las escalas son del tipo Likert 1-5.

- Incertidumbre percibida del entorno (PEU)

Se define este constructo como una función de los dos anteriores, del tipo $PEU=F(C,D)$.

El tamaño muestral es de 338 ejecutivos, habiéndose realizado el trabajo de campo de enero a abril de 2001.

El trabajo empírico utiliza el Modelo de Rasch, previa conceptualización de los parámetros utilizados, la linealización de la medición conjunta y determinación de la probabilidad de Rasch. El modelo de computación utilizado es Winsteps.

Los resultados obtenidos permite determinar la fiabilidad y la separabilidad de ítems y encuestados así como la validez de la aplicación del Modelo de Rasch.

Se presentan las estadísticas de los ítems ordenados por su medida, tanto en el caso del constructo dinámica como en complejidad. Se asigna un valor escalar del 1 al 25 a los distintos ítems según su ordenación en cada constructo.

A partir de la ordenación realizada para los constructos se determina la función PEU, cuyo ajuste lineal con una $R=0,848$ permite confirmar las hipótesis teóricas de Daft y otros sobre la relación entre complejidad y dinámica. Este ajuste lleva a descartar los planteamientos de Duncan sobre la contribución equilibrada de la complejidad y dinámica en PEU.

Las conclusiones que se destacan se refieren a la validez del Modelo de Rasch para el análisis de las percepciones de los directivos, dentro de la aproximación cognitiva de Dirección Estratégica.

La función propuesta, permite un diagnóstico del entorno (PEU), con la posibilidad de análisis individuales, base del planteamiento destacado por la teoría de los puntos de referencia estratégicos.

Finalmente, se ha logrado constatar la validez de planteamientos teóricos existente en la doctrina sobre PEU.

“Agrupaciones de empresas según el número de vínculos interorganizativos establecidos mediante el Modelo de Rasch”. **XV Jornadas Hispanos – Lusas** de Sevilla 2005. Publicado en C. Barroso y J. L. Galán [Eds.] (2005): **Cities in Competition. Notes on strategy, planning and internationalization**. Universidad de Sevilla, pp. 171 / 181.

Ana M^a García Pérez*; Ángeles Sanfiel Fumero*; Pedro Álvarez Martínez** y Juan Ramón Oreja Rodríguez*

*Universidad de La Laguna; ** Universidad de Extremadura.

El objetivo de este trabajo es el establecimiento de agrupaciones de empresas en función de su propensión a establecer vínculos interorganizativos, a la vez que constatar hipótesis referidas a la relación entre el tamaño de la empresa y el mercado con el número de vínculos interorganizativos que se establecen.

El constructo planteado es “vínculos interorganizativos”. El instrumento de medida utiliza cinco ítems, con un escalamiento dicotómico.

La aplicación del instrumento de medida se realizó entre septiembre de 2001 y enero de 2002 a 537 empresas de la Industria Agroalimentaria (IAA) canaria, resultando un tamaño muestral de 236 respuestas válidas.

El trabajo empírico utiliza el Modelo de Rasch. Se presentan los parámetros y la linealización de ambos en la medición conjunta. Se desarrolla el Modelo dicotómico de Rasch, indicando que el programa de computación utilizado para la estimación de los parámetros y las estadísticas referidas a los mismos es Winsteps.

Los resultados alcanzados permiten destacar tanto la fiabilidad y la separabilidad de los ítems, como la validez del estudio a partir del estudio de los ajustes que presenta las estadísticas incluidas en la ordenación de los ítems (tipos de vínculos interorganizativos) por su medida.

A partir del mapa que presenta el análisis conjunto de empresas y tipos de vínculos interorganizativos, se establece los cluster de empresas en función del número de vínculos interorganizativos utilizados.

El estudio de los distintos cluster destaca las características de las empresas que los componen tipificándolas en función de su tamaño y tipo de mercado al que destinan su producción.

La jerarquización de los ítemes (tipos de vínculos interorganizativos) permite, por otra parte, contrastar el alcance de la validez de propuestas teóricas previas existentes sobre la priorización de las diferentes tipos de vínculos interorganizativos por parte de las empresas.

Finalmente, se contrasta las hipótesis sobre la relación entre el número de vínculos interorganizativos considerados por los cluster de empresas establecidos, en base a las características de las empresas que los integran (tamaño y mercados).

Como conclusión se destaca la contratación de las propuestas teóricas previas sobre ordenación de vínculos interorganizativos, la influencia del tamaño y mercado en la decisión de utilizar un determinado número de vínculos interorganizativo, así como la utilidad del Modelo de Rasch para la realización de este estudio.

TRABAJOS DE MARKETING

La satisfacción del turista en el destino Marbella. Medida y análisis mediante el Modelo Rasch. CEDMA. Málaga. 1999.

José L. Santos Arrebola
Universidad de Málaga.

El objetivo de este libro es medir la satisfacción de los turistas que visitan Marbella.

El constructo utilizado en el trabajo empírico es “grado de satisfacción del turista”. El instrumento de medida aplicado consta de 32 ítemes. La escala utilizada es de cuatro categorías.

La muestra estimada, con un error del 3 por 100 y 95% de intervalo de confianza es de 150 turista, para una población de 4000 posibles ocupantes de los hoteles de 4 y 5 estrellas de Marbella.

El Modelo de Rasch se presenta integrado en la Teoría de Respuesta al Ítem (IRT). Los parámetros utilizados se refieren a la satisfacción de los turistas y a los atributos del destino Marbella. Se conceptualiza la linealización en la medición conjunta de los parámetros, así como la expresión matemática del Modelo de Rasch. El programa de computación para la estimación de los parámetros y las correspondientes estadísticas es Bigsteps.

Los resultados que presenta se refieren tanto a los ítemes como los turistas. En el caso de los ítemes se jerarquizan de acuerdo a sus medidas, de acuerdo al grado de satisfacción (más satisfacción / menor medida; menor satisfacción / mayor medida).

Los turistas se jerarquizan de acuerdo a su grado de satisfacción, agrupando a los encuestados en dos grandes núcleos a partir del corte de la mediana de la medida. La información de los turistas, referida al hotel, su categoría, duración de la estancia y características socio-económicas del mismo permite el establecimiento de perfiles de los distintos grupos.

El análisis de los desajustes, tanto referidos a los atributos del destino como a los propios turistas, proporciona una abundante información para diagnosticar el grado de satisfacción de los turistas del destino Marbella.

En conclusión, el autor destaca la determinación del grado de satisfacción de los turistas del destino Marbella, el perfil de los turistas, de acuerdo con su grado de satisfacción y la posibilidad de un análisis detallado (a nivel del turista individual, grupo de turistas y atributo del destino) de las causas de (in)satisfacción. Finalmente explicita la importancia del Modelo de Rasch para la realización de este tipo de trabajo.

“Industrial Marketing Applications of Quantum Measurement Techniques” **Industrial Marketing Management** 30, pp. 13/22. 2001

Pedro Álvarez y Clementina Galera
Universidad de Extremadura

El objetivo de este trabajo es presentar una nueva metodología para analizar la conveniencia de un determinado bien de equipo agrícola (tractor).

El constructo diseñado es “conveniencia del tractor”. El instrumento de medida está compuesto de 9 ítems.

Participan en el trabajo de campo 231 propietarios de tractores de diferentes marcas.

Se presenta bajo la expresión “Quantum Measurement Techniques” el Modelo de Rasch. Se define la variable latente y los parámetros que la integra, formulando la expresión matemática del Modelo de Rasch.

Los resultados obtenidos de la aplicación del modelo de computación utilizado se centran en:

- Mapa conjunto de la jerarquización de los tipos de tractores y sus atributos de conveniencia.
- Fiabilidad y separabilidad de sujetos e ítems.
- Medición / calibración de ítems. Se clasifican los ítems (atributos de conveniencia) de acuerdo a su relevancia en dos grupos de acuerdo a sus medidas.
- Validez de los ajustes de los ítems y análisis de los desajustes de cada tipo de tractor en función de los atributos de conveniencia de los mismos.
- Análisis de la jerarquía de los tractores de acuerdo a sus medidas.
- Validez de los ajustes de los tractores y análisis de los desajustes de cada atributo de conveniencia en cada tipo de tractor.

Se presenta una conclusión general sobre la utilidad de esta técnica de medición en la medición de la conveniencia de los distintos tipos de tractores de acuerdo a los atributos especificados en el instrumento de medida.

“Application of the Rasch Model to Customer Satisfaction in the Destination of Marbella”

Journal of Quality Assurance in Hospitality & Tourism, vol. 3(1/2) 2002

José L. Santos-Arrebola

Universidad de Málaga

El objetivo de este trabajo se centra en determinar la satisfacción de los turistas que visitan Marbella.

El constructo diseñado es “satisfacción de los turistas en Marbella”, referida a los visitantes que se alojan en hoteles de cuatro y cinco estrellas de ese destino turístico. El instrumento de medida utilizado consta de treinta y dos ítems, correspondientes a atributos de Marbella. El escalonamiento es de uno a cuatro.

El cuestionario se cumplimentó por ciento cincuenta turistas alojados en los hoteles indicados, siendo el periodo de ejecución del trabajo de campo las dos primeras semanas de agosto de 1996.

En el trabajo empírico se ha utilizado el Modelo de Rasch, indicándose que se integra dentro de la Teoría de Respuesta al Ítem, con un anexo en donde se incluye el desarrollo de la expresión matemática del modelo. También se incluye la explicación de la linealización en la medición conjunta de los parámetros (sujetos e ítems).

Es de destacar la indicación de que se ha utilizado el Modelo de Rasch también en el estudio piloto, cumplimentado por diez sujetos, de tal forma que le ha permitido la calibración del cuestionario, pasando de ochenta ítems previos al utilizado de treinta y dos.

Los resultados obtenidos permiten presentar una ordenación jerárquica, por la medida, de los ítems. Ello permite analizar cada ítem por el grado de satisfacción que generan al conjunto de los visitantes. Que complementado por el estudio de los desajustes, puede llegar al detalle de qué turistas está satisfecho de cada ítem.

También se presenta la ordenación por la medida de los turistas, en función de su grado de satisfacción. Se complementa con el estudio de los desajustes, que permite determinar qué turistas está desajustado y en qué ítem se produce el residual.

Las conclusiones que se presentan se refieren a la innovación derivada de la aplicación del Modelo de Rasch, que permite individualizar el análisis a nivel de ítem o sujeto, lo cual no es posible con los tradicionales métodos estadísticos multivariados.

TRABAJO DE ECONOMÍA FINANCIERA

“Las matemáticas en la investigación económica financiera”

IX Jornadas de la Asociación española de profesores universitarios de Matemáticas Aplicadas para la Economía y la Empresa (ASEPUMA) Las Palmas de Gran Canaria 2001

Pedro Álvarez Martínez

M^a Isabel Escalona Fernández

Antonio Pulgarín Guerrero

Universidad de Extremadura

El objetivo de esta comunicación es determinar la aportación de las matemáticas en el ámbito de la investigación de Economía Financiera.

El constructo que se delimita es el “servicio que los métodos cuantitativos ofrece a la Economía Financiera”. El instrumento de medición diseñado contiene 24 áreas de Economía Financiera.

La información utilizada procede de la base de datos EcoLit, con un total de 34.103 publicaciones revisadas en el período de 1990 a 1999. De éstas al final se han decantado por un total de 58 descriptores cuantitativos.

Previo al trabajo empírico se han categorizado los datos disponibles, de acuerdo a un escalonamiento de diez categorías. Al máximo y mínimo de cada área se le asigna el nivel máximo y mínimo de servicio respectivamente. Los niveles intermedios se determinan por interpolación logarítmica.

En el trabajo empírico se ha aplicado el Modelo de Rasch. Los parámetros definidos se han estimado por medio del programa de computación Bigsteps. Se explica previamente los principios de la medición conjunta y la linealización y se expone la formulación matemática de la probabilidad de Rasch.

Los resultados obtenidos se refieren a la jerarquización de las áreas o temáticas de Economía Financiera en función de la utilización de métodos cuantitativos. Del análisis de los estadísticos se aprecia los desajustes existentes.

También se obtienen la jerarquización, en función de la medida, de los descriptores cuantitativos utilizados. El estudio de los estadísticos de ajuste permite determinar las situaciones anormales.

Como conclusión se destaca la importancia del Método de Rasch para determinar las áreas temáticas de Economía Financiera que más demandan el servicio de los métodos cuantitativos, lo cual permite configurar los correspondientes programas de estudio.

3.- LA PRÁCTICA DE LA MEDICIÓN OBJETIVA: UNA APLICACIÓN DEL MODELO DE RASCH EN ADE.

3.1.- Programas informáticos: Ministep / Winsteps

Existen una amplia variedad de programas informáticos para tratar la información mediante el Modelo de Rasch. En Bond y Fox (2001) se presenta una valoración y descripción de los principales programas y sus URL:

- Quest y ConQuest (<http://www.acer.edu.au/>),
- Winsteps/ Ministep y Facets/Minifac (<http://www.winsteps.com/>),
- RUMM (<http://www.rummlab.com.au/>)

El programa que hemos utilizado para la aplicación que se presenta es Winsteps /Ministep disponible gratuitamente en la Web del Centro de Investigación y Documentación Educativa (CIDE) del Ministerio de Educación y Ciencia (<http://www.mec.es/cide/>). El itinerario a seguir en el CIDE es: Investigación Educativa/ Recursos para la educación/ Recursos metodológicos/Ministep. El enlace nos lleva a la Web de Winsteps y a las instrucciones para su instalación gratuita (en la fecha de la finalización del presente documento de trabajo: abril de 2005)

Free by Download MINISTEP

MINISTEP is a reduced version of WINSTEPS. It has complete WINSTEPS functionality, but is limited to 30 items and 200 persons (cases). **You may copy, distribute and use it without charge or time-limit.** Of course, we hope you will want to purchase the full version of **WINSTEPS** (which can analyze 10,000 items and 1,000,000 persons).



- 1.
2. "Save as" "c:\windows\desktop\MinistepInstall.exe"
3. Click on "**MinistepInstall.exe**" on your desktop to install MINISTEP, the free student/evaluation version of **Winsteps**
MINISTEP will start automatically.
4. Delete "MinistepInstall.exe"
5. Click on "Ministep" icon on desktop to run MINISTEP
6. *Note:* Winsteps looks for wordpad.exe on your system. Some computers have more than one copy and Winsteps finds one that doesn't work. If so, go to Initial settings on the Edit menu and paste in the path to the correct WordPad, usually in your Accessories folder. Or paste in the path to your favorite word processor or text editor.


Estando disponible en la misma fecha una versión on-line del manual de instrucciones, que hemos utilizado para la comprensión del programa:

WINSTEPS-MINISTEP User Manual On-Line

Go to [WINSTEPS-MINISTEP User Manual](#)

Aunque, también está disponible en formato PDF:

For **WINSTEPS User Manual only** in PDF Format, **mostly same as "Help" file**.

1. 
2. Use Acrobat to read.

3.2.- ENTRADAS

La aplicación que presento a continuación corresponde a una revisión actual, por medio del Modelo de Rasch, de parte un artículo publicado en 1995: “El cambio organizativo en la sucursal bancaria. **ESIC-MARKET** 88, abril-junio, pp. 63 /81. Específicamente en el epígrafe titulado EL CAMBIO CULTURAL definíamos dos constructos: “Valoración de los trabajos a desarrollar por el personal de oficinas” y “Tipos de capacidades en el personal de oficina”, nos centraremos en el primero.

Los datos provenían de una determinada zona comercial de una empresa bancaria, constituyendo el colectivo de directores de las distintas sucursales los que cumplimentaron un cuestionario que entre otras preguntas, contenía específicamente una referida al constructo indicado con un escalamiento en la respuesta a los 9 ítems propuestos de 1 a 5.

La información obtenida en 1994, se trató a nivel estadístico básico, permitiendo una jerarquización de acuerdo con las medias obtenidas.

3.2.1.- FICHERO DE CONTROL: ÓRDENES SINTÁCTICAS BÁSICAS.

Al replicar el trabajo, replanteamos el tratamiento de la información obtenida en 1994, mediante la aplicación del Modelo de Rasch, lo que nos permite obtener unos resultados más amplios que los presentados en el artículo de referencia.

El primer paso para el tratamiento de la información obtenida de la encuesta es el diseño de un programa de control para la ejecución del programa. Este programa de control se debe realizar en Wordpad y en fuente Courier New (Occidental). Contiene un conjunto de órdenes sintácticas para la ejecución del programa.

El programa de control básico que hemos diseñado y que presentamos a continuación es un ejemplo sobre el cual se puede trabajar, modificando los aspectos necesarios.

NOTA: No se pretende presentar un manual completo de instrucciones para el uso de este programa, el cual está disponible en la Web indicada, sino sólo proporcionar los primeros elementos para la introducción el uso del mismo.

A.- Estructura del programa de control

```
;FICHERO DE CONTROL DEL ESTUDIO DE LA VALORACIÓN DEL TRABAJO A
DESARROLLAR POR EL PERSONAL DE OFICINAS 1994

&INST

; Formato de entrada de datos
NAME1=1 ; Columna de inicio de la
información de Oficinas
NAMLEN=30 ; Máxima longitud de la información
de Oficinas
ITEM1=32 ; Columna inicio información ítems
NI=9 ; Número de ítems del test
XWIDE=1 ; Número de columnas por respuesta

; Datos
DATA=CYTDATT.txt ; Fichero de datos del estudio
; Manejo de datos
CODE=12345 ; Códigos de respuestas válidas

; Tablas resultantes
PERSON=Empleado oficina ; Determinación del encuestado
ITEM=VALORACION TRABAJO ; Determinación del Ítem
TITLE= "Valoración Trabajos de Empleados Oficinas"
TFILE=*
1 ; Mapas de medición conjunta
3 ; Resumen estadísticas del análisis
7 ; Ajustes de los empleados oficinas
11 ; Ajustes de los ítems
13 ; Medidas de los ítems
17 ; Medidas de los empleados oficinas
22 ; Escalograma de Guttman
*
UMEAN=100 ; Valor de referencia de la media
UASCALE=4.55
UDEDECIMALS=1 ; Número de decimales
```


COMANDOS E INSTRUCCIONES GENERALES

Comando de inicio:	& INST
Explicación de las órdenes:	;
Comando de finalización de las órdenes:	& END
Comando de finalización de las etiquetas de ítems:	END NAMES
Inicio y finalización de información	*

ENTRADA DE DATOS

DATA= *.txt	Nombre del fichero de datos. Véase epígrafe 3.1.2
NAME1=	Primera columna donde se inicia la información identificativa de las personas, normalmente 1
NAMLEN=	Número total de columnas que incluye información identificativa de las personas, por. ej. 30
ITEM1=	Primera columna donde se inicia la información de los ítems, por ej. 32
NI=	Número de ítems, por ej. 9
XWIDE=	Columnas por ítem, por ej. 1

SELECCIÓN DE DATOS

CODE=	Códigos de datos válidos, por ej. 01 (dicotómico) ó 12345 (politómico escala 1 a 5)
-------	--

TABLAS DE RESULTADOS

ITEM=	Título o etiqueta de los ítems (ver el ejemplo)
CLFILE=	Título o etiquetas de las categorías de las escalas de los ítems (por ej. Muy Bajo / Bajo / Medio / Alto / Muy Alto, en una escala 1 a 5)
PERSON=	Título o etiqueta de los sujetos (ver el ejemplo)
TITLE=	Título o etiqueta del trabajo

FTITLE= Identificación de las tablas solicitadas

MEDIDAS

UMEAN= Valor asignado a la media de la escala de los ítems (logits, por defecto 0)

UDECIMALS= Número de decimales en los resultados

USCALE= Valor de un logit en la escala

3.1.2.- FICHERO DE DATOS: DEL EXCEL AL WORDPAD

Los datos que constituyen el input de la aplicación del Winsteps, puede tener almacenados en diferentes fichero, tales como los de SPSS, Word o Excel, en cualquier caso el fichero de datos que utiliza el programa debe reunir determinados requisitos.

Se utiliza el Wordpad, o sustituto alternativo que aconseja la empresa editora del programa para su ejecución en Winsteps / Ministep.

El fichero debe guardarse con formato: solo texto.

Todos los datos se deben presentar en Courier New (Occidental).

Los datos, de los sujetos como de los ítems, deben ocupar unas determinadas posiciones, las cuales se expresan en el correspondiente programa de control.

En el manual de instrucciones de Winsteps / Ministep se proporcionan instrucciones más detalladas.

En el trabajo que estamos replicando los datos inicialmente estaban en un fichero de Excel.

OFICINA	ITEMES								
	1	2	3	4	5	6	7	8	9
O1	5	3	5	3	4	5	3	3	4
O2	5	3	5	3	4	5	3	3	4
O3	4	3	3	3	5	3	3	3	4
O4	5	4	5	3	5	4	2	2	5
O5	5	3	5	3	4	5	4	4	5
O6	5	4	4	4	5	5	3	3	4
O7	5	3	4	2	5	4	3	3	4
O8	4	3	4	2	4	5	3	3	5
O9	5	3	5	3	5	5	5	3	3
O10	5	4	5	4	5	5	4	5	5
O11	4	4	3	3	4	5	4	4	5
O12	5	4	4	3	5	5	1	4	5
O13	4	3	4	3	4	5	4	4	4
O14	5	4	5	4	5	5	4	4	5

Estos datos se pasaron a Wordpad, como **CYTDATT.txt** (Documento de texto), quedando de la siguiente forma:

```

O1          535345334
O2          535345334
O3          433353334
O4          545354225
O5          535345445
O6          544455334
O7          534254334
O8          434245335
O9          535355533
O10         545455455
O11         443345445
O12         544355145
O13         434345444
O14         545455445

```

Quedando preparado para su utilización por el programa de control.

3.3. - RESULTADOS

3.3.1. - FICHERO DE RESULTADOS: CREACIÓN Y GESTIÓN

La ejecución del programa Winsteps requiere un fichero de resultados, que es conveniente tenerlo creado inicialmente. Para ello, se puede seguir el mismo procedimiento como el realizado para la creación de los ficheros de control y datos en Wordpad.

En este caso el fichero guardado sin contenido se ha denominado **CYTREST.txt** (Documento de texto).

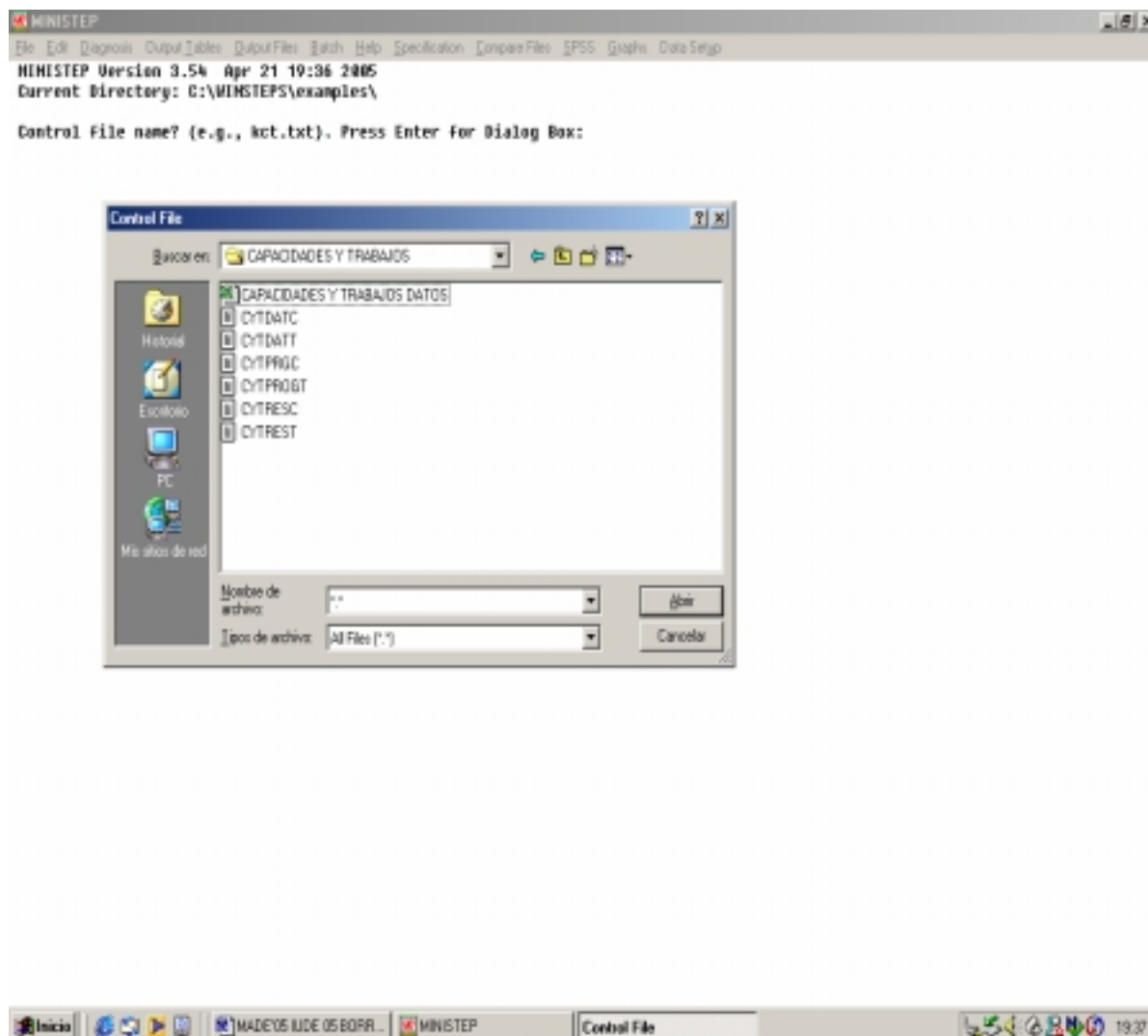
En este fichero se guardarán la información de las tablas solicitadas en el programa de control, así como datos relativos a la ejecución y órdenes del fichero de control utilizado.

3.3.2.- EJECUCIÓN DEL PROGRAMA

Al pulsar el icono de Ministep Winsteps aparece una pantalla de diálogo, en cuya parte superior se sitúa una Barra de Menú, poco utilizable a nivel básico, siempre que se trabaje con un fichero de control bien diseñado que no proporcione problemas. En caso contrario, las instrucciones que proporciona el programa nos llevará a esta Barra de Menú.

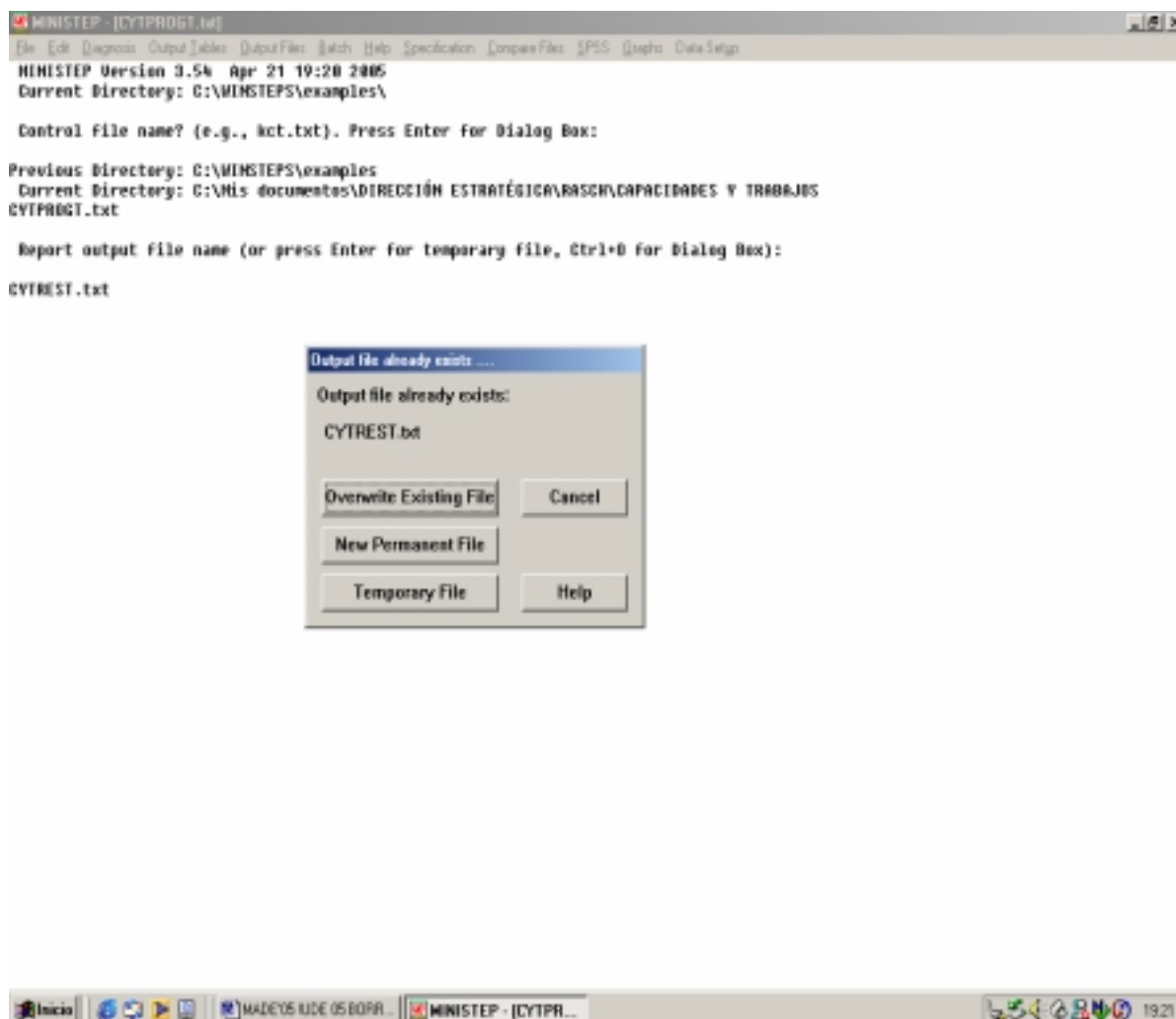
Se presenta una descripción genérica de las distintas pestañas, que puede ser completada en el correspondiente manual de instrucciones.

FILE	Archivo, con una lógica similar al Windows.
EDIT	Edición. Despliegue y edición de ficheros de entrada, de resultados y tablas.
DIAGNOSIS	Diagnosis. Tablas para entender, evaluar y mejorar el sistema de medición
OUTPUT TABLES	Presenta / produce todas las tablas de resultados generadas por Winsteps
OUTPUT FILES	Genera resultados inicialmente válidos como entrada para otros software
BATCH	Facilita la ejecución de Winsteps en modo Bach
HELP	Despliega el fichero de ayuda
SPECIFICATION	Permite la introducción de especificaciones tras el análisis, una cada vez, en la forma de especificación = valor
PLOTS	Utiliza el programa Excel de Windows para desplegar y comparar análisis
SPSS	Remodela los ficheros SPSS.sav para poder utilizarlos como ficheros de control y de datos
GRAPH MENU	Gráficos “en forma de mapas de bits” para la presentación del test, ítems y categorías
DATA SETUP	Proporciona medios inmediatos para establecer ficheros de control y de datos



En la pantalla de diálogo, se solicita primero el nombre del programa de control. En este caso se pulsa **intro** y aparece una cuadro de ficheros del tipo del Word, que permite buscar entre ficheros del usuario para determinar el fichero de control a utiliza, en este caso: **CYTPROGT.txt**

Posteriormente solicita el fichero de resultados, ofertando también un cuadro de ficheros si se pulsa simultáneamente la combinación de teclas "**Ctrl o**". Una vez que se indique un fichero existente, en este caso **CYTREST.txt**, el programa recuerda que existe y si se quiere sobrescribir. Por dos veces una en castellano y otra en inglés (que es la imagen que se presenta).



Una vez que se acepta, vuelve a solicitar si existe alguna especificación más y en caso negativo se ejecuta el programa. En la pantalla aparecen todas iteraciones de ajuste del mismo en la estimación de los parámetros, información que a este nivel no es necesario utilizar.

Cuando termine la ejecución se puede directamente pulsar **X (salir)**.

Una primera descripción del gráfico, referido a ítems y sujetos, podría ser:

En la parte derecha del gráfico, se destaca la presencia de dos grandes grupos de ítems: uno situado en la parte baja de la media de la escala ($M= 100$) con medidas en un entorno entre 99.1 y 97.7 (algo más que una vez la desviación S) referido a actividades comerciales y otro situado más arriba con medidas en un entorno entre 101.8 y 102.7 (entre una vez y dos veces la desviación S / T), referido a las actividades administrativas.

Los encuestados nos indican que las actividades que más frecuentemente deberían realizar son las comerciales (Menor medida => mayor frecuencia: cuentan con una mayor valoración personal) y las que con menos frecuencia deberían realizar serían las actividades administrativas (Mayor medida => menor frecuencia (raras): cuentan con una menor valoración personal).

A partir de esta información se puede realizar un análisis de la posición relativa de cada ítem y su significado.

En la parte de la izquierda del gráfico, se destaca la distribución de sujetos encuestados identificados por el número de la oficina. La situación que presentan, concentrados alrededor de la media (M y dos veces la desviación T) entre 100.8 y 105.2 es de empleados que prefieren realizar las actividades comerciales (todas están por debajo de la oficina inferior: O3) y en menor medida las actividades administrativas.

En las actividades administrativas vemos que diferentes ítems discriminan grupos de oficinas:

- Control de efectivo en caja y gestión de datos: O1, O13, O2, O4
- Control de balance: O11, O12
- Control de depósitos: O6, O9, O5, O14, O10

El análisis de esta situación se puede complementar con el análisis del Escalograma de Guttman, donde se advierte dentro de la jerarquía que se presenta las puntuaciones obtenidas por los diferentes ítems.

A partir de esta información se puede realizar un análisis de la posición relativa de cada empleado de las oficinas y su significado.

Si se hubiera recogido información complementaria de cada empleado se podría determinar los perfiles comunes debido a la pertenencia a cada grupo de empleados de oficinas y su significado.

También sería posible contrastar hipótesis causales de características del empleado con respecto a las posiciones relativas de los ítems y las probabilidades de asumir cada uno de ellos.

3.3.4.- ESTADISTICOS DE FIABILIDAD

TABLE 3.1 Valoración Trabajos de Empleados Oficina CYTREST.txt Apr 21 19:39 2005
 INPUT: 14 Empleado ofis, 9 VALORACION TS MEASURED: 14 Empleado ofis, 9 VALORACION TS, 5
 CATS

SUMMARY OF 14 MEASURED Empleado ofis								
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	36.0	9.0	102.48	.61	1.00	-.1	.96	.0
S.D.	2.9	.0	1.15	.07	.73	1.4	.68	1.2
MAX.	42.0	9.0	105.18	.79	2.40	2.2	2.20	2.0
MIN.	31.0	9.0	100.82	.50	.08	-2.7	.07	-1.3
REAL RMSE	.70	ADJ.SD	.91	SEPARATION	1.29	Emplea	RELIABILITY	.62
MODEL RMSE	.62	ADJ.SD	.97	SEPARATION	1.57	Emplea	RELIABILITY	.71
S.E. OF Empleado ofi MEAN = .32								

Empleado ofi RAW SCORE-TO-MEASURE CORRELATION = .99

CRONBACH ALPHA (KR-20) Empleado ofi RAW SCORE RELIABILITY = .61

SUMMARY OF 9 MEASURED VALORACION TS								
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	56.0	14.0	100.00	.49	1.02	.0	.96	.0
S.D.	9.0	.0	1.94	.06	.48	1.1	.53	1.1
MAX.	66.0	14.0	102.71	.60	2.12	2.1	2.29	2.3
MIN.	43.0	14.0	97.70	.42	.37	-1.7	.34	-1.8
REAL RMSE	.54	ADJ.SD	1.86	SEPARATION	3.47	VALORA	RELIABILITY	.92
MODEL RMSE	.50	ADJ.SD	1.87	SEPARATION	3.77	VALORA	RELIABILITY	.93
S.E. OF VALORACION T MEAN = .69								

UMEAN=100.000 USCALE=1.000

VALORACION T RAW SCORE-TO-MEASURE CORRELATION = -1.00

Se presenta dos de fiabilidades, una referida a sujetos y otra a ítemes.

La interpretación se refiere a una medida en una escala con un rango entre 0 y 1.

3.3.5.- CALIBRACIÓN DE LOS ÍTEMES

TABLE 13.1 Valoración Trabajos de Empleados Oficinas CYTREST.txt Apr 21 19:39 2005
 INPUT: 14 Empleado ofis, 9 VALORACION TS MEASURED: 14 Empleado ofis, 9 VALORACION TS, 5
 CATS

Empleado ofi: REAL SEP.: 1.29 REL.: .62 ... VALORACION T: REAL SEP.: 3.47 REL.: .92

VALORACION T STATISTICS: MEASURE ORDER

ENTRY	RAW		MODEL	INFIT	OUTFIT	PTMEA					VALORACION T
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.		
4	43	14	102.7	.4	.37	-1.7	.34	-1.8	.76		Control depositos
7	46	14	102.2	.4	2.12	2.1	2.29	2.3	.40		Control Balance
2	48	14	101.8	.4	.55	-1.1	.55	-1.1	.55		Control efectivo caja
8	48	14	101.8	.4	.82	-.3	.85	-.2	.68		Gestion datos
3	61	14	99.1	.5	1.17	.6	1.05	.3	.53		Identificacion necesidades clientes
9	62	14	98.8	.5	1.23	.7	1.18	.5	.34		Atraer clientes oficinas
5	64	14	98.3	.5	1.04	.2	.95	.1	.25		Venta servicios
1	66	14	97.7	.6	.74	-.6	.63	-.3	.46		Conocimiento clientela
6	66	14	97.7	.6	1.13	.5	.77	-.1	.52		Resolucion problemas
MEAN	56.0	14.0	100.0	.5	1.02	.0	.96	.0			
S.D.	9.0	.0	1.9	.1	.48	1.1	.53	1.1			

En esta tabla tenemos una ordenación jerárquica de la valoración de los distintos tipos de trabajos a realizar por el personal de las oficinas bancarias. En la posición baja se encuentra “Resolución de problemas” cuya frecuencia de admisión es alta y en cambio en Control de depósitos su admisión es más rara por parte de los empleados. Ello puede orientar a los responsables de oficinas sobre el tipo de trabajo que los empleados consideran conveniente.

La información que se obtiene en relación a los ítems puede permitir la calibración del cuestionario, si se destaca indicios de falta de validez (desajustes) de algún ítem. Así con una variación en los valores de MNSQ y ZSTD de INFIT / OUTFIT que se ha indicado en la tabla de medición de los sujetos, se aprecia el desajuste del ítem Control de Balance, que podría aconsejar su supresión.

No obstante, en situaciones de análisis comparativos entre factores discriminantes de actuaciones o atributos de la variable latente, lo que estamos recogiendo es un desorden que necesita ser analizado.

3.3.6.- DESAJUSTES EN ÍTEMES

TABLE 11.1 Valoración Trabajos de Empleados Oficin CYTREST.txt Apr 21 19:39 2005
 INPUT: 14 Empleado ofis, 9 VALORACION TS MEASURED: 14 Empleado ofis, 9 VALORACION TS, 5
 CATS

TABLE OF POORLY FITTING VALORACION TS (Empleado ofis IN ENTRY ORDER)
 NUMBER - NAME -- POSITION ----- MEASURE - INFIT (MNSQ) OUTFIT

7	Control Balance				102.2		2.1	A	2.3
RESPONSE:	1:	3	3	3	2	4	3	3	5
Z-RESIDUAL:									2
RESPONSE:	11:	4	1	4	4				
Z-RESIDUAL:			-3						
9	Atraer clientes oficinas				98.8		1.2	B	1.2
RESPONSE:	1:	4	4	4	5	5	4	4	3
Z-RESIDUAL:									-2
RESPONSE:	11:	5	5	4	5				
Z-RESIDUAL:									

Tal como se resaltaba en la tabla de medida de ítems el Control de Balance esta desajustado. El análisis detallado de las causas de desajuste permite considerar las acciones necesarias para corregir dichas las causas del desajuste, en el caso de que sea necesario.

Los valores **residuales positivos** indican que han puntuado al ítem por encima del valor que esperaba el modelo. En el ítem 7 Control de Balance, el empleado de la oficina 9 ha puntuado muy alto ese ítem cuando el modelo ajustado esperaba un valor inferior.

Los valores **residuales negativos** indican que han sido puntuados en la realidad más bajo de lo que el modelo esperaba. En el caso de ítem 7 Control de Balance el empleado de la oficina 12 ha puntuado muy bajo ese ítem cuando el modelo ajustado esperaba un valor algo superior. En el mismo sentido se aprecia el desajuste puntual destacado por el modelo en el ítem 9 Atraer clientes oficinas, dado que el empleado de la oficina 9 ha puntuado ese ítem con un valor más bajo del estimado por el modelo.

Como se ha indicado ya la utilidad de este análisis es poder bajar al detalle de los desajustes, por si es necesaria su corrección o, simplemente, confirmar su eliminación.

3.3.7.- MEDIDA DE SUJETOS

TABLE 17.1 Valoración Trabajos de Empleados Oficinas CYTREST.txt Apr 21 19:39 2005
 INPUT: 14 Empleado ofis, 9 VALORACION TS MEASURED: 14 Empleado ofis, 9 VALORACION TS, 5
 CATS

Empleado ofi: REAL SEP.: 1.29 REL.: .62 ... VALORACION T: REAL SEP.: 3.47 REL.: .92

Empleado ofi STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEA CORR.	Empleado ofi
10	42	9	105.2	.8	.38	-1.2	.22	-.3	.81	O10
14	41	9	104.6	.7	.08	-2.7	.07	-1.2	.97	O14
5	38	9	103.1	.6	.75	-.4	.79	-.2	.81	O5
6	37	9	102.7	.6	.80	-.3	.72	-.4	.79	O6
9	37	9	102.7	.6	2.39	2.2	2.14	1.9	.58	O9
11	36	9	102.4	.6	1.67	1.3	1.65	1.3	.45	O11
12	36	9	102.4	.6	2.40	2.2	2.20	2.0	.76	O12
1	35	9	102.0	.6	.44	-1.2	.44	-1.2	.91	O1
2	35	9	102.0	.6	.44	-1.2	.44	-1.2	.91	O2
4	35	9	102.0	.6	1.80	1.4	1.79	1.4	.80	O4
13	35	9	102.0	.6	.82	-.2	.86	-.1	.68	O13
7	33	9	101.4	.5	.43	-1.2	.43	-1.3	.92	O7
8	33	9	101.4	.5	.54	-.9	.53	-.9	.89	O8
3	31	9	100.8	.5	1.01	.2	1.13	.4	.57	O3
MEAN	36.0	9.0	102.5	.6	1.00	-.1	.96	.0		
S.D.	2.9	.0	1.1	.1	.73	1.4	.68	1.2		

La primera información que obtenemos en esta tabla es la ordenación, de acuerdo a la medida, de los empleados de las diferentes oficinas en función de su valoración de los trabajos a desarrollar por el personal de las oficinas.

Las mediciones más bajas que corresponden a las oficinas O3, O8, ..., corresponden a empleados que no han valorado como convenientes todos los ítemes y, como hemos visto en la tabla de medición conjunta, no valoran con puntuaciones altas los trabajos administrativos.

Con los niveles altos de medición se encuentran los empleados de las oficinas O10 y más abajo O14, ..., estos encuestados han valorado de forma alta no solo las actividades comerciales sino prácticamente todas las actividades administrativas.

Complementando la tabla de medición conjunta nos permite concluir respecto al perfil de los empleados y la posibilidad de posicionar en el futuro a otros encuestados de acuerdo con su valoración de estos ítemes.

Respecto a los estadísticos de validez, se puede analizar los ajustes derivados de INFIT / OUTFIT:

En MNSQ del Infit/Outfit el valor esperado es 1. Se admiten como aceptables valores entre 0.50 y 1.50. Valores superiores a 1.5 e inferiores a 2 indican desajustes que son improductivos para la construcción de la medida, pero no distorsionan. Valores superiores a 2, indican desajustes que sí distorsionan.

En ZSTD del Infit /Outfit los valores incluidos en entre -1.9 y + 1.9 son aceptables. Los valores superiores a 2 están desajustados.

Finalmente el valor de PTMA CORR. Se refiere a la correlación punto biserial que mide la relación entre las puntuaciones de los ítems de cada empleado de oficinas y las mediciones obtenidas. El valor que debe aparecer ha de ser positivo. En caso contrario existe un claro desajuste del encuestado. Menor importancia tendrá el nivel del valor obtenido.

Este análisis, en el caso de tratar de determinar qué encuestados están afectando al diseño del instrumento de medida, aconsejará su eliminación de la base de datos o no.

No obstante, en situaciones de análisis comparativos (benchmarking) entre encuestados, empresas, sujetos,..., lo que estamos recogiendo es un desorden que necesita ser analizado, explicado y, en todo caso, susceptible de acción correctora.

3.3.8.- DESAJUSTE DE SUJETOS

TABLE 7.1 Valoración Trabajos de Empleados Oficina CYTREST.txt Apr 21 19:39 2005
INPUT: 14 Empleado ofis, 9 VALORACION TS MEASURED: 14 Empleado ofis, 9 VALORACION TS, 5
CATS

TABLE OF POORLY FITTING Empleado ofis (VALORACION TS IN ENTRY ORDER)

NUMBER	NAME	POSITION	MEASURE	INFIT (MNSQ)	OUTFIT
12	O12		102.4	2.4	A 2.2
	RESPONSE:	1: 5 4 4	3 5	5 1	4 5
	Z-RESIDUAL:			-3	
9	O9		102.7	2.4	B 2.1
	RESPONSE:	1: 5 3 5	3 5	5 5	3 3
	Z-RESIDUAL:			2	-2
4	O4		102.0	1.8	C 1.8
	RESPONSE:	1: 5 4 5	3 5	4 2	2 5
	Z-RESIDUAL:			-2	
11	O11		102.4	1.7	D 1.7
	RESPONSE:	1: 4 4 3	3 4	5 4	4 5
	Z-RESIDUAL:			-2	

Las medidas de los sujetos se complementan con el análisis detallado del origen del desajuste entre el valor real y el del modelo. Ello permite considerar las acciones necesarias para corregir las causas del desajuste, en el caso de que sea necesario.

Así, los desajustes de los empleados de las oficinas recogidos en la tabla 17.1 se complementa con la expresión de qué ítem es el que genera el desajuste. Los valores **residuales positivos** indican que han puntuado al ítem por encima del valor que esperaba el modelo. Como en los casos del empleado de la oficina 9 al referirse a su preferencia hacia el control del balance.

Los valores **residuales negativos** indican que han sido puntuados más bajo de lo que el modelo esperaba. Como es el caso del empleado de la oficina 12, en relación con su puntuación al control del balance; el empleado de la oficina 9, en referencia a la atracción de clientes a la oficina; el empleado de la oficina 4, con respecto a la gestión de datos y el empleado de la oficina 11, respecto a la identificación de las necesidades de los clientes

3.3.9.- ESCALOGRAMA DE GUTTMAN

TABLE 22.1 Valoración Trabajos de Empleados Oficin CYTREST.txt Apr 21 19:39 2005
INPUT: 14 Empleado ofis, 9 VALORACION TS MEASURED: 14 Empleado ofis, 9 VALORACION TS, 5
CATS

GUTTMAN SCALOGRAM OF RESPONSES:

```

Emplead|VALORACION T
        |165932874
        |-----
    10 +555554544
    14 +555554444
     5 +554553443
     6 +555444334
     9 +555353353
    11 +454534443
    12 +555544413
     1 +554453333
     2 +554453333
     4 +545554223
    13 +454443443
     7 +545443332
     8 +454543332
     3 +435433333
        |-----
        |165932874

```

Este escalograma de Guttman jerarquiza las oficina y los ítemes simultáneamente, proporcionando información complementaria a las tablas anteriores.

De arriba a bajo se aprecia las mayores puntuaciones, en este caso del empleado de la oficina 10.

De izquierda a derecha se ordenan los ítemes en orden decreciente. Así el ítem con mayor puntuación obtenida en la aplicación del cuestionario es el 1, le sigue el 6, etc.

Desde la perspectiva de la jerarquización de Guttman hay ítemes o sujetos que, una vez ajustado el modelo, presentan valores que desajustan. Se puede destacar el caso del empleado de la oficina 9 que presenta desajuste con respecto a los ítemes 4, 7 y 9, tanto con puntuaciones más alta y más bajas de las esperadas por su situación el escalograma. Lo cual le hace poco representativo del grupo analizado.

También hay desajustes en los empleados de las oficinas 11, 12 y 4 que presentan valores más bajos de los esperados.

BIBLIOGRAFÍA

- Álvarez, P. y C. Galera (2001): "Industrial Marketing Applications of Quantum Measurement Techniques" **Industrial Marketing Management** 30, pp. 13/22. 2001
- Álvarez, P. y J. García del Junco (1995): "Correlación entre los perfiles profesional y académico de los ejecutivos españoles" en la **Revista de Organización y Dirección**. Madrid. Nº 15 julio – septiembre.
- Álvarez, P.; M^a I. Escalona y A. Pulgarín (2001): "Las matemáticas en la investigación económica financiera". **IX Jornadas de la Asociación española de profesores universitarios de Matemáticas Aplicadas para la Economía y la Empresa (ASEPUMA)** Las Palmas de Gran Canaria 2001
- Andersen, E.B. (1973): "A goodness of fit test for the Rasch model". **Psychometrika**, 38, 1, 123/140.
- Andersen, E.B. (1977): "Sufficient Statistics and Latent Trait Models" **Psychometrika**, 42, 69/81
- Andrich, D. (1978a): "Application of a Psychometric Rating Model to Ordered Categories which are scored with Successive Integers". **Applied Psychological Measurement**, 2 (4), pp. 581/594, (citado por Bond y Fox, 2001).
- Andrich, D. (1978b): "A rating formulation for ordered response categories". **Psychometrika**, 43, pp. 357/374
- Andrich, D. (1982): "An Index of Person Separation in Latent Trait Theory, the Traditional KR.20 Index, and the Guttman Scale Response Pattern" **Educational Research and Perspectives** 9, pp. 95/104
- Andrich, D. (1988): **Rasch Models for measurement**. Newbury Park CA: Sage
- Andrich, D. (1989): "Distinction between Assumptions and Requirements in Measurement in the Social Science", en J.A. Keats, R. Taft, R.A. Heath et S.H. Lovibond [Eds.]: **Mathematical and Theoretical Systems**, pp. 7/15. London: Elsevier Science Publishers, (citado por Leplège, 2003).
- Andrich, D. (1996): "Measurement Criteria for Choosing among Models with Graded Responses", en von Eye, A. y C.C. Clogg [Eds.]: **Categorical variables in developmental research: Methods of analysis**. San Diego, CA: Academic Press, pp. 3/ 35, (citado por Bond y Fox, 2001).
- Andrich, D. (2005): "Adminstrating, Analyzing and Improving Tests". En Andrich, D. e I. Marais (2005): "Instrument Design with Rasch IRT and data Analysis I" **Readings EDU 435/635**. School of Education. Murdoch University
- Andrich, D. y C. Hagquist (2004) : "Detection of Differential Item Functioning Using Analysis of Variance". **Paper presented at the Second International Conference on Measurement in Health, Education, Psychology and Marketing : Developments with Rasch Models**. Murdoch University. Perth
- Andrich, D. ; B.E. Sheridan y G. Luo (2004) : **RUMM2020 : Rasch Unidimensional Measurement Models (version 4.1)[Computer Software]**. Perth, Western Australia : RUMM Laboratory
- Andrich, D. e I. Marais (2005): "Instrument Design with Rasch IRT and Data Analysis I" **Unit Materials EDU 435/635**. School of Education. Murdoch University.
- Barnes, L. L.B. y S. L. Wise (1991): "The Utility of a Modified One-Parameter IRT Model with Small Samples" **Applied Measurement in Education**, 4 (2), pp. 143/157
- Berger, P. L. y T. Luckman (2003): **La construcción social de la realidad**. Amorrortu. Buenos Aires.
- Birnbaum, A. (1968): "Some Latent Trait Models," en Lord, F. M. y M.R. Novick[Eds.]
- Bond, T. G. y C. M. Fox (2001): **Applying the Rasch Model Fundamental Measurement in the Human Sciences**. New Jersey: Lawrence Erlbaum Ass. Inc.
- Boudon, R. (1999): "Attitudes", en **Encyclopaedia Universalis**, CD-ROM version 5, (citado por Leplège, 2003)

- Bock, R. D. (1972): "Estimating Item Parameters and Latent Ability When the Response are Scored in Two or More Nominal Categories," **Psychometrika**, 37, 29/51
- Brannic, M. (2005): Item Response Theory, <http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm>
- Campbell, N. R. (1920): **Physics: The Elements**. Cambridge: Cambridge University Press, (citado por Leplège, 2003)
- Campbell, D.T. y D. W. Fiske (1959): "Convergent and discriminate validation by the multitrait-multimethod matrix", **Psychological Bulletin**, Vol. 56, pp. 81/105
- Carretero-Dios, H y C. Pérez (2005): "Normas para el desarrollo y revisión de estudios instrumentales," **International Journal of Clinical and Health Psychology**, vol. 5, n° 3, pp. 521/551
- Churchill, G. A. (1979): "A Paradigm for Developing Better Measures of Marketing Constructs," **Journal of Marketing Research**, 16 (February), pp. 64/73
- Churchill, G. A y J. P. Peter (Research Design Effects on the Reliability of Rating Scales: A Meta – analysis," **Journal of Marketing Research**, 21 (November), pp. 360/375
- Cliff, N. (1992): Abstract measurement theory and the revolution that never happened. **Psychological Science** 3 (3), pp. 186 / 190 (citado por Wright, 1997).
- Cortada, N. (2004): "Teoría de Respuesta al Ítem: Supuestos básicos" **Evaluar** n° 4, pp. 95/110.
- Cronbach, L. J. (1946): "Response Sets and Test Validity," **Educational and Psychological Measurement**, 6, pp. 475/494 (citado por Andrich y Marais, 2005)
- Cronbach, L. J. (1951): "Coefficient Alpha and the Internal Structure of Tests," **Psychometrika**, 16 (September), pp. 297/334
- Cronbach, L. J. (1971): "Test Validation," en Thorndike, R.L. [Ed.]: **Educational Measurement**, 2nd ed., Washington D.C.: American Council on Education, pp. 443/507 (citado por Peter, 1981)
- Cronbach, L. J.; N. Rajaratnam y G.C. Gleser (1963):" Theory of Generalizability: A Liberalization of Reliability Theory," **British Journal of Statistical Psychology**, 16 (November), pp. 137/163 (citado por Peter, 1979).
- Cronbach, L. J.; G.C. Gleser; H. Nanda y N. Rajaratnam (1972): **The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles**. New York: John Wiley & Sons, Inc. (citado por Peter, 1979).
- Dickes, P., J. Tournois, A. Flieller y J.L. Kop (1994): **La psychométrie**. Paris: PUF (Le psychologue), (citado por Leplège, 2003).
- Dotson L. y G. F. Summers (1976): "Cómo elaborar escalas técnicas de Guttman" en Summers, G. F. [Ed.], pp. 248/261
- Duncan, O.D. (1984): **Notes on Social Measurement**. Russell Sage Foundation, New York
- Engelhard, G. (2001): Historical View of the Influence of Measurement and Reading Theories on the Assessment of Reading". **Journal of Applied Measurement** 2(1), pp. 1/26, (citado por Leplège, 2003).
- Everitt, B. S. (1984): **An Introduction to Latent Variable Models**. Londres: Chapman and Hall.
- Fischer, G. H. (1973): "Linear logistic test model as an instrument in educational research". **Acta psicológica**, 37, pp. 359/374.
- Fischer, G.H. e I.W. Molenaar [Eds] (1995): **Rasch models, Foundations, recent developments, and applications**. New York: Springer-Verlag
- Fisher, R.A. (1920). A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error and by the Mean Square Error. **Monthly Notices of the Royal Astronomical Society**, (53), 758/770. <http://www.rasch.org/memo66.htm>
- Fisher, W.P. jr. (1997): "Blind Guides to Measurement," **Rasch Measurement Transactions** 11(2) <http://www.rasch.org/rmt/rmt112g.htm>
- Fisher, W. P. jr (2005): "Meaningfulness, Measurement and Item Response Theory (IRT)," **Rasch Measurement Transactions** 19 (2) Autumn, pp.1018/1020 <http://www.rasch.org/rmt/rmt192d.htm>

- Frisbie D.A. (1988): "Reliability of Scores from Teacher – Made Tests". **Educational Measurement: Issues and Practice. National Council on Measurement in Education**, 7 (1) 25/35
- García del Junco, J. y M. del M. Ros (1997): "Ética y Dirección de Empresas en Europa" **ESIC-MARKET** N° 95, pp. 189/198
- García, A. M.; A. Sanfiel; P. Álvarez y J.R. Oreja (2005): "Agrupaciones de empresas según el número de vínculos interorganizativos establecidos mediante el Modelo de Rasch". **XV Jornadas Hispanos – Lusas de Sevilla 2005**. Publicado en C. Barroso y J.L. Galán [Eds.] (2005): **Cities in Competition. Notes on strategy, planning and internationalization**. Universidad de Sevilla, pp. 171 / 181.
- Glass, G. V. y J. C. Stanley (1970): **Statistical Methods in Education and Psychology**. Englewood Cliffs, N. J.: Prentice Hall, Inc.
- Gleser, G.C.; L.J. Cronbach y N. Rajaratnam (1965): "Generalizability of Score Influenced by Multiple Source of Variance," **Psychometrika**, 30 (December), pp. 395/418
- Guttman, L. (1944): "A Basis for Scaling Qualitative data". **American Sociological Review**, pp. 139 –150. Ver en Guttman, L. (1976a): "Bases para elaborar escalas con datos cualitativos", en Summers, G.F.[Ed.], pp. 213/229
- Guttman, L. (1947): "The Cornell Technique for Scale and Intensity Analysis," **Educational and Psychological Measurement**, vol. 7, pp. 247/280. Ver un extracto en Guttman, L. (1976b): "La técnica Cornell para el análisis de escalas y de intensidad," en Summers, G.F. [Ed.], pp. 230/247
- Guttman, L. (1950): "The Basis for Scalogram Analysis", en Stouffer et al. [Eds.]: **Measurement and Prediction**, vol. 4, pp. 60 – 90. Princenton, N.J. : Princenton University Press, (citado por Leplège, 2003).
- Hagquist, C. y D. Andrich (2004): "Is the Sense of Coherence-instrument Applicable on Adolescents? A Latent Trait Analysis Using Rasch modelling. **Personality and Individual Differences**, 36, pp. 955/968
- Hambleton, R.K. [Ed.] (1980): **Applications of Item Response Theory**. Vancouver, B.C.: Educational Research Institute of British Columbia (citado por Hambleton et al, 1991).
- Hambleton, R.K. y L.L. Cook (1977): "Latent trait models and their use in the analysis of educational test data" **Journal of Educational Measurement** 14(2), pp. 75/96 (citado por Fisher, 1977)
- Hambleton, R.K. y H. Swaminathan (1985): **Item Response Theory: Principles and Applications**. Boston: Kluwer (citado por Hambleton et al, 1991).
- Hambleton, R. K.; H. Swaminathan y H. J. Rogers (1991): **Fundamentals of Item Response Theory**. Newbury Park, Ca: Sage Publication, Inc.
- Kaplan, A. (1964): **The Conduct of Inquiry: Methodology for Behavioral Science**, Scranton, PA: Chandler Publishing Company (citado por Glass y Stanley, 1970).
- Kaiser, H. F. (1960): "Review of Measurement and Statistics by Virginia Senders". **Psychometrika** 25: 411/413
- Krantz, D. H. (1964): "Conjoint Measurement: the Luce – Tukey Axiomatization and Some Extensions", **Journal of Mathematical Psychology** 1, pp. 82 – 100, (citado por Leplège, 2003).
- Krantz, D. H., R. D. Luce, P. Suppes y A. Tversky (1971): **Foundations of Measurement**, vol. 1. New York: Academic Press, (citado por Leplège, 2003).
- Kuhn, T. (1977): "The Function of Measurement in Modern Physical Science", en T. Kuhn: **The Essential Tension**. Chicago, IL. : University of Chicago Press. Original publicado en 1961, *Isis* 52, pp. 161 – 193, (citado por Leplège, 2003).
- Kuder, G. F. & M.W. Richardson (1937): "The theory of the estimation of test reliability," **Psychometrika**, 2, pp. 151/160 (citado por Andrich y Marais, 2005).

- IIC (2003): "Las personas: el activo intangible de mayor valor para las organizaciones" **Jornada de Difusión Científico Técnica sobre Innovación en Capital Humano**, organizada por el IIC. Bajo el lema " <http://www.iic.uam.es/jornada2003.htm>
- Lazarsfeld, P. F. (1961): "Note on the History of Quantification in Sociology. Trends, Sources and Problems", en Wolf, H. [Eds.]: **Quantification, a History of the Meaning of Measurement in the Natural and Social Sciences**, pp. 147 – 203. New York: Bobbs – Merrill, (citado por Leplège, 2003).
- Lazarsfeld, P. F. (1977): **Análisis de la estructura latente**. Instituto de Estudios Políticos, Madrid.
- Leplège, A. (2003): Note technique: invariance et modèles de mesure dans les sciences sociales, introduction à l'oeuvre de Georg Rasch. **Information sur les sciences social** vol 42 –no 4, pp. 535 / 555
- Linacre, J. M. (1989): **Many – facet Rasch measurement**. Chicago: MESA Press
- Linacre, J.M. (1994): "Sample size and Item Calibration Stability," **Rasch Measurement Transactions**, 7 (4), p. 328 <http://www.rasch.org/rmt/rmt74m.htm>
- Linacre, J.M. (2002a): "What do Infit and Outfit, Mean-Square and Standardized mean?" **Rasch Measurement Transactions** vol. 16 (2), pp. 878
- Linacre, J. M. (2002b): "Facets, Factors, Elements and Levels" **Rasch Measurement Transactions** vol. 16 (2), pp. 880 <http://www.rasch.org/rmt/rmt162h.htm>
- Linacre, J. M. (2005): **Winsteps. Rasch measurement computer program**. Chicago: Winsteps.com. <http://www.winsteps.com/winman/index.htm>
- Loevinger, J. (1957): "Objective Tests as Instruments of Psychological Theory," **Psychological Reports**, 3, pp. 635/694 (citado por Carretero-Dios y Pérez, 2005)
- Lord, F. M. (1953): "On the Statistical Treatment of Football Numbers" **American Psychologist** 8: 750 / 751.
- Lord, F. M. (1980): **Applications of Item Response Theory to Practical Testing Problems**. Hillsdale, N.J. :Lawrence Erlbaum Associates.
- Lord, F. M. (1983): "Small N Justifies Rasch Model," en Weiss D.J. [Ed.]: **New Horizons in Testing: Latent Trait Theory and Computerized Adaptive Testing**. New York, NY: Academic Press, Inc., pp. 51/61 (citado por Fisher, 1997 y 2005)
- Lord (1984): "Standard Errors of Measurement at Different Score Levels" **Journal of Education Measurement**, 21, pp. 239/243
- Lord, F. M. y M.R. Novick [Eds.] (1968): **Statistical Theories of Mental Test Scores**. Reading, M.A.: Addison-Wesley.
- Luce, R. D. y Tukey, J. M. (1964): "Simultaneous Conjoint Measurement. A New Type of Fundamental Measurement", **Journal of Mathematical Psychology**, 1, 1 – 27, (citado por Leplège, 2003).
- Martin, O. (1997): **La mesure de l'esprit – Origine et développement de la psychométrie 1900 – 1950**. Paris : L' Harmattan, (citado por Leplège, 2003).
- Masters, G. N. (1982): A Rasch Model for Partial Credit Scoring. **Psychometrika**, 47 (2), pp.149 – 174, (citado por Bond y Fox, 2001).
- Masters, G.N. y B.D. Wright (1984): "The essential process in a family of measurement models". **Psychometrika**, (49), 529/544 (citado por Wright, B. D., 1997b)
- Messick, S. (1989): Meaning and Values in Test Validation: The Science and Ethics of Assessment. **Educational Researcher**, 18 (2) 5/11
- Michell, J. (1999): **Measurement in Psychology: A Critical History of a Methodological Concept**. Cambridge: Cambridge University Press, (citado por Leplège, 2003).
- Montero, I. y J. R. Oreja (2005 a): " Diseño de un instrumento de medida para el análisis comparativo de los recursos culturales tangibles". **PASOS Revista de Turismo y Patrimonio Cultural** Vol. 3(2), pp. 245/255. <http://www.pasosonline.org/Publicados/3205/PS030205.pdf>

- Montero, I. y J.R. Oreja (2005 b): "Actitudes de los artistas hacia las galerías de arte según su nivel profesional en el mercado del arte", ponencia admitida en las **XVI Jornadas Luso-Espanholas**. Universidad de Évora. Portugal.
- Morán, J. C. y P. Álvarez (2001): "Medida del desarrollo humano para los países de América Latina" **El Trimestre Económico** (Méjico) Vol. LXVIII (2) abril – junio N° 270, pp. 195 / 208
- Nunnally, J. C. (1967; 1978): **Psychometric Theory**, New York: McGraw - Hill
- Oreja, J. R. (1995): "El cambio organizativo en la sucursal bancaria", **ESIC-MARKET**, 88, abril-junio, pp. 63 / 81.
- Pardo, M (2004): "Internet como herramienta de los recursos humanos," http://www.cibersociedad.net/congres2004/grups/fitxacom_publica2.php?grup=37&id=507&idioma=es
- Peter, J. P. (1979): "Reliability: A Review of Psychometric Basics and Recent Marketing Practices" **Journal of Marketing Research**, 16 (February), pp. 6/17
- Peter, J. P. (1981): "Construct Validity: A Review of Basic Issues and Marketing Practices" **Journal of Marketing Research**, vol. XVIII (May), pp. 133/145
- Peterson, R.A. (1994): "A Meta-analysis of Cronbach's Coefficient Alpha," **Journal of Consumer Research**, 21 (September), pp. 381/391
- Perline, R., B.D. Wright y H. Wainer (1979): "The Rasch Model as an Additive Conjoint Measurement", **Applied Psychological Measurement** 3 (2), pp. 237 – 255, (citado por Leplège, 2003).
- Rasch, G. (1961): "On General Laws and the Meaning of Measurement in Psychology," en J. Neyman [Ed.]: **Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability**. IV, 321/334, Berkeley, CA: University of California Press <http://www.rasch.org/memo1960.pdf>
- Rasch, G. (1980): **Probabilistic Models for Some Intelligence and Attainment Tests** (Expanded Ed.) Chicago: University of Chicago Press, en originalmente publicado en 1960, Copenhagen: Denmark's paedagogiske Institut.
- RMT (2005): "Rasch Model vs. 1-PL IRT," **Rasch Measurement Transactions** 19:3 Winter, p. 1032. <http://www.rasch.org/rmt/rmt193.pdf>
- Ryan, J.P. (1983): "Introduction to Latent Trait Analysis and Item Response Theory," en W.E. Hathaway [Ed.]: **Testing in the Schools. New Directions for Testing and Measurement**, 19, San Francisco: Jossey-Bass, pp.49/65
- Salzberger, T. (2004): "Determining the Frame of Reference of a Scale Measuring the Attitude towards Advertising", **2nd International Conference on Measurement in Health, Education, Psychology and Marketing: Developments with Rasch and Unfolding Models**, School of Education, Murdoch University, Western Australia
- Samejima, F. (1969): "Estimation of Latent Ability Using a Response Pattern of Graded Scores." **Psychometrika Monographs**, 34 (17)
- Samejima, F. 1997. Graded Response Model. In W. J. Van der Linden, & R. K. Hambleton (Eds.), **Handbook of Modern Item Response Theory**. NY: Springer.
- Santos, J. L. (1999): **La satisfacción del turista en el destino Marbella. Medida y análisis mediante el Modelo Rasch**. CEDMA. Málaga. 1999.
- Santos-Arrebola J. L. (2002): "Application of the Rasch Model to Customer Satisfaction in the Destination of Marbella". **Journal of Quality Assurance in Hospitality & Tourism**, vol. 3(1/2) 2002
- Sarabia, F. J. (1999): **Metodología para la investigación en Marketing y Dirección de Empresas**. Madrid: Pirámide
- Senders, V.L. (1958): **Measurement and Statistics**. Oxford University Press. New York
- Siegel, S. (1956): **Nonparametric Statistics**. McGraw Hill. New York

- Spearman, C. (1904): "General Intelligence' Objectively determined and measured", **American Journal of Psychology**, 15, pp. 201 – 293, (citado por Peter, 1979).
- Spearman, C. (1910): "Correlation Calculated from Faulty Data". **British Journal of Psychology**, 3 (October), pp. 271/295 (citado por Peter, 1979).
- Stankov, L. y A. Cregan (1993): Quantitative and Qualitative Proprieties of an Intelligence Test: Series Completion", **Learning and Individual Differences** 5, pp. 137 – 169, (citado por Leplège, 2003).
- Stenner, A. J. (1990): "Objectivity: Specific and General". **Rasch Measurement Transactions**, 4 (3), p. 111 <http://www.rasch.org/rmt/rmt43a.htm>
- Stenner, A. J. (1994): "Specify objectivity – local and general" **Rasch Measurement Transactions**, 8 (3), p. 374 <http://www.rasch.org/rmt/rmt83e.htm>
- Stevens, S. S. (1959): Measurement, Psychophysics and Utility, Cap. 2 en C.W. Churchman y P. Ratoosh [Eds.]: **Measurement: Definitions and Theories**. New York: John Wiley (citado por Wright, 1997).
- Summers, G. F. (1976) [Ed.]: **Medición de actitudes**. Trillas. México
- Styles, I. y D. Andrich (1993): "Linking the Standard and Advanced Forms of the Raven's Progressive Matrices in Both the Pencil-and-paper and Computer-adaptive-testing Formats" **Educational and Psychological Measurement**, 53 winter, pp. 905/925
- Thissen, D. y M. Orlando (2001): "Item Response Theory for Items Scored in Two Categories," en Thissen, D. y H. Wainer [Eds.]: **Test Scoring**. Hillsdale, N.J.: Erlbaum.
- Thurstone, L. L. (1925): "A Method of Scaling Psychological and Educational Test". **Journal of Educational Psychology**, 16, 433/451 (citado por Andrich, 1989).
- Thurstone (1927): "A Law of Comparative Judgment" **Psychological Review** 34, pp. 273/286. Cita y revisión en Li, Z.; K.E. Cheng; Y.Wang; S.R. Hiltz y M.Turoff (2001): **Thurstone's Law of Comparative Judgment for Group Support** <http://web.njit.edu/~zx18078/Publication/BB004.PDF>
- Thurstone, L.L. (1928): Attitudes can be measured. **American Journal of Sociology**, vol. XXXIII (enero, 1928), pp. 529/554 (citado por Andrich y Hagquist 2004). Hay una versión en español en: Thurstone, L. L. (1976): "Las actitudes pueden medirse", en G. F. Summers (1976) [Ed.]: **Medición de actitudes**. Trillas. México, pp. 157 – 173.
- Thurstone, L.L. (1959): **The Measurement of Values**. Chicago, IL. University of Chicago Press (Citado por Leplège)
- Thurstone, L. L. y E. J. Chave (1976): "Los valores escalares", en G. F. Summers (1976) [Ed.]: **Medición de actitudes**. Trillas. México, pp. 174/181. Es una reproducción de **The Measurement of Attitudes**, (septiembre, 1929), pp. 45 / 56
- Torgerson, W. S. (1958): **Theory and Methods of Scaling**. New York: John Wiley (citado por Wright, 1997).
- Traub, R.E. y G.L. Rowley (1991): "Understanding Reliability". **Educational Measurement: Issues and Practice, National Council on Measurement in Education**, 10 (1) 37/45
- Tristán, A. (2001): Análisis de Rasch para todos. México:CENEVAL
- Weiss, D. J. (1995): "Improving Individual Differences Measurement with Item Response Theory and Computerized Adaptive Testing," en Lubinski, D. J. Y R.V. Dawis [Eds]: **Assessing individual differences in human behavior: New concepts, methods, and findings**. Palo Alto, CA: Davies-Black Publishing pp. 49/79
- Wright, B. D. (1992a): "Rasch vs. Birnbaum," **Rach Measurement Transactions**, vol. 5 (4), p. 178 <http://www.rasch.org/rmt/rmt54c.htm>
- Wright, B. D. (1992b): "IRT in the 1990s: Which Models Work Best," **Rach Measurement Transactions**, vol. 6 (1), pp. 196/200. <http://www.rasch.org/rmt/rmt61a.htm>
- Wright, B. D. (1997a): S.S. Stevens Revisited. **Rasch Measurement Transactions** vol. 11 (1), pp. 552 – 553. <http://www.rasch.org/rmt/rmt111n.htm>

- Wright, B. D. (1997b): “Fundamental Measurement for Outcome Evaluation” **IOM Research Memoranda Memo66** <http://www.rasch.org/memo66.htm>
- Wright, B. D. y M. H. Stone (1979): **Best test design**. Chicago: MESA Press. Hay versión en castellano Wright, B. D. y M. H. Stone (1998): **Diseño de mejores pruebas, utilizando la técnica Rasch**. México:CENEVAL
- Wright, B.D. y J. M. Linacre, J.M. (1997). **BIGSTEPS: Rasch Computer Program for All Two Facet Problems**. Chicago: MESA Press
- Wright, B. D. y G. N. Maters (1982): **Rating Scale Analysis**. Chicago: MESA Press
- Yánes – Estévez, V.; J. R. Oreja – Rodríguez y P. Álvarez (2004): “Computing the Perceived Environmental Uncertainty Functions by Rasch Model” **WSEAS Transactions on Business and Economics**. Issue 4, vol. 1 Octubre 2004, pp. 281 –285