

MEDICIONES, POSICIONAMIENTOS Y DIAGNÓSTICOS COMPETITIVOS

JUAN RAMÓN OREJA RODRÍGUEZ

© 2015 Fundación FYDE-CajaCanarias (Edición)

© 2015 El autor

Prohibida la reproducción total o parcial de este libro ni su incorporación a sistema informático alguno, ni su transmisión en cualquier forma o por cualquier medio sin el permiso escrito del autor.



Edita:

Fundación FYDE-CajaCanarias

Avda. Reyes Católicos, 31. 2º

38005 Santa Cruz de Tenerife

Tfno.: 922 22 78 35

Fax: 922 22 77 18

e-mail: administracion@fyde-cajacanarias.es

<http://www.FYDE-CajaCanarias.es>

Colección E-Book nº 9 de FYDE-CajaCanarias

ISBN: 978-84-606-6783-4

Depósito Legal: TF 152-2015

INDICE	9
PRESENTACIÓN	
CAPÍTULO 1º.- LA MEDICIÓN OBJETIVA: CONCEPTOS Y TEORÍAS	15
1.1.- INTRODUCCIÓN A LA MEDICIÓN OBJETIVA	16
1.1.1.- CONCEPTOS BÁSICOS DE MEDICIÓN.	16
1.1.2.- ESCALAS DE MEDIDA.	25
1.1.3.- CATEGORIZACIÓN DE DATOS	28
1.2.- LAS TEORÍAS DEL TEST	30
1.2.1.- LA TEORÍA CLÁSICA DEL TEST (TCT)	30
1.2.2.- LOS MODELOS DE VARIABLES LATENTES	40
1.2.2.1.- LA TEORÍA DE LA RESPUESTA AL ÍTEM (TRI)	41
1.2.2.2.- LA TEORÍA DE LA MEDICIÓN DE RASCH (TMR)	49
2.- DIAGNÓSTICO DE LOS SUPUESTOS BÁSICOS DE LA MEDICIÓN OBJETIVA	61
2.1.- ASUNCIONES TMR DEL CONSTRUCTO	62
2.1.1.- DIMENSIONALIDAD DEL CONSTRUCTO	62
2.1.2.- FIABILIDAD Y VALIDEZ GLOBAL DE LAS MEDIDAS DE LAS DOS FACETAS: SUJETOS ECONÓMICOS, EMPRESARIALES Y TURÍSTICOS vs. ÍTEMS O FACTORES DE POSICIONAMIENTO.	68
2.2.- ANÁLISIS DE LAS CATEGORÍAS DE LAS ESCALAS DE MEDICIÓN DE LOS CUESTIONARIOS.	75
2.2.1.- LA DISTRIBUCIÓN DE LAS RESPUESTAS A LAS CATEGORÍAS ORDENADAS DE UN ÍTEM	75
2.2.2.- LOS UMBRALES ENTRE LAS CATEGORÍAS ORDENADAS DE UN ÍTEM. EL UMBRAL RASCH-ANDRICH PARA ÍTEMS POLITÓMICOS CON CATEGORÍAS ORDENADAS.	76
2.2.3.- ESTRATEGIAS DE CATEGORIZACIÓN ANTE UMBRALES DESORDENADOS	77
3.-MEDICIÓN Y POSICIONAMIENTO COMPETITIVO INDIVIDUAL DE DOS FACETAS ANALIZADAS DE FORMA CONJUNTA.	83
3.1.- ANÁLISIS CONJUNTO LAS DOS FACETAS: EL MAPA DE WRIGHT.	84

3.1.1.- ESTRUCTURA DEL MAPA DE WRIGHT	84
3.1.2.- POSICIONAMIENTO DE DOS FACETAS	86
3.1.3.- LOGITS Y PROBABILIDADES DE ÉXITO	89
3.2.- MEDIDAS DE LA FACETA: SUJETOS (EMPRESAS, PRODUCTOS Y ENTES ECONÓMICOS, EMPRESARIALES Y TURÍSTICOS). POSICIONAMIENTOS, AJUSTES Y DESAJUSTES.	91
3.2.1.- MEDIDAS DE LOS SUJETOS Y NIVELES DE AJUSTES. POSICIONAMIENTOS COMPETITIVOS DE LOS SUJETOS.	91
3.2.2.- MEDIDAS Y POSICIONAMIENTOS DE SUJETOS QUE NO HAN PARTICIPADO EN EL ANÁLISIS CONJUNTO PREVIO, PERO QUE PERTENECEN A LA POBLACIÓN EN ESTUDIO	95
3.2.3.- ANÁLISIS DE LOS DESAJUSTES DE LAS MEDIDAS DE LOS SUJETOS.	97
3.3.- CALIBRACIONES DE LOS ÍTEMS, AJUSTES Y DESAJUSTES. JERARQUIZACIÓN DE LOS ÍTEMS. POSICIONAMIENTO COMPETITIVO.	99
3.3.1.- CALIBRACIONES DE LOS ÍTEMS Y NIVELES DE AJUSTE. JERARQUIZACIÓN DE LOS ÍTEMS.	99
3.3.2.- ANÁLISIS DE LOS DESAJUSTES EN LAS CALIBRACIONES DE LOS ÍTEMS.	102
3.3.3.- DISCUSIÓN DE LOS DESAJUSTES EN LAS CALIBRACIONES DE LOS ÍTEMS Y SUS ORÍGENES.	103
4.- MEDICIÓN Y POSICIONAMIENTO COMPETITIVO DE GRUPOS DE LAS DOS FACETAS.	105
4.1.-MEDICIÓN PARA GRUPOS DE LOS DOS FACETAS.	106
4.1.1.-POSICIONAMIENTO DE GRUPOS DE SUJETOS. VALIDACIÓN DE SUS DIFERENCIAS.	106
4.1.2.- JERARQUIZACIÓN DE GRUPOS DE ÍTEMS. VALIDACIÓN DE SUS DIFERENCIAS.	109
4.2.- COMPORTAMIENTO DIFERENCIAL DEL ÍTEM (DIF) PARA GRUPOS DE SUJETOS.	112
4.2.1.- SESGO Y FUNCIONAMIENTO DIFERENCIAL DE ÍTEMS.	112
4.2.2.- ASPECTOS OPERATIVOS DEL DIF: DETECCIÓN Y GRÁFICAS.	114

4.3.- ANÁLISIS LONGITUDINAL DE LAS DOS FACETAS.	119
4.3.1.- TIPOLOGÍA DE AGRUPAMIENTOS DE LAS DOS FACETAS PARA ANÁLISIS TEMPORALES.	119
4.3.2.- APROXIMACIÓN STACK.	121
4.3.3.- APROXIMACIÓN RACK.	124
5.-BENCHMARKING DE DIAGNÓSTICO.	127
5.1.- EL ESCALOGRAMA DE GUTTMAN.	127
5.1.1.- BASE CONCEPTUAL Y CONTENIDO	127
5.1.2.- INTEGRACIÓN DE LOS DESAJUSTES EN LAS MEDIDAS DE LAS FACETAS	129
5.1.3.- INFORMACIÓN SOBRE LAS DESVIACIONES DE LAS OBSERVACIONES RESPECTO A LAS EXPECTATIVAS.	131
5.2.- FICHAS DE DIAGNÓSTICO COMPETITIVO (PKMAP / KIDMAP)	131
5.2.1.- BASE TEÓRICA.	131
5.2.2.- REPRESENTACIONES GRÁFICAS SEGÚN TIPOLOGÍA DE CATEGORÍAS	134
5.3.- BENCHMARKING DE DIAGNÓSTICO	145
5.3.1.- CONSTRUCCIÓN DE LA MATRIZ INFORMATIVA	145
5.3.2.- INTERPRETACIÓN DE LAS FORTALEZAS Y DEBILIDADES. PROPUESTAS DE PUNTOS DE REFERENCIA ESTRATÉGICA DETERMINANTES DE OPORTUNIDADES Y AMENAZAS.	146
BIBLIOGRAFÍA	149

ÍNDICES DE TABLAS, CUADROS, FIGURAS Y GRÁFICOS

INDICE DE TABLAS

TABLA 1.1. EJEMPLO DE ESCALOGRAMA DE GUTTMAN PARA UN TEST CON ESCALAMIENTO DE 1 A 5	23
TABLA 2.1. VARIANZA DE LOS RESIDUALES ESTANDARIZADOS (EN UNIDADES DE AUTOVALOR)	64
TABLA 2.2 IDENTIFICACIÓN DE ÍTEMS Y CARGAS FACTORIALES	67
TABLA 2.3 FIABILIDAD GLOBAL DE SEPARACIÓN Y VALIDEZ GLOBAL DE LAS MEDIDAS DE LOS SUJETOS E ÍTEMS	73
TABLA 2.4 RESUMEN DE LA ESTRUCTURA DE 9 CATEGORÍAS	79
TABLA 2.5 RESUMEN DE LA ESTRUCTURA DE 6 CATEGORÍAS	81
TABLA 3.1 MEDIDAS, AJUSTES Y POSICIONAMIENTOS COMPETITIVOS DE LOS SUJETOS	92
TABLA 3.2 CONVERSIÓN DE LAS PUNTUACIONES TOTALES DE LOS SUJETOS EN MEDIDAS	95
TABLA 3.3 DESAJUSTES DE LAS RESPUESTAS DE LOS SUJETOS	97
TABLA 3.4 ANÁLISIS DE DESAJUSTES DE LOS SUJETOS MEDIANTE EL ESCALOGRAMA DE GUTTMAN	98
TABLA 3.5 CALIBRACIONES, AJUSTES Y JERARQUIZACIÓN DE LOS ÍTEMS O FACTORES DE POSICIONAMIENTO COMPETITIVO.	100
TABLA 3.6 DESAJUSTES EN LOS ÍTEMS	102
TABLA 4.1 MEDICIONES DE GRUPOS DE SUJETOS	107
TABLA 4.2 MEDICIONES DE GRUPOS DE ÍTEMS	110
TABLA 4.3 DIF GRUPOS DE SUJETOS, SEGÚN CRITERIO TAMAÑO, ENTRE SÍ	115
TABLA 4.4 DIF GRUPOS DE SUJETOS, SEGÚN CRITERIO TAMAÑO, CON RESPECTO AL CONJUNTO DE ÍTEMS	116
TABLA 4.5 RESUMEN DE INFORMACIÓN SOBRE DIF SIGNIFICATIVOS	118
TABLA 4.6 MEDIDAS DEL DINAMISMO DE LAS EMPRESAS EN LOS AÑOS 2000-2003	122
TABLA 4.7 CALIBRACIONES DEL DINAMISMO DE LOS ÍTEMS DEL ENTORNO EN LOS AÑOS 2000 Y 2003	124
TABLA 5.1 ESCALOGRAMA DE GUTTMAN	129
TABLA 5.2 ESCALOGRAMA DE GUTTMAN PARA DOS SEGMENTOS DE VEHÍCULOS	130
TABLA 5.3 ESCALOGRAMA DE GUTTMAN PARA UN GRUPO ESTRATÉGICO DE VEHÍCULOS	145
TABLA 5.4 MATRIZ DE BENCHMARKING DE DIAGNÓSTICO PARA UN GRUPO DE VEHÍCULOS	146

ÍNDICE DE CUADROS

CUADRO 1.1 RESUMEN DE LAS CARACTERÍSTICAS DE LAS ESCALAS DE STEVENS	25
CUADRO 1.2 ESTIMACIÓN DE PARÁMETROS EN LOS MODELOS TRI	47
CUADRO 2. 1 RELACIÓN ENTRE SEPARACIÓN, FIABILIDAD Y NIVELES	70
CUADRO 2.2 INTERPRETACIÓN DEL PARÁMETRO MNSQ (ESTADÍSTICO DE AJUSTE)	72
CUADRO 2.3 INTERPRETACIÓN DEL PARÁMETRO ZSTD (ESTADÍSTICO DE AJUSTE)	72
CUADRO 2.4 RECOMENDACIONES PARA LA OPTIMIZACIÓN DEL USO DE CATEGORÍAS EN ESCALAS ORDENADAS	78
CUADRO 3.1 PROBABILIDAD DE ÉXITO EN EL CASO DE ÍTEMS DICOTÓMICOS	89
CUADRO 3.2. PAUTAS DE RESPUESTAS A LOS ÍTEMS DICOTÓMICOS LIGADOS A LOS (DES)AJUSTES DE LOS SUJETOS	93
CUADRO 3.3. PAUTAS DE RESPUESTAS A LOS ÍTEMS POLITÓMICOS LIGADAS A LOS (DES) AJUSTES DE LOS SUJETOS	94
CUADRO 4.1 INDICADORES DE UN DIF (BOX Y FOX)	116
CUADRO 4.2 INDICADORES DE UN DIF (ETS)	117

ÍNDICE DE FIGURAS

FIGURA 1.1 CONTINUO LINEAL (UNIDIMENSIONALIDAD)	53
FIGURA 2.1 RELACIONES ENTRE FIABILIDAD, SEPARACIÓN Y NIVELES (3 NIVELES O ESTRATOS)	70
FIGURA 2.2 CURVAS DE LAS CATEGORÍAS DE UN ÍTEM Y UMBRALES	77
FIGURA 4.1 APROXIMACIÓN STACK	119
FIGURA 4.2 APROXIMACIÓN RACK	120
FIGURA 5.1 PKMAP MEORD (ASCII=NO)	135
FIGURA 5.2 PKMAP MRD (ASCII=WEBPAGE)	136
FIGURA 5.3 PKMAP MEORP	138
FIGURA 5.4 JERARQUIZACIÓN DE ÍTEMS Y CATEGORÍAS EN PKMAP	139

ÍNDICE DE GRÁFICOS

GRÁFICO 1.1. CURVAS CARACTERÍSTICAS DE ÍTEMS CON DISTINTOS NIVEL DE DIFICULTAD (B)	24
GRÁFICO 1.2. PROBABILIDAD DE RESPUESTA AFIRMATIVA DE UN ENCUESTADO $P[X_{Ni}] = 1$, EN FUNCIÓN DE LA DIFERENCIA ENTRE PARÁMETROS $\beta_N - \delta_i$	54
GRÁFICO 1.3 CURVAS DE LAS CATEGORÍAS DE UN ÍTEM Y SUS UMBRALES	58
GRÁFICO 2.1 PESOS DE LAS VARIANZAS DE LOS RESIDUALES ESTANDARIZADOS	65
GRÁFICO 2.2 RESIDUALES ESTANDARIZADOS DEL PRIMER CONTRASTE	66
GRÁFICO 2.3 ANÁLISIS DE LOS NIVELES O ESTRATOS	74
GRÁFICO 2.4 UMBRALES DE LAS 9 CATEGORÍAS	80
GRÁFICO 2.5 UMBRALES DE LAS 6 CATEGORÍAS	82
GRÁFICO 3.1. MAPA DE WRIGHT. HUECOS	86
GRÁFICO 3.2. MAPA DE WRIGHT. LA MEDIA DE LA MEDIDA DE LOS SUJETOS ESTÁ POR ENCIMA DE LA MEDIA DE LAS CALIBRACIONES DE LOS ÍTEMS	87
GRÁFICO 3.3. . MAPA DE WRIGHT. LA MEDIA DE LA MEDIDA DE LOS SUJETOS ESTÁ POR ENCIMA DE LA MEDIA DE LAS CALIBRACIONES DE LOS ÍTEMS	88
GRÁFICO 3.4 PROBABILIDAD DE ÉXITO EN EL LOGRO DE UN OBJETIVO	90
GRÁFICO 3.5 CURVA LOGÍSTICA QUE RELACIONA LA PUNTUACIÓN TOTAL Y LAS MEDIDAS DE UN SUJETO EN UN CUESTIONARIO	96
GRÁFICO 4.1 POSICIONAMIENTO COMPETITIVO DE GRUPOS DE SUJETOS	108
GRÁFICO 4.2 JERARQUIZACIÓN COMPETITIVA DE GRUPOS DE ÍTEMS	111
GRÁFICO 4.3 DIF (TAMAÑO) SIZE (DIF.)	118
GRÁFICO 4.4. ANÁLISIS STACK DEL DINAMISMO DEL ENTORNO ENTRE 2000 Y 2003	123
GRÁFICO 4.5 ANÁLISIS RACK DEL DINAMISMO DEL ENTORNO ENTRE 2000 Y 2003	125
GRÁFICO 5.1 PKMAP DE U-FIAT GRANDE PUNTO	141
GRÁFICO 5.2 PKMAP DE U-FORD FIESTA	142
GRÁFICO 5.3 PKMAP DE U-MITSUBISHI COLT	143
GRÁFICO 5.4 PKMAP DE U-TOYOTA YARIS 2006	144

PRESENTACIÓN

Este trabajo tiene como antecedente el Documento de Trabajo del Instituto Universitario de la Empresa (IUDE) de la Universidad de La Laguna, Serie Estudios 2005/47, titulado: **Introducción a la medición objetiva en Economía, Administración y Dirección de empresas: El Modelo de Rasch**, que ha sido modificado y ampliado en su versión actual.

En la Presentación de su primera versión se indicaba:

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 un 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 permite 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 son el resultado de cruzar los buscadores bibliográficos con la información disponible sobre grupos de investigación que trabajan con el Modelo de Rasch.

El conocimiento de los programas de computación permite el tratamiento de la información y generan 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 si no 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*

En su primera revisión en diciembre de 2005, se indicaba:

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

El documento de trabajo IUDE 2005/47 ha sido ampliamente utilizado por los miembros del IUDE y del Departamento de Economía y Dirección de Empresas de la Universidad de La Laguna, así como los alumnos que han cursado la asignatura de Análisis Competitivo de la Empresa de la Licenciatura y el Grado de Administración y Dirección de Empresas de la Universidad de La Laguna.

Este documento de trabajo ha aportado soporte metodológico a los participantes, provenientes de universidades nacionales, europeas e iberoamericanas, en los Workshops de Modelos de Rasch en Administración de Empresas que el Instituto Universitario de la Empresa ha venido organizando desde 2006 e incluido anualmente por Winsteps y el *Institute for Objective Measurement, Inc.* en su programación periódica de eventos de Rasch (*Coming Rasch-related Events*). Estando previsto celebrar en 2015 el Xº Workshop anual.

Las ponencias presentadas en estos Workshop anuales del IUDE se han publicado, hasta la terminación de esta publicación, en un total de 7 E-Books. La edición de estas publicaciones ha sido realizada en colaboración con la Fundación FYDE-CajaCanarias, que también edita este E-Book.

La inicial aplicación de la Teoría de la Medición de Rasch a la Administración de Empresas se ha generalizado en nuestro contexto a Economía, Empresa y Turismo incluyendo en esta publicación algunos trabajos que se han ido realizando en estos diez años de desarrollo docente e investigador en la Universidad de Laguna.

Fruto de estos trabajos se presenta esta publicación en donde hemos incluido las experiencias teórico - práctica en Mediciones, Posicionamientos y Diagnósticos Competitivos que reflejan la problemática de los diferentes apartados analizados. Su finalización ha contado con la revisión que han realizado del texto completo las profesoras de la Universidad de La Laguna D^a Esperanza Gil Soto y D^a Isabel Montero Muradas.

El Instituto Universitario de la Empresa de la Universidad de La Laguna ha publicado este libro, debiendo destacar el apoyo que he recibido del Director el profesor D. Jaime Febles Acosta y del PAS D. Carlos Ledesma González.

Agradezco a todos el apoyo recibido para la preparación, publicación y edición de este libro.

Juan Ramón Oreja-Rodríguez

La Laguna

19 de marzo de 2015

CAPITULO 1º

LA MEDICIÓN OBJETIVA: CONCEPTOS Y TEORÍAS

1.- LA MEDICIÓN OBJETIVA: CONCEPTOS Y TEORÍAS

1.1.- INTRODUCCIÓN A LA MEDICIÓN OBJETIVA

1.1.1.- Conceptos básicos de medición

En la metodología de la Economía, Empresa y Turismo 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), este planteamiento se debe reconducir en los términos expresados por Campell: *Medir es asignar número a observaciones de acuerdo a leyes científicas*. 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 cuando el físico británico Campbell negó la posibilidad de estimación cuantitativa de los eventos sensoriales (Campbell, 1940).

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”

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 de la 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 ítems, 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.

Antecedentes epistemológicos

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.

Planteamientos previos

THURSTONE

Fue un pionero de la aplicación de los métodos estadísticos al estudio de los procesos psíquicos (Thurstone, 1928; Tourstone y Chave, 1929/1976). 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:

1. La existencia de un continuo,
2. 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 explícitos.*

“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 está 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 produce 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:

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

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)

- *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 pueden 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 del 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 categorías ordenadas de 1 a 5, en la Tabla 1.1)

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 incluyan ítems con dificultad relativa que permitan verificar la progresión de los sujetos a lo largo del constructo.

TABLA 1.1
EJEMPLO DE UN ESCALOGRAMA DE GUTTMAN PARA UN TEST CON
ESCALAMIENTO DE 1 A 5

	Ítems										Total
	5	10	4	1	8	6	2	3	9	7	
Encuestados											
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

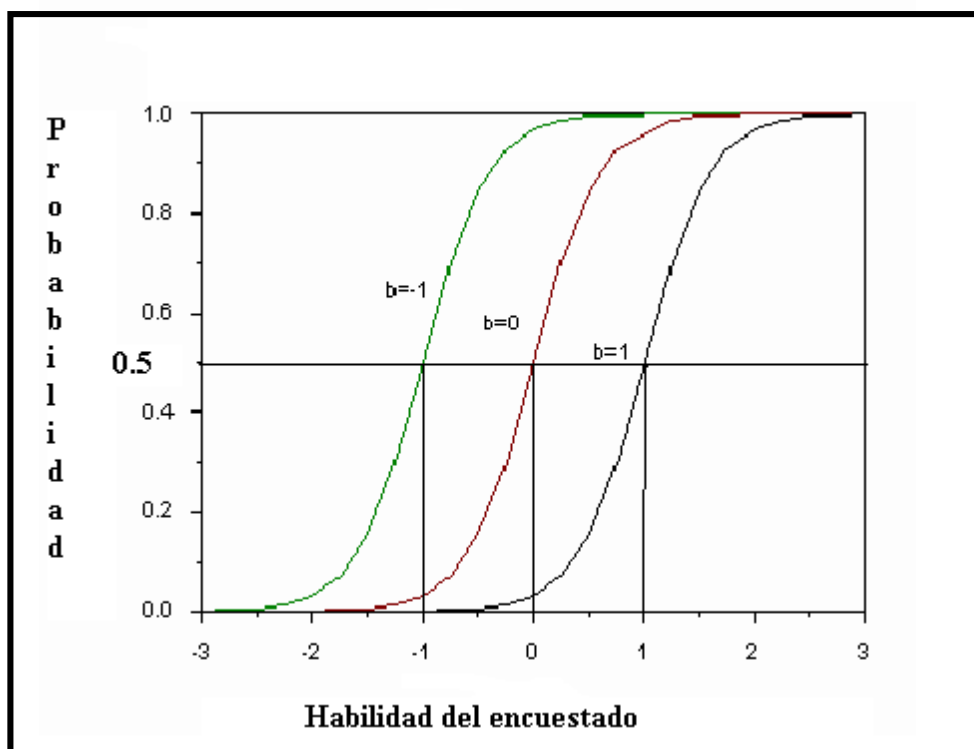
La principal limitación de este planteamiento es que no se tiene en cuenta la incertidumbre derivada de las opiniones, valores y juicios de los encuestados.

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 ítems 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 ítems de los encuestados. Una actitud, habilidad, atributo puede representarse como una variable latente (Lazarsfeld, 1961).

GRÁFICO N° 1.1
CURVAS CARACTERÍSTICAS DE ÍTEMS
CON DISTINTO NIVEL DE DIFICULTAD (B)



Fuente: Brannick (2005) y elaboración propia.

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 (CCI) 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 CCI, 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 (véase Gráfico 1.1).

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

1.1.2.- Escalas de medidas

Stevens (1959) indicaba que medir es la asignación de un número a eventos u objetos de acuerdo con una regla, definición que ha sido ya matizada en el contexto de la medición objetiva.

CUADRO N° 1.1

RESUMEN DE LAS CARACTERÍSTICAS DE LAS ESCALAS DE STEVENS

ESCALA	OPERACIÓN	EJEMPLO	ESTADISTICO CENTRAL	ESTADISTICO DISPERSIÓN	ASOCIACIÓN	TEST
NOMINAL	Igualdad	Contar ciudades	Moda	-	-	X^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		

Fuente: Elaboración propia, a partir de Stevens (1959)

En su trabajo trata de precisar qué tipo de aritmética era válida para los diferentes números (véase Cuadro nº1.1) que resultaban de las mediciones (Wright, 1997 a). Glaass y Stanley (1970), de acuerdo a los planteamientos de Stevens (1959), delimitaron las escalas nominales, ordinales, intervalo y ratio, como:

Medición Nominal. La aplicación de una escala de medida nominal se lleva a cabo para agrupar objetos en clases de acuerdo a un determinado atributo de los sujetos (objetos) a los que se les aplica el instrumento de medida del constructo delimitado. La operación que se aplica es la igualdad de conceptos y se procede a contar los existentes en cada categoría nominal.

A pesar de definirla como medición, no es posible considerarla como tal.

Nominal 3 875 23 12

Medición Ordinal. Mediante la aplicación de una escala ordinal se trata de determinar la presencia de diferentes niveles del constructo en el sujeto (objetos). La operación que se aplica es la desigualdad, determinando relaciones de orden entre los niveles del constructo en el sujeto (objeto).

No obstante, las distancias entre los números asignados no son iguales. Por lo tanto no se puede aplicar la adición, sustracción y multiplicación de los números que se han obtenido en aplicación de la escala ordinal.

Ordinal 1 2 3 4 5 6 7

Medición Intervalo. Los números obtenidos mediante la aplicación de la escala Intervalo permiten distinguir entre los distintos niveles del constructo disponible por el sujeto (objeto) (característica de la medición ordinal) sino que además se puede conocer la presencia de distancias iguales entre los sujetos (objetos). La operación que se destaca en la medición intervalo es la distancia. Por ello, los números asignados en el proceso de medición intervalo tienen las propiedades de orden y distancia, además la diferencia entre los números tiene significado (Glass y Stanley, 1970).

En la medición intervalo se asignan números de tal manera que las diferencias iguales en los números se corresponden a diferencias iguales en los niveles del constructo disponible por el sujeto (objeto).

Esta medición cuenta con una unidad de medida definida y el origen de la medición es arbitrario (no indica la ausencia del constructo).

Intervalo -3 -2 -1 0 1 2 3

Medición Ratio. En este caso el origen no es arbitrario sino que denota la total ausencia del constructo en el sujeto (objeto). Dispone de unidades de medida que reflejan la cantidad del constructo que tiene el sujeto (objeto).

Las diferencias iguales entre los números obtenidos reflejan diferencias iguales en la cantidad del constructo que tiene el sujeto (objeto).

Dado que el origen no es arbitrario sino absoluto los ratios entre las cantidades del constructo disponible por el sujeto A (objeto) pueden interpretarse como que el sujeto analizado tiene una, dos, veces las propiedades del constructo en relación al sujeto B.

En el caso de las medidas intervalo esta propiedad no tiene significado.

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 medición 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 (1970:14) 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 como tales y no como 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 significación a los números en 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”

1.1.3.- Categorización de datos

Los datos se pueden presentar en forma categórica (ajustados a una escala de categorías) y no categórica (no ajustados).

En algunos casos es necesario transformar los datos no categóricos en categóricos (escala ordinal) para el procesamiento de la información. Para ello es necesario conocer su distribución y proceder a asignar a tramos de datos no categóricos una categoría.

Los datos no categóricos pueden ser categorizados con diferentes métodos, Álvarez (2005) explica su método de la siguiente forma: “Se asignan códigos a los datos no categóricos de acuerdo al siguiente plan: A la magnitud más pequeña de cada agente se le asigna la categoría mínima o nivel de presencia “1”. Luego se le asigna la más alta categoría o nivel de presencia (1+max), para la más grande magnitud de cada agente. De esta forma, una presencia de (n+3) en el constructo de un agente coincide con una

presencia de (n+3) de otro ítem, aunque en su escala original cuantitativa fueran obviamente diferentes. Lógicamente, una categoría de (n+7) de un agente implica más presencia que (n+3) de otro agente del constructo. Aunque la escala original para la cantidad (n+7) puede ser menor que la original establecida en (n+3), la categorización efectuada fuerza a los diferentes agentes a un marco común. En general, categorías del agente referidas como A significa más presencia que categorización del agente B cuando $B < A$, por lo que A y B pueden compararse en términos de puntuaciones brutas, independientemente de sus unidades originales”.

El método que más frecuentemente hemos aplicado en nuestros trabajos de investigación y docentes es el de homogeneización de puntuaciones obtenidas al cumplimentar cuestionarios, números índices y los datos contables utilizados por el World Economic Forum e integrados en los informes anuales del World Economic Forum Global Competitiveness (Schwab, 2014) que, de forma generalizada, permitiría lograr una categorización politómica de escalas ordenadas de rango (1, N) aplicando la formulación, ajustada al nivel de la escala utilizada más cercano, explicada en el anexo metodológico de Salas et al (2014):

A) Para polaridades¹ positivas

$$(N - 1) * \frac{(puntuación\ del\ ítem_i - Min)}{(Max - Min)} + 1$$

B) Para polaridades¹ negativas

$$-(N - 1) * \frac{(puntuación\ del\ ítem_i - Min)}{(Max - Min)} + N$$

Siendo, en la escala: 1: nivel bajo; N: nivel alto

¹ Orientación de la escala con respecto a la variable latente

1.2.- LAS TEORÍAS DEL TEST

Bajo el término test se engloban los instrumentos de medida (cuestionarios) que permiten hacer operativo el constructo.

Una teoría de test modeliza las puntuaciones obtenidas en la aplicación del instrumento de medida al sujeto (objeto). Esta modelización permite la obtención de medidas referidas al nivel de disponibilidad del constructo por el sujeto y medidas de los ítems utilizados en el instrumento.

La diferencia entre las distintas teorías se centra en el modelo utilizado para la estimación de las medidas.

1.2.1.- Teoría Clásica del Test (TCT)

La mayoría de los trabajos de investigación en Economía, Empresa y Turismo 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 test 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 en la TCT, a partir de esos criterios son (Andrich y Marais, 2005: 26/27):

Ecuación general de la TCT: $X_v = T_v + E_v$ [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

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, Empresa y Turismo se está desarrollando un cuerpo científico construido a partir de la consideración de constructos. Éstos no pueden ser utilizados para desarrollos 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 en 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 medio 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 que los resultados obtenidos por la aplicación de un determinado instrumento de medida estén relacionados, dentro de lo esperado, con los resultados de otros instrumentos de medida del constructo relevantes.

Validez predictiva, apreciada en el caso de que los resultados obtenidos en un determinado instrumento de medida sean consistentes con los que se obtendrían 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 (divergente) 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 la 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.

En la estimación de un coeficiente de fiabilidad del test se parte de la idea de que existen dos test (test paralelos) para medir el mismo constructo.

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]

Cada test tiene su error de medida, pero la puntuación verdadera es la misma. A partir de la ecuación [1] se puede determinar:

$$X_{v1} = T_v + E_{v1}$$

$$X_{v2} = T_v + E_{v2}$$

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^2_t (varianza de la puntuación verdadera). La correlación entre estos dos test sería:

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

O, lo que es lo mismo:

$$S^2_t / S^2_x = r_{xx}. \quad [3]$$

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]
-------------------------------	-----------------------------------

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 Richarson (1937) desarrollaron la denominada fórmula KR-20, cuya expresión es:

$$KR - 20 = \left(\frac{k}{k-1} \right) \left[1 - \left(\frac{\sum p_i q_i}{S_x^2} \right) \right]$$

En donde:

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

$q_i = 1 - p_i$: (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 Cronbach:

$$\alpha = \frac{k}{k-1} \left[1 - \left(\frac{\sum_{j=1}^k S_i^2}{S_x^2} \right) \right]$$

La variación de la i sería: $\{1, \dots, k\}$

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, Empresa y Turismo. El rango de variación del coeficiente α es $\{0, \dots, 1\}$.

Las recomendaciones de Nunnally (1967; 1978), ampliamente difundidas en Economía, Empresa y Turismo por los trabajos de Churchill (1979) sobre los mínimos de fiabilidad aceptables (modificados entre las dos ediciones indicadas) para los estudios preliminares, básicos e investigaciones aplicadas (0,70/0,80/0,90) no han sido justificados teóricamente y han dado lugar a 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 test alternativos. La fiabilidad se determina, a partir de la consideración de que 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 test alternativos.

La crítica de la TCT

La principal ventaja de 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 dependen de la muestra utilizada en el instrumento de medida.
- 2) Las estimaciones del concepto analizado dependen 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 (Véase Fórmula [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:

1. 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).
2. No es posible asegurar directamente la validez de un test.

Se 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 utilizadas 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 (1979/1998), Lord (1980), Hambleton (1980), Hambleton y Swaminathan (1985) han planteado las características deseables de una alternativa a la TCT que serían (Hambleton et al, 1991):

- 1.- Puntuaciones de los ítems que no sean dependientes de la muestra utilizada.
- 2.- Puntuaciones de las percepciones de los encuestados respecto al constructo analizado que no sean dependientes 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

La Teoría de la Generalizabilidad

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 Generalizabilidad. 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 Generalizabilidad 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 ítems que representan los atributos a medir.

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

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 en 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 consistentes entre todas las posible facetas.

1.2.2.- Los Modelos de Variables Latentes

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 (variables) Latentes.

Los Análisis de Rasgos (variables) 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 ítems del test o la propia administración del test. Por tanto, si un constructo es unidimensional dependerá del propósito de la medición. Los tipos de ítems 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 perciben medidas con un alto nivel de error o la presencia de otras dimensiones en las que los ítems se 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 (variable) Latente se centran en la interacción de los sujetos con los ítems y los modelos estadísticos que recogen esta relación que se expresan gráficamente por las curvas características del ítem.

Hay dos grandes aproximaciones en los Modelos de Variables Latente. La diferencia entre ambas aproximaciones se puede establecer 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 de la variable latente y de las propiedades de los ítems. Como indican Thissen y Orlando (2001) los modelos de esta Teoría miden lo que son los ítems no lo que debería ser.

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

1.2.2.1.- La Teoría de la Respuesta al Ítem (TRI)

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. La relación 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 ítems 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 dependen del test.
- Las estimaciones de los parámetros de los ítems no dependen 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 considera al modelo de Rasch como un modelo TRI de un parámetro. No obstante, se pueden destacar algunas diferencias entre ambos (Linacre, 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:

$$P_i(\theta) = c + (1 - c_i) \frac{e^{1.7a_i(\theta - b_i)}}{1 + e^{1.7a_i(\theta - b_i)}}$$

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

e : 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)

En los diferentes modelos según el número de sus parámetros las expresiones serían:

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

$$P_i(\theta) = \frac{e^{1.7(\theta - b_i)}}{1 + e^{1.7(\theta - b_i)}}$$

El significado de los parámetros es el indicado 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 con los modelos de Rasch de la Teoría de la Medición de Rasch (Linacre, 2005)

$P_i(\theta)$ 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.

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) (Birnbbaum, 1968)

$$P_i(\theta) = \frac{e^{1.7a_i(\theta - b_i)}}{1 + e^{1.7a_i(\theta - b_i)}}$$

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 refieren 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) (Birnbbaum, 1968)

$$P_i(\theta) = c + (1 - c_i) \frac{e^{1.7a_i(\theta - b_i)}}{1 + e^{1.7a_i(\theta - b_i)}}$$

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 acomodó 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 (Cuadro 1.2).

CUADRO 1.2.
ESTIMACIÓN DE PARÁMETROS EN LOS MODELOS TRI

P_{ij}	δ_1	δ_2	δ_j	δ_{11}	$R_j = \sum P_i.$
β_1	P_{11}	P_{12}	P_{1j}	P_{111}	R_1
β_2	P_{21}	P_{22}	P_{2j}	P_{211}	R_2
β_i	P_{i1}	P_{i2}	P_{ij}	P_{i11}	R_i
β_{15}	P_{151}	P_{152}	P_{15j}	P_{1511}	R_{15}
$S_i = \sum P_{.j}$	S_1	S_2	S_j	S_{11}	

Fuente: Elaboración propia, a partir de Leplège, 2003

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

Siguiendo a Hambleton et al. (1991), la valoración del ajuste de los datos y el modelo se debe basar en tres 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 test 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, desde la Teoría de la Medición de Rasch (véase por ej. Wright, 1992a y 1992b; Ryan, 1983; Linacre (2005).

Se destaca que son esencialmente diferentes al utilizar procedimientos distintos. En ese contexto, incluso Ryan (1983) distingue entre los modelos de Rasgo Latente, que se refieren a la medición de las características subyacentes o latentes de los sujetos (Modelos de Rasch), y los de Respuesta al Ítem que describen y resumen estadísticamente las puntuaciones observadas.

1.2.2.2- La Teoría de la Medición de Rasch

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 Medición 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 suficientes 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 aun 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.

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 (si/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

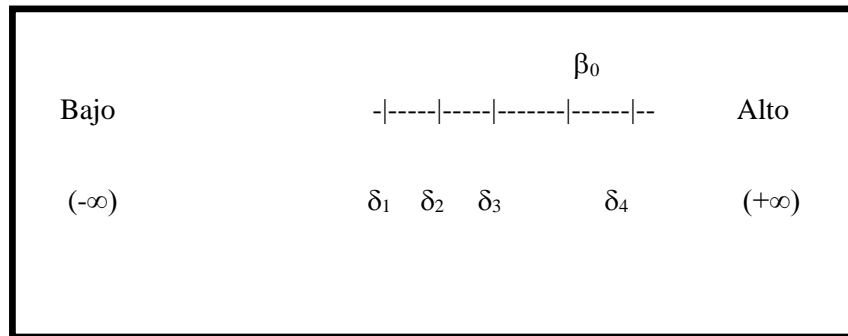
En el modelo de Rasch (1960/1980) se considera una variable latente en la cual dos diferentes entidades interactúan. Por una parte las unidades de investigación por medio de los sujetos económicos encuestados y por otra los ítems del instrumento de medida diseñado para llevar a cabo la investigación: 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 ítems 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 : habilidad de los sujetos y δ : dificultad de los ítems 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 ítems $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ en la escala lineal representativa de la variable latente. Los extremos de esta escala se establecen como: bajo y alto.

FIGURA 1.1
CONTINUO LINEAL (UNIDIMENSIONALIDAD)



Fuente: Elaboración propia

Los ítems δ_1 , δ_2 , δ_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:

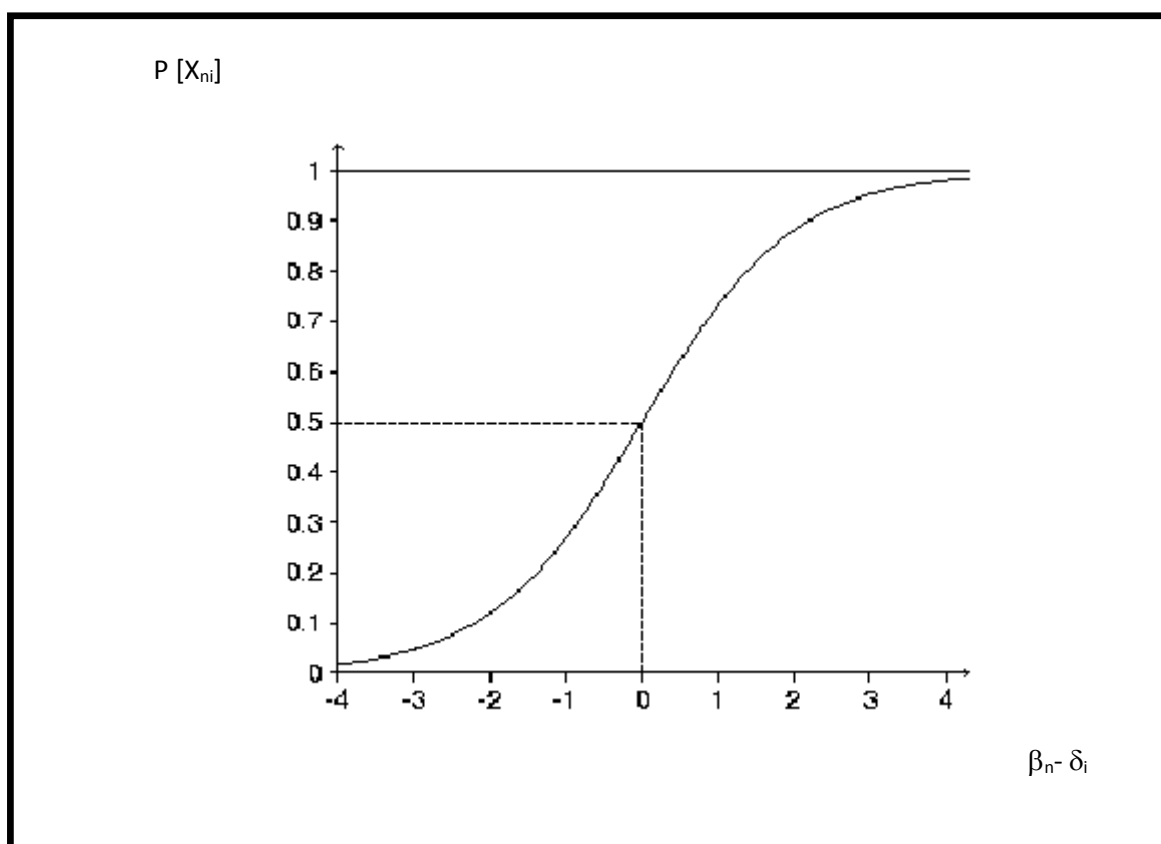
Si $\beta_n > \delta_i$, $(\beta_n - \delta_i) > 0$, entonces $P[X_{ni}=1] > 0.5$

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 $P[X_{ni}=1] = 0.5$

GRÁFICO 1.2

**PROBABILIDAD DE RESPUESTA AFIRMATIVA UN ENCUESTADO $P[X_{ni}]$
EN FUNCIÓN DE LA DIFERENCIA ENTRE PARÁMETROS $\beta_n - \delta_i$**



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 (odds ratio):

$$\frac{P}{1 - P}$$

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:

$$L \left[\frac{P}{1-P} \right] = \beta - \delta \quad (1)$$

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:

$$\frac{P}{1-P} = e^{(\beta - \delta)} \quad (2)$$

Cuyo campo de variación es:

$$0 \leq e^{(\beta - \delta)} \leq \infty$$

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

$$P[X_{ni} = 1 ; \beta_n, \delta_i] = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \quad (3)$$

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

$$0 \leq \frac{e^{(\beta - \delta)}}{1 + e^{(\beta - \delta)}} \leq 1$$

Los parámetros β_n y δ_i en el caso del programa de computación Winsteps se estiman mediante el método de máxima verosimilitud (Linacre, 2014 a).

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.

En este modelo cada ítem tiene su propia escala de medida.

La función de respuesta

$$P [X_{ni} = x] = \frac{1}{\gamma} e^{\sum_{k=0}^x (\beta_n - \delta_{ik})}$$

Siendo:

n: 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,

$$\gamma = \sum_{h=0}^n e^{\sum_{k=0}^x (\beta_h - \delta_{ik})} \quad \forall x \in \{0, 1, \dots, m\}$$

Modelo de Rasch para Categorías Ordenadas

El desarrollo de este modelo de Rasch para ítems politómicos para Categorías Ordenada 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 para ítems que comparten la misma escala de medida.

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

$$P[X_{ni} = x] = \frac{1}{\gamma} e^{(x(\beta_n - \delta_i) - \sum_{k=1}^x \tau_{ki})}$$

Siendo:

n : sujeto encuestado

i : ítem contestado

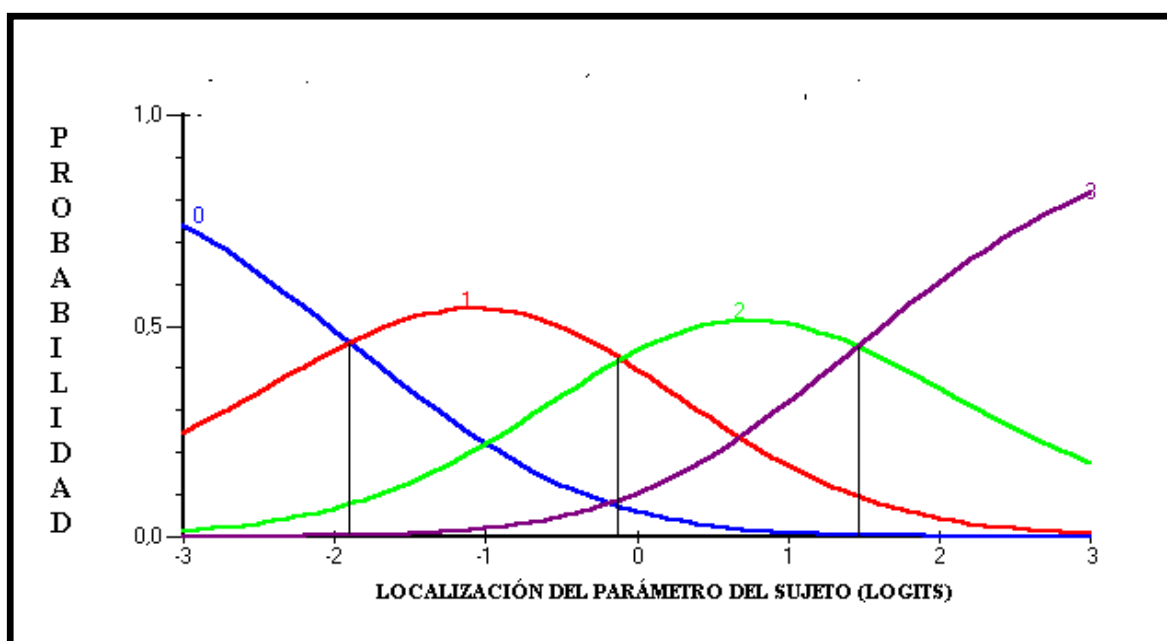
γ : Suma de todos los posibles numeradores, que surgen de acuerdo con el número de categorías de los ítems.

τ : Umbrales entre las categorías de los ítems.

Los umbrales (Andrich y Marais, 2005) 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 (véase Gráfico 1.3.)

GRÁFICO 1.3.
CURVAS DE LAS CATEGORÍAS DE UN ÍTEM Y SUS UMBRALES



Fuente: Elaboración propia con el programa RUMM2020 (Andrich et al, 2004).

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 categoría 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 (FACETS)

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 ítems 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 (ítems).

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

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 son amplias en Economía, Empresa y Turismo², siendo desarrollado por Linacre (1989) y contando con un programa de computación propio como es el Facets de la empresa Winsteps (Linacre, 2013).

Otros desarrollos

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

² Véase la propuesta de diseños de medición realizada por parte de Lunz y Linacre, 1998 como uno de los modernos métodos para la investigación empresarial y la aplicación de este modelo a la valoración por medio de expertos de la situación competitiva de destinos turísticos en Parra-López y Oreja-Rodríguez, 2014.

CAPÍTULO 2º

DIAGNÓSTICO DE LOS SUPUESTOS BÁSICOS DE LA MEDICIÓN OBJETIVA

2.- DIAGNÓSTICO DE LOS SUPUESTOS BÁSICOS DE LA MEDICIÓN OBJETIVA

2.1.- ASUNCIONES TMR DEL CONSTRUCTO

2.1.1.- Dimensionalidad del constructo

El modelo de Rasch permite la obtención de medidas de intervalo para constructos psicométricamente unidimensionales. La presencia de tensiones de multidimensionalidad en los constructos analizados puede derivarse tanto de aspectos aleatorios de los datos, de perturbaciones e intensificación de las relaciones entre ítems o a la presencia de ítems que representan otros constructos.

Conceptualmente se ha pasado de un planteamiento dicotómico sobre su existencia (se es unidimensional o no) a un planteamiento de continuo lineal ¿Cuál es el grado de multidimensionalidad que puede ser admitido en el análisis?

El análisis de la unidimensionalidad se ha llevado a cabo básicamente a partir de la consideración de Análisis de Componentes Principales de los Residuales de los ítems (PCAR) que permitiría detectar otros factores de dimensionalidad que influyen en las medidas una vez que ha sido determinado y removido el “factor Rasch”. En sus trabajos relacionados con la ejecución del programa de computación Winsteps, Linacre (2009) destaca la varianza explicada total, así como la derivada de los sujetos e ítems³.

³ La interpretación de los resultados obtenidos en la ejecución de PCAR del programa Winsteps en Linacre (2014b): “Table 23.0 Variance components for ítems” en Winsteps Help for Rasch Analysis / Contents / Output Table Index

Su discusión conceptual se puede seguir en Linacre (2014b): “Dimensionality: contrasts & variances” y “Dimensionality: when is a test multidimensional?” en Winsteps Help for Rasch Analysis / Contents /Special Topics

La “regla de oro” propuesta por Linacre (2009) para la determinación de la presencia de la unidimensionalidad es:

- Varianza explicada por los ítems > 4 veces la varianza del primer contraste:
Bueno
- Varianza explicada por las medidas $> 50\%$: Bueno
- Varianza no explicada.
- Autovalor explicado por el primer contraste < 3 Bueno; 1.5 Excelente
- Varianza explicada por el primer contraste $< 5\%$ Excelente

No obstante, no siempre está tan clara la presencia de la unidimensionalidad, por lo que hacen falta análisis complementarios para determinar el alcance de la influencia multidimensional en las medidas obtenidas en el modelo de Rasch.

No hay un valor mínimo de varianza explicada para todos los trabajos empíricos en el contexto del modelo de Rasch, dependerá de la base de datos. Linacre (2009b) indica que para pequeñas distribuciones de sujetos e ítems la varianza explicada se puede predecir que será inferior al 50 %. Así se especifica que para el Test de Cubo de Knox⁴ la varianza explicada es del 71.1%; para el estudio de “Liking for Science” (example0.txt) es del 50.8%; “Datos NSF” (interest.txt): 37.5%; “Datos NSF” (agree.txt): 30.0%; Test CAT (exam5.txt):29.5%

El proceso del PCAR (aplicado a los residuales no a los datos originales) procede a descomponer la matriz de correlaciones para encontrar componentes (o factores latentes) con los que los ítems puedan tener una alta correlación (carga factorial).

Después de encontrar el primer componente (dimensión de Rasch), el efecto de ese componente se retira de la matriz de correlación y el PCAR busca la matriz revisada del

⁴ Linacre (2014 a): exam1.txt. En los ficheros de ejemplos del Winsteps

segundo componente, continuando este proceso en el software de computación Winsteps hasta lograr la presentación de cinco componentes.

Un caso de análisis de la dimensionalidad.

La dimensionalidad de los puntos de referencia externos de la empresa canaria (Oreja - Rodríguez, 2010).

En este análisis se ha delimitado un constructo: Puntos de Referencia Externos, representado por 28 ítems en su instrumento de medida.

En la ejecución del modelo de Rasch para una muestra de 387 empresas se ha procedido a determinar el nivel de unidimensionalidad del mismo. Los resultados se presentan en la Tabla 2.1⁵

TABLA 2.1
VARIANZA DE LOS RESIDUALES ESTANDARIZADOS
(EN UNIDADES DE AUTOVALOR)

WINSTEPS Table 23.0					
INPUT: 387 PERSONAS 28 ITEMS MEDIDAS: 386 PERSONAS 28 ITEMS 140 CATS 3.68.1					
		-- Empirical --		Modeled	
Total raw variance in observations	=	40.1	100.0%		100.0%
Raw variance explained by measures	=	12.1	30.2%		30.0%
Raw variance explained by persons	=	3.2	7.9%		7.8%
Raw Variance explained by items	=	9.0	22.4%		22.2%
Raw unexplained variance (total)	=	28.0	69.8%	100.0%	70.0%
Unexplned variance in 1st contrast	=	3.3	8.2%	11.7%	
Unexplned variance in 2nd contrast	=	2.2	5.6%	8.0%	
Unexplned variance in 3rd contrast	=	2.0	4.9%	7.0%	
Unexplned variance in 4th contrast	=	1.7	4.3%	6.2%	
Unexplned variance in 5th contrast	=	1.6	4.0%	5.7%	

Fuente: Elaboración propia

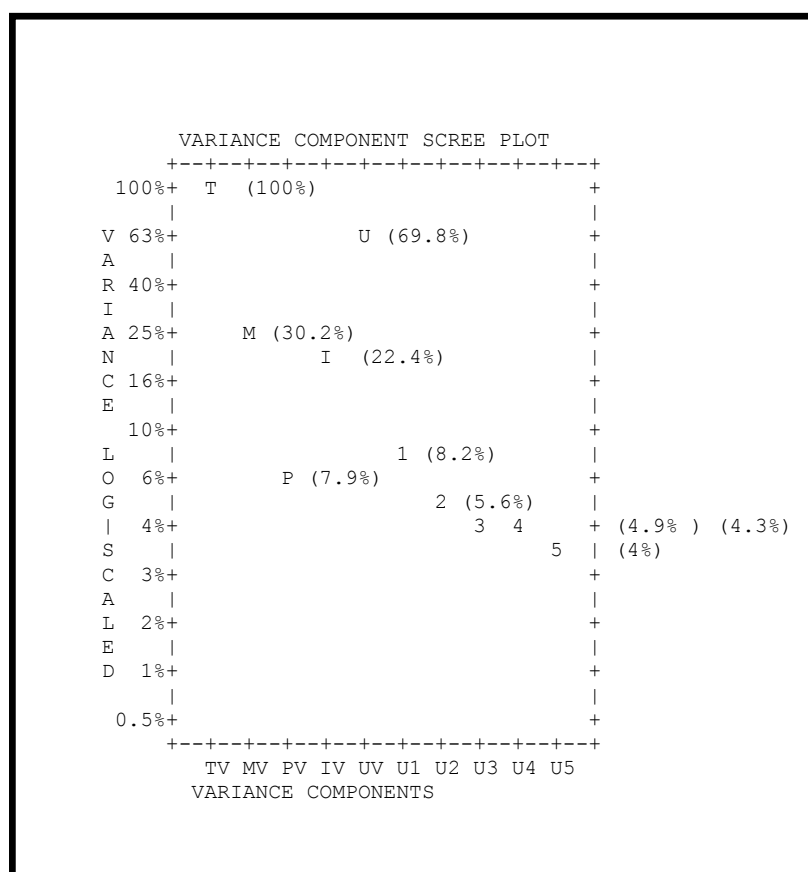
⁵ Linacre (2014b): Table nº 23.0 de Winsteps en Winsteps Help for Rasch Analysis / Contents / Output Table Index /

La dimensión Rasch explica el 30.2% de la varianza de los datos. Los resultados empíricos se aproximan a los modelados, que correspondería al caso de que los datos se ajustasen perfectamente al modelo de Rasch.

La segunda dimensión es el “primer contraste” que explica el 8.2% de la varianza. Este nivel de explicación es algo mayor que el 4% que suele observarse en los datos en las simulaciones de ajuste al modelo de Rasch⁶. La varianza explicada por los ítems 22.4% es inferior a tres veces la varianza explicada por el primer contraste (8.2%) por lo que se puede pensar que las medidas del constructo está influidas por una segunda dimensión.

El peso porcentual de las distintas medidas y contrastes se recoge en el Gráfico 2.1.

GRÁFICO 2.1
PESOS DE LAS VARIANZAS DE LOS RESIDUALES ESTANDARIZADOS



Fuente: Elaboración propia

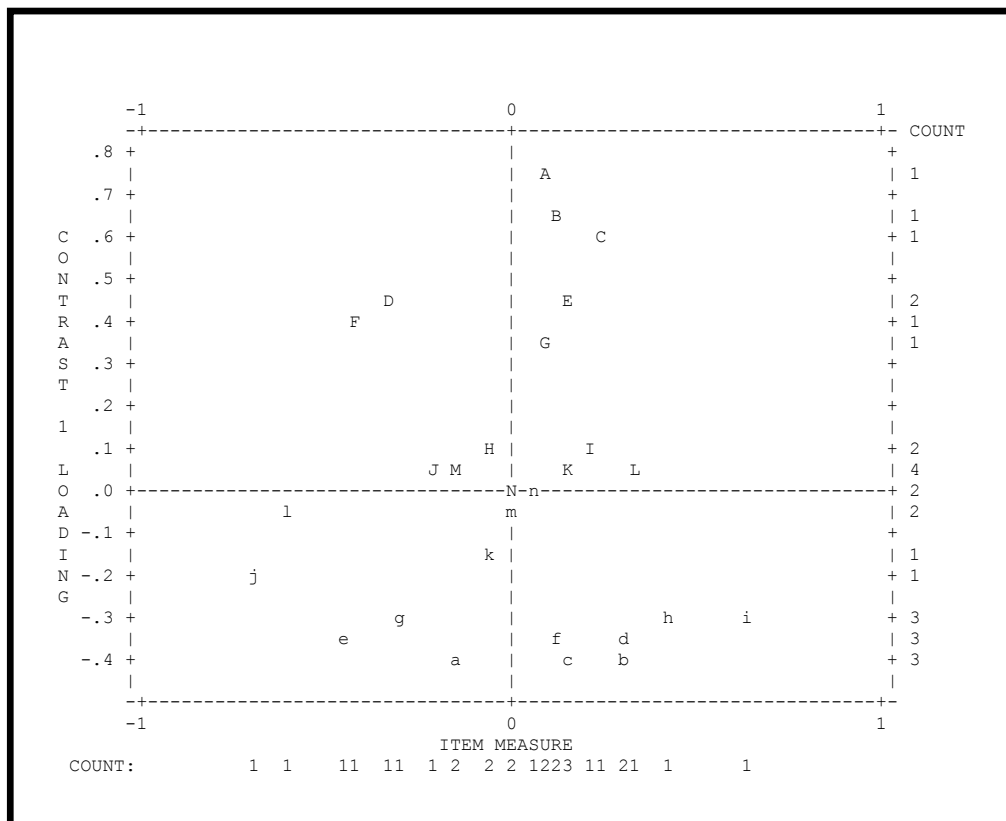
⁶ Linacre (2014b): Dimensionality, contrast & variances en Winsteps Help for Rasch Analysis / Contents / Special Topics

El autovalor del primer contraste es 3.3, lo que indica que tiene aproximadamente la fuerza de tres ítems, algo mayor que el número de ítems que se pueden considerarse para considerar una “dimensión”, desde la perspectiva de las simulaciones aleatorias, lo que nos lleva a considerar la necesidad de efectuar análisis complementarios para determinar si estamos en presencia de una dimensión secundaria o simplemente ante una pequeña perturbación en los datos.

Los componentes de la varianza están expresados en una escala logarítmica en el eje y.

Se puede deducir la presencia de otra dimensión a partir de estos datos. Para ello se procederá a analizar el Gráfico 2

GRÁFICO 2. 2
RESIDUALES ESTANDARIZADO DEL PRIMER CONTRASTE



Fuente: Elaboración propia

El eje x es la dimensión de Rasch (DR) en logits (medida de los ítems).

El eje y es la dimensión “Contraste”. Los números que se expresan son las cargas factoriales (correlaciones con la DR).

En lo más alto hay tres ítems A B C con las cargas factoriales más altas (+).

En lo más bajo del cuadro está presente un clúster de ítems (a b c d...) con las cargas factoriales más bajas (-)

TABLA 2.2
IDENTIFICACIÓN DE ÍTEMS Y CARGAS FACTORIALES

CON-	TRAST	LOADING	MEASURE	INFINIT	OUTFIT	ENTRY	NUMBER	ITEM
			MNSQ	MNSQ				
1		.75	.09	.83	.83	A	11	Política administración autonó
1		.65	.12	.84	.83	B	12	Política administracion insula
1		.62	.24	.88	.88	C	10	Política administración centra
1		.46	-.32	.87	.86	D	9	Fiscalidad
1		.44	.15	.89	.90	E	13	Política administración munic
1		.41	-.42	.92	.92	F	8	Costes laborales
1		.36	.09	.86	.89	G	7	Cambio de la normativa laboral
1		.10	-.07	1.13	1.14	H	18	Lentitud en la Admón.
1		.09	.21	1.09	1.09	I	15	Acceso a cursos de formación e
1		.07	-.21	.92	.92	J	3	Tipos de interés
1		.06	.16	.98	1.06	K	5	Paro
1		.06	.35	.99	1.02	L	14	Asociacionismo empresarial
1		.05	-.16	1.16	1.17	M	4	Cambio tecnológico
1		-.42	-.15	1.09	1.07	a	22	Hábitos de consumo
1		-.41	.31	1.07	1.09	b	24	Presión de los distribuidores
1		-.39	.16	1.13	1.12	c	23	Presión de los proveedores
1		-.37	.32	1.09	1.10	d	26	Los productos sustitutivos
1		-.35	-.45	1.14	1.14	e	27	Los competidores actuales
1		-.33	.11	.92	.92	f	21	Protección al consumidor
1		-.29	-.32	1.06	1.06	g	28	La amenaza de los nuevos compe
1		-.28	.43	.99	1.03	h	19	Ley del comercio (horario come
1		-.28	.65	.88	.89	i	20	Cambio de tipo de comercio
1		-.18	-.69	1.20	1.24	j	25	Presión de los clientes finale
1		-.15	-.06	1.06	1.05	k	17	Disponibilidad de medios de co
1		-.06	-.60	.92	.89	l	2	Inflación
1		-.04	-.01	1.05	1.12	m	6	Tipo de cambio de la peseta
1		-.02	.07	1.05	1.04	n	16	Asesoramiento externo
1		-.01	.00	1.10	1.13	N	1	Pertenencia a la UE

Fuente: Elaboración propia

En la Tabla 2.2 se presenta la identificación de esos ítems y sus cargas factoriales. Los ítems que parecen configurar una segunda dimensión se refieren a las políticas de índole superior a la “política administrativa municipal”, en contraste con los ítems con cargas negativas (verdes) y un grupo neutral con cargas positivas (azules).

2.1.2.- FIABILIDAD Y AJUSTE GLOBAL DE LAS MEDIDAS DE LA DOS FACETAS: SUJETOS ECONÓMICOS, EMPRESARIALES Y TURÍSTICOS vs. ITEMS O FACTORES DE POSICIONAMIENTO.

Se desarrollan los conceptos de fiabilidad de separación, separación y ajuste (validez), de acuerdo al manual del programa Winsteps (Linacre, 2014b)⁷.

Fiabilidad de separación y separación

Fiabilidad (índice de separación) significa: “reproductibilidad de la localización relativa de las medidas”.

Se expresa 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

Cuanto más alto sea el valor obtenido, mejor es la separación existente y más precisa es la medición.

Alta fiabilidad indica que hay una alta probabilidad de que las medidas de los sujetos (o de los ítems) que han sido estimadas con alto nivel de medición tengan más altas medidas que los sujetos (o ítems) estimados con bajas medidas.

Para mejorar el nivel de fiabilidad es necesario muestras amplias y/o bajos niveles de error de las medidas. Se logran los más altos niveles de fiabilidad de los sujetos (test) si se dispone de una muestra de sujetos con un amplio rango de habilidades y / o un instrumento de medida con muchos ítems (o amplias escalas de medición).

Se logran los más altos niveles de fiabilidad de los ítems si se dispone de un cuestionario con un amplio rango de dificultad de los ítems o una amplia muestra de sujetos. Frecuentemente, aparece un bajo nivel de fiabilidad de las medidas de los ítems si el

⁷ La interpretación de la fiabilidad global de separación y del ajuste global de las medidas de sujetos e ítems obtenidas en la ejecución del programa Winsteps está disponible en Linacre (2014b): “Table 3.1 Summaries of persons and items” en Winsteps Help for Rasch Analysis / Contents / Output Table Index

Su discusión conceptual se puede seguir en Linacre (2014b): “Reliability and separation of measures”, “Dichotomous mean-square fit statistics”, “Polytomous mean-square fit statistics “ en Winsteps Help for Rasch Analysis / Contents /Special Topics

tamaño de la muestra de los sujetos es muy pequeño para poder garantizar la reproductibilidad de la jerarquía de la dificultad de los ítems.

En el caso de que haya datos ausentes se reduce la precisión de las mediciones, por lo que la fiabilidad es menor. Linacre (2011) sugiere suprimir los sujetos con datos ausentes al estimar la fiabilidad.

Relación entre fiabilidad y separación

La determinación de la fiabilidad se efectúa a partir de la varianza “verdadera” y el error de la varianza. Dado que son difíciles de conceptualizar directamente, ya que son valores cuadráticos, es más sencillo utilizar sus raíces cuadradas: la desviación estándar “verdadera” (TSD) y la raíz cuadrada de la media de los errores estándar al cuadrado (RMSE).

El coeficiente de separación se define como el ratio entre la TSD y el RMSE. Este ratio se establece en unidades de RMSE, lo que le hace fácilmente interpretable.

El índice de fiabilidad de la separación (sujetos o ítems) es equivalente al coeficiente KR-20 o a la Alpha de Cronbach.

La relación entre ambos índices es:

$$\text{Fiabilidad de separación: } \frac{\text{Coeficiente de separación}^2}{1 + \text{Coeficiente de separación}^2}$$

$$\text{Coeficiente de Separación: } \sqrt{\frac{\text{Fiabilidad de separación}}{1 - \text{Fiabilidad de separación}}}$$

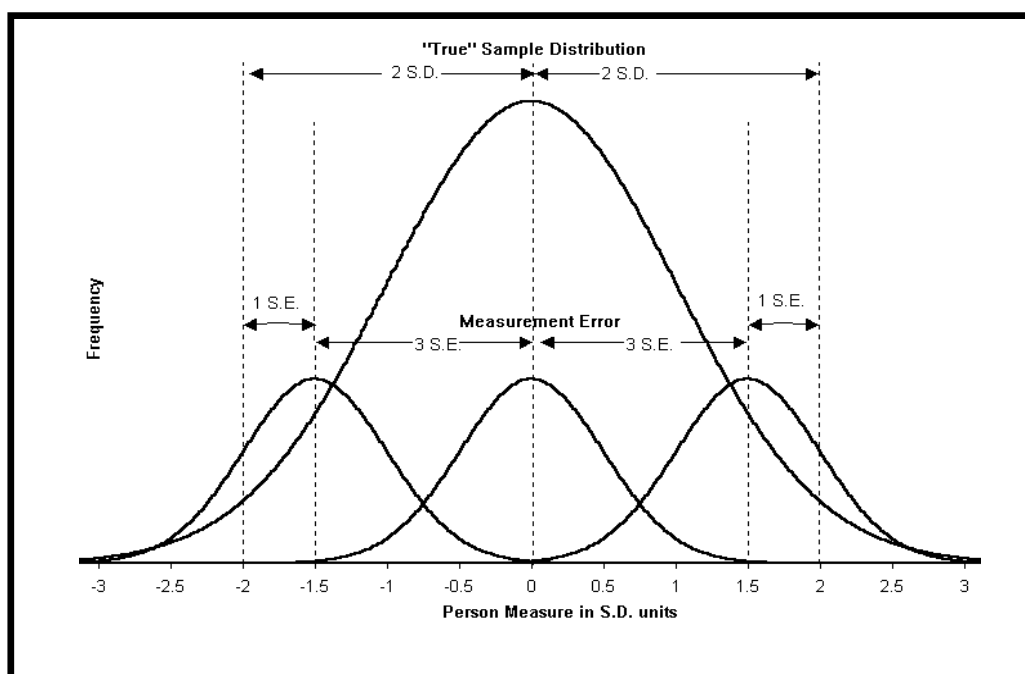
Se puede establecer una equiparación entre el coeficiente de Separación y el número de niveles que se observan en el posicionamiento (Wright, 1996; Wright y Master, 2002). El Cuadro 2.1 y la Figura 2.1 se presentan las relaciones entre fiabilidad, separación y niveles (3 niveles o estratos).

CUADRO 2.1
RELACIÓN ENTRE SEPARACIÓN, FIABILIDAD Y NIVELES
(3 NIVELES O ESTRATOS)

Error RMSE	True SD	True Variance	Observed Variance	Signal- to-Noise Ratio	Separation = True SD / RMSE	Strata (Niveles) = (4*Sep.+1) / 3	Reliability (Fiabilidad) = True Variance / Observed Variance
1	0	0	1	0	0	0	0
1	1	1	2	1	1	1.67	0.5
1	2	4	5	2	2	3	0.8
1	3	9	10	3	3	4.33	0.9
1	4	16	17	4	4	5.67	0.94

Fuente: Linacre (2011)

FIGURA 2.1
RELACIONES ENTRE FIABILIDAD, SEPARACIÓN Y NIVELES
(3 NIVELES O ESTRATOS)



Fuente: Wright, 1996; Wright y Masters, 2002

Ajuste (validez)

Siguiendo a Bond y Fox (2007: 57) y Linacre (2014b) se puede establecer que 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.

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.

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.

El estadístico INFIT correspondería al ajuste interno.

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.

En ambos casos se expresa en dos formas:

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 medio esperado es 1.

CUADRO 2.2
INTERPRETACIÓN DEL PARÁMETRO MNSQ (ESTADÍSTICO DE AJUSTE)

>2.0	Desajustes que generan distorsiones para el sistema de medición. Puede estar causadas por una o dos observaciones.
1.5 - 2.0	Improductivos para la construcción de la medida, pero utilizables
0.5 - 1.5	Aceptables para la construcción de las medidas
<0.5	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..

Fuente: Linacre (2002)

Estandarizado como test de hipótesis: ZSTD

Se presentan como valores normalizados z , con una distribución normal con media 0 y varianza 1.

CUADRO 2.3
INTERPRETACIÓN DEL PARÁMETRO ZSTD (ESTADÍSTICO DE AJUSTE)

(>+3)	Valores muy desajustados
(>+2)	Valores 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
+/- 1.9	Valores aceptables ($p < 0.05$)
(<-2)	Valores sobre ajustados. Indican menos variabilidad en los datos que los que predice el Modelo de Rasch, evidenciando generalmente dependencia de datos.

Fuente: Linacre (2002)

Análisis de un caso de fiabilidad, separación y ajustes globales (véase Tabla 2.3) (INFIT, OUTFIT, MNSQ Y ZSTD).

Análisis del posicionamiento competitivo de vehículos a motor en el segmento de mercado de utilitarios y ciudadanos (Montero y Oreja, 2006).

TABLA 2.3
FIABILIDAD GLOBAL DE SEPARACIÓN Y VALIDEZ GLOBAL DE LAS MEDIDAS DE LOS SUJETOS E ÍTEMS

TABLE 3.1 Posicionamiento vehículos 2006 U_C.xls									
INPUT: 21 PERSON 9 ITEM REPORTED: 21 PERSON 9 ITEM 4 CATS WINSTEPS 3.71.0.1									

SUMMARY OF 21-MEASURED PERSON									
	TOTAL			MODEL	INFIT		OUTFIT		
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD	

MEAN	32.3	9.0	.27	.55	.99	.0	1.01	.0	
S.D.	4.7	.0	1.41	.05	.46	1.0	.53	1.0	
MAX.	41.0	9.0	2.92	.75	2.04	1.9	2.37	1.9	
MIN.	21.0	9.0	-3.46	.53	.37	-1.7	.36	-1.7	

REAL RMSE	.60	TRUE SD	1.28	SEPARATION	2.13	PERSON RELIABILITY .82			
(PEOR ALT.)									
MODEL RMSE	.55	TRUE SD	1.30	SEPARATION	2.36	PERSON RELIABILITY .85			
(MEJOR ALT.)									
S.E. OF PERSON MEAN = .32									

PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00									
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .83									
SUMMARY OF 9 MEASURED ITEM									
	TOTAL			MODEL	INFIT		OUTFIT		
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD	

MEAN	75.3	21.0	.00	.36	.97	-.2	1.01	.0	
S.D.	9.0	.0	1.14	.02	.41	1.3	.44	1.3	
MAX.	92.0	21.0	1.16	.40	1.76	2.2	1.81	2.3	
MIN.	66.0	21.0	-2.11	.34	.58	-1.6	.57	-1.5	

REAL RMSE	.39	TRUE SD	1.07	SEPARATION	2.76	ITEM	RELIABILITY .88		
(PEOR ALT.)									
MODEL RMSE	.36	TRUE SD	1.08	SEPARATION	3.00	ITEM	RELIABILITY .90		
(MEJOR ALT.)									
S.E. OF ITEM MEAN = .40									

UMEAN=.0000 USCALE=1.0000									
ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00									
189 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 337.40 with 158 d.f. p=.0000									
Global Root-Mean-Square Residual (excluding extreme scores): .6034									

Fuente: Elaboración propia

2.2.- ANÁLISIS DE LAS CATEGORÍAS DE LAS ESCALAS DE MEDICIÓN DE LOS CUESTIONARIOS.

2.2.1.- LA DISTRIBUCIÓN DE LAS RESPUESTAS A LAS CATEGORÍAS ORDENADAS DE UN ÍTEM.

En un instrumento de medida, diseñado para obtener información sobre un constructo psicométricamente unidimensional e integrado por una muestra de ítems de ese constructo, las respuestas de los encuestados evidencian sus percepciones o el nivel del constructo de que disponen y se recogen en puntuaciones ordinales de una escala de medida.

La escala de medida permite el tratamiento de la información proporcionada por el encuestado a partir del uso de las puntuaciones asignadas por cada encuestado a las distintas categorías de la escala en su respuesta a las cuestiones planteadas.

Estas respuestas a las categorías asumen que cada número superior de la escala representa un nivel cualitativo más alto de disponibilidad del constructo por parte del encuestado. La escala se presenta como la división del constructo (variable latente) dentro de categorías ordenadas que cualitativamente va avanzando a lo largo de dicha variable latente.

En los modelos de Rasch de escalas ordenadas (politómicos) el esquema que se asume es el siguiente:

Categoría Inferior (1) Categoría (2) Categoría (3) Categoría Superior (por ej. 4)

$-\infty$ -----1-----+-----2-----+-----3-----+-----4----- ∞

Las categorías inferior y superior llegan hasta (- / +) ∞ , por lo que son infinitamente amplias.

2.2.2.- LOS UMBRALES ENTRE LAS CATEGORÍAS ORDENADAS DE UN ÍTEM. EL UMBRAL RASCH-ANDRICH PARA ÍTEMS POLITÓMICOS CON CATEGORÍAS ORDENADAS.

El modelo desarrollado por Rasch inicialmente es dicotómico, aunque había previsto la posibilidad politómica. Andrich (1978 a y b, 1988) percibe que las relaciones entre categorías adyacentes tienen la forma de dicotomías de Rasch, estableciendo una modificación de la fórmula inicial e introduciendo el umbral entre categorías adyacentes, en su modelo Rasch Rating Scale Model (Modelo de Rasch de escalas ordenadas).

$$\log \left(\frac{P_{nij}}{P_{ni(j-1)}} \right) = B_n - D_i - F_j$$

En donde:

P_{nij} Probabilidad de la categoría observada j (de la respuesta de la persona n en el ítem i). Mientras $P_{ni(j-1)}$ es la probabilidad de la categoría observada $j-1$ (de la respuesta de la persona n en el ítem i).

B_n Habilidad de la persona n

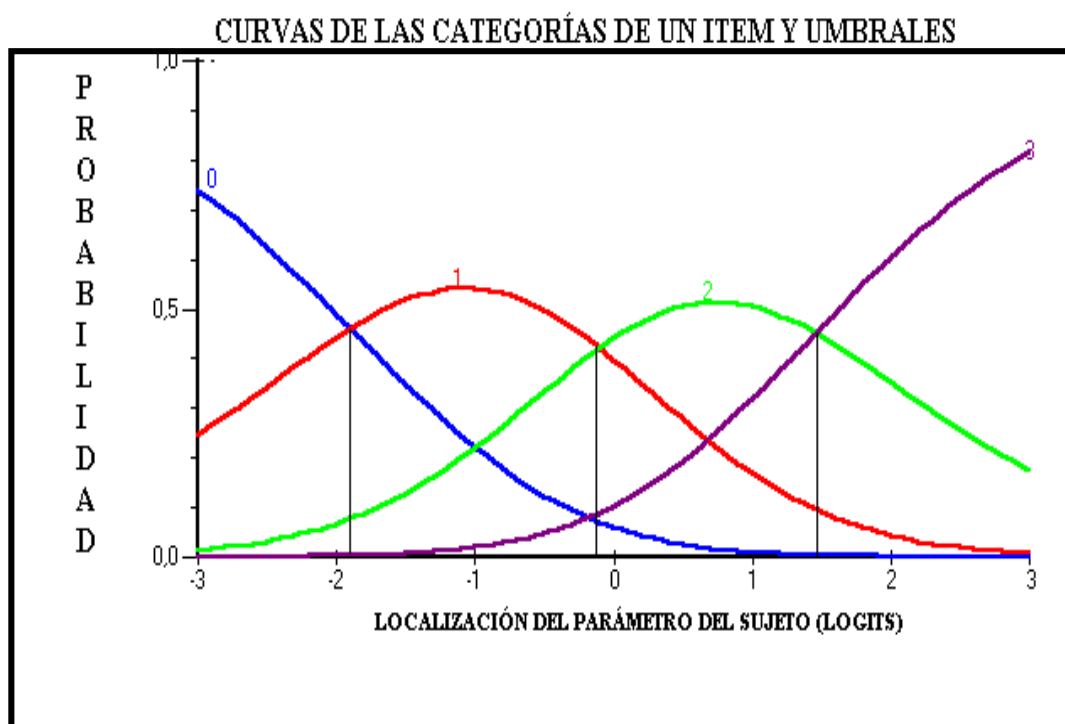
D_i Dificultad del ítem i

F_j Umbral de Andrich (diferencial de dificultad de la categoría observada j en relación a la anterior $j-1$)

En los umbrales (τ_j), puntos de cortes de las curvas características correspondientes a las distintas categorías de respuestas en los ítems, es donde la probabilidad de respuesta de una categoría con la siguiente es equiprobable.

El número de umbrales es igual al número de categorías menos uno (véase Figura 2.2).

FIGURA 2.2



Fuente: Elaboración propia con el programa RUMM2020

2.2.3.- ESTRATEGIAS DE CATEGORIZACIÓN ANTE UMBRALES DESORDENADOS.

En el diagnóstico de las categorías utilizadas mediante el análisis de la estructuras de las categorías es recomendable seguir las recomendaciones de Linacre (2002) que se establecen en forma de Guía para optimizar la efectividad de las categorías en una escala ordenada.

Esta guía (Linacre 2002) establece 8 recomendaciones:

Previa: Escala orientada con la variable latente.

- 1.- Frecuencias de las categorías: Al menos 10 observaciones por categoría.
- 2.- Distribución de las observaciones de forma regular.
- 3.- La medida media avanza monótonamente con las categorías.
- 4.- OUTFIT / MNSQ menor de 2.

5.- Avance de los umbrales

6.- Las categorías implican medidas y las medidas implican categorías.

7.- Los umbrales avanzan en al menos 1.4 logits

8.- Los umbrales avanzan hasta 5.0 logits

La importancia de estas recomendaciones ha sido establecida por Linacre (2002) como esenciales o útiles, dependiendo de su impacto en las asunciones básicas (véase Cuadro 2.4)

CUADRO 2.4
RECOMENDACIONES PARA LA OPTIMIZACIÓN DEL USO DE
CATEGORÍAS EN ESCALAS ORDENADAS

Recomendación	Estabilidad de las medidas	Precisión del Ajuste	Descripción de la muestra	Inferencia para la siguiente muestra
Previo	Esencial	Esencial	Esencial	Esencial
1	Esencial	Útil		Útil
2	Útil			Útil
3	Útil	Esencial	Esencial	Esencial
4	Útil	Esencial	Útil	Útil
5				Útil
6		Útil		Útil
7				
8	Útil			

Fuente: Elaboración propia a partir de Linacre (2002)

En el estudio de Montero-Muradas y Oreja-Rodríguez, 2014) se estableció inicialmente una categorización de rango [1-9] en las escalas aplicables a los ítems del constructo analizado. La Tabla obtenida en la ejecución de Winsteps (Linacre 2014 a)⁸ es la 3.2.

⁸ Linacre (2014b): Table 3.2 Summary of dichotomous, rating scale or partial credit structures en Winsteps Help for Rasch Analysis / Contents / Output Table Index /
Linacre (2014b): Category boundaries y Thresholds Disordered rating or partial credit structures (Categories or Rasch-Andrich thresholds)en Winsteps Help for Rasch Analysis / Contents / Special Topics

EJECUCIÓN PREVIA. ANÁLISIS DE LA ESTRUCTURA DE LA CATEGORIZACIÓN DE DATOS.

TABLA 2.4
RESUMEN DE LA ESTRUCTURA DE 9 CATEGORÍAS

TABLE 3.2 WEF_P5_BD RACK 2008_13 CBI 20140929.xl
INPUT: 21 PERSON 6 ITEM REPORTED: 21 PERSON 6 ITEM 9 CATS WINSTEPS 3.81.0

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	ANDRICH	CATEGORY		
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	THRESHOLD	MEASURE
1	1	3	2	-4.81	-5.95	1.57	1.38	NONE	(-8.43)
2	2	24	19	-4.35	-4.53	1.98	1.61	-7.32	-5.02
3	3	10	8	-3.57	-2.57	.69	1.55	-2.68	-2.08
4	4	6	5	-.86	-.63	.31	.27	-1.07	-.91
5	5	12	10	.78	.57	.69	.62	-.66	.07
6	6	18	14	1.25	1.42	.68	.59	.59	1.22
7	7	21	17	2.62	2.57	.78	1.07	1.81	2.56
8	8	25	20	4.25	4.06	.48	.72	3.11	4.72
9	9	7	6	5.66	5.89	.85	.99	6.22	(7.35)

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE	50% CUM.	COHERENCE	ESTIM	OBSERVED-EXPECTED						
LABEL	MEASURE	S.E.	AT CAT. ---ZONE---	PROBABLT	M->C C->M	RMSR	DISCR	RESIDUAL DIFFERENCE					
1	NONE		(-8.43)	-INF	-7.35	0%	0%	1.0838		-1.4%	.0	1	
2	-7.32	.65	-5.02	-7.35	-3.04	-7.33	64%	75%	.6001	.64	-.4%	-.1	2
3	-2.68	.44	-2.08	-3.04	-1.44	-2.84	38%	30%	.7106	.21	.3%	.0	3
4	-1.07	.56	-.91	-1.44	-.43	-1.31	67%	33%	.6428	1.48	.6%	.0	4
5	-.66	.49	.07	-.43	.62	-.49	42%	42%	.7647	1.69	.6%	.1	5
6	.59	.36	1.22	.62	1.86	.60	50%	61%	.6192	1.20	.5%	.1	6
7	1.81	.33	2.56	1.86	3.44	1.83	67%	67%	.6462	1.54	.2%	.0	7
8	3.11	.33	4.72	3.44	6.33	3.28	69%	72%	.4551	1.36	-.3%	-.1	8
9	6.22	.50	(7.35)	6.33	+INF	6.26	20%	14%	.7255	.89	-1.1%	-.1	9

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

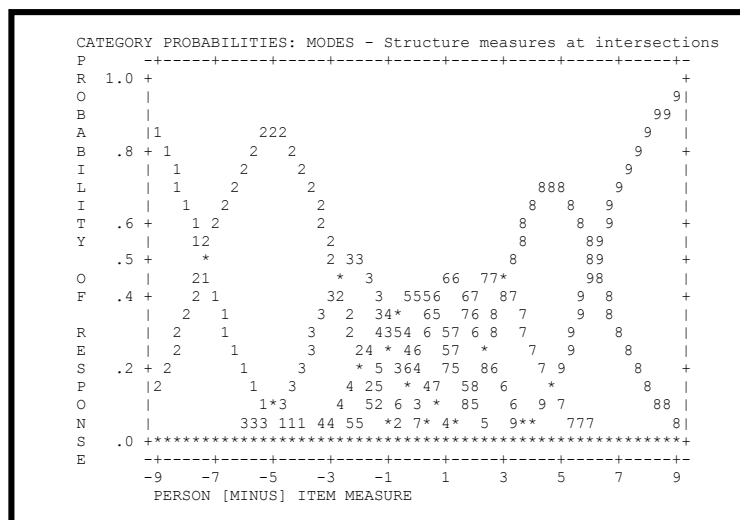
Fuente: Elaboración propia

Considerando que se cumple la recomendación Previa de la orientación de las escalas y las categorías (Esencial para lograr, Estabilidad de las medidas, Precisión del Ajuste, Descripción de la muestra e Inferencia para la siguiente muestra), se pasa a analizar el resto de las recomendaciones esenciales. En la Tabla 2.4 se observa el incumplimiento de la Recomendación 1 (Esencial para la estabilidad de las medidas), al existir en las categorías 1, 4 y 9 menos de 10 observaciones por categorías. Se pasa de la codificación inicial: 123456789 a una recodificación del tipo: 112234566, integrando las categorías de menor frecuencia y así sucesivamente hasta lograr el cumplimiento de las recomendaciones esenciales. En este proceso es necesario tener en cuenta el sentido

asignado a las categorías dentro del análisis de los ítems del constructo, con objeto de integrar aspectos conceptualmente próximos.

GRÁFICO 2.4

UMBRALES DE LAS CATEGORÍAS



Fuente: Elaboración propia

La medida observada media avanza de forma monótonamente con las categorías (Recomendación 3, esencial para lograr la Precisión del Ajuste, Descripción de la muestra e Inferencia para la siguiente muestra) y todos los OUTFIT (MNSQ) son inferiores a 2 (Recomendación 4, esencial para lograr la precisión del ajuste).

Tras un proceso de análisis de las diferentes categorizaciones disponibles y el cumplimiento de las recomendaciones establecidas por Linacre (2002), incluyendo la eliminación de sujetos que distorsionan las medidas, se considera que 6 categorías permitirían cumplir los requisitos establecidos.

EJECUCIÓN CON 6 CATEGORÍAS RECODIFICADAS Y OBTENCIÓN DE SUS ESTRUCTURAS

TABLA 2.5
RESUMEN DE LA ESTRUCTURA DE 6 CATEGORÍAS

TABLE 3.2 WEF_P5_BD RACK 2008_13 CBI 20140929.xl
INPUT: 19 PERSON 6 ITEM REPORTED: 19 PERSON 6 ITEM 6 CATS WINSTEPS 3.81.0

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	ANDRICH	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	THRESHOLD	MEASURE	
1	1	27	24	-10.00	-10.3	1.76	1.92	NONE	(-8.20)	1
2	2	14	12	-5.97	-5.16	1.14	1.62	-7.09	-4.37	3
3	3	10	9	.64	.32	.96	.84	-1.64	-.37	5
4	4	15	13	1.99	2.24	.84	.73	.97	1.91	6
5	5	19	17	3.98	4.12	.64	.43	2.90	3.93	7
6	6	29	25	6.95	6.76	.47	.60	4.86	(6.06)	8

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY	STRUCTURE	SCORE-TO-MEASURE	50% CUM.	COHERENCE	ESTIM					
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----	PROBABLT	M->C	C->M	RMSR	DISCR	
1	NONE		(-8.20)	-INF	-7.11	82%	85%	.3433	1	
2	-7.09	.50	-4.37	-7.11	-1.80	-7.10	62%	57%	.5822	3
3	-1.64	.69	-.37	-1.80	.85	-1.70	43%	30%	.7138	5
4	.97	.47	1.91	.85	2.91	.91	53%	60%	.5740	6
5	2.90	.40	3.93	2.91	5.15	2.91	65%	79%	.4691	7
6	4.86	.37	(6.06)	5.15	+INF	4.98	92%	83%	.3090	8

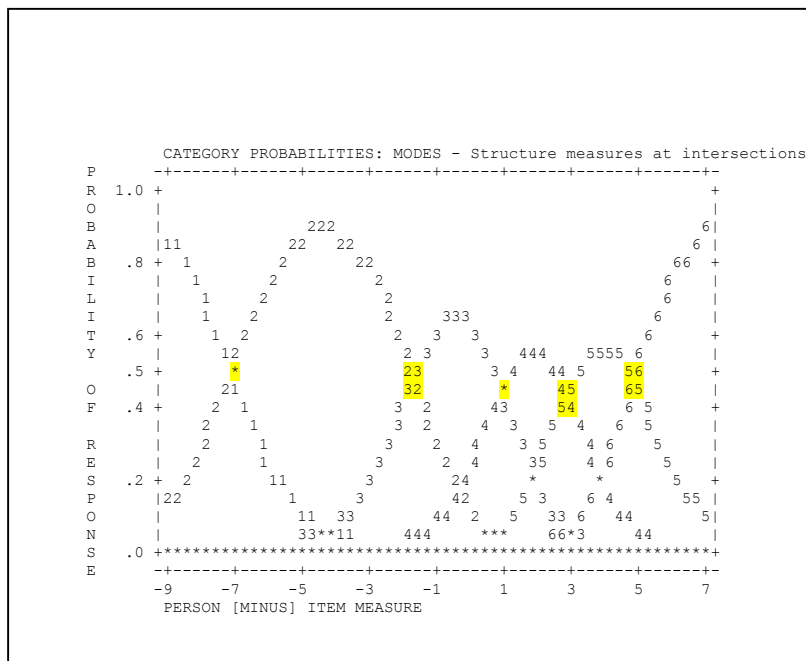
M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

Fuente: Elaboración propia

La nueva Tabla 2.5 obtenida, una vez recodificadas las categorías que no cumplían las normas de Linacre (2002) refleja el cumplimiento de los requisitos esenciales Previo (la escala está orientada con la variable latente), Recomendación 1 (todas las categorías tienen más de 10 observaciones), Recomendación 3 (la medida media avanza monótonamente con las categorías), Recomendación 4 (todos los MNSQ/OUTFIT de las categorías son menores que 2).

GRÁFICO 2.5

UMBRALES DE LAS 6 CATEGORÍAS



Fuente: Elaboración propia

En el mismo sentido se logra el cumplimiento de las recomendaciones útiles (excepto la 8, que se aproxima al nivel de 5 logits establecido).

CAPÍTULO 3ª

MEDICIÓN Y POSICIONAMIENTO COMPETITIVO

INDIVIDUAL DE DOS FACETAS ANALIZADAS DE FORMA

CONJUNTA

3.- MEDICIÓN Y POSICIONAMIENTO COMPETITIVO INDIVIDUAL DE DOS FACETAS ANALIZADAS DE FORMA CONJUNTA.

3.1.-ANÁLISIS CONJUNTO DE LAS DOS FACETAS: EL MAPA DE WRIGHT

En el análisis conjunto de los posicionamientos individuales se ha utilizado el mapa de Wright (Linacre, 2014b)⁹.

3.1.1.- ESTRUCTURA DEL MAPA DE WRIGHT

La *Table 1* del programa Winsteps (ver gráfica 3.1) recoge el continuo lineal de la variable latente (constructo definido), presentado gráficamente en vertical. Aunque el campo de variación es desde $-\infty$ a $+\infty$, frecuentemente la escala de unidades logits que recoge el cuadro varía entre límites más finitos (por ej. +3 y -3).

La escala de unidades logits se encuentra ubicada en la parte izquierda del cuadro, siendo común para las medidas de sujetos e ítems.

El continuo lineal está segmentado alrededor de la media M , por una vez la desviación típica S y dos veces la desviación típica T .

La tabla está dividida en dos partes. La parte de la izquierda, en donde se localizan los sujetos, posicionados de acuerdo a sus medidas¹⁰ y, la parte de la derecha, en donde se localizan los ítems de acuerdo a sus calibraciones¹¹.

El continuo lineal, referido a los sujetos, varía desde la parte inferior señalada con el indicativo menos (less) hasta la parte superior señalada como más (more). El significado relativo de estos términos se corresponde con la polaridad de la definición del constructo analizado. Los sujetos disponen de más o menos variable latente (constructo) y, en función a ello, se localizan a lo largo del continuo lineal.

El continuo lineal, referido a los ítems, varía desde la parte superior, con el indicativo raro (rare) hasta la parte inferior, con el indicativo frecuente (frequ). En la zona superior se localizan los ítems más raros o difíciles, mientras que en la parte inferior se ubican los

⁹ Linacre (2014b): *Table 1* Wright maps of the latent variable en Winsteps Help for Rasch Analysis / Contents / Output Table Index /

Linacre (2014b): Excel item map- Wright map for Word en Winsteps Help for Rasch Analysis / Contents / Special Topics

¹⁰ Linacre (2014b): *Table 17* en Winsteps Help for Rasch Analysis / Contents / Output Table Index /

¹¹ Linacre (2014b): *Table 13* en Winsteps Help for Rasch Analysis / Contents / Output Table Index /

ítems más fáciles o de mayor frecuencia de aparición. Dado el significado de la variable, esta parte de la tabla tiene la polaridad invertida. Siendo los ítems más frecuente los que presentan calibraciones negativas, mientras que los más raros o difíciles, presentan medidas positivas.

La media de las calibraciones de los ítems se considera que es eje (arbitrario) del continuo lineal.

3.1.2.- POSICIONAMIENTO DE DOS FACETAS

El posicionamiento de sujetos e ítems se presentan frecuentemente con forma normal a lo largo del continuo lineal (Gráfica 3. 1 y 3. 2). No obstante, también son frecuentes las excepciones.

Los ítems pueden presentar huecos entre sí, que afecta sujetos, lo que evidencia una pobre definición del constructo y la necesidad de emplear ítems que cubran dichos huecos (Gráfica 3.1). De la misma forma, pueden aparecer varios ítems con el mismo nivel de dificultad, un análisis nos debe permitir determinar su posible duplicidad o no, desde la perspectiva del constructo utilizado.

Los ítems también pueden presentar posiciones significativas en los extremos de su distribución. El ítem superior, sobre el cual aún hay sujetos lo vamos a denominar “techo” (en la gráfica 3.2, sería el ítem 27.6) y el inferior, por debajo del cual aún hay sujetos “suelo” (en la gráfica 3.2, sería el ítem 27.2). En ambos casos evidencia la necesidad de más ítems que discriminen el posicionamiento de los sujetos.

Comparación de distribuciones

A) La media de las medidas de los sujetos está por encima de la media de las calibraciones de los ítems. Gráfica 3.2.

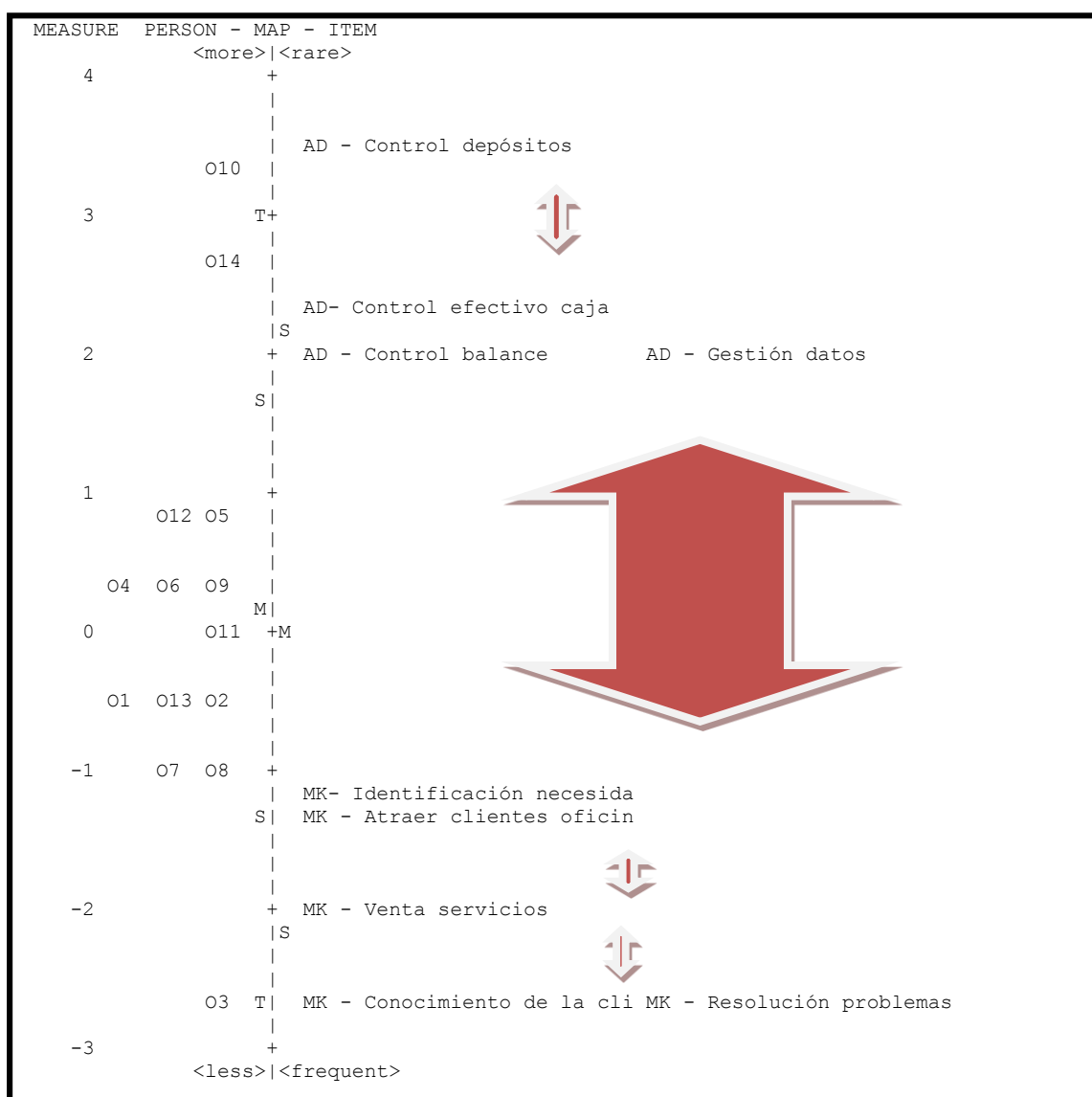
En estos casos se considera que el conjunto de sujetos disponen de un nivel de variable latente mayor que el indicado por la media de los ítems. Su significado dependerá de la dispersión de los sujetos y del nivel alcanzado por los sujetos mejor posicionados.

En la gráfica 3.2, las empresas perciben un alto nivel de riesgo, derivado principalmente por las empresas que se encuentran posicionadas por encima del ítem techo 27.6 Falta de información.

C) La media de las medidas de los sujetos está por debajo de la media de las calibraciones de los ítems. Gráfica 3.3.

En estos casos se considera que el conjunto de los sujetos no disponen de un nivel de variable latente que iguale al indicado por la media de las calibraciones de los ítems. Su significado dependerá de la dispersión de los sujetos y del nivel que luzcan los sujetos peor posicionados.

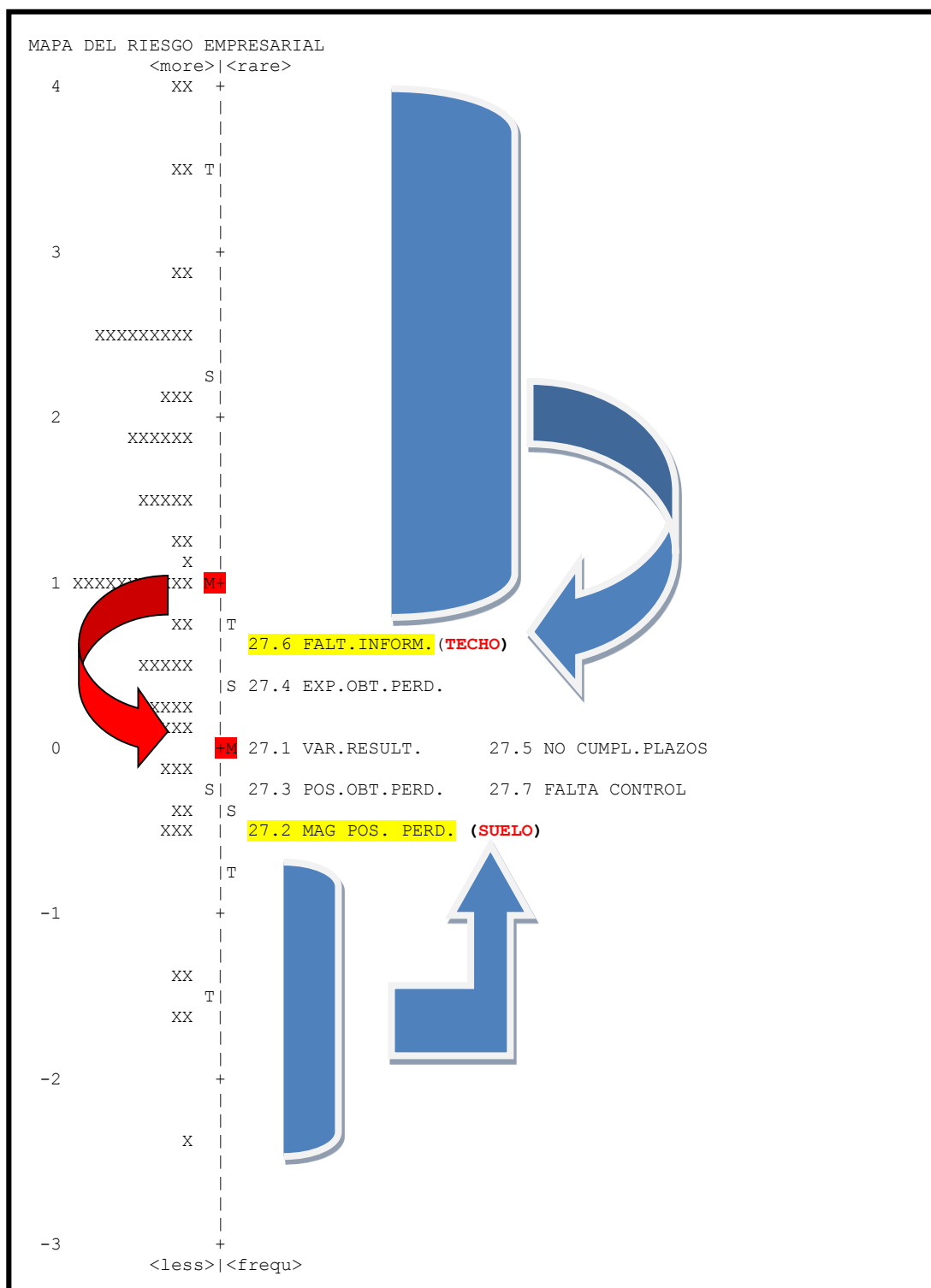
GRÁFICA 3.1
MAPA DE WRIGHT. HUECOS



Fuente: Elaboración propia

GRÁFICA 3.2

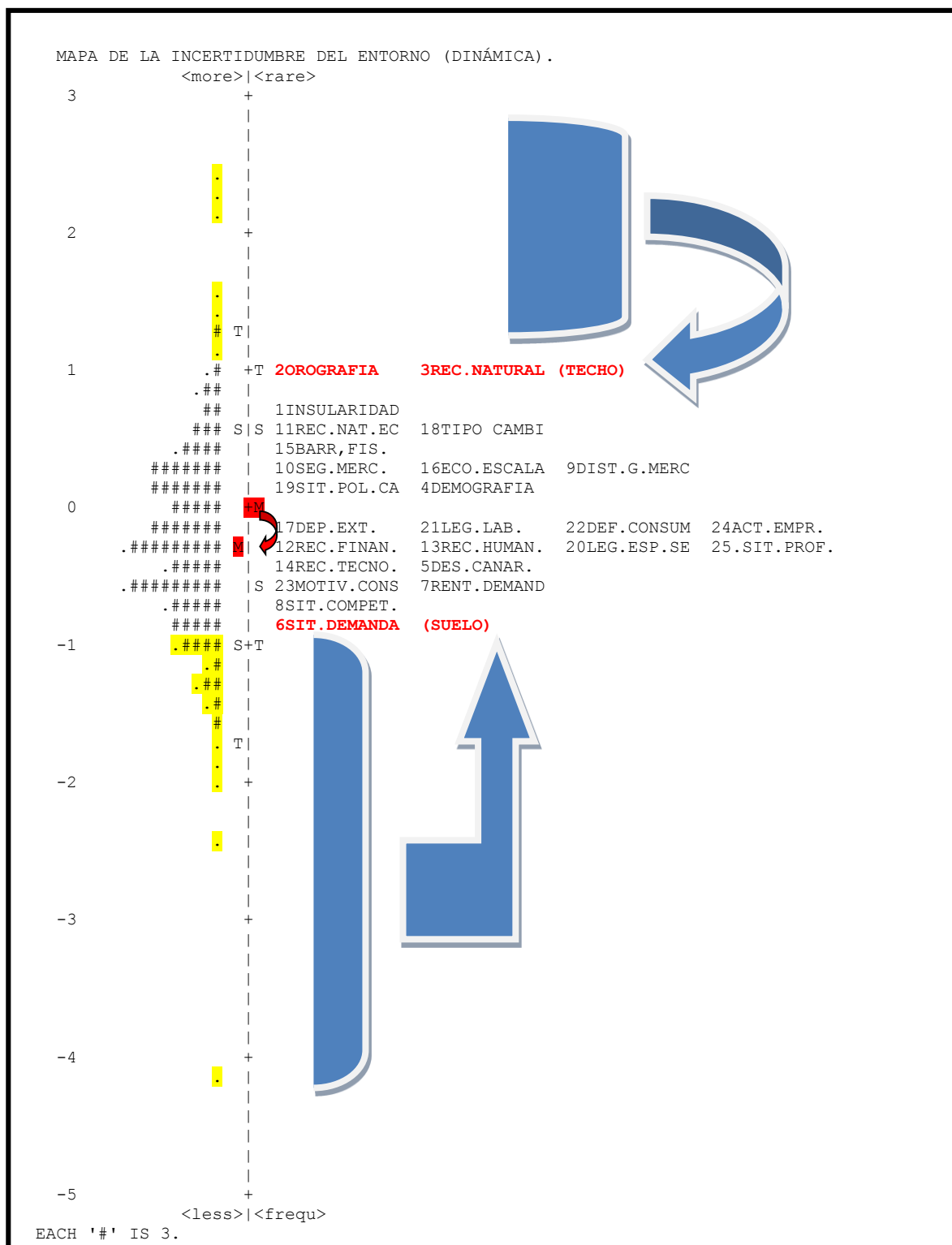
MAPA DE WRIGHT. LA MEDIA DE LAS MEDIDAS DE LOS SUJETOS ESTÁ POR ENCIMA DE LA MEDIA DE LAS CALIBRACIONES DE LOS ÍTEMS.



Fuente: Elaboración propia

GRÁFICA 3.3.

MAPA DE WRIGHT. LA MEDIA DE LAS MEDIDAS DE LOS SUJETOS ESTÁ POR DEBAJO DE LA MEDIA DE LAS CALIBRACIONES DE LOS ÍTEMS.



Fuente: Elaboración propia

En la gráfica 3.3, las empresas perciben un nivel relativamente bajo de incertidumbre (dinamismo) del entorno general, derivado de la distribución de sujetos, principalmente los situados más abajo de los ítems suelo Situación de la demanda.

3.1.3.- LOGITS Y PROBABILIDADES DE ÉXITO

Mediciones en logits (unidades log-odds)

Linacre (2014b) indica que el *logit* es una unidad de medición aditiva, definida en el contexto de un cuestionario homogéneo. La comparación de medidas logits resultantes de cuestionarios vinculados al mismo constructo, con distintos ítems, exige una previa equiparación de las medidas obtenidas.

Se puede determinar la probabilidad de acaecimiento de ítem analizados teniendo en cuenta la distancia entre las medidas de los sujetos (β_n) y las calibraciones de los ítems (δ_i).

CUADRO 3.1

PROBABILIDAD DE ÉXITO EN EL CASO DE ÍTEMS DICOTÓMICOS

Logits ($\beta_n - \delta_i$)	Probabilidad de éxito	Logits ($\beta_n - \delta_i$)	Probabilidad de éxito
5.0	99%	-5.0	1%
4.6	99%	-4.6	1%
4.0	98%	-4.0	2%
3.0	95%	-3.0	5%
2.2	90%	-2.2	10%
2.0	88%	-2.0	12%
1.4	80%	-1.4	20%
1.1	75%	-1.1	25%
1.0	73%	-1.0	27%
0.8	70%	-0.8	30%
0.5	62%	-0.5	38%
0.4	60%	-0.4	40%
0.2	55%	-0.2	45%
0.1	52%	-0.1	48%
0.0	50%	-0.0	50%

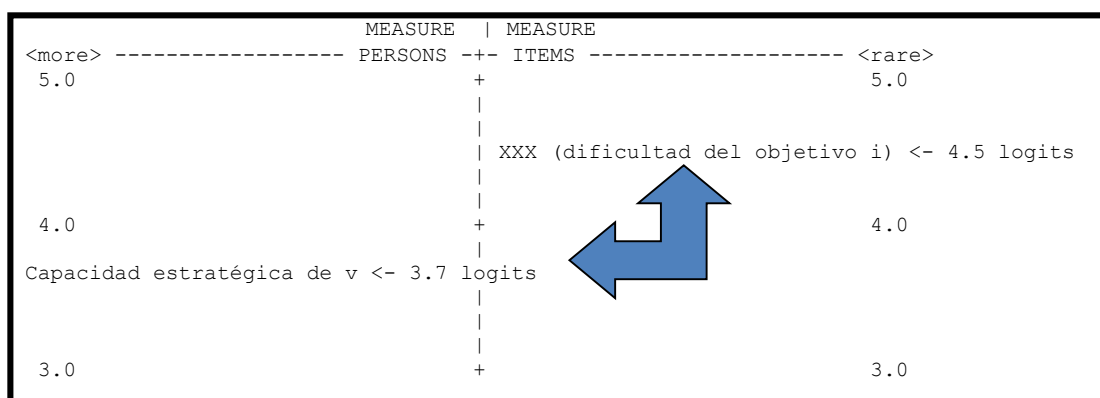
Fuente: Linacre 2014b

La determinación para casos dicotómicos o politómicos se pueden consultar en Linacre (2014b)¹².

Ejemplo de Probabilidad de éxito con datos dicotómicos:

En la gráfica 3.4, la distancia entre la capacidad estratégica de una empresa n y un objetivo i determina la probabilidad de éxito en el logro de un objetivo.

GRÁFICA 3.4
PROBABILIDAD DE ÉXITO EN EL LOGRO DE UN OBJETIVO



Fuente: Elaboración propia

Si estamos considerando la capacidad estratégica de una empresa, cuya medida obtenida es $\beta_n = 3.70$ logits, ante el logro (o no) de un determinado objetivo, cuya dificultad ha sido calibrada como $\delta_i = 4.5$ logits.

La diferencia entre ambas medidas sería: $(\beta_n - \delta_i) = (3.7 - 4.5) = -0.8$ logits.

La probabilidad de éxito de la empresa n al tratar de lograr el objetivo i es del 30%.

Para ítems dicotómicos, el cálculo de probabilidades se puede programar en una hoja Excel, mediante la siguiente fórmula:

$$\text{Probabilidad de error (0)} = 1 / (1 + e^{(\beta_n - \delta_i)})$$

$$\text{Probabilidad de acierto (1)} = 1 / (1 + e^{(\delta_i - \beta_n)})$$

La suma de ambas probabilidades es 1.

¹² Linacre (2014b): Probabilities from measures en Winsteps Help for Rasch Analysis / Contents / /Special Topics

3.2.- MEDIDAS DE LA FACETA: SUJETOS (EMPRESAS, PRODUCTOS Y ENTES ECONÓMICOS, EMPRESARIALES Y TURÍSTICOS). POSICIONAMIENTO, AJUSTES Y DESAJUSTES.

3.2.1.- MEDIDAS DE LOS SUJETOS Y SUS NIVELES DE AJUSTE. POSICIONAMIENTOS COMPETITIVOS DE LOS SUJETOS.

La medida β_n , en logits, nos indica la localización de la habilidad β (denominación genérica) del sujeto n en el continuo lineal. Está referida al origen común del constructo unidimensional analizado. Esta medición es independiente de la muestra de ítems, relacionados con el constructo analizado, utilizada en el cuestionario.

A la vez la localización determinada por la medida obtenida a nivel individual de cada sujeto fija el posicionamiento, en el contexto del continuo lineal recogido en el mapa de Wright, de cada uno de los sujetos.

Para el análisis de los ajustes se utilizan los estadísticos OUTFIT e INFIT (MNSQ Y ZSTD) de acuerdo a Linacre (2014b) ¹³, reflejando la desviación de la observación obtenida respecto a la expectativa generada por el modelo probabilístico utilizado (véase 3.1)

La correlación PT-Measure. Es la correlación entre las observaciones de las personas (de sus respuestas a los ítems) y las medidas de las personas. Se esperan sean positivas.

En la Tabla 3.1 se recoge la salida de la *Table 17.1* de Winsteps. En la columna β_n se presentan las mediciones obtenidas de los diferentes sujetos analizados (en este caso vehículos de los segmentos ciudad y utilitarios), esa medida dispone de valores de ajustes bajo las columnas MNSQ/INFIT y MNSQ/OUTFIT, todos ellos dentro del rango ± 2 , siendo 1 el valor esperado. Las especificidades presentes, derivadas de la aplicación de la información de los Cuadros 2.2 y 2.3 determinan situaciones de análisis específico. Todas ellas están identificadas en rojo y señalan la necesidad de análisis complementarios (véase el epígrafe 3.2.3).

¹³ Linacre (2014b): *Table 17.1* Person statistics in measure order en Winsteps Help for Rasch Analysis / Contents / Output Table Index

TABLA 3.1

MEDIDAS, AJUSTES Y POSICIONAMIENTOS COMPETITIVOS DE LOS SUJETOS

TABLE 17.1 Posicionamiento vehiculos 2006 U_C.xl													
INPUT: 21 PERSON 9 ITEM REPORTED: 21 PERSON 9 ITEM 4 CATS WINSTEPS 3.71.0.1													

PERSON: REAL SEP.: 2.13 REL.: .82 ... ITEM: REAL SEP.: 2.76 REL.: .88													

PERSON STATISTICS: MEASURE ORDER													

nENTRY	TOTAL	TOTAL	β_n	MODEL	INFIT	OUTFIT	PT-MEASURE	EXACT	MATCH				
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	PERSON

5	41	9	2.92	.62	1.37	.9	2.37	1.7	-.23	.39	22.2	64.2	U Mercedes Cl. A
8	37	9	1.62	.54	.96	.1	.87	-.1	.60	.54	66.7	61.0	U Peugeot 207
9	37	9	1.62	.54	.99	.1	.89	-.1	.58	.54	66.7	61.0	U Renault ClioII
10	37	9	1.62	.54	.67	-.7	.61	-.8	.74	.54	66.7	61.0	U Seat Ibiza
1	36	9	1.33	.53	.88	-.1	.97	.1	.48	.56	55.6	58.1	U Fiat Punto Gr.
2	36	9	1.33	.53	.69	-.6	.65	-.7	.81	.56	55.6	58.1	U Ford Fiesta
7	36	9	1.33	.53	1.17	.5	1.23	.6	.52	.56	33.3	58.1	U Mitsub. Colt
12	36	9	1.33	.53	.68	-.7	.63	-.8	.67	.56	77.8	58.1	U Toyota Yaris06
6	34	9	.77	.53	1.80	1.6	1.74	1.5	.15	.59	33.3	54.8	U Mini
3	32	9	.22	.53	1.42	1.0	1.39	.9	.63	.61	55.6	53.7	U Honda Jazz
11	32	9	.22	.53	.45	-1.4	.44	-1.4	.79	.61	66.7	53.7	U Suzuki Swift
21	32	9	.22	.53	.63	-.8	.64	-.7	.67	.61	66.7	53.7	C Toyota Aygo
4	31	9	-.06	.53	.77	-.4	.78	-.4	.77	.61	55.6	57.8	U Kia Rio
13	31	9	-.06	.53	.37	-1.7	.36	-1.7	.84	.61	77.8	57.8	C Citroën C1
15	31	9	-.06	.53	.74	-.5	.75	-.4	.79	.61	55.6	57.8	C Fiat Panda
18	31	9	-.06	.53	.37	-1.7	.36	-1.7	.84	.61	77.8	57.8	C Peugeot 107
20	29	9	-.62	.54	2.04	1.9	2.07	1.9	.49	.61	33.3	61.4	C Smart Fortwo
16	27	9	-1.21	.55	1.70	1.5	1.57	1.3	.02	.60	66.7	57.7	C Ford Ka
19	26	9	-1.51	.55	1.06	.3	1.08	.3	.55	.59	55.6	54.7	C Renault Twingo
17	25	9	-1.82	.57	.71	-.6	.73	-.6	.83	.59	66.7	54.0	C Kia Picanto
14	21	9	-3.46	.75	1.26	.6	1.10	.4	.31	.56	66.7	78.5	C Chevr. Matiz

MEAN	32.3	9.0	.27	.55	.99	.0	1.01	.0			58.2	58.7	
S.D.	4.7	.0	1.41	.05	.46	1.0	.53	1.0			15.3	5.2	

Fuente: Elaboración propia

Junto a los ajustes, se aprecia la columna PT-MEASURE que contiene valores positivos, excepto el sujeto U Mercedes CL A que los presenta negativos. El análisis de los datos indica una correlación negativa entre las observaciones y la medición. En el análisis detallado realizado en el epígrafe 3.2.3 se destacan los datos que originan esta situación.



La presencia de niveles de desajustes en MNSQ superiores a +/-2, pueden implicar que la mediciones obtenidas están distorsionadas, por lo que tras un análisis de las causas de los desajustes puede ser conveniente la supresión del sujeto del análisis. En el mismo sentido un valor negativo PT-MEASURE puede implicar, tras el correspondiente análisis de las causas, la supresión del sujeto del análisis.

Bajo la columna Person está indicada la ordenación jerárquica de los sujetos, cuyo código de entrada se encuentra en la columna **n** (Entry Number). Esta ordenación está determinada por el orden decreciente de las medidas recogidas en la columna **β_n**

(Measure). Este posicionamiento competitivo coincide con el que aparece en la parte de la izquierda del mapa de Wright.

CUADRO 3.2.

PAUTAS DE RESPUESTAS A LOS ÍTEMS DICOTÓMICOS LIGADAS A LOS (DES) AJUSTES DE LOS SUJETOS

Person Responses to Items: Easy--Items--Hard	Diagnosis of Pattern	Person OUTFIT MNSQ	Person INFIT MNSQ
111:0110110100:000	Modelled/Ideal	1.0	1.1
111:1111100000:000	Guttman/Deterministic	0.3	0.5
000:0000011111:111	Miscoder	12.6	4.3
011:1111110000:000	Carelessness/Sleeping	3.8	1.0
111:1111000000:001	Lucky Guessing	3.8	1.0
101:0101010101:010	Response set/Miskey	4.0	2.3
111:1000011110:000	Special knowledge	0.9	1.3
111:1010110010:000	Imputed outliers *	0.6	1.0
Right/Transition/Wrong			
 high - low - high	OUTFIT sensitive to outlying observations	>>1.0 unexpected outliers	>>1.0 disturbed pattern
 low - high - low	INFIT sensitive to pattern of inlying observations	<<1.0 overly predictable outliers	<<1.0 Guttman pattern
* as when a tailored test (such as a Binet intelligence test) is scored by imputing all "right" responses to unadministered easier items and all "wrong" responses to unadministered harder items. The imputed responses are indicated by italics			

Fuente: Linacre y Wright, 1994 y Linacre (2014b)¹⁴

Las respuestas de los sujetos a los diferentes ítems pueden ser analizadas con objeto de generar tipologías susceptibles de permitir un diagnóstico de las pautas implícitas en ellas, de acuerdo al análisis del Escalograma de Guttman (véase epígrafe 5.1) en función de los

¹⁴ Linacre (2014b): Dichotomous mean-square fit statistics en Winsteps Help for Rasch Analysis / Contents // Special Topics

niveles de valores obtenidos en los estadísticos MNSQ/INFIT y MNSQ/OUTFIT (véase Cuadros 3.2 y 3.3).

CUADRO 3.3.

**PAUTAS DE RESPUESTAS A LOS ÍTEMS POLITÓMICOS LIGADAS A LOS
(DES) AJUSTES DE LOS SUJETOS**

Response String Easy.....Hard	INFIT MNSQ	OUTFIT MNSQ	Point-measure correlation	Diagnosis
I. modeled:				
33333132210000001011	0.98	0.99	0.78	Stochastically
31332332321220000000	0.98	1.04	0.81	monotonic in form
33333331122300000000	1.06	0.97	0.87	strictly monotonic
33333331110010200001	1.03	1	0.81	in meaning
II. overfitting (muted):				
3322222221111111100	0.18	0.22	0.92	Guttman pattern
3333322221111100000	0.31	0.35	0.97	high discrimination
3222222221111111110	0.21	0.26	0.89	low discrimination
32323232121212101010	0.52	0.54	0.82	tight progression
III. limited categories:				
3333333332222222222	0.24	0.24	0.87	high (low) categories
2222222221111111111	0.24	0.34	0.87	central categories
3333322222222211111	0.16	0.2	0.93	only 3 categories
IV. informative-noisy:				
3222222201111111130	0.94	1.22	0.55	noisy outliers
33233332212333000000	1.25	1.09	0.77	erratic transitions
33133330232300101000	1.49	1.4	0.72	noisy progression
33333333300000000000	1.37	1.2	0.87	extreme categories
V. non-informative:				
2222222222222222222	0.85	1.21	0	one category
12121212121212121212	1.5	1.96	-0.09	central flip-flop
01230123012301230123	3.62	4.61	-0.19	rotate categories
03030303030303030303	5.14	6.07	-0.09	extreme flip-flop
03202002101113311002	2.99	3.59	-0.01	random responses
VI. contradictory:				
11111122233222111111	1.75	2.02	0	folded pattern
1111111112222222222	2.56	3.2	-0.87	central reversal
2222222223333333333	2.11	4.13	-0.87	high reversal
0011111112222222233	4	5.58	-0.92	Guttman reversal
0000000003333333333	8.3	9.79	-0.87	extreme reversal

Fuente: Smith, 1996 y Linacre, 2014b

3.2.2.- MEDIDAS Y POSICIONAMIENTO DE SUJETOS QUE NO HAN PARTICIPADO EN EL ANÁLISIS CONJUNTO PREVIO, PERO QUE PERTENECEN A LA POBLACIÓN EN ESTUDIO.

El principio de invarianza del modelo de Rasch indica que si los datos se ajustan al modelo las medidas obtenidas de los sujetos e ítems son independiente de la muestra de ítems o sujetos que, respectivamente, se hayan utilizados. Todo ello siempre que los ítems pertenezcan al conjunto de ítems delimitados por el constructo definido y los sujetos al colectivo analizado.

En este caso, un sujeto que no haya sido considerado en el análisis inicial, perteneciente al colectivo de sujetos analizados, puede obtener sus medidas y su posicionamiento a partir de la conversión de la puntuación total (estimador suficiente) obtenida por el sujeto al cumplimentar el instrumento de medida vinculado al constructo.

En la Tabla 3.2 se presenta la conversión de las puntuaciones en medidas, así como la formulación que lo hace posible, de acuerdo a Linacre (2014).

TABLA 3.2.
CONVERSIÓN DE LAS PUNTUACIONES TOTALES
DE LOS SUJETOS EN MEDIDAS

TABLE 20.1 Posicionamiento vehículos 2006 U_C.xl									
INPUT: 21 PERSON 9 ITEM REPORTED: 21 PERSON 9 ITEM 4 CATS WINSTEPS 3.73									

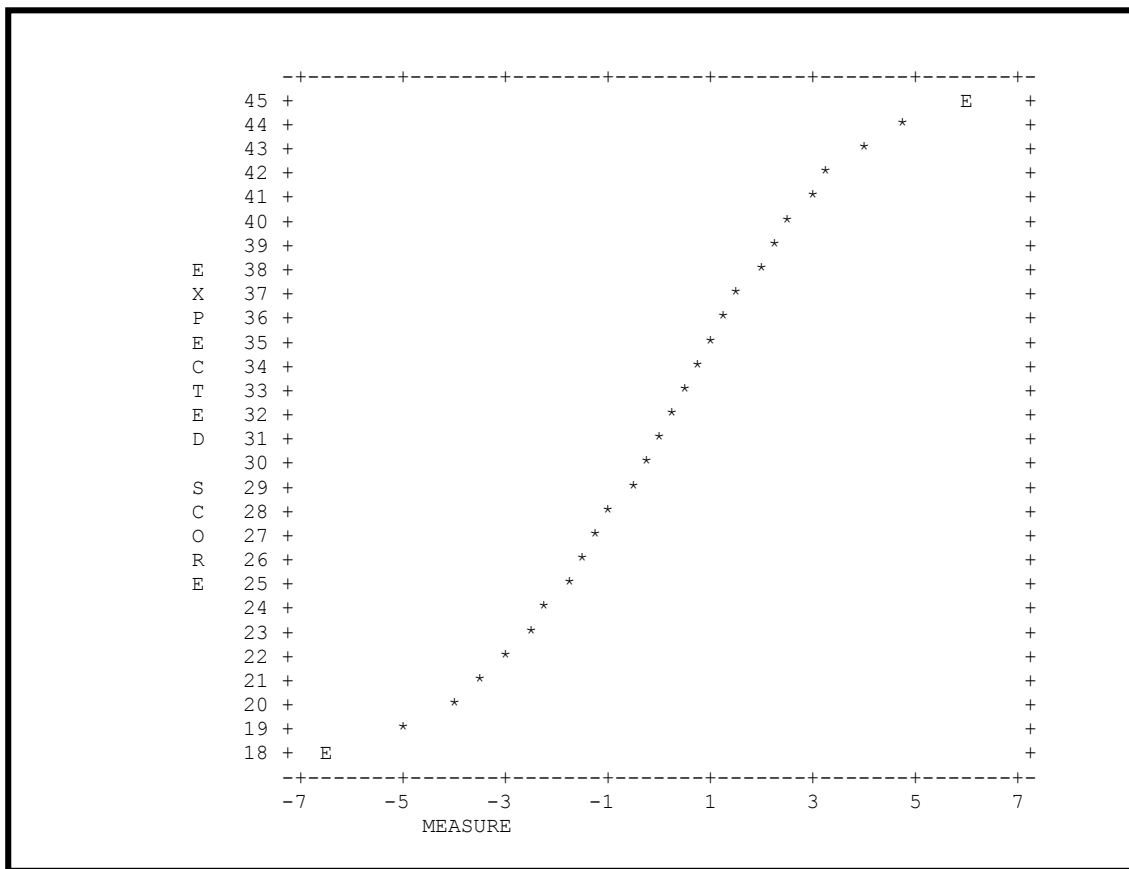
TABLE OF MEASURES ON TEST OF 9 ITEMS									
SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	
18	-6.50E	1.90	28	-.91	.54	38	1.92	.55	
19	-5.09	1.13	29	-.62	.54	39	2.23	.57	
20	-4.12	.88	30	-.34	.53	40	2.56	.59	
21	-3.46	.75	31	-.06	.53	41	2.93	.63	
22	-2.95	.68	32	.22	.53	42	3.35	.68	
23	-2.53	.63	33	.50	.53	43	3.89	.79	
24	-2.16	.59	34	.77	.53	44	4.71	1.06	
25	-1.83	.57	35	1.05	.53	45	5.99E	1.86	
26	-1.51	.56	36	1.33	.53				
27	-1.21	.55	37	1.62	.54				

CURRENT VALUES, UMEAN=.0000 USCALE=1.0000									
TO SET MEASURE RANGE AS 0-100, UMEAN=52.0040 USCALE=8.0067									
TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UMEAN=32.0411 USCALE=2.1618									
Predicting Score from Measure: Score = Measure * 2.7159 + 13.5232									
Predicting Measure from Score: Measure = Score * .3566 + -4.8231									

Fuente: Elaboración propia

La representación gráfica de la conversión de las puntuaciones en medidas se recoge en el Gráfico 3.5.

GRÁFICO 3.5
CURVA LOGÍSTICA QUE RELACIONA LA PUNTUACIÓN TOTAL Y LAS
MEDIDAS DE UN SUJETO EN UN CUESTIONARIO



Fuente: Elaboración propia

Este proceso de conversión de las puntuaciones totales en medidas y posicionamientos, supone una gran ventaja en el análisis competitivo. No sólo los sujetos que no han participado en la muestra inicial que permite obtener la validez de contenido del constructo, sino los que participando han sido excluidos debido a sus desajustes (véase epígrafe 3.2.3) pueden ser posicionados en el análisis, a la vez que denotan sus desajustes susceptibles de un análisis específico.

3.2.3.- ANÁLISIS DE LOS DESAJUSTES EN LAS MEDIDAS DE LOS SUJETOS.

El análisis de los desajustes de las medidas de los sujetos, que representan la separación de las observaciones con respecto al valor esperado, en forma de residuales medidos por los estadísticos MNQS/INFIT y MNSQ/OUTFIT, permite la obtención de informaciones competitivas referentes al sujeto que luce tales residuales.

La principal información que se obtiene es en qué ítems el sujeto presenta el desajuste¹⁵. En la Tabla 3.3 la información de cada sujeto que presenta desajustes de acuerdo con la *Table 17.1* de Winsteps (véase Tabla 3.1), en donde los ítems se presentan por orden de entrada en la línea RESPONSE (1:).

TABLA 3.3
DESAJUSTES DE LAS RESPUESTAS DE LOS SUJETOS

TABLE 7.1 Posicionamiento vehiculos 2006 U_C.xls ZOU294WS.TXT Mar 14 9:52 2011											
INPUT: 21 PERSON 9 ITEM REPORTED: 21 PERSON 9 ITEM 4 CATS WINSTEPS 3.71.0.1											

TABLE OF POORLY FITTING PERSON (ITEM IN ENTRY ORDER)											
NUMBER - NAME -- POSITION ----- MEASURE - INFIT (MNSQ) OUTFIT											
Por ejemplo:											
5	U Mercedes Clase A				2.92		1.4	A	2.4		
RESPONSE:	1:	4	4	5	4	4	5	5	5	5	
Z-RESIDUAL:											
6	U Mini				.77		1.8	C	1.7		
RESPONSE:	1:	4	4	3	4	2	4	4	5	4	
Z-RESIDUAL:											
16	C Ford Ka				-1.21		1.7	D	1.6		
RESPONSE:	1:	2	3	3	3	2	4	3	4	3	
Z-RESIDUAL:											
3	U Honda Jazz				.22		1.4	E	1.4		
RESPONSE:	1:	5	3	2	4	5	4	3	3	3	
Z-RESIDUAL:											

Fuente: Elaboración propia

¹⁵ Linacre (2014b): *Table 7.1* Person misfitting responses en Winsteps Help for Rasch Analysis / Contents / Output Table Index

Los desajustes, recogidos en la línea Z-RESIDUAL, pueden ser positivos o negativo. En el caso de desajustes positivos indican que la respuesta observada por el sujeto al cumplimentar el cuestionario está por encima de la expectativa generada por el modelo. Si el desajuste es negativo, la respuesta observada es inferior a la que se esperaba.

Su significado dependerá del sentido de la cuestión planteada, pero se puede advertir que el modelo esperaba valores diferentes a los obtenidos, por lo que las observaciones de los sujetos a esos ítems deben ser analizadas.

El análisis complementario se puede realizar mediante el Escalograma de Guttman (véase epígrafe 5.1). En la Tabla 3.4 se presenta dos versiones del escalograma de Guttman (Tables 22.1 y 2 de Winsteps) que reflejan el significado de los desajustes. Así el vehículo Mercedes Clase A tiene una características reflejadas por los expertos en los cuestionarios que permiten posicionarle como el mejor coche de los segmentos analizados, por su puntuación total.

TABLA 3.4
ANÁLISIS DE DESAJUSTES DE LOS SUJETOS MEDIANTE EL
ESCALOGRAMA DE GUTTMAN

TABLE 22.1_2 Posicionamiento vehículos 2006 U_C.xl	
INPUT: 21 PERSON 9 ITEM REPORTED: 21 PERSON 9 ITEM 4 CATS WINSTEPS 3.71.0.1	

GUTTMAN SCALOGRAM OF RESPONSES:	GUTTMAN SCALOGRAM OF ZONED RESPONSES: @=0, A=1, B=2, C=3, D=4, E= 5, F=6
PERSON ITEM	PERSON ITEM
614258397	614258397
5 +54445555 U Mercedes Clase A	5 +5DD5EEE U Mercedes Clase A
8 +555433444 U Peugeot 207	8 +55E4CC444 U Peugeot 207
9 +554533444 U Renault Clio III	9 +554ECC444 U Renault Clio III
10 +554445334 U Seat Ibiza	10 +55444ECC4 U Seat Ibiza
1 +455443443 U Fiat Grande Punto	1 +D5E44C44C U Fiat Grande Punto
2 +554445333 U Ford Fiesta	2 +55444ECCC U Ford Fiesta
7 +454455333 U Mitsubishi Colt	7 +D544EECCC U Mitsubishi Colt
12 +553443444 U Toyota Yaris 2006	12 +55C44C444 U Toyota Yaris 2006
6 +444425344 U Mini	6 +DD44BE3DD U Mini
3 +454353233 U Honda Jazz	3 +4ED3E3B33 U Honda Jazz
11 +454343333 U Suzuki Swift	11 +4ED3D3333 U Suzuki Swift
21 +543433334 C Toyota Aygo	21 +E43D3333D C Toyota Aygo
4 +453344332 U Kia Rio	4 +4E33DD33B U Kia Rio
13 +543433333 C Citroën C1	13 +E43D33333 C Citroën C1
15 +543443332 C Fiat Panda	15 +E43DD333B C Fiat Panda
18 +543433333 C Peugeot 107	18 +E43D33333 C Peugeot 107
20 +453222443 C Smart Fortwo	20 +4E3BBBDD3 C Smart Fortwo
16 +423324333 C Ford Ka	16 +4B33BD333 C Ford Ka
19 +433243322 C Renault Twingo	19 +4C3BD33B2 C Renault Twingo
17 +442233322 C Kia Picanto	17 +44BB33C22 C Kia Picanto
14 +233232222 C Chevrolet Matiz	14 +B3C2C2222 C Chevrolet Matiz
614258397	614258397

Fuente: Elaboración propia

No obstante, tiene en uno de los ítems, el nº de entrada 1, una puntuación de 4 inferior a los vehículos posicionados tras él que la tienen de 5 (*Table 22.1* de Winsteps en la Tabla 3.4). Sería conveniente analizar las causas y tomar las decisiones competitivas que procedan. Si ampliamos el estudio a la versión recogida por la *Table 22.2* de Winsteps (Tabla 3.4) se puede apreciar la amplitud de las diferencias percibidas respecto tanto a los ítems consecutivos como respecto a los sujetos cercanos en su posicionamiento competitivo. En este caso se señalan con letras los desajustes relativos que se encuentran, estando incluidos en los mismos los desajustes especialmente reflejados en la Tabla 3.3.

3.3.- MEDIDAS DE LOS ÍTEMS, AJUSTES Y DESAJUSTES. JERARQUIZACIÓN DE LOS ÍTEMS O FACTORES DE POSICIONAMIENTO COMPETITIVO.

3.3.1.- CALIBRACIÓN DE ÍTEMS Y SUS NIVELES DE AJUSTE. JERARQUIZACIÓN DE LOS ÍTEMS.

La medida δ_i , en logits, es la calibración del ítem i , referido al origen del continuo lineal del constructo analizado, por lo que marca su localización en el mismo. Si los datos se ajustan al modelo, esta calibración es independiente de la muestra utilizada en la administración del cuestionario. La denominamos genéricamente como dificultad del ítem i . La media de las medidas de los ítems se configura como el origen del continuo lineal (0.00) del análisis conjunto de las dos facetas.

TABLA 3.5
CALIBRACIONES, AJUSTES Y JERARQUIZACIÓN
DE LOS ÍTEMS O FACTORES DE POSICIONAMIENTO

TABLE 13.1 Posicionamiento vehiculos 2006 U_C.xl														
INPUT: 21 PERSON 9 ITEM REPORTED: 21 PERSON 9 ITEM 4 CATS WINSTEPS 3.71.0.1														

PERSON: REAL SEP.: 2.13 REL.: .82 ... ITEM: REAL SEP.: 2.76 REL.: .88														
ITEM STATISTICS: MEASURE ORDER														

i	ENTRY	TOTAL	TOTAL	δ_i	MODEL	INFIT	OUTFIT	PT-MEASURE	EXACT MATCH					
	NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	ITEM

	7	66	21	1.16	.36	.75	-.8	.73	-.8	.77	.63	61.9	57.5	PRECIO
	3	68	21	.91	.35	.72	-.9	.71	-.9	.64	.64	66.7	57.8	EQUIPAMIENTO
	9	68	21	.91	.35	.58	-1.6	.57	-1.5	.78	.64	66.7	57.8	SEGURIDAD
	2	73	21	.30	.35	.69	-1.1	.71	-1.0	.80	.65	57.1	57.7	COMODIDAD
	5	73	21	.30	.35	1.76	2.2	1.81	2.3	.34	.65	38.1	57.7	HABITABILIDAD
	8	73	21	.30	.35	1.56	1.7	1.49	1.5	.56	.65	47.6	57.7	PRESTACIONES
	4	74	21	.18	.34	.67	-1.2	.74	-.9	.68	.66	61.9	57.2	ESTABILIDAD
	1	91	21	-1.96	.39	1.22	.8	1.53	1.4	.62	.65	61.9	62.2	CONSUMO
	6	92	21	-2.11	.40	.79	-.6	.81	-.4	.71	.65	61.9	62.8	MOTOR

	MEAN	75.3	21.0	.00	.36	.97	-.2	1.01	.0			58.2	58.7	
	S.D.	9.0	.0	1.14	.02	.41	1.3	.44	1.3			8.9	2.1	

Fuente: Elaboración propia

La información referida a las calibraciones o medidas de la dificultad de los ítems o factores de posicionamiento, su jerarquización, así como ajustes y desajustes se presenta en la Tabla 3.5¹⁶.

Dada la estructura del modelo, hay una relación inversa entre las puntuaciones totales (TOTAL SCORE) y las medidas δ_i , de tal manera que a la mayor puntuación le corresponde la menor medida. Esta situación la denominaremos como polaridad invertida.

El posicionamiento competitivo de los ítems o jerarquía de los factores de posicionamiento competitivos evidencia el diferente peso o frecuencia de presencia en el constructo. Por una parte se aprecia que no todos los ítems tienen la misma presencia en las respuestas de los sujetos, variando su puntuación total.

¹⁶ Linacre (2014b): *Table 13.1 Items statistics in measure order* en Winsteps Help for Rasch Analysis / Contents / Output Table Index

Los ítems más frecuentes en las respuestas se denominan “fáciles” y se encuentran ubicados en la parte baja del mapa de Wright y tienen en la Tabla 3.5 (*Table* 13.1 de Winsteps) medidas negativas. El más frecuente es el que tiene la mayor medida negativa. El menos frecuente o más “raro” es el que tiene mayor calibración o medida positiva, debido a su escasa presencia en las repuestas a los ítems por parte de los sujetos.

Para el análisis de los ajustes de los ítems se utilizan los estadísticos OUTFIT e INFIT (MNSQ Y ZSTD), reflejando la diferencia entre los valores observados en los ítems y los valores esperados derivados del modelo de Rasch al que se ajustan los datos. Los criterios de aceptación de los ajustes y desajustes recogidos en el epígrafe 2.1.2 son de aplicación también para los ítems.

En la Tabla 3.5 se puede apreciar que los desajustes de los sujetos no han generado niveles de desajustes en ítems que pudieran distorsionar las mediciones.

La correlación PT-MEASURE es la correlación entre las observaciones en los ítems (por las personas encuestadas) y las medidas de los ítems. Se esperan sean positivas. Como lo son en este caso (Tabla 3.5).

3.3.2.-ANÁLISIS DE LOS DESAJUSTES EN LAS CALIBRACIONES DE ÍTEMS

Los mismos sujetos que ya aparecían como desajustados en determinados ítems vuelven a presentarse bajo la lógica inversa. En este caso se parte de la información detallada por ítems con indicación del sujeto que genera el desajuste. En la Tabla 3.6 se recoge la información de la *Table 11.1* generada por el programa Winsteps¹⁷.

TABLA 3.6
DESAJUSTES EN LOS ÍTEMS

TABLE 11.1 Posicionamiento vehiculos 2006 U_C.xl										
INPUT: 21 PERSON 9 ITEM REPORTED: 21 PERSON 9 ITEM 4 CATS WINSTEPS 3.71.0.1										

TABLE OF POORLY FITTING ITEM (PERSON IN ENTRY ORDER)										
NUMBER - NAME -- POSITION ----- MEASURE - INFIT (MNSQ) OUTFIT										
5 HABITABILIDAD										
					.30		1.8	A	1.8	
RESPONSE:	1:	4	4	5	4	4	2	5	3	3 4
Z-RESIDUAL:				2			-2			
RESPONSE:	11:	4	4	3	3	4	2	3	3	4 2
Z-RESIDUAL:										2
RESPONSE:	21:	3								
Z-RESIDUAL:										
8 PRESTACIONES										
					.30		1.6	B	1.5	
RESPONSE:	1:	3	5	3	4	5	5	5	3	3 5
Z-RESIDUAL:							2			
RESPONSE:	11:	3	3	3	2	3	4	3	3	3 2
Z-RESIDUAL:										
RESPONSE:	21:	3								
Z-RESIDUAL:										
1 CONSUMO										
					-1.96		1.2	C	1.5	
RESPONSE:	1:	5	5	5	5	5	4	5	5	5 5
Z-RESIDUAL:							-3			
RESPONSE:	11:	5	5	4	3	4	2	4	4	3 5
Z-RESIDUAL:							-2			
RESPONSE:	21:	4								
Z-RESIDUAL:										

Fuente: Elaboración propia

Las desviaciones se presentan en la línea Z-RESIDUAL y pueden tener valores positivos y negativos. Los valores positivos indican que la observación del sujeto para ese ítem tiene un valor mayor de lo esperado por el modelo de Rasch al que se han ajustado los

¹⁷ Linacre (2014b): *Table 11.1 Item misfitting responses* en Winsteps Help for Rasch Analysis / Contents / Output Table Index

datos. Los valores negativos indican que la observación del sujeto para ese ítem tiene un valor inferior al esperado por el modelo de Rasch al que se han ajustado los datos.

En la línea RESPONSE se ubican los sujetos identificados por su número de entrada tras los códigos 1: (sujetos del 1 al 10), 11: (sujetos del 11 al 20) y sucesivamente.

Al igual que en los sujetos, el análisis de los desajustes se puede complementar mediante el uso del Escalograma de Guttman (epígrafe 5.1).

El análisis de los desajustes determina una selección de casos de especial observación pues denotan circunstancias fuera de lo esperado, generando una información preciosa para el análisis competitivo.

3.3.3.- DISCUSIÓN DE LOS DESAJUSTES EN LAS CALIBRACIONES DE LOS ÍTEMS Y SUS ORÍGENES.

El objetivo principal de TMR es la obtención del ajuste de los datos al modelo de Rasch con objeto de utilizar el modelo como estándar de dichos datos generando valores esperados ajustables a los observables. Ello puede llevar a la supresión de ítems o sujetos que desajustan distorsionando la medición.

El paso previo de la depuración es la revisión de los datos originales observados, con objeto de eliminar cualquier problema derivado con la administración del instrumento de medida. Superada esta fase inicial se deben considerar los desajustes de los sujetos, que pueden evidenciar tanto su no pertenencia al colectivo analizado o su pertenencia con algún tipo de distorsión específica en algún ítem del constructo.

Si el desajuste presenta un alto nivel de distorsión generado por un determinado sujeto que afecta a las calibraciones de los ítems se procedería a eliminarlo. El objetivo es que los ítems reduzcan al mínimo su nivel de desajuste. En caso de que el proceso de depuración de los sujetos no permita la reducción de los desajustes de los ítems se eliminaría el ítem. El límite de la depuración debe establecerse de acuerdo al marco conceptual utilizado, de tal forma que se pueda obtener una clara validez de contenido del constructo delimitado.

La información que se puede obtener en este proceso es importante para el análisis competitivo, pues puede determinar qué sujetos pueden considerarse pertenecientes o no a un determinado colectivo, qué anomalías presenta o que factor de competitividad está distorsionando el constructo delimitado, debido a la inclusión de ítems que no pertenecen al mismo.

Una utilidad secundaria derivada de este proceso, que se obtiene en una primera aproximación al ejecutar el programa Winsteps, es conocer la estructura competitiva de sujetos y factores de competitivos previa a los procesos de depuración. Esta información solo tiene una utilidad descriptiva o exploratoria de los posicionamientos de los sujetos y las jerarquías de los ítems, antes de iniciar la depuración de los desajustes que lleva a la obtención de mediciones objetivas de constructos cuyo contenido sea válido.

CAPÍTULO 4º

POSICIONAMIENTOS COMPETITIVOS DE GRUPOS

4.- POSICIONAMIENTO COMPETITIVO DE GRUPOS DE FACETAS

El análisis competitivo exige la determinación de la posición competitiva de agrupaciones de entes empresariales, sean empresas u organizaciones, productos,..., y sujetos económicos en general así como de la jerarquización de los factores de competitividad y de las referencias estratégicas consideradas en el análisis, de acuerdo a diferentes criterios competitivos.

Mediante el uso de TMR, al obtener mediciones objetivas de sujetos e ítems individuales en forma de mediciones intervalos aditivas es factible analizar competitivamente a esos grupos de sujetos e ítems tanto desde la perspectiva de su posición competitiva, decisiva para los análisis competitivos intergrupales e intragrupal, como desde la del comportamiento diferencial de los sujetos, de acuerdo al criterio de agrupación elegido, ante los diferentes ítems que integran el instrumento de medida, representativos de los factores de competitividad reflejados en el constructo competitivo delimitado.

4.1.- MEDICIÓN DE GRUPOS DE FACETAS

4.1.1.- POSICIONAMIENTOS DE GRUPOS DE SUJETOS. VALIDACIÓN DE SUS DIFERENCIAS.

Una vez obtenidas las mediciones individuales y los posicionamientos de los distintos sujetos, se puede proceder al posicionamiento de grupos de sujetos, integrados de acuerdo a un determinado criterio de clasificación¹⁸ competitivo.

Para obtener el posicionamiento de los grupos se obtienen previamente las medidas de cada grupo, que el programa Winsteps proporciona junto a los análisis estadísticos correspondientes (véase Tabla 4.1).

¹⁸ Linacre (2014b): *Table 28.1 Person subtotal summaries on one line in Winsteps Help for Rasch Analysis / Contents / Output Table Index*

En el caso que se analiza en la Tabla 4.1 se refiere a la obtención de la medición, previa a su posicionamiento, de dos segmentos de mercado de vehículos pequeños, agrupados según su tipología de ciudadanos y urbanos (codificados como C y U).

Se destaca la medida respectiva de cada grupo, así como la fiabilidad de separación (reproductividad) de las medidas de los grupos de sujetos obtenidos.

TABLA 4.1.

MEDICIONES DE GRUPOS DE SUJETOS

TABLE 28.1 Posicionamiento vehiculos 2006 U_C.xl								
INPUT: 21 PERSON 9 ITEM REPORTED: 21 PERSON 9 ITEM 4 CATS WINSTEPS 3.81.0								

Subtotal specification is: PSUBTOTAL=\$S1W1								
ALL PERSON SCORES ARE NON-EXTREME								

	PERSON	MEAN	S.E.	OBSERVED	MEDIAN	MODEL	MODEL	
	COUNT	MEASURE	MEAN	S.D.		SEPARATION	RELIABILITY	CODE

	21	.27	.32	1.41	.22	2.36	.85	*
	9	-.95	.40	1.12	-.62	1.70	.74	C
	12	1.19	.23	.78	1.33	1.03	.51	U

SUBTOTAL RELIABILITY=.91								
UMEAN=0 USCALE=1								

	PERSON	MEAN DIFFERENCE		Welch-2sided				
	CODE	CODE	MEASURE	S.E.	t	d.f.	Prob.	

	C	U	-2.14	.46	-4.65	13	.000	

	ANOVA - PERSON							
	Source	Sum-of-Squares	d.f.	Mean-Squares	F-test	Prob>F		

	\$S1W1	23.52	1.00	23.52	24.14	.0002		
	Error	18.51	19.00	.97				
	Total	42.03	20.00	2.10				

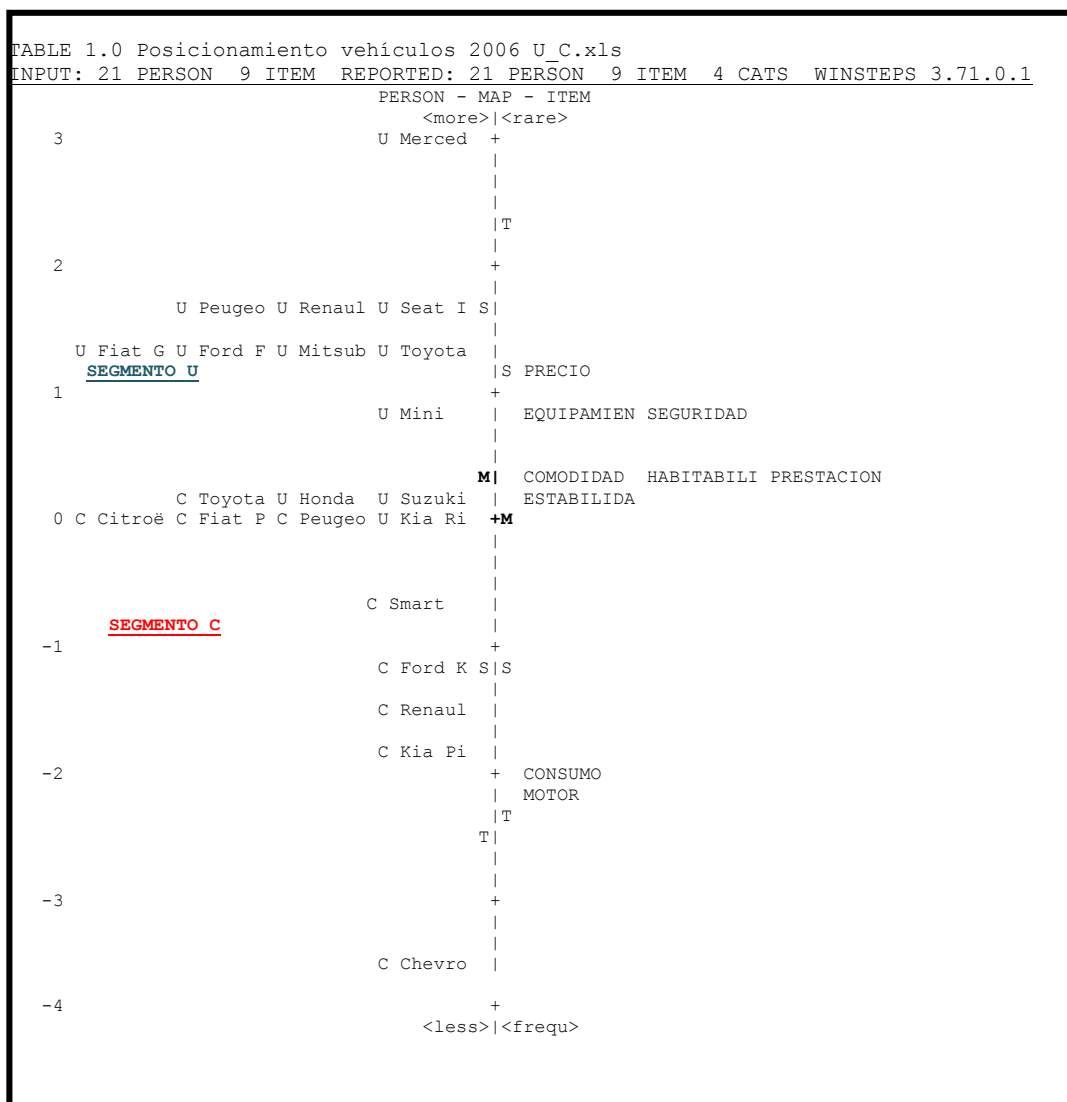
	Fixed-Effects Chi-Square: 21.6285 with 1 d.f., prob. .0000							

Fuente: Elaboración propia

El análisis del valor de p (significación del 5%), probabilidad obtenida en el test de significación de Welch – 2 colas, nos lleva a concluir que las medidas de los grupos de vehículos C y U son significativamente distintas, para cualquier nivel de significación.

GRÁFICO 4.1

POSICIONAMIENTO COMPETITIVO DE GRUPOS DE SUJETOS



Fuente: Elaboración propia

Esta situación se ratifica, también en la *Table 28.1* de Winsteps (Tabla 4.1) por medio del análisis ANOVA de un factor para las medidas de los subtotales (segmentos). Según indica Linacre (2014b) la cuestión es determinar si hay diferencia estadísticamente significativa entre las medidas obtenidas de los grupos (segmentos C y U) respecto a la medida global. La prob>F (p-valor) recoge la cola derecha de la probabilidad de valor del F-test con grados de libertad: Grupos analizados y Error.

Una probabilidad menor que 0.05 indica que hay diferencias significativas entre las medias (de las medidas).

A su vez, el test de homogeneidad de las muestras obtenidas de los grupos de sujetos analizados se lleva a cabo mediante una Chi-cuadrada de efecto fijo. La hipótesis nula es que todas las medidas de los grupos son iguales, exceptuando el efecto del error muestral. Una probabilidad menor de 0.05 indica diferencias significativas entre las medidas de los grupos analizados.

En el caso analizado (Tabla 4.1) los grupos (segmentos C y U) son competitivamente distintos, siendo la competitividad del segmento U superior a la del segmento C.

Las medidas de los diferentes grupos de vehículos (Segmentos C y U) se posicionan en el mapa de Wright (véase Gráfico 4.1 que recoge la *Table 1.0* de Winsteps).

4.1.2.- JERARQUIZACIÓN GRUPOS DE ÍTEMS

Una vez obtenidas las calibraciones de los ítems de forma individual y la jerarquización de los mismos, se puede proceder a la jerarquización de grupos de ítems, integrados de acuerdo a un determinado criterio de clasificación¹⁹ competitivo dentro del constructo delimitado.

Para obtener la jerarquización de los grupos de ítems se obtienen previamente las medidas de cada grupo, que el programa Winsteps proporciona junto a los análisis estadísticos correspondientes (véase Tabla 4.2).

En la Tabla 4.2 se analiza un caso de agrupaciones de tareas en un diseño de puestos de trabajo. Tareas Administrativas y de Marketing (codificadas como A y M). En esta tabla se destaca la medida respectiva de cada grupo, así como la baja fiabilidad de separación (reproductividad) de las medidas de los grupos de ítems obtenidos.

¹⁹ Linacre (2014b): *Table 27.1* Item subtotal summaries on one line en Winsteps Help for Rasch Analysis / Contents / Output Table Index

TABLA 4.2

MEDICIONES DE GRUPOS DE ÍTEMS

TABLE 27.1 VALORACIÓN DEL TRABAJO EN OFICINAS BA								
INPUT: 14 PERSON 9 ITEM REPORTED: 14 PERSON 9 ITEM 3 CATS WINSTEPS 3.81.0								

Subtotal specification is: ISUBTOTAL=\$S1W1								
ALL ITEM SCORES ARE NON-EXTREME								

ITEM	MEAN	S.E.	OBSERVED	MEDIAN	MODEL	MODEL		
COUNT	MEASURE	MEAN	S.D.		SEPARATION	RELIABILITY	CODE	

	9	.00	.79	2.25	-1.09	3.69	.93	*
	4	2.41	.37	.63	2.10	.31	.09	A
	5	-1.93	.32	.63	-1.93	.47	.18	M

SUBTOTAL RELIABILITY=.98								
UMEAN=0 USCALE=1								

ITEM	MEAN DIFFERENCE		Welch-2sided					
CODE CODE	MEASURE	S.E.	t	d.f.	Prob.			

A	M	4.34	.48	8.96	6	.000		

ANOVA - ITEM								
Source	Sum-of-Squares	d.f.	Mean-Squares	F-test	Prob>F			

\$S1W1	41.76	1.00	41.76	80.95	.0002			
Error	3.61	7.00	.52					
Total	45.37	8.00	5.67					

Fixed-Effects Chi-Square: 80.3063 with 1 d.f., prob. .0000								

Fuente: Elaboración propia

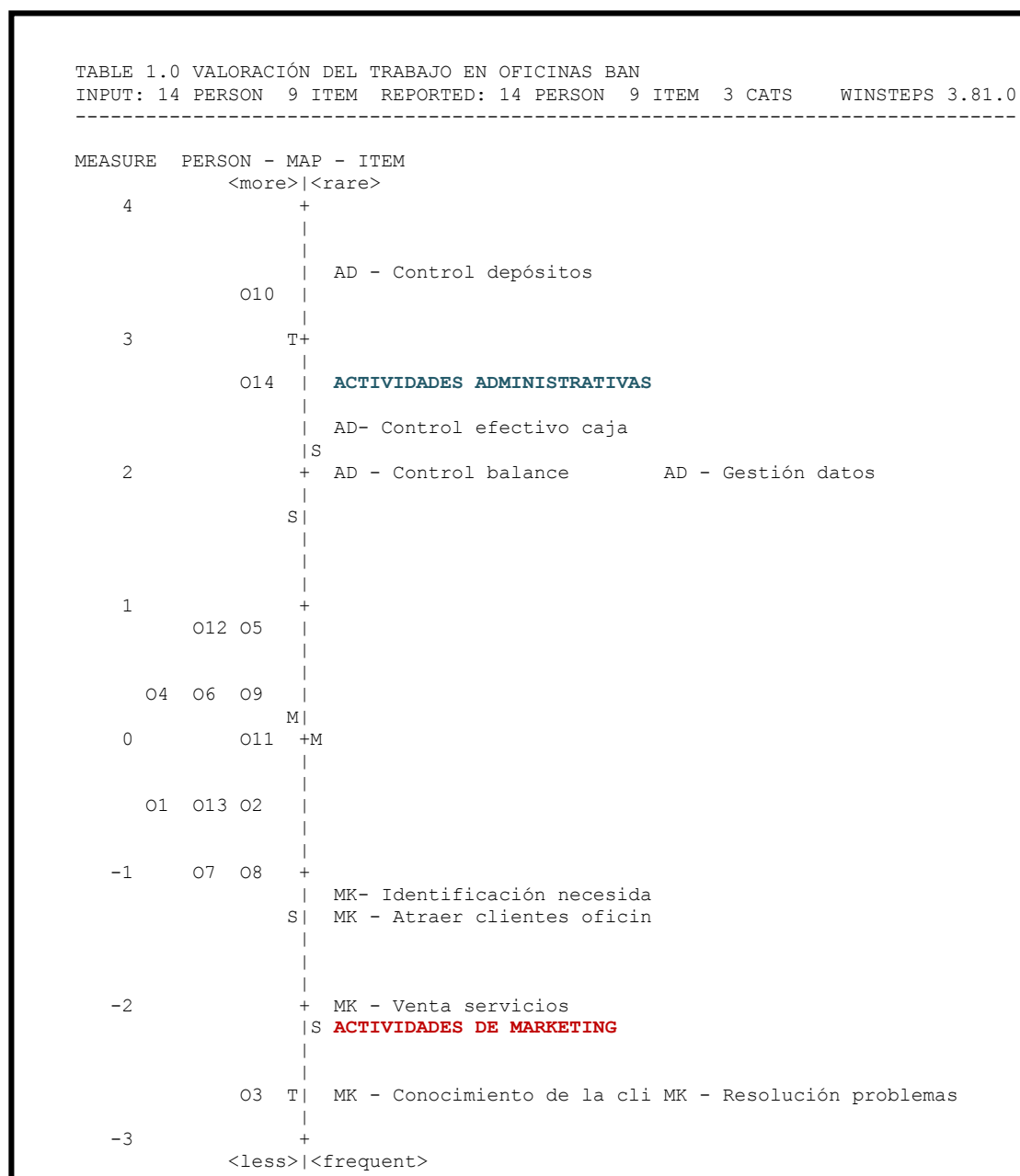
El análisis del valor de p (significación del 5%), probabilidad obtenida en el test de significación de Welch – 2 colas, nos lleva a concluir que las medidas de los grupos de ítems vinculadas a las tareas de Administración y Marketing son significativamente distintas.

Esta situación se ratifica, también en la *Table 27.1* de Winsteps (Tabla 4.2) por medio del análisis ANOVA de un factor para las medidas de los subtotales (Grupos de ítems A y M). Según indica Linacre (2014b) la cuestión es determinar si hay diferencia estadísticamente significativa entre las medidas obtenidas de los grupos (A y M) respecto a la medida global. La prob>F (p-valor) recoge la cola derecha de la probabilidad de valor del F-test con grados de libertad: Grupos analizados y Error.

Una probabilidad menor que 0.05 indica que hay diferencias significativas entre las medias (de las medidas).

GRÁFICO 4.2

JERARQUIZACIÓN COMPETITIVA DE GRUPOS DE ÍTEMS



Fuente: Elaboración propia

A su vez, el test de homogeneidad de las muestras obtenidas de los grupos de ítems analizados se lleva a cabo mediante una Chi-cuadrada de efecto fijo. La hipótesis nula es que todas las medidas de los grupos son iguales, exceptuando el efecto del error muestral. Una probabilidad menor de 0.05 indica diferencias significativas entre las medidas de los grupos analizados.

En el caso analizado (Tabla 4.1) los grupos de ítems, asociados a las tareas Administrativas y de Marketing, son distintos, siendo la presencia del grupo de ítems de Marketing mayor que el grupo de ítems asociados a las tareas de Administración.

Una vez obtenidas las medidas de los diferentes grupos de ítems (A y M) se procede a su jerarquización en el mapa de Wright (véase Gráfico 4.2 que recoge la *Table 1.0* de Winsteps).

4.2.- COMPORTAMIENTO DIFERENCIAL DEL ÍTEM (DIF)

En este epígrafe se ha seguido el trabajo de Oreja et al (2011) sobre el funcionamiento diferencial de ítems y el planteamiento práctico del manual de Winsteps de Linacre (2014b)²⁰.

4.2.1.- SESGO Y FUNCIONAMIENTO DIFERENCIAL DE ÍTEMES

La operatividad de un constructo necesita de un instrumento de medición configurado por los distintos componentes o factores del constructo. Los ítems se presentan como los componentes básicos de dicho instrumento de medición al reflejar el concepto a medir.

Los ítems como elementos básicos del cuestionario deben generarse partiendo de la premisa metodológica de que serán interpretados, cada uno de ellos, de la misma forma (es decir, realizan la misma función) por los distintos sujetos a los que se les administrará dicho cuestionario. Por lo tanto, todos los ítems de un cuestionario deben analizarse para

²⁰ Linacre (2014b): DIF – DPF – bias interactions concepts y DIF for two persons groups en Winsteps Help for Rasch Analysis / Contents / Special Topics

constatar que son apropiados para su aplicación a todo el colectivo de sujetos y a sus diferentes subpoblaciones o grupos.

En caso de que algún grupo, caracterizado por algún aspecto específico, dentro de los encuestados interprete de forma diferente alguno de los ítems estaremos en presencia de un sesgo (*Bias* en terminología inglesa). Un sesgo o funcionamiento diferencial en un ítem (DIF: Differential Item Functioning) implica que el ítem desempeña una función diferente en la medición del constructo.

La presencia de un DIF se aprecia por la existencia de distintos niveles de dificultad del ítem afectado para los colectivos analizados. Ello se debe a que hay una diferencia entre los grupos en su probabilidad de respuesta correcta a un ítem en particular. El impacto acumulativo de la presencia de sesgos en diferentes ítems repercutirá en la puntuación total, generando un sesgo en el test o *Test Bias* (Smith, 2004:391).

Si un grupo de sujetos está obteniendo de forma significativa una mejor calibración que otro grupo en la dificultad en un ítem (después de ajustarse por la puntuación total de los sujetos) estaremos, según indica Linacre (2010), en presencia de un DIF.

Según este autor, su significado podría ser que, entre otros:

1. Un grupo está actuando en su nivel actual de capacidad o habilidad, mientras que el otro está operando mejor de lo usual.
2. Un grupo está comportándose a su nivel usual de capacidad o habilidad, mientras que el otro está operando a un nivel inferior a lo usual.
3. **El ítem tiene su dificultad usual para un grupo, pero es más difícil de lo usual para el otro.**
4. El ítem tiene su dificultad usual para un grupo pero es más fácil de lo usual para el otro.

Aunque los procedimientos estadísticos no permiten determinar qué alternativa determina la presencia del DIF, se suele considerar políticamente correcta²¹ la alternativa 3.

Se pueden presentar dos tipos de DIF: DIF uniforme y DIF no uniforme (NUDIF). Consideramos que estamos ante un DIF uniforme cuando el impacto del comportamiento diferencial en el funcionamiento es similar en todos los niveles de capacidad o habilidad de los sujetos de los distintos grupos afectados. En el caso contrario estaremos ante el caso de un DIF no uniforme (NUDIF). En este segundo caso, el DIF puede tener diferentes tamaños para los diferentes niveles de habilidad dentro del grupo en el que se ha detectado.

4.2.2.- ASPECTOS OPERATIVOS, DETECCIÓN Y GRÁFICAS.

Como paso previo a la detección de un posible DIF, es conveniente analizar los grupos de sujetos y de ítems presentes en el constructo. De esta manera se podrá valorar la información que se deriva de sus medidas y la significación de sus diferencias entre sí, la fiabilidad de separación y los ajustes.

Posteriormente se procede al análisis de los posibles comportamientos diferenciales de cada uno de los grupos con respecto a los ítems del constructo.

Se pueden utilizar dos alternativas disponibles en el programa de computación Winsteps, dependiendo del número de grupos de sujetos y de la estrategia de comparación.

En la Tabla 4.3 ²² se presenta la comparación entre los distintos grupos de sujetos (empresas de diferentes tipos de tamaño: Mediana, pequeña y micro-empresas) entre sí, con respecto a los dos primeros ítems del constructo delimitado.

²¹ Desde esa perspectiva se considera que sí existe una discriminación contra el grupo que presenta un mayor nivel de dificultad. En ese sentido los responsables del cuestionario deben demostrar que no se está penalizando a un determinado sujeto por pertenecer a un grupo específico (sexo, edad, raza,...).

²² Linacre (2014b): *Table 30.1 Differential item functioning DIF pairwise en Winsteps Help for Rasch Analysis / Contents / Output Table Index*

TABLA 4.3

DIF DE GRUPOS DE SUJETOS, SEGÚN CRITERIO TAMAÑO, ENTRE SÍ,

TABLE 30.1 BD OBJ_SRP_EXT PYMES05.xlsx
 INPUT: 174 PERSON 25 ITEM MEASURED: 174 PERSON 25 ITEM 5 CATS WINSTEPS 3.70.0.2

DIF class specification is: DIF=@TAMAÑO

PERSON	DIF	DIF	PERSON	DIF	DIF	DIF	JOINT	Welch	MantelHanzl	ITEM	
CLASS	MEASURE	S.E.	CLASS	MEASURE	S.E.	CONTRAST	S.E.	t	d.f.	Prob.	Size Number Name
MED	-.02	.19	MIC	-.42	.14	.40	.24	1.69	65	.0954	.2617 +. 1 Desarrollo interno
MED	-.02	.19	PEQ	-.22	.12	.20	.23	.86	62	.3927	.8110 +. 1 Desarrollo interno
MED	1.10	.18	MIC	1.33	.13	-.24	.22	-1.06	65	.2924	.2191 +. 2 Absorción
MED	1.10	.18	PEQ	1.26	.12	-.17	.21	-.78	64	.4354	.2958 1.79 2 Absorción

Fuente: Elaboración propia

La hipótesis contrastada, por medio de la t de Welch es que el ítem tiene la misma dificultad para los dos grupos comparados ($p < 0.05$).

En la Tabla 4.4 se presenta la *Table 30.2* de Winsteps²³ que permite comparaciones de cada agrupación con respecto al conjunto del colectivo analizado, con referencia a cada uno de los ítems.

La hipótesis contrastada, por medio de la t de Welch, es que este ítem tiene la misma dificultad (determinada como la dificultad media) que todos los grupos del colectivo comparados ($p < 0.05$).

TEST DE SIGNIFICACIÓN

Una vez apreciados los síntomas de la existencia de un DIF (diferentes niveles de dificultad en el mismo ítem para los distintos grupos analizados), antes de proceder al ajuste del cuestionario, se debería comprobar que el DIF es estadísticamente significativo. Si así fuera, sería aconsejable reestructurar el cuestionario ajustando el ítem afectado a los diferentes subgrupos detectados. Con ello se procederá a obtener medidas ajustadas a la nueva situación.

²³ Linacre (2014b): *Table 30.2* Differential item functioning DIF bias/interaction analysis en Winsteps Help for Rasch Analysis / Contents / Output Table Index /

TABLA 4.4

DIF DE GRUPOS SUJETOS, SEGÚN CRITERIO TAMAÑO, CON RESPECTO AL CONJUNTO DE ÍTEMS.

TABLE 30.2 BD OBJ_SRP_EXT PYMES05.xlsx
 INPUT: 174 PERSON 25 ITEM MEASURED: 174 PERSON 25 ITEM 5 CATS WINSTEPS 3.70,0.2

DIF class specification is: DIF=@TAMAÑO

PERSON CLASS	OBSERVATIONS COUNT	BASELINE AVERAGE	BASELINE EXPECT	DIF MEASURE	DIF SCORE	DIF MEASURE	DIF MEASURE	DIF MEASURE	ITEM S.E.	t	Prob.	Number	Name
MED	28	2.64	2.86	-.26	-.22	-.02	.24	.19	1.25	.2228	1	Desarrollo interno	
MIC	65	2.78	2.65	-.26	.13	-.42	-.16	.14	-1.16	.2490	1	Desarrollo interno	
PEQ	75	2.73	2.77	-.26	-.04	-.22	.04	.12	.36	.7196	1	Desarrollo interno	
MED	28	1.39	1.21	1.26	.19	1.10	-.17	.18	-.95	.3523	2	Absorción	
MIC	61	1.03	1.10	1.26	-.06	1.33	.07	.13	.51	.6150	2	Absorción	
PEQ	71	1.18	1.20	1.26	-.01	1.26	.00	.12	.00	1.000	2	Absorción	

Fuente: Elaboración propia

Los factores que influyen en la significación estadística del DIF (Linacre, 2010) son:

- 1.- El tamaño del efecto DIF
- 2.- El tamaño de los grupos de clasificación

Bond y Fox (2007) sugieren tres indicadores para detectar un DIF basados en el estudio de grupos, que se recoge en el Cuadro 4.1

CUADRO 4. 1

INDICADORES DE UN DIF (BOX Y FOX)

Un valor de t superior a + 2.0 ó inferior a - 2.0
Un contraste del DIF de tamaño $> 0.5 $ logits
Una significación inferior a 0.05 ($p < 0.05$)

Fuente: Elaboración propia, a partir de Box y Fox (2007)

Estos indicadores han sido completados por las categorías propuestas por el Educational Testing Service (ETS) (véase Cuadro 4. 2) según recoge Linacre (2014b), citando a Zwick et al (1999).

CUADRO 4. 2
INDICADORES DE UN DIF (ETS)

Categorías ETS de DIF	Tamaño del efecto DIF (logits)	Significación estadística del DIF
C= de moderado a grande	$ DIF \geq 1.5 / 2.35 = 0.64$ logits	$P(DIF \leq 1 / 2.35 = 0.43 \text{ logits}) \leq 0.05$
B= de reducido a moderado	$ DIF \geq 1 / 2.35 = 0.43$ logits	$P(DIF = 0 \text{ logits}) \leq 0.05$
A= insignificante		

Fuente: Elaboración propia, a partir de Linacre (2014b) y Zwick et al (1999)

Nota: Se suele considerar que resultados DIF generados por grupos con tamaño inferiores a los 30 sujetos son poco robustos.

Análisis de los resultados del DIF

Se ha utilizado un caso de análisis de la percepción de puntos de referencia en función del tamaño de la empresa (García-Pérez; Oreja – Rodríguez y Yanes – Estévez, 2010).

El objetivo de este caso es identificar las posibles diferencias en la importancia que conceden las PYMES a un conjunto de objetivos estratégicos en función de su tamaño.

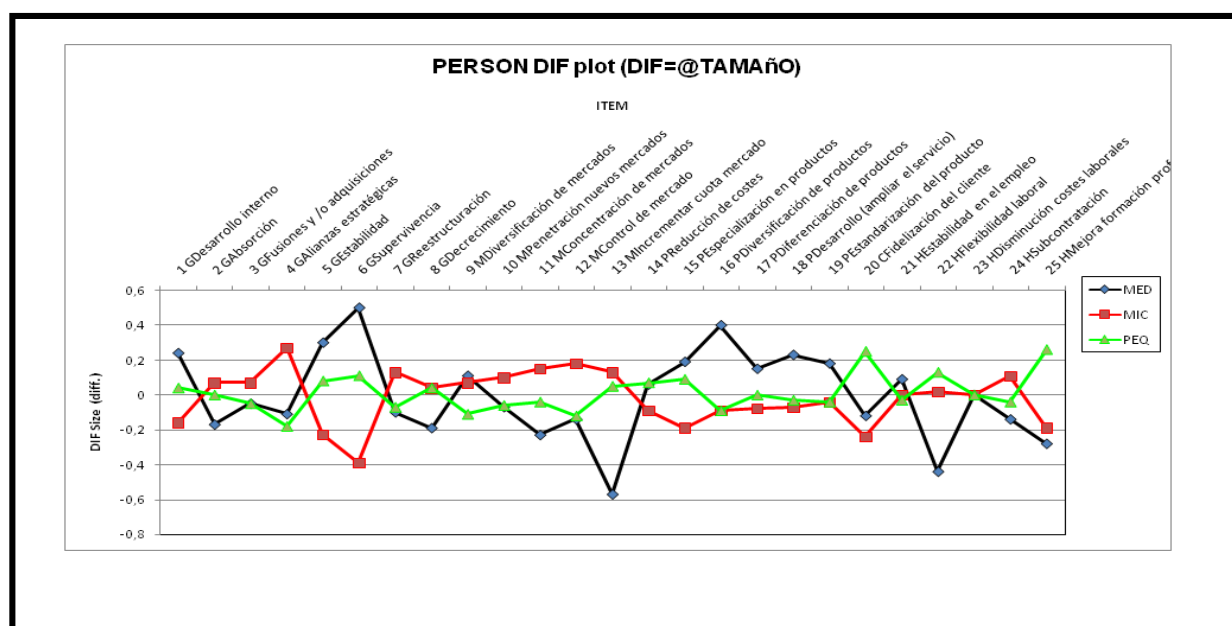
El colectivo analizado es una muestra de empresas en Canarias. La característica empresarial considerada con objeto de estudiar la posible presencia de DIF es el tamaño de la empresa. Las empresas se han agrupado en los siguientes conjuntos clasificatorios según su tamaños: Microempresas (MIC), pequeñas (PEQ) y medianas (MED).

TABLA 4.5
RESUMEN DE INFORMACIÓN SOBRE DIF SIGNIFICATIVOS

NIVEL DIF	COLECTIVO	TAMAÑO DIF	OBJETIVOS
A	MIC	0.27 logits	Alianzas
A	PEQ	0.26 logits	Mejora formación profesional
A	MIC	-0.39 logits	Supervivencia
A	MED	0.40 logits	Diversificación
B	MED	0.50 logits	Supervivencia

Fuente: Elaboración propia

GRAFICO 4.3
DIF (TAMAÑO) SIZE (DIF.)



Fuente: Elaboración propia

En la Tabla 4.5 se han destacado los ítems que presentan un nivel significativo de DIF ($P < 0.05$). Ahora bien de acuerdo al criterio ETS (Cuadro 4.2) recogido en Linacre (2014b) se puede profundizar en la implicación del DIF según su tamaño, por lo que se les ha asignado la etiqueta A (insignificante), B (moderado) y C (grande) al nivel del DIF. Ninguno de los colectivos refleja un gran DIF (nivel C).

En la Gráfica 4.3 se presenta el análisis gráfico de los niveles de dificultades por ítems de los diferentes grupos de empresas, agrupados de acuerdo a su tamaño: Medianas, pequeña y micro-empresas). Las diferencias entre los valores extremos pueden señalar la presencia de DIF, cuyo análisis nos indicará si son significativos.

4.3.- ANÁLISIS LONGITUDINAL DE FACETAS

4.3.1.- TIPOLOGÍA DE AGRUPAMIENTOS DE FACETAS PARA ANÁLISIS TEMPORALES.

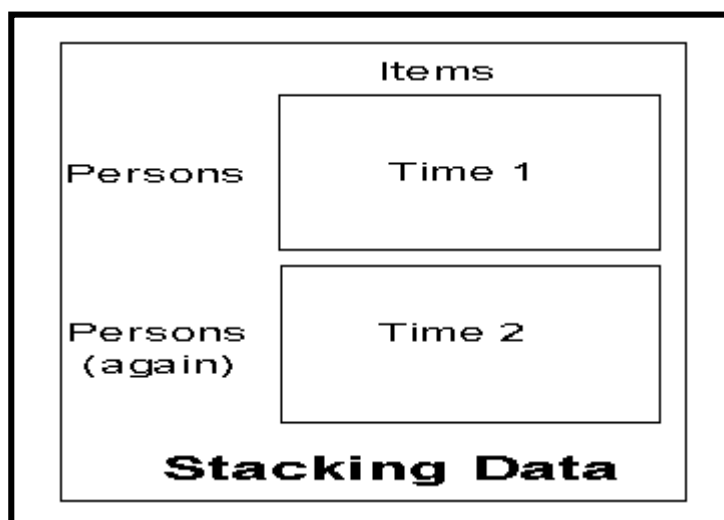
Siguiendo el trabajo de Yanes-Estévez y Oreja-Rodríguez (2008) para el análisis longitudinal del dinamismo del entorno de la empresa en Canarias se presentan las agrupaciones de datos Stack y Rack para análisis temporal.

El planteamiento teórico parte de Wright (2003) que presenta las aproximaciones Rack and Stack como sistemas de comparación de las medidas del mismo sujeto obtenidas a lo largo de los periodos distintos.

En la aproximación Stack se trataría de valorar dos situaciones del mismo sujeto presentes en dos momentos diferentes.

FIGURA 4.1

APROXIMACIÓN STACK



Fuente: Wright (2003)

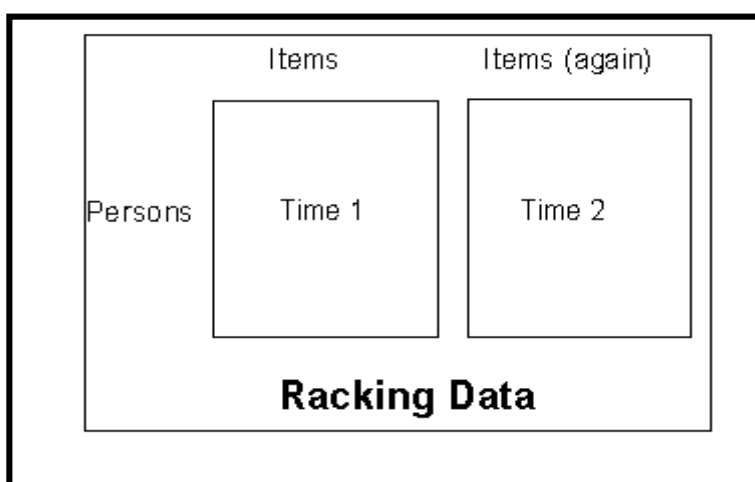
El sujeto se identifica como diferente en cada uno de los momentos (no es el mismo sujeto, uno es el inicial y otro el resultante de la acción de los ítems en el periodo de referencia) mientras que los ítems se calibran de forma conjunta en el mismo marco de referencia.

Se podría apreciar el cambio de los sujetos entre esos dos periodos diferentes.

En la aproximación Rack se consideran los cambios que han sufrido los ítems en los periodos analizados. Se calibran en un marco común los ítems para su comparación en el contexto de la jerarquización de los sujetos.

FIGURA 4.2

APROXIMACIÓN RACK



Fuente: Wright (2003)

Se podría analizar el cambio de los ítems (factores de competitividad) entre esos dos periodos diferentes.

ANÁLISIS EMPÍRICOS

Se utiliza el trabajo empírico disponible en Yanes-Estévez y Oreja-Rodríguez (2008) sobre el análisis del dinamismo del entorno empresarial de la empresa en Canarias.

4.3.2.- APROXIMACIÓN STACK

Esta perspectiva de análisis temporal de datos a partir de TMR, compara las percepciones del dinamismo desde el punto de vista de las empresas, analizando el cambio en los sujetos.

El análisis *Stack* compara las puntuaciones de un mismo sujeto en dos momentos diferentes del tiempo respecto a los mismos ítems. Para ello, se estructura la información obtenida, a partir de los datos de encuestas a un conjunto de empresas referidas al concepto dinamismo en los años 2000 y 2003, en una matriz de datos con formato como el indicado en la Figura 4.1.

En la aplicación del programa de computación Winsteps (Linacre 2014 a), a partir de dicha base de datos se obtiene las medidas de los sujetos recogidos en la Tabla 4.6.

Mediante este análisis se consigue también situar en el mismo constructo lineal a las observaciones de ambos años de las empresas para que compartan el mismo modelo y sean comparables.

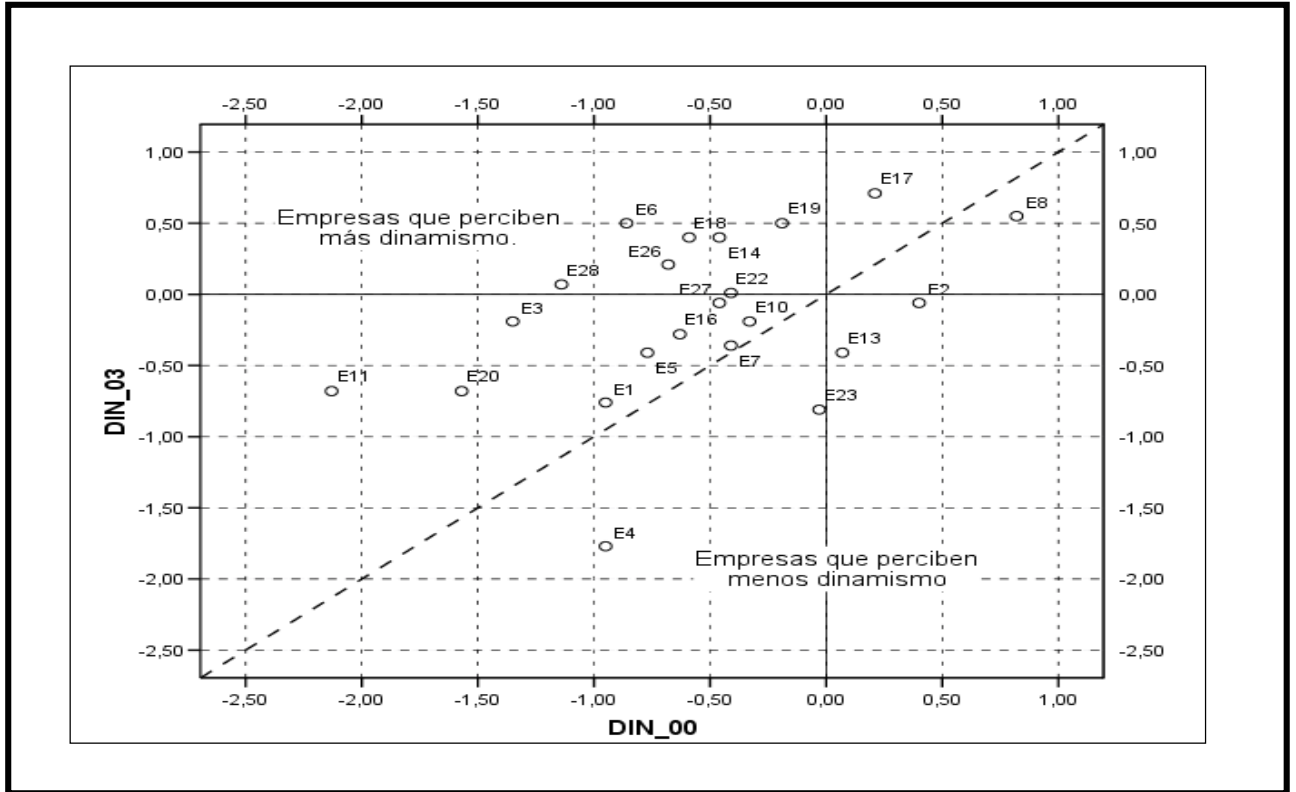
Las medidas de los sujetos (Tabla 4.6) son las que se representan en el Gráfico 4.4, cuyos ejes nos indican la medida asignada cada año a las percepciones del dinamismo en esas empresas para extraer las conclusiones del análisis Stack.

TABLA 4.6.
MEDIDAS DE DINAMISMO DE LAS EMPRESAS
EN LOS AÑOS 2000-2003.

EMPRESAS	MEDIDA DINAMISMO 2000	MEDIDA DINAMISMO 2003
E1	-,95	-,76
E2	,40	-,06
E3	-1,35	-,19
E4	-,95	-1,77
E5	-,77	-,41
E6	-,86	,50
E7	-,41	-,36
E8	,82	,55
E10	-,33	-,19
E11	-2,13	-,68
E13	,07	-,41
E14	-,46	,40
E16	-,63	-,28
E17	,21	,71
E18	-,59	,40
E19	-,19	,50
E20	-1,57	-,68
E22	-,41	,01
E23	-,03	-,81
E26	-,68	,21
E27	-,46	-,06
E28	-1,14	,07

Fuente: Yanes y Oreja (2008).

GRÁFICO 4.4
ANÁLISIS STACK DEL DINAMISMO DEL ENTORNO
ENTRE 2000 Y 2003.



Fuente: Yanes y Oreja (2008).

En el análisis del Gráfico 4.4²⁴, se toma de diagonal como referencia para identificar las diferencias entre las percepciones de los dos años.

El número de empresas situadas en cada grupo permite concluir que ha aumentado el dinamismo que se percibe del entorno por parte de estas empresas.

El análisis Stack puede identificar individualmente la trayectoria que ha seguido cada empresa.

²⁴ El Gráfico 4.4 ha sido elaborado con SPSS a partir de las medidas de las empresas.

4.3.3.- APROXIMACIÓN RACK

El análisis *Rack* tiene como objeto de estudio la evolución temporal de los ítems. Se compara cómo son percibidos en dos momentos temporales diferentes.

TABLA 4.7

**CALIBRACIONES DE DINAMISMO DE LOS ÍTEMS DEL ENTORNO
EN LOS AÑOS 2000 Y 2003.**

ÍTEMS DEL ENTORNO	CODIGO	CALIBRACIÓN DINAMISMO	
		2000	2003
Recursos Naturales	DRnat	1,51	,85
Orografía	DOrog	1,23	,81
Recursos Naturales	DRENA	1,19	,46
Tipo de cambio	DTipo	,49	,49
Barreras físicas	DBarf	,56	-,12
Economías de escala	DEeces	,49	-,12
Situación política	DSpol	,43	-,12
Segmentación de mercados	DSegm	,19	,19
Demografía	DDemo	,15	,09
Actitud ante la empresa	DActe	-,05	-,30
Legislación laboral	DLLab	,19	-,22
Defensa del consumidos	DDfco	-,05	-,22
Distancia a grandes mercados	DGram	,64	,05
Formación profesional	DForp	,02	-,44
Legislación sectorial	DLsec	-,02	-,51
Recursos financieros	DRecf	-,36	-,51
Recursos humanos	DRrhh	-,15	-,58
Recursos tecnológicos	DRect	-,62	-,66
Desarrollo de Canarias	DDesa	-,36	-,90
Motivación del consumidor	DMtco	-,15	-,87
Insularidad	DInsu	1,13	,94
Dependencia exterior	DDext	,04	-,05
Renta de la demanda	DRnta	,61	-,58
Situación de la demanda	DDmda	,85	-1,32
Competencia	DComp	,43	-,94

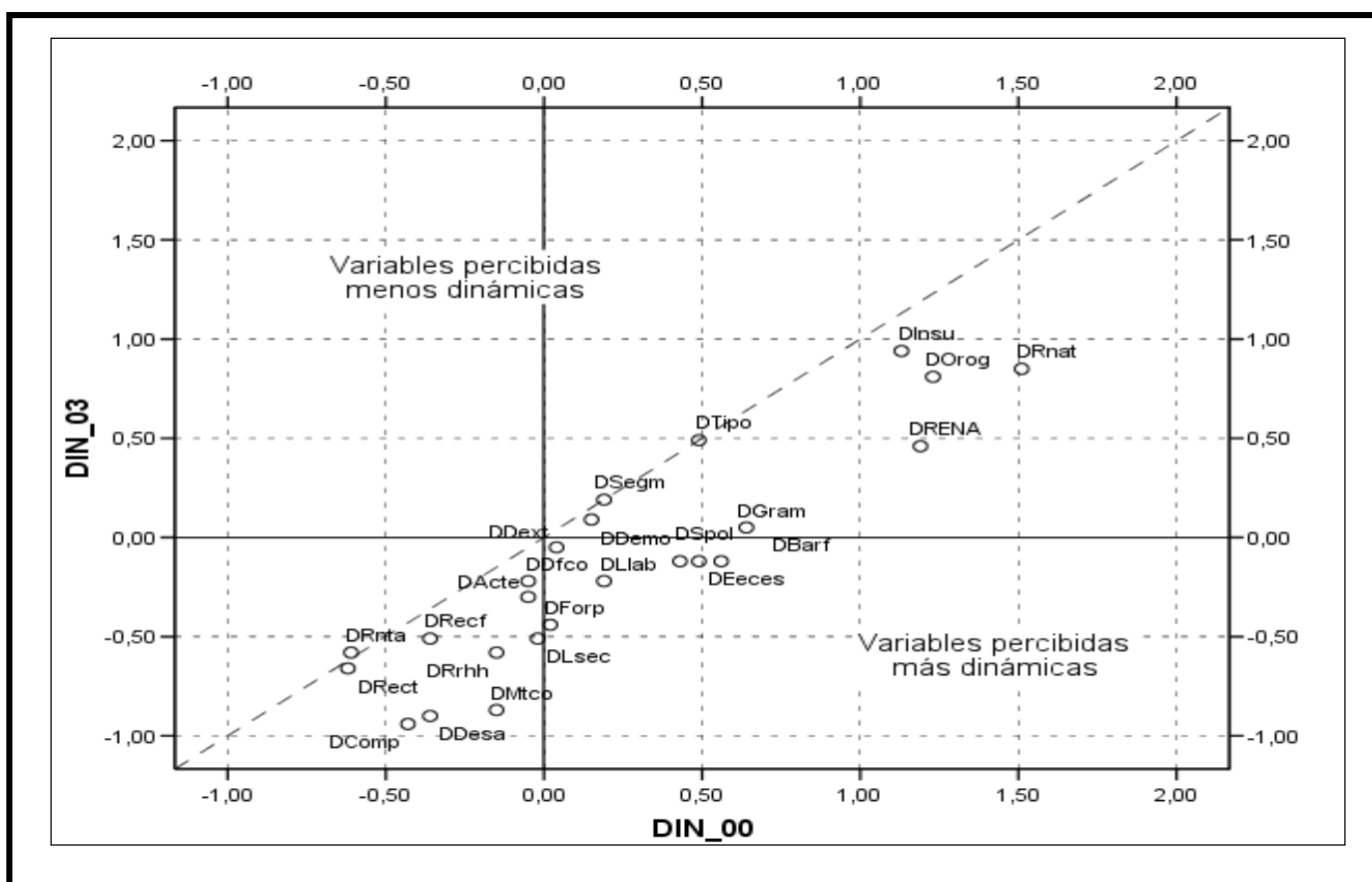
Fuente: Yanes y Oreja (2008).

A partir del proceso de encuestación se obtendrían dos observaciones de los ítems por cada sujeto, una de cada año (en este caso 2000 y 2003). Estas puntuaciones se conforman en una base de datos con la estructura recogida en la Figura 4.2.

Mediante la aplicación del programa de computación Winsteps (Linacre, 2014) se logra situar en el mismo constructo lineal a las observaciones de ambos años para que sean comparables.

Las medidas que se obtienen para los ítems, recogidas en la Tabla 4.7 se representan en un Gráfico 4.5, que nos permite conocer cómo ha evolucionado la percepción de cada ítem entre 2000 y 2003.

GRÁFICO 4.5.
ANÁLISIS RACK DEL DINAMISMO DEL ENTORNO
ENTRE 2000 Y 2003.



Fuente: Yanes y Oreja (2008).

Para interpretar los resultados del Gráfico 4.5²⁵ se toma la diagonal como referencia ya que cualquier desviación respecto a la misma implica que no se mantiene el mismo

²⁵ El Gráfico 4.5 ha sido elaborado con SPSS a partir de las medidas de las variables del entorno.

coeficiente en los dos años y que hay diferencias en cuánto de dinámico se percibe el ítem correspondiente.

Así, razonamos que las variables que se sitúan en la diagonal del gráfico o muy próximas, son consideradas igual de dinámicas en ambos años.

Las variables situadas por encima de la diagonal han sido percibidas como menos dinámicas.

Las situadas por debajo de la diagonal han sido percibidas como más dinámicas.

CAPÍTULO 5º

BENCHMARKING DE DIAGNÓSTICO.

5.- BENCHMARKING DE DIAGNOSTICO

5.1.- ESCALOGRAMA DE GUTTMAN

5.1.1.- BASE CONCEPTUAL Y CONTENIDO

El paradigma establecido por el psicómetra Louis Guttman (1916-1987) establece que las observaciones ideales de un test serían aquellas en que el sujeto tiene éxito en todos los ítems hasta una cierta dificultad y falla los que están por encima de la misma. Ese límite se establece en la medida de la habilidad del sujeto.

Cuando los sujetos y los ítems están ordenados de acuerdo a este paradigma dispondremos de un conjunto de datos presentados según el escalograma de Guttman.

En la Tabla 5.1²⁶ se recoge el Escalograma de Guttman (*Table 22* en Winsteps) que señala que la ordenación de los sujetos varía desde las altas medidas a las bajas en las filas, mientras que en las columnas los ítems se presentan desde los que tienen una medida baja (los más fáciles) hasta los que tienen una medida alta (los más difíciles). Se destaca en el escalograma la información de la diagonal de transición, situado en el ejemplo entre dos rayas rojas. En ítems politómicos con más categorías, esta área será más amplia.

Los recuadrados en verde muestra el avance probabilístico desde las respuestas de en las altas categorías para los ítems fáciles a las respuestas a las bajas categorías para los ítems difíciles.

La presencia de observaciones “fuera de lugar” nos guiará en la búsqueda de las causas determinante en el proceso comparativo. Como es el caso del recuadro rojo, que en zonas de cero, aparecen 1 y 2.

En las zonas OUT, las respuestas inesperadas influyen más intensamente en la determinación de los niveles que alcanzan los estadísticos OUTFIT. De igual forma, en la zona IN, las respuestas inesperadas influyen más intensamente en la determinación de los niveles que alcanzan los estadísticos INTFIT.

²⁶ Linacre (2014b): *Table 22 Scalograms* en Winsteps Help for Rasch Analysis / Contents / Output Table Index/

Linacre (2014b): Sorted observed data matrix – Guttman scalogram y Guttman patterns and Guttman spiltts en Winsteps Help for Rasch Analysis / Contents / Special Topics /

TABLA 5.1
ESCALOGRAMA DE GUTTMAN

GUTTMAN ESCALOGRAM OF RESPONSES		
KID	ACT	
	1111112 1 221 121 22	
	8920311251427634569784035	
2	+222222222222222222222222222222	M Roesser, Lawrence F.
41	+222222222222222222222222222222	F Fatake, Natasha
34	+222222222222222222222222222222	F Foster, Ruth
17	+222222222222222222222222222222	F Schuttner, Gail
50	+222222222222222222222222222222	M Doepner, Tom
45	+222222222222222222222222222222	F Maloughlin, Billy
7	+222222222222222222222222222222	M Wright, Benjamin
16	+222222222222222222222222222222	F Huff, Maury Baby
48	+222222222222222222222222222222	M Charrelle, Bernie
28	+222222222222222222222222222222	F Seiler, Karen
59	+222222222222222222222222222222	F Clapp, LB
18	+222222222222222222222222222222	M Ernst, Richard Max
39	+222222222222222222222222222222	M Spurrell, Rocky J.
23	+222222222222222222222222222222	M Troom, Jeff
56	+222222222222222222222222222222	M Hatwood, Clint
37	+222222222222222222222222222222	F Clapp, Chastie
40	+222222222222222222222222222222	M Badenov, Boris
70	+222222222222222222222222222222	M Kant, Clark
33	+222222222222222222222222222222	F Roesser, Bea
38	+222222222222222222222222222222	Moose, Bullwinkle
43	+222222222222222222222222222222	M Man, Spider
74	+222222222222222222222222222222	M Fataleasca, Ray
60	+222222222222222222222222222222	M Squilly, Michael
64	+222222222222222222222222222222	M Cosleone, Vito
11	+222222222222222222222222222222	M Haich, Paul Fred
22	+222222222222222222222222222222	F Hogan, Kathleen
58	+222222222222222222222222222222	F Clapp, Decent
3	+222222222222222222222222222222	M Roesser, Toby G.
51	+222222222222222222222222222222	M Reiss, Steve
63	+222222222222222222222222222222	M Backes, Selwin
8	+222222222222222222222222222222	M Lambert, Md., Ross W
69	+222222222222222222222222222222	M Blofeld, William
71	+202102222222222222222222222222	M Stoller, Dave
19	+222222222222222222222222222222	M Fontanilla, Nanea
24	+222222222222222222222222222222	F Torer, Amy Elizabeth
42	+222222222222222222222222222222	F Leades, Fearless
31	+222222222222222222222222222222	M Sabile, Jack
65	+222222222222222222222222222222	M Cosleone, Michael
1	+222222222222222222222222222222	M Roesser, Marc Daniel
55	+222222222222222222222222222222	M Baboo, Baboo
61	+222222222222222222222222222222	M Pinhead, Zippy
67	+222222222222222222222222222222	M Allen, Peter
10	+222222222222222222222222222222	M Squilly, Ray Of
54	+222222222222222222222222222222	F Malaprop, Mrs.
44	+22122212121210212111100000	M Haich, Daniel Seb
73	+0212020102221112211012200	M Eisan, Moen L.
4	+222222222222222222222222222222	M Cianci, Buddy
15	+1211111111111111111111111111	M Balser, Ed
27	+2221211111300111110110120	M Sandberg, Ryne
35	+2221211111111111111111111111	M Roesser, Michael T.
37	+222222222222222222222222222222	F Dyson, Stephanie Nina
52	+2211111111111111111111111111	M Denny, Don
20	+222222222222222222222222222222	M Rinzler, James
32	+21212222112211111110100000	M Aisenhead, John
46	+222222222222222222222222222222	M Stultz, Maxwell
75	+222222222222222222222222222222	M Anguiano, Rob
26	+1111111111111111111111111111	M Roesser, Jack
36	+222222222222222222222222222222	M Muller, Jeff
55	+222222222222222222222222222222	M Furling, Linda
6	+222222222222222222222222222222	M Mainan, Raymond
9	+212222222222222222222222222222	M Amirault, Zippy
13	+222222222222222222222222222222	M Driskel, Michael
47	+221222222222222222222222222222	M Roesser, Tr Cat
14	+222222222222222222222222222222	M Scholz, Matthew
49	+1110111111111111111111111111	M Lieberman, Benjamin
5	+2211111100111111111111111111	M Van Dam, Andy
12	+1110201100102001110010012	F Hwa, Nancy Marie
30	+222222222222222222222222222222	M Boudet, Gerard
29	+221101100100000000010121	F Roesser, Rebecca A.
72	+211010001110000010001230	M Bond, James
53	+2110011110000000001120001	M Lieberman, Daniel
	1111112 1 221 121 22	F Woodgren, Jan Swede
	8920311251427634569784035	M Lindman, Alan
		M Jackson, Solomon
		M Sabol, Andrew

Fuente: Linacre (2014b)

5.1.2- INTEGRACIÓN DE LOS DESAJUSTES EN LAS MEDIDAS DE LAS FACETAS

La información obtenida de las medidas de los sujetos y sus desajustes, así como las calibraciones de los ítems y sus desajustes se integran en el Escalograma de Guttman.

Así en el caso de los desajustes de los sujetos (coches) recogidos en la Tabla 3.3 y de los desajustes de los ítems recogidos en la Tabla 3.6 se puede apreciar su integración mediante los Escalogramas de Guttman obtenidos en las *Tables* 22.1 y 22.2 de Winsteps, integradas en la Tabla 5.1.

TABLA 5.2
ESCALOGRAMAS DE GUTTMAN PARA DOS SEGMENTOS DE AUTOMÓVILES

<i>Table 22.1 de Winsteps</i>		<i>Table 22.2 de Winsteps</i>	
GUTTMAN SCALOGRAM OF RESPONSES:		GUTTMAN SCALOGRAM OF ZONED RESPONSES: @=0, A=1, B=2, C=3; D=4, E=5	
PERSON	ITEM	PERSON	ITEM
	614258397		614258397
	-----		-----
5	+544445555 U Mercedes Clase A	5	+5DDDD5EEE U Mercedes Clase A
8	+555433444 U Peugeot 207	8	+55E4CC444 U Peugeot 207
9	+554533444 U Renault Clio III	9	+554ECC444 U Renault Clio III
10	+554445334 U Seat Ibiza	10	+55444ECC4 U Seat Ibiza
1	+455443443 U Fiat Grande Punto	1	+D5E44C44C U Fiat Grande Punto
2	+554445333 U Ford Fiesta	2	+55444ECCC U Ford Fiesta
7	+454455333 U Mitsubishi Colt	7	+D544EECCC U Mitsubishi Colt
12	+553443444 U Toyota Yaris 2006	12	+55C44C444 U Toyota Yaris 2006
6	+444425344 U Mini	6	+DD44BE3DD U Mini
3	+454332333 U Honda Jazz	3	+4ED3E3B33 U Honda Jazz
11	+454343333 U Suzuki Swift	11	+4ED3D3333 U Suzuki Swift
21	+543433334 C Toyota Aygo	21	+E43D3333D C Toyota Aygo
4	+453344332 U Kia Rio	4	+4E33DD33B U Kia Rio
13	+543433333 C Citroën C1	13	+E43D33333 C Citroën C1
15	+543443332 C Fiat Panda	15	+E43DD333B C Fiat Panda
18	+543433333 C Peugeot 107	18	+E43D33333 C Peugeot 107
20	+453222443 C Smart Fortwo	20	+4E3BBBDD3 C Smart Fortwo
16	+423323333 C Ford Ka	16	+4B33BD333 C Ford Ka
19	+433243322 C Renault Twingo	19	+4C3BD33B2 C Renault Twingo
17	+442233322 C Kia Picanto	17	+44BB33C22 C Kia Picanto
14	+233232222 C Chevrolet Matiz	14	+B3C2C2222 C Chevrolet Matiz
	-----		-----
	614258397		614258397

Fuente: Elaboración propia

En la parte de la izquierda de la Tabla 5.2 (*Table 22.1 de Winsteps*) se corresponde a los conceptos estrictos del planteamiento de Guttman. Integrando en las filas de mayor a menor medida a los sujetos en función de su posicionamiento. En las columnas se recogen

los ítems o factores de competitividad, clasificadas de mayor (frecuentes) a menor frecuencia (raros) de acuerdo a las calibraciones obtenidas.

5.1.3.- INFORMACIÓN SOBRE LAS DESVIACIONES DE LAS OBSERVACIONES RESPECTO A LAS EXPECTATIVAS.

En la parte de la derecha de la Tabla 5.2 se incluye la información de la *Table 22.2* de Winsteps que presenta una referencia al valor esperado de cada sujeto /ítem. Así las observaciones obtenidas se expresan con una marca, en forma de letra, según esté o no conforme con el valor esperado.

Las observaciones que están alrededor de sus expectativas (± 0.5 puntos de las categorías) conservan el valor de la categoría: {1, 2,...} utilizada en el análisis. Aquellas que estén fuera de este rango aparecen con letras: “A, B,...”

5.2.- FICHAS DE DIAGNÓSTICO COMPETITIVO (PKMAP / KIDMAP)

La presentación de las Fichas de Diagnostico Competitivo se basa en el trabajo de Oreja-Rodríguez y Montero –Muradas (2012), referido a la revisión teórica de los PKMAP / KIDMAP, habiendo sido ampliado mediante un análisis empírico desarrollado de acuerdo al manual de Winsteps de Linacre (2014b).

5.2.1.- BASE TEÓRICA

Los PKMAPS / KIDMAPS se obtienen a partir de la aplicación del modelo de Rasch como representaciones gráficas con un informe individual de los resultados de la aplicación del instrumento de medida utilizado. Se presentan las respuestas de las personas a los ítems, en sus diferentes categorías de la escala de medida aplicada.

Estos informes de diagnóstico de sujetos analizados con el modelo Rasch fueron desarrollados por Wright et al (1980), proporcionando una descripción del comportamiento individual de un conjunto de ítems en un formato gráfico. La información proporcionada es una detallada relación, a nivel individual, entre el sujeto y los ítems constatados. De la información representada se puede diagnosticar la presencia de posibles puntos fuertes y débiles de los sujetos analizados (Lee y Yeoh, 1998).

Masters (1994) indica que con los KIDMAP se abrió una vía para presentar con detalle los resultados educativos, mediante la indicación de los ítems en los cuales han tenido éxito o han errado los alumnos. Los KIDMAP proporcionan, de forma visual, información clara de los sujetos, dicha información podría ocultarse entre el conjunto de tablas y gráficos que un proceso de aplicación del Modelo de Rasch proporciona (Doig, 1990).

Desde 1989 se han venido aplicando los cuadros de diagnóstico KIDMAP a millones de estudiantes a lo largo del mundo. Master (1994) destaca que la primera gran ejecución de los KIDMAP se llevó a cabo en los ochenta, distribuyendo entre los padres cientos de miles de informes, por parte de Los Angeles Independent School District (EEUU). En Australia se inicia la aplicación en febrero de 1982, en el Western Australia Institute for Educational Research, para posteriormente extenderse a las escuelas del estado de Victoria.

En 1989 se viene aplicando el Modelo de Rasch a todo el sistema educativo australiano (Doig, 1990). Como indica Doig (1990) fueron pioneros proyectos como el desarrollado por Australian Council for Education Research, dirigido por Dr. Jan Loban, que llevó a cabo exámenes de “*Numeracy y Literacy*” a niños de 6º año, con un resultado de 56000 estudiantes evaluados. Después de un proceso de calibración de las cuestiones de los exámenes y de su aplicación a los estudiantes, los informes desarrollados para informar a los padres se conformaron de acuerdo a los KIDMAP.

El estado australiano de Nueva Gales del Sur (New South Wales) lo aplica para cada estudiante en el NSW Basic Skills Testing Program en el periodo 1989 – 1993 (Masters, 1994).

Desde las primeras aplicaciones hasta la actualidad la reacción ante los KIDMAP ha sido positiva ya que frente a los tradicionales sistemas evaluativos (aprobado / suspenso) se presentan información sobre la situación del estudiante y permite evaluar sus posibles pautas de mejora. A su vez, la integración de los distintos informes de los estudiantes de un centro, permite evaluar los recursos necesarios para la mejora de dicho centro educativo.

Su divulgación ha generado múltiples alternativas de utilización, incluso desarrollándolos a nivel manual, como instrumento de autoevaluación comparativa de los alumnos en los propios centros (por ej. Doig, 1990; Lee y Yeoh, 1998; Linacre, 1999).

La aplicación de los KIDMAP ha avanzado a campos diferentes de los Educativos. Recientemente ha sido presentada a la comunidad científica de KIDMAP aplicados a la gestión hospitalaria, en relación a la percepción de cómo se sienten tratados los pacientes en distintos centros hospitalarios. Estas aplicaciones, en conjunción con la utilización de modernas tecnologías de la información, como Internet, ha derivado hacia los conocidos Web-KIDMAP (Chien et al 2006 y 2009; Chien et al 2009).

En el desarrollo de estos informes de diagnóstico de los sujetos analizados es necesario la utilización de un programa de computación. Varios han sido los tradicionalmente utilizados. El más conocido es Quest (Adams y Khoo, 1998), en nuestro caso utilizaremos el PKMAP de Winsteps (véase epígrafe 5.2.2.).

El programa utilizado genera una página para cada sujeto analizado que muestra que puntuación ha asignado a cada ítem y, dada la puntuación total, podría haber sido la respuesta esperada. La comparación de la respuesta real y la esperada clarifica las implicaciones de la habilidad estimada del sujeto e identifica cualquier incongruencia en sus pautas de respuestas (Wright et al, 1980).

La construcción básica se centra en cuatro cuadrantes que englobaría el cruce de dos alternativas. Por una parte el nivel de dificultad de los ítems y por otro el acierto o no (o su nivel de logro, dependiendo del tipo de escala utilizada).

Estimada la posición del sujeto y establecido su intervalo de precisión se puede destacar la información de los cuatro cuadrantes (véase casuística en el epígrafe 3.2). Atención en todos los casos a los desajustes que evidencia el sujeto.

Un cuadrante se referirá a los ítems fáciles (los que están por debajo de su medida, incluido el error estándar de la medición) acertados (logrados a un nivel adecuado por el sujeto). Es lo que se esperaba del sujeto.

En la misma vertical se referirá a los ítem más difíciles, los que están por encima de su medida, pero acertados por el sujeto. Esto no se esperaba de él. Está por encima de sus

posibilidades. ¿Cuál ha sido la razón? Simplemente los ha adivinado o es una pauta que evidencia una especial preparación o punto fuerte.

Un tercer cuadrante se refiere a los ítems fáciles (cuyo nivel de dificultad está por debajo del margen de precisión de la medida del sujeto, y que no han sido acertados (logrados a nivel adecuado). Se espera que los ítems fáciles sean acertados por el sujeto, con un alto nivel de probabilidad, pero no ha sido así. Razones: Descuido o una pauta que evidencia puntos débiles.

Encima del anterior cuadrante, en la vertical de los ítems no acertados (o logrados), pero por encima del nivel de precisión de la medida del sujeto, en el espacio de los ítems difíciles, nos encontramos con ítems difíciles no acertados por el sujeto. Es lo lógico, su nivel de probabilidad de acierto es pequeño.

5.2.2.- REPRESENTACIONES GRÁFICAS SEGÚN TIPOLOGÍA DE CATEGORÍAS.

En el manual de Linacre (2014b), las posibles salidas de la *Table 36* de Winsteps (PKMAP) se configuran de la siguiente forma, según el modelo de Rasch utilizado.

MODELO DE ESCALAS ORDENADAS DE RASCH DICOTÓMICO (MEORD)

El informe de diagnóstico PKMAP ha sido realizado usando un modelo de escalas ordenadas dicotómicas de Rasch. (Figura 5.1).

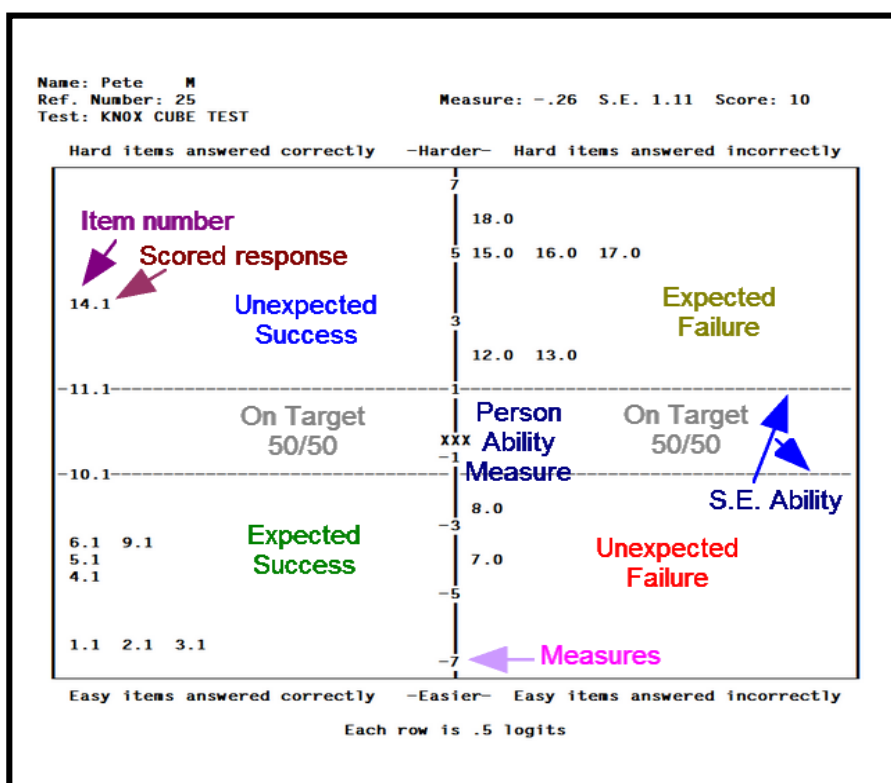
Este sería el informe de diagnóstico típico en el caso de aplicar un modelo dicotómico de Rasch. Junto a la identificación del ítem (ítem number) se indica la puntuación de la respuesta (scored response). Así: 11.1, indica respuesta acertada del ítem 11. El informe identifica al sujeto: Pete, M. su número de referencia 25 y el tipo de test llevado a cabo: KNOX CUBE TEST. Así como su nivel de habilidad y puntuación.

En la base y techo del gráfico se establecen los límites de los niveles de dificultad de los ítems (fáciles abajo y difíciles arriba). De derecha a izquierda, nos encontramos con los tipos de respuestas. Correcta e incorrectas. La combinación de estas cuatro alternativas establece un espacio que puede concretarse en cuatro cuadrantes.

El primer cuadrante analizado, bajo izquierda, recoge el área de éxito esperado. Ítems fáciles correctamente acertados. Su límite superior se establece alrededor de la medida del sujeto. En esa vertical un segundo cuadrante, alto izquierda, tiene un límite inferior en el nivel de precisión del sujeto y recoge los éxitos inesperados. Como es el caso del ítem 14 (14.1).

FIGURA 5.1

PKMAP MEORD (ASCII = NO)



Fuente: Linacre (2014b)

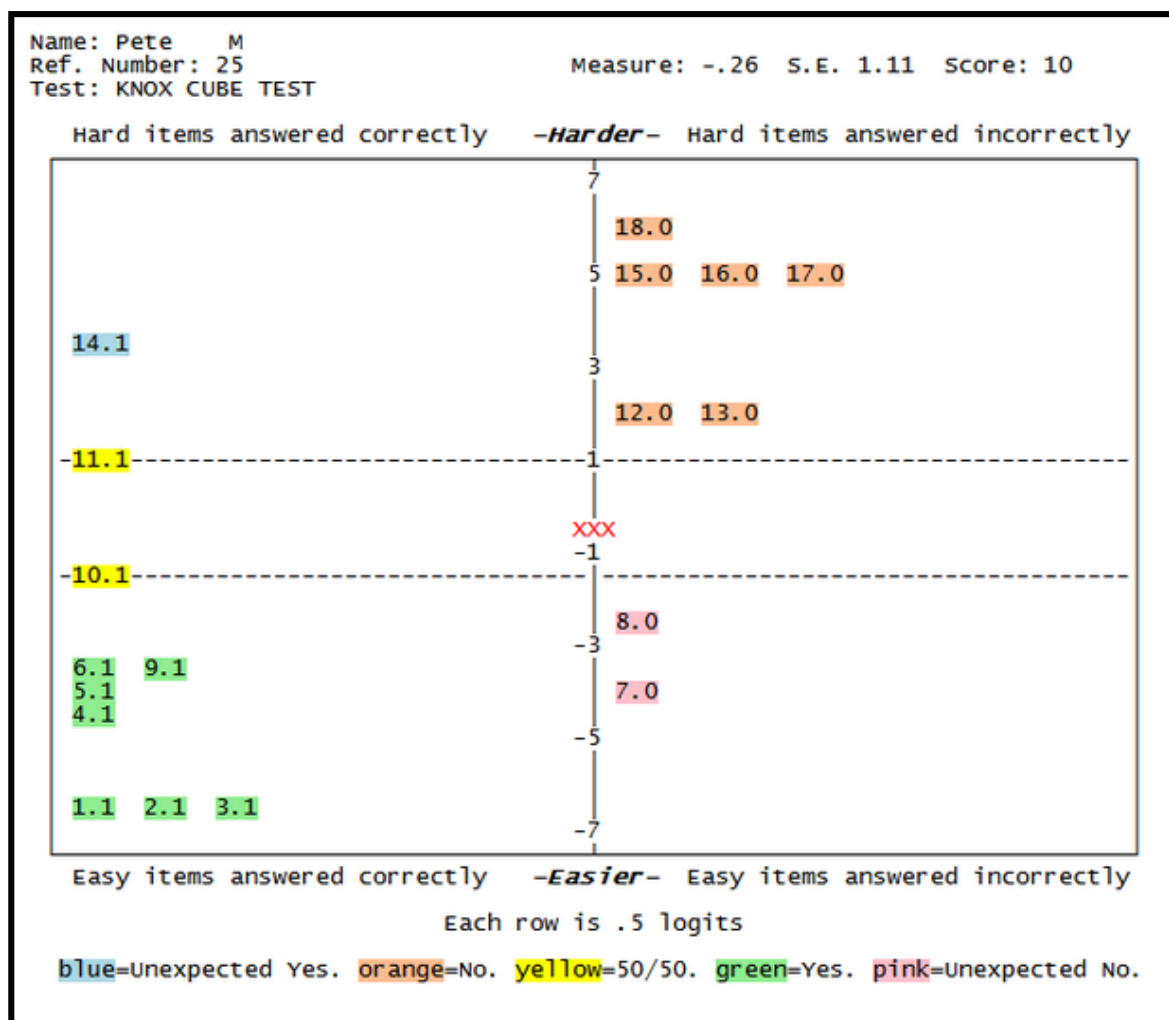
Abajo a la derecha, se ubicaría el tercer cuadrante considerado. Se refiere a los fallos inesperados del sujeto. Ítems fáciles no acertados. Sería el caso de ítem 8 (8.0). En esta vertical, por encima de intervalo de precisión de la medida del sujeto, se ubicaría el cuarto cuadrante, referido a fallos esperados. Son ítems difíciles para el sujeto y se esperaba que no los acertarse. Es el caso del ítem 18 (18.0).

La Figura 5.2 presenta una versión del PKMAP para modelos dicotómicos de Rasch preparada para una salida tipo “webpage” (véase, por ejemplo, los trabajos de Chien et al

2009; Chien et al, 2009). Lo que implica su presentación con un corte de página (por cada informe) en el proceso de impresión.

FIGURA 5.2

PKMAP MRD (ASCII = WEBPAGE)



Fuente: Linacre (2014b)

MODELO DE ESCALAS ORDENADAS RASCH POLITÓMICO (MEORP)

El informe de diagnóstico PKMAP ha sido realizado usando un modelo de escalas ordenadas politómicas de Rasch.

En la Figura 5.3 se recoge en vertical el continuo alto y bajo. Los extremos son en la parte inferior: lo más bajo (Lower) y en la parte alta (Higher): lo más alto. Se refiere al

nivel de habilidad necesaria en el sujeto para alcanzar el nivel de dificultad del ítem en la categoría señalada.

De izquierda a derecha, las alternativas son de alcanzada y no alcanzada. El informe recoge la puntuación asignada a cada ítem por el sujeto, comparándola con lo que se esperaba debería haber sido asignado de acuerdo a la habilidad del sujeto.

En el primer cuadrante, abajo a la izquierda, denominado por Winsteps como *Expected Level*, se refiere a ítems establecidos como de fácil nivel para el sujeto y que han sido alcanzadas por el mismo, al señalar una categoría esperada. Su límite superior es el margen de precisión de su medida de habilidad. En este caso estaría el ítem 9, con respecto a la puntuación en el mismo con la categoría 2 de su escala de medida (9.2).

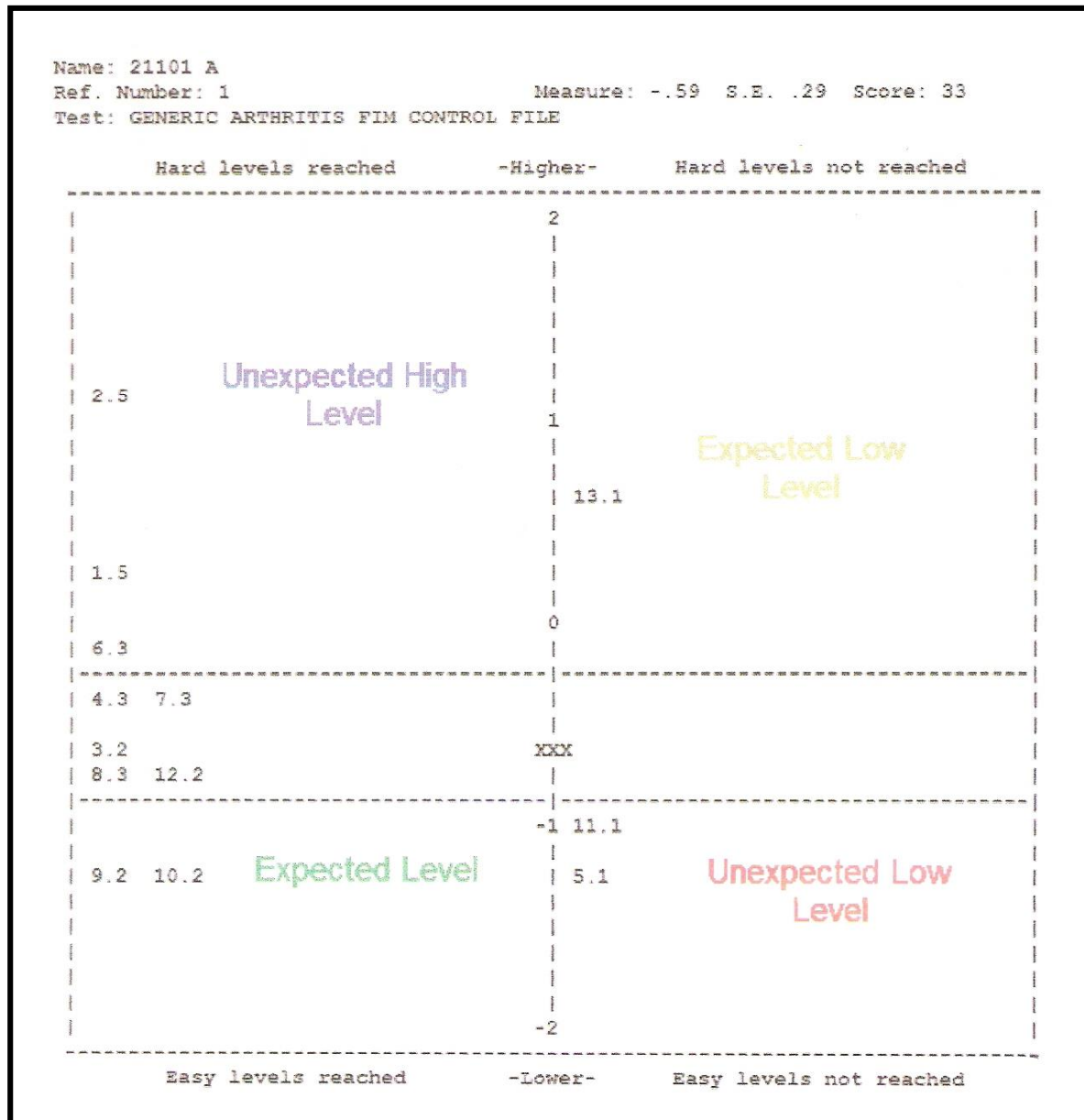
Encima del intervalo de precisión de la medida del sujeto y en la vertical del nivel de dificultad de los ítems se encuentra el segundo cuadrante *Unexpected High Level*, donde no se espera que el sujeto alcance el nivel que se recoge. Así el hecho de que el sujeto haya indicado la categoría 5 en el ítem 2, es algo inesperado. Puede suponer un punto fuerte del mismo.

En la parte derecha del informe de diagnóstico PKMAP se ubican las cumplimentaciones del cuestionario que evidencia un bajo nivel de logro del sujeto. En el tercer cuadrante, bajo derecha, tiene como límite superior el intervalo de precisión de su medida de habilidad, es denominado por Winstep como *Unexpected Low Level*. El sujeto ha señalado en la escala de medición una puntuación para el ítem baja. Siendo una zona de ítems y categorías fáciles ésto no se esperaba de él. Debería haber señalado un nivel más alto, en la escala utilizada, habida cuenta de su nivel de habilidad. Sería el caso del ítem 5, al que el sujeto le ha asignado una puntuación 1 (en una escala de 5 puntos).

En la vertical de la derecha y por encima del tercer cuadrante se encuentra el cuarto, cuyo límite inferior el intervalo de precisión de la medida del sujeto. Winsteps le denomina *Expected Low Level*. En este cuadrante se encuentran ítems, cuyas categorías que están por encima del nivel de habilidad del sujeto y no se espera que el mismo le asigne una alta puntuación en la escala de medida de los mismos. Es algo esperado. Así el ítem 13 no se esperaba que el sujeto indicase un alto nivel de puntuación y efectivamente se le ha asignado una puntuación de uno (la más baja en la escala de 1 a 5).

FIGURA 5.3

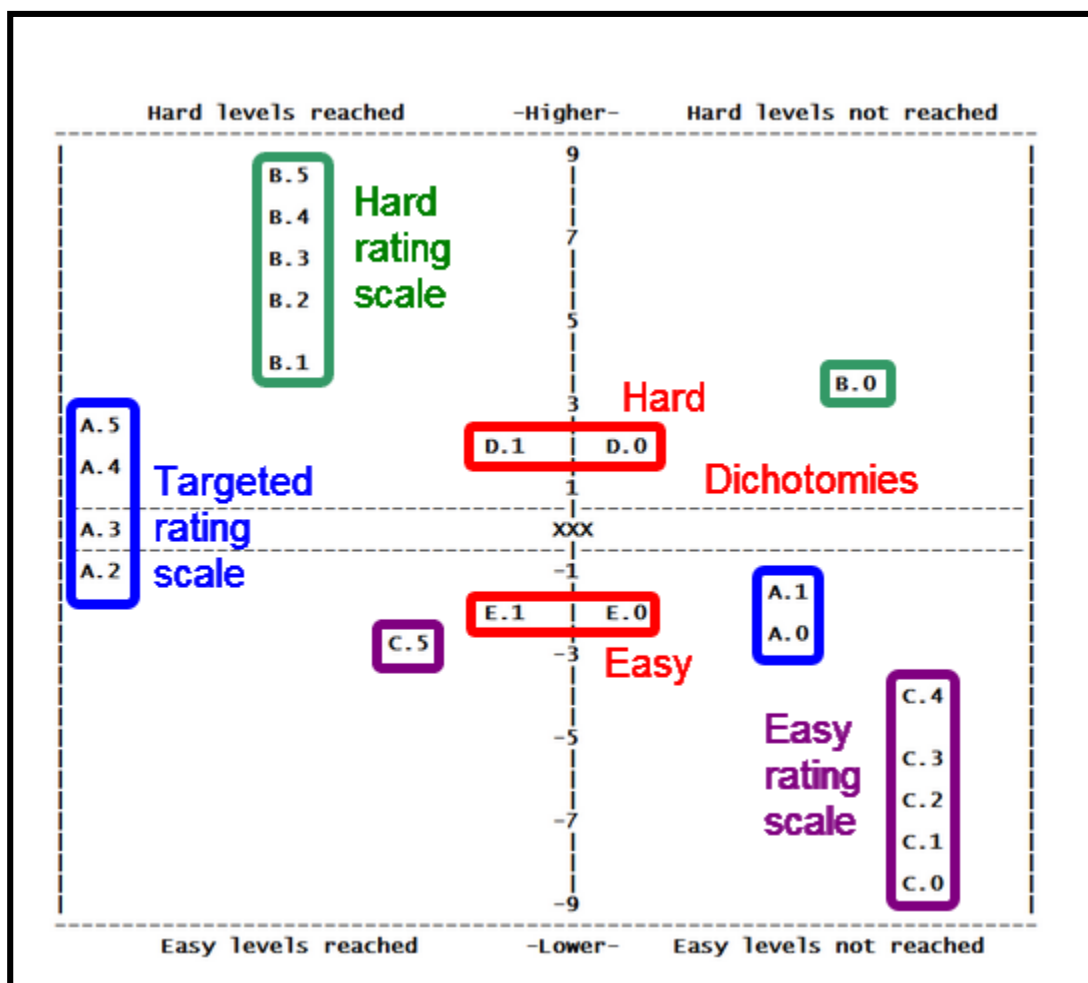
PKMAP MEORP



Fuente: Linacre (2011)

FIGURA 5.4

JERARQUIZACIÓN DE ÍTEMS Y CATEGORÍAS EN PKMAP



Fuente: Linacre (2011)

En la Figura 5.4 se recogen dos grupos de ítems con todas sus categorías posibles, indicando los distintos niveles de habilidad que el sujeto debe desplegar para alcanzar el nivel de dificultad que lucen los ítems, para esa categoría. Un primer grupo de ítems politómicos: A, B y C, con seis categorías (0, 1, 2, 3, 4, 5) y otro de dicotómicos: D y E, con las categorías (0,1).

Al presentar en el PKMAP (Figura 5.4) la dificultad de los ítems politómicos según las categorías alcanzadas es necesario considerar qué nivel de habilidad debería desplegar el sujeto para alcanzar dichas categorías. En el caso de que lo logre, o no, sería conveniente investigar las causas determinantes.

El ítem A presenta su categoría 3 al nivel de la habilidad del sujeto. Las categorías 4 y 5 exigen el desarrollo de un nivel de habilidad del sujeto superior, dado que las dificultades de dichas categorías entrarían en el nivel difícil. Si las alcanza el sujeto, ¿cuál es la razón? La categoría 2 se encuentra con un nivel de dificultad inferior a la habilidad disponible por el sujeto. Si el sujeto la alcanza, es lo normal. Las categorías 0 y 1 del ítem A se consideran fáciles de alcanzar, dado que el sujeto tiene un nivel de habilidad superior. Aunque en este caso no lo ha hecho, ¿por qué?

El ítem B tiene todas sus categorías dentro del nivel de difícil. Exige un nivel de habilidad del sujeto superior al que dispone. No sorprende que no alcance la categoría 0, pero sí que alcance el resto de categorías, ¿cuál puede ser la razón?

El ítem C tiene todas sus categorías con un nivel de dificultad inferior a la habilidad del sujeto, lo normal es que alcance cualquiera de ellas. Ese es el caso de la categoría 5 pero, ¿por qué no ha alcanzado las restantes?

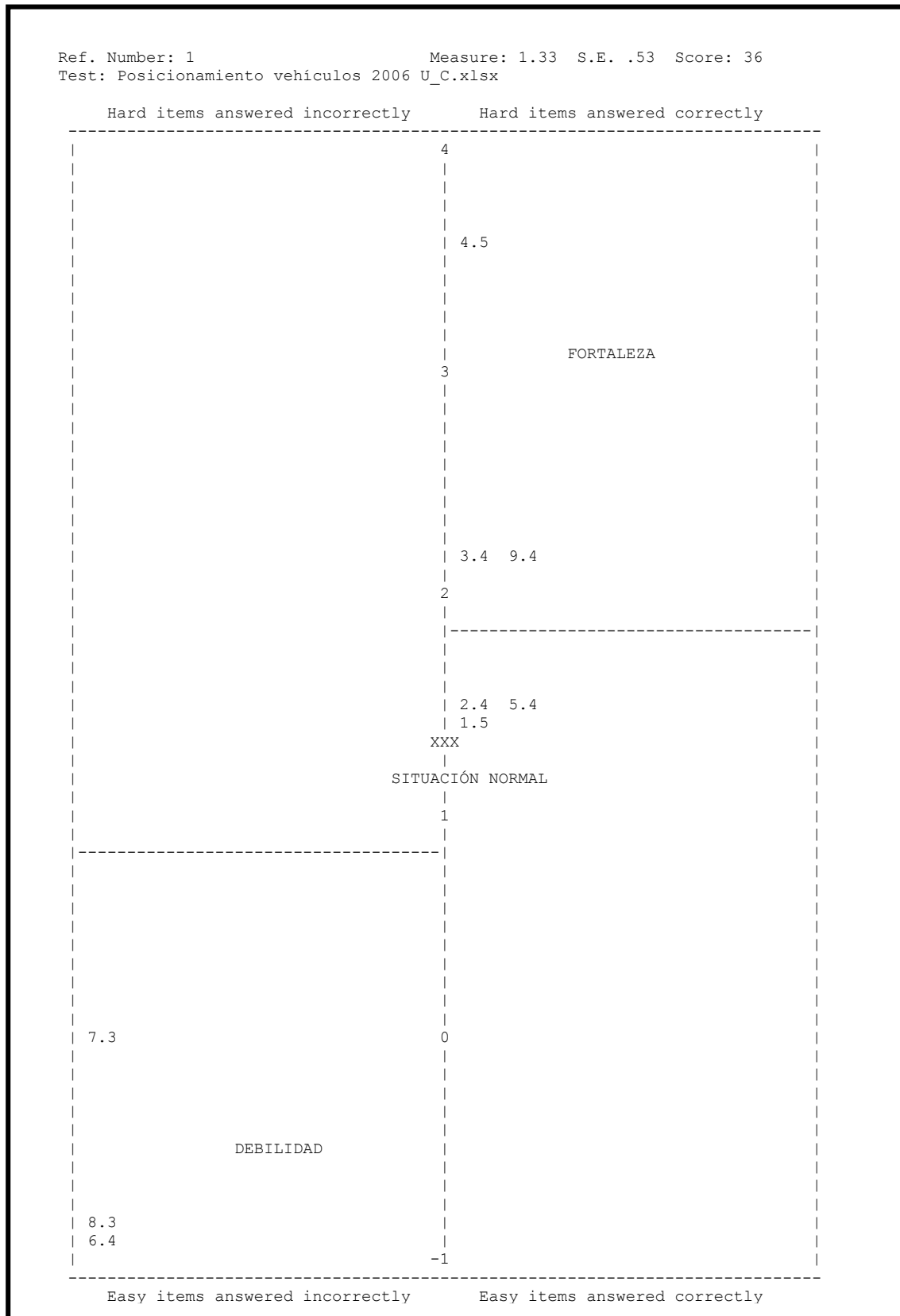
Si se plantea, también en la misma Figura 5.4, la dificultad de los ítems dicotómicos según las categorías alcanzadas, lo que refleja el nivel de habilidad que debería desplegar el sujeto para alcanzar dichas categorías. En el caso de que lo logre, o no, sería conveniente investigar las causas determinantes. Los ítems dicotómicos D y E tienen distinto nivel de dificultad.

El ítem D está por encima de la habilidad del sujeto, lo más probable es que no lo alcance (no lo acierte= 0), pero y ¿si lo alcanza (lo acierta =1)?, ¿cuál ha sido la causa?

El ítem E tiene un nivel de dificultad por debajo del nivel de habilidad del sujeto, para él es fácil. Es probable que pueda acertarlo (=1), sería lo normal. Pero cuál es la causa de que no lo acierte (=0)

La aplicación de los PKMAPS a un grupo de coches urbanos analizados en el escalograma de Guttman nos lleva a la obtención de las siguientes fichas de diagnóstico (Gráficos 5.1, 2, 3 y 4):

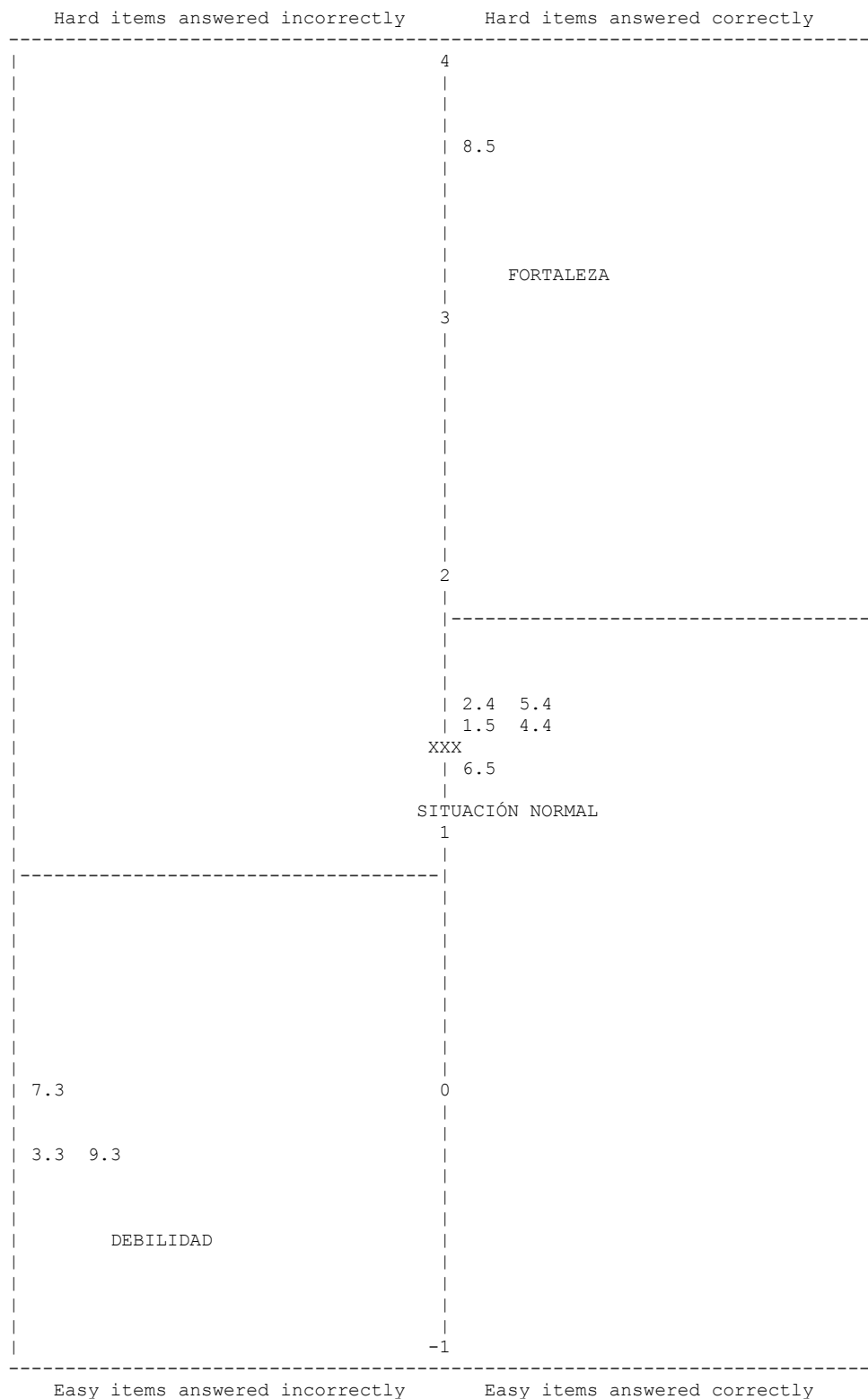
GRÁFICO 5.1
PKMAP DE U - FIAT GRANDE PUNTO



Fuente: Elaboración propia

GRÁFICO 5.2 PKMAP DE U-FORD FIESTA

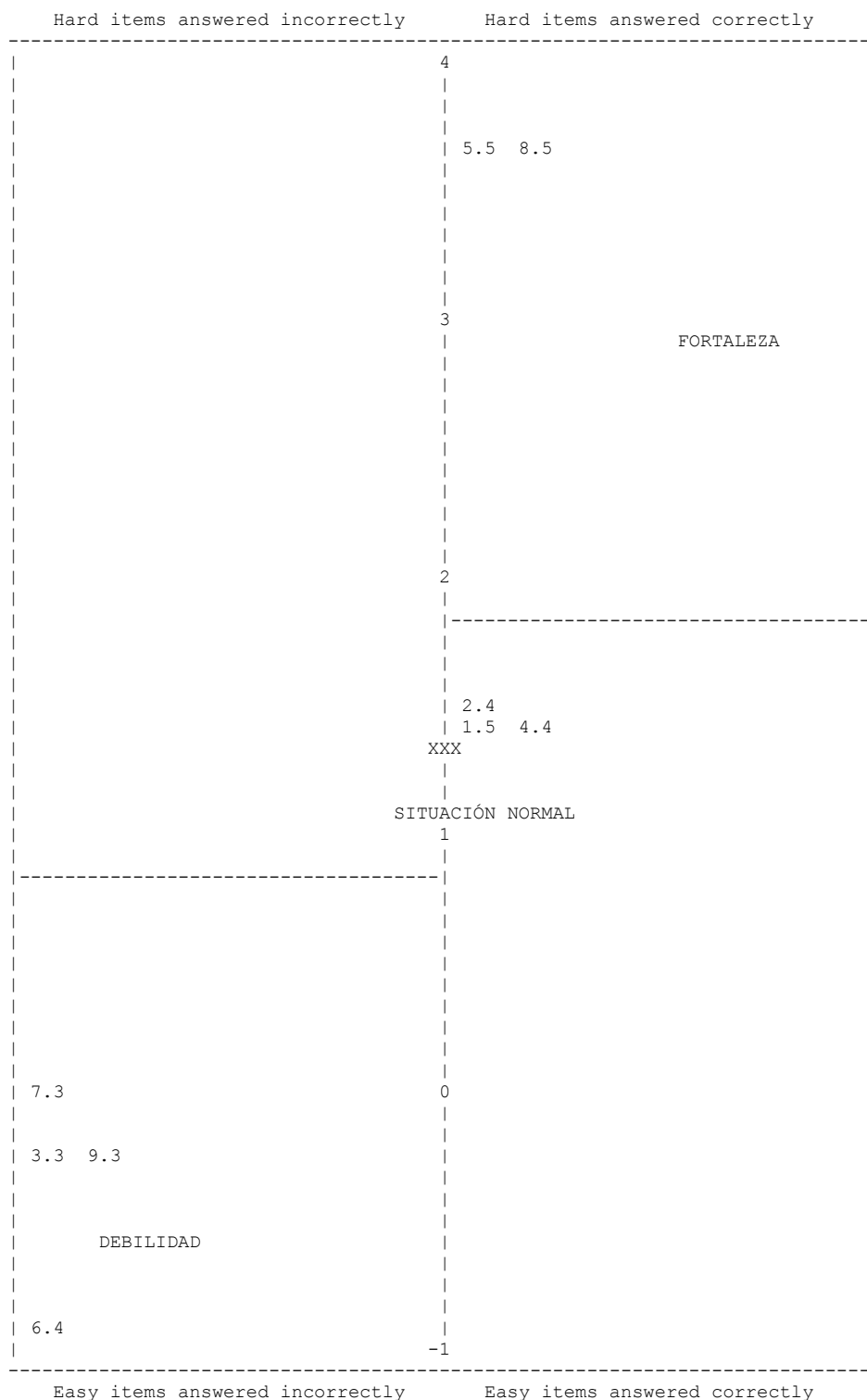
Ref. Number: 2 Measure: 1.33 S.E. .53 Score: 36
Test: Posicionamiento vehículos 2006 U_C.xlsx



Fuente: Elaboración propia

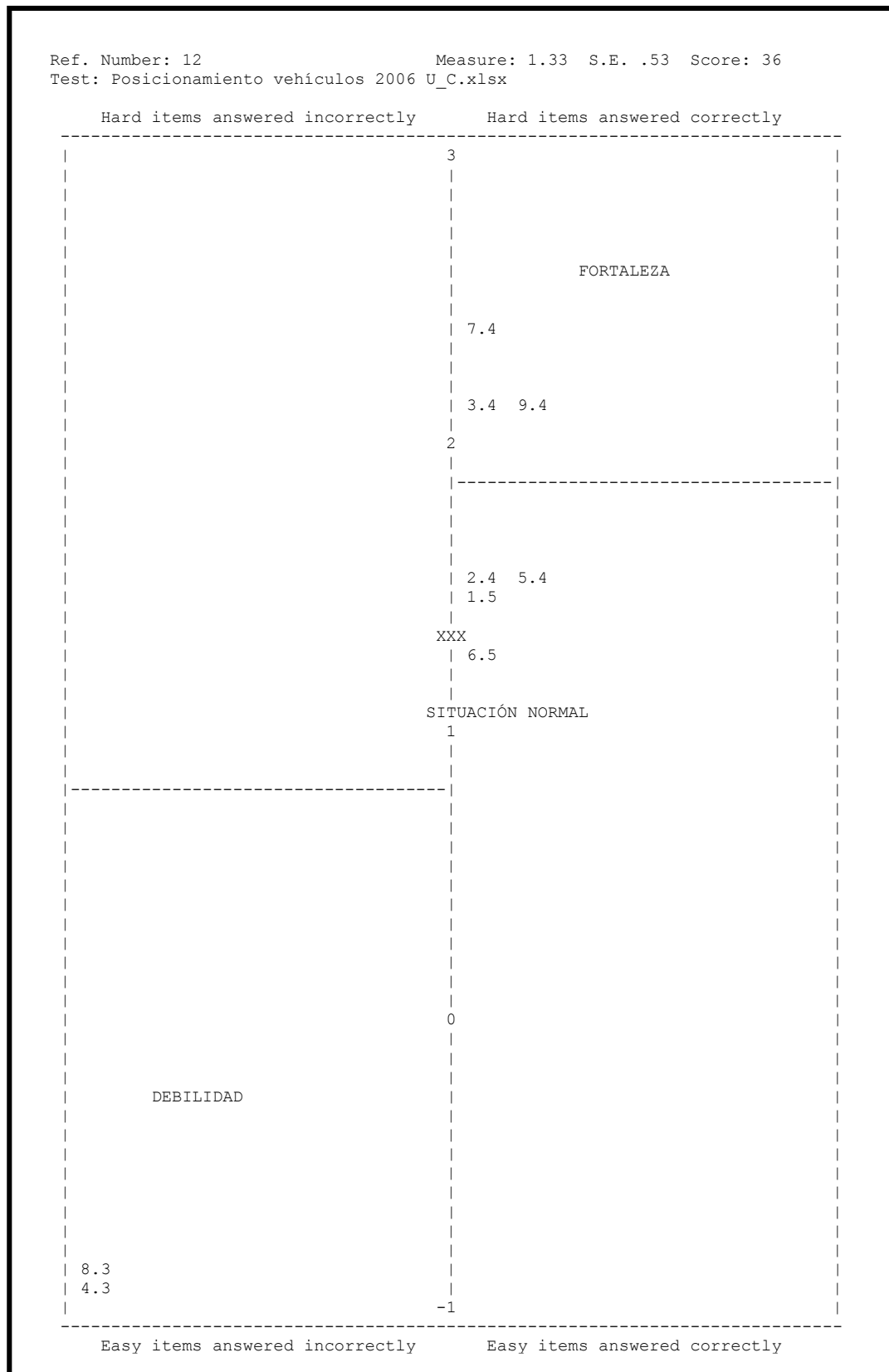
GRÁFICO 5.3 PKMAP DE U-MITSUBISHI COLT

Ref. Number: 7 Measure: 1.33 S.E. .53 Score: 36
Test: Posicionamiento vehículos 2006 U_C.xlsx



Fuente: Elaboración propia

GRÁFICO 5.4 PKMAP DE U-TOYOTA YARIS 2006



Fuente: Elaboración propia

5.3.- BENCHMARKING DE DIAGNÓSTICO

5.3.1.- CONSTRUCCIÓN DE LA MATRIZ INFORMATIVA:

Se parte de la información disponible en los escalogramas de Guttman (Figura nº 10), en donde se integran la información general del posicionamiento de los sujetos, en las filas y la información de los ítems, en columnas.

TABLA 5.3

ESCALOGRAMA DE GUTTMAN PARA UN GRUPO ESTRATÉGICO DE VEHÍCULOS

614258397	614258397
1 +455443443 U Fiat Grande Punto	1 +D5E44C44C U Fiat Grande Punto
2 +554445333 U Ford Fiesta	2 +55444ECCC U Ford Fiesta
7 +454455333 U Mitsubishi Colt	7 +D544EECCC U Mitsubishi Colt
12 +553443444 U Toyota Yaris 2006	12 +55C44C444 U Toyota Yaris 2006
614258397	614258397

Fuente: Elaboración propia

El análisis de la información de las desviaciones percibidas en el Escalograma de Guttman se destaca, por ejemplo, que en el coche nº 1 hay un valor bajo (D=4) en el ítem 6. Un valor alto (E=5) en el ítem 4. Dos valores bajos (C=3) en los ítems 8 y 7.

A la vez en el PKMAP de este coche aparecen como fortalezas un valor 5 en el ítem 4 y dos valores 4, en los ítems 3 y 9.

Como situación normal se situarían los ítems 2, 5 con un valor 4 y 1 con un valor 5.

Como debilidad se situarían los ítems 7, 8 con un valor 3 y 6 con un valor 4

Establecida esta información en el conjunto de coches analizados se puede establecer el cuadro del Benchmarking de Diagnóstico (Figura nº 11):

TABLA 5.4

**MATRIZ DEL BENCHMARKING DE DIAGNÓSTICO GRUPO
ESTRATÉGICO DE VEHÍCULOS**

	MOTOR	CONSUMO	ESTABILIDAD	COMODIDAD	HABITABILIDAD	PRESTACIONES	EQUIPAMIENTO	SEGURIDAD	PRECIO	
1	4	5	5	4	4	3	4	4	3	Fiat Grande Punto
2	5	5	4	4	4	5	3	3	3	Ford Fiesta
7	4	5	4	4	5	5	3	3	3	Mitsubishi Colt
12	5	5	3	4	4	3	4	4	4	Toyota Yaris 2006
	6	1	4	2	5	8	3	9	7	

Fuente: Elaboración propia

**5.3.2.- INTERPRETACIÓN DE LAS FORTALEZAS Y DEBILIDADES.
PROPUESTAS DE PUNTOS DE REFERENCIA ESTRATÉGICA
DETERMINANTES DE OPORTUNIDADES Y AMENNAZAS**

En este contexto nos lleva a interpretar que los coches 1 y 7 se encuentran con una fuerte situación de debilidad en el ítem 1 (motor). Estando en una situación normal, pero superior a sus competidores los coches 2 y 12. Estas situaciones de debilidad debe generar en estos coches el establecimiento de una referencia de mejora de sus motores, amenazando a los que están en mejor situación, que deben considerar las acciones que se derivan de la fijación de estos objetivos como una amenaza.

Con respecto al ítem 1 (consumo) todos los coches están en una situación normal, al nivel exigible en el mercado.

En el ítem 4 (estabilidad). El coche 1 se encuentra en una situación de fortaleza, con niveles superiores tanto dentro del contexto de su perfil como con respecto a sus competidores. En situación normal se encuentran los coches 2 y 7 y en situación de debilidad el coche 12. Estos pueden establecer como referencias estratégicas a lograr los niveles de sus competidores, considerando la oportunidad de mejora en estos ítems, mientras que los que se encuentran en niveles superiores de normalidad o de fortaleza pueden percibir la existencia de una amenaza en el desarrollo de las acciones derivadas del intento de mejora de los peor posicionados.

En el ítem 2 (comodidad) todo los coches se encuentran en una situación de normalidad relativa.

En el ítem 5 (habitabilidad) se destaca tanto dentro de su perfil como en relación a sus competidores el coche 7. Esa fortaleza en el mercado puede generar la acción competitiva de los demás coches pudiendo ser percibida como una amenaza por este coche, mientras que para los demás sería una oportunidad mejorar esta prestación en su posicionamiento en el mercado.

En el ítem 8 (prestaciones) se destacan dos situaciones extremas. Por una parte los coches 2 y 7 con puntuaciones máximas, que resaltan fortalezas como valores bajos en los coches 1 y 12. De esta forma se establecen los perfiles competitivos y las oportunidades de estos coches y las amenazas respecto a los primeros indicados.

En los ítems 3 (equipamiento) y 9 (seguridad) la situación es normal en los coches 1 y 12, pero de debilidad en los coches 2 y 7. Estos pueden amenazar con sus referencias estratégicas a los que se encuentran en situación normal.

En el ítem 7 (precio), la situación es de debilidad en los coches 1, 2 y 7, mientras que en el 12 es de normalidad. Los anteriores pueden establecer como referencia estratégica la de este último y amenazar su nivel con sus acciones estratégicas.

En resumen, por coche, se tendrá que el coche 1 luce una debilidad en el ítem más frecuente, el 6 (motor) lo que le posiciona de forma competitiva muy desfavorablemente en el mercado. Frente a esa situación puede presentar su fortaleza en el ítem 4 (estabilidad) que es de orden inferior. En el resto de ítem posee una situación superior a lo normal (3 y 9) o normal (1, 2, 5). Mientras que en los ítems 7 y 8 tiene también una situación de debilidad.

El coche 2, tiene una situación de fortaleza en un ítem de frecuencia media 8 (prestaciones), situación normal en los ítems 6, 1, 4, 2, 5. Aunque se podría considerar que la situación normal en el mercado del ítem 6 es, en término de este grupo, una situación competitivamente superior. Presenta en cambio debilidad en el resto de los ítems.

El coche 7, presenta fortaleza en dos ítems de frecuencia media 5 y 8 (habitabilidad y prestaciones). Situación normal en ítems de frecuencia superior a la indicada: 1,4, 2. En

cambio tiene una fuerte debilidad en el ítem 7. Un ítem importante competitivamente en este segmento de mercado. Presenta debilidades en los ítems 3, 9 y 7.

El coche 12, se sitúa con una situación superior a lo normal en los ítems 3, 9 y 7. En situación normal en el segmento en los ítems 6, 1, 2 y 5. Aunque en estos casos se podría considerar la situación relativa del ítem 1 como fortaleza dada las situaciones de debilidades de los coches 1 y 7. Las debilidades se presentan en los ítems 4 y 8 (estabilidad y prestaciones).

BIBLIOGRAFÍA

- Adams, RJ y ST Khoo (1998): “Quest – version 2:1. The interactive test analysis system” **Rasch Measurement Transactions** 11 (4), p.598. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt114e.htm>
- Alvarez (2005): “Several Noncategorical Measures Define Air Pollution Construct”, en Bezruzko, N. (2005): **Rasch Measurement in Healt Sciences**. Maple Grove, Minnesota:Jam Press,pp. 277-292
- 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. (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. (2005): “Administrrating, 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. e I, Marais (2005): “Instrument Design with Rasch IRT and Data Analysis I” **Unit Materials EDU 435/635**. School of Education. Murdoch University.
- Andrich, D. ; B.E. Sheridan y G. Luo (2004) : **RUMM2020 : Rasch Unidimensional Measurement Models (version 4.1) [Computer Software]**. Perth, Western Australia : RUMM Laboratory
- 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.] (1968): **Statistical Theories of Mental Test Scores**. Reading, M.A.: Addison-Wesley
- 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
- Bond, T. G. y C. M. Fox (2001): **Applying the Rasch Model Fundamental Measurement in the Human Sciences**. New Jersey: Lawrence Erlbaum Ass. Inc.
- Bond, T. G. y C. M. Fox (2007): **Applying the Rasch Model Fundamental Measurement in the Human Sciences**. New Jersey: Lawrence Erlbaum Ass. Inc. Second Edition.
- Boudon, R. (1999): “Attitudes”, en **Encyclopaedia Universalis**, CD-ROM version 5, (citado por Leplège, 2003)
- Brannick, M. (2005): **Item Response Theory**, Online Tutorial disponible el 4 de febrero de 2015 en <http://luna.cas.usf.edu/~mbrannick/files/pmet/irt.htm>

- Campbell, (1940) **Final Report of the Committee appointed to consider and report upon the possibility of Quantitative Estimates of Sensory Events**. British Association for the Advancement of Science. (Citado por Reese, 1943).
- 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, nº 3, pp. 521/551
- Chien, TW; WC Wang; HY Wang y HJ Lin (2009): “Online assessment of patient’s views on hospital performance using Rasch Model’s KIDMAP diagram” **BMC Health Services Research** 9:135. (Disponible el 4 de febrero de 2015 en Open Access. Online <http://www.biomedcentral.com/1472-6963/9/135>).
- Chien, TW; WC Wang; NS Chen; HJ Lin (2006): “Improving hospital indicator management with the web-KIDMAP module: This is an example”. **Journal of Taiwan Association for Medical Informatics**, 15 (4): 15-26
- Chien, TW; WC Wang; SB Lin; CY Lin; HR Guo y SB Su (2009): “KIDMAP, a web based system for gathering patients’ feedback on their doctors” **BMC Health Services Research** 9:38. (Disponible el 4 de febrero de 2015 en Open Access. Online <http://www.biomedcentral.com/1471-2288/9/38>).
- 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 (1984): “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, 1997a).
- Cortada, N. (2004): “Teoría de Respuesta al Ítem: Supuestos básicos” **Evaluar** nº 4, pp. 95/110.

- 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. (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.; 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).
- 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).
- Doig, B. (1990): "Rasch down under – Kidmap". **Rasch Measurement Transactions** 4(1): 96-100
- 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. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/memo66.htm>
- Fisher, W. P. jr (2005): “Meaningfulness, Measurement and Item Response Theory (IRT),” **Rasch Measurement Transactions** 19 (2) Autumn, pp.1018/1020 Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt192d.htm>
- Fisher, W.P. jr. (1997): “Blind Guides to Measurement,” **Rasch Measurement Transactions** 11(2) Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt112g.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-Pérez, A., V. Yanes-Estévez y J.R. Oreja-Rodríguez (2010): “Puntos de referencia estratégicos, Riesgo y Elecciones Estratégicas” en Llorens-Montes, F.J. y M del M. Fuentes-Fuentes (2010): **Competitividad y Nuevas Tecnologías en la Empresa del Siglo XXI**. ACEDE 2010. XX Congreso Nacional. Granada, Septiembre 2010. Resumen Sesión Comunicación Oral (Estrategias Corporativas y Competitivas I), p. 40. Comunicación Oral en el CD-ROM (Actas del Congreso).
- 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. Citado por 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).
- 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 H. Swaminathan (1985): **Item Response Theory: Principles and Applications**. Boston: Kluver (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.
- 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)
- Kaiser, H. F. (1960): “Review of Measurement and Statistics by Virginia Senders”. **Psychometrika** 25: 411/413
- Kaplan, A. (1964): **The Conduct of Inquiry: Methodology for Behavioral Science**, Scranton, PA: Chandler Publishing Company (citado por Glass y Stanley, 1970).
- 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).
- 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.

- Lee, YJ y OC Yeoh (1998): “KIDMAP construction by hand” **Rasch Measurement Transactions** 12(2):640. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt122t.htm>
- 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 Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt74m.htm>
- Linacre, J.M. (1999): “Spanish – Languaje KIDMAP” **Rasch Measurement Transactions**, 12 (4): 674. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt124k.htm>
- Linacre, J.M. (2002a): “What do Infit and Outfit, Mean-Square and Standardized mean?” **Rasch Measurement Transactions** vol. 16 (2), pp. 878. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt162f.htm>
- Linacre, J.M. (2002b): “Facets, Factors, Elements and Levels” **Rasch Measurement Transactions** vol. 16 (2), pp. 880 Disponible el 4 de febrero de 2015 en. <http://www.rasch.org/rmt/rmt162h.htm>
- Linacre, J. M. (2002c): “Optimizing rating Scale Category Effectiveness”. **Journal of Applied Measurement**, Vol. 3(1), pp. 85-106
- Linacre, J. M. (2005): “Rasch dichotomous model vs. One-parameter Logistic Model,” **Rasch Measurement Transactions** 19:3. Winter, p. 1032. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt193.pdf>
- Linacre, J.M. (2009a): **Winsteps® Rasch Measurement Computer Program User’s Guide**. Beaverton, Oregon: Winsteps. com. (Versión 3.68.0)
- Linacre, J. M. (2009b): “Investigating Dimensionality” en Linacre, J. M. (2009): **Practical Rasch Measurement. Futher Topics**. On line Course Statistics.com. Arlington:Virginia: The Institute for Statistics Education.

- Linacre, J. M. (2013): **Facets computer program for Many-Facet Rasch Measurement, version 3.71.2**. Beaverton, Oregon: Winsteps.com
- Linacre, J. M. (2014a): **Winsteps® Rasch Measurement Computer Program**. Beaverton, Oregon: Winsteps. com. (Versión 3.81.0)
- Linacre, J. M. (2014b): **Winsteps® Rasch Measurement Computer Program User's Guide**. Beaverton, Oregon: Winsteps. com. (Versión 3.81.0)
- Linacre, J.M. y B.D. Wright (1994): "Dichotomous Mean – Square Fit Statistics" en **Rasch Measurement Transactions** 8:2 p. 360, Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt82a.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, F. M. (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.
- Lunz, M. E. y J. M. Linacre (1998): "Measurement Designs Using Multifacet Rasch Modelling", en Marcoulides, G. A. [Ed] (1998): **Modern Methods for Business Research**. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers. Capítulo 3º. pp. 47/77
- 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. (1994): “Rasch KIDMAP – A History”, en **Rasch Measurement Transactions**, 8(2), p. 366. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt82k.htm>
- 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-Muradas, I. y Oreja-Rodríguez, J. R. (2006). “El posicionamiento del automóvil todoterreno en España. Un análisis conjunto multiatributo”, en Febles, J. y J. R. Oreja [Coord.] (2006). **Modelos de Rasch en Administración de Empresas**. Fundación FYDE CajaCanarias, Colección E-BOOKS.nº 1 Santa Cruz de Tenerife, pp. 134-145. Disponible el 4 de febrero de 2015 en <http://www.fyde-cajacanarias.es/inic/download.php?idfichero=616>
- Montero – Muradas, I. y J. R. Oreja-Rodríguez (2014): “Análisis competitivo de la oferta internacional de Ferias y Exhibiciones durante el periodo de 2004_2011” en Oreja-Rodríguez, J. R. y J. Febles-Acosta (2014): **Modelos de Rasch en Administración de Empresas. Actuaciones y respuestas frente a la crisis**. Santa Cruz de Tenerife: FYDE-CajaCanarias -Colección E-BOOKS nº 8, pp. 49 / 71 Disonible el 4 de febrero de 2015 en <http://www.fyde-cajacanarias.es/inic/download.php?idfichero=651>
- Nunnally, J. C. (1967; 1978): **Psychometric Theory**, New York: McGraw - Hill

- Oreja-Rodríguez, J. R. (2005): “Introducción a la medición objetiva en Economía, Administración y Dirección de Empresas: El Modelo de Rasch”. **IUDE. Serie Estudios** nº 2005/47. Santa Cruz de Tenerife: Instituto Universitario de la Empresa de la Universidad de La Laguna. Disponible el 4 de febrero de 2015 en http://iude.webs.ull.es/investigacion/publicaciones/pdf_docs_trabajo/iude-0547.pdf
- Oreja-Rodríguez, J. R. (2010): “Análisis de la dimensionalidad en el modelo de Rasch. Estudio comparativo de dos casos”, en Oreja-Rodríguez, J. R. y J. Febles-Acosta [Coord.] (2010): **Modelos de Rasch en Administración de Empresas: Nuevos Desarrollos**. Santa Cruz de Tenerife: FYDE-CajaCanarias -Colección E-BOOKS nº 4, pp. 23/47. Disponible el 4 de febrero de 2015 en <http://www.fyde-cajacanarias.es/inic/download.php?idfichero=636>
- Oreja-Rodríguez, J. R. ; A. García-Pérez y V. Yanes-Estévez (2011): “Introducción al estudio del funcionamiento diferencial de ítems. Análisis de un caso” en Oreja-Rodríguez, J. R. y J. Febles-Acosta [Coord.] (2011): **Modelos de Rasch en Administración de Empresas: Avances sectoriales**. Santa Cruz de Tenerife: FYDE-CajaCanarias -Colección E-BOOKS nº 5, pp. 1- 20. Disponible el 4 de febrero de 2015 en <http://www.fyde-cajacanarias.es/inic/download.php?idfichero=645>
- Oreja-Rodríguez, J. R. e I. Montero -Muradas (2012): El uso del KIDMAP como soporte del Benchmarking de diagnóstico en el Análisis Competitivo de las Empresas: Estudios de casos de empresas participantes en una cadena de suministros agroalimentaria. en Oreja-Rodríguez, J. R. y J. Febles – Acosta (2012): **Modelos de Rasch en Administración de Empresas. Técnicas de Diagnóstico y Posicionamiento**. Santa Cruz de Tenerife: FYDE-CajaCanarias -Colección E-BOOKS nº 6, pp. 31 – 60. Disponible el 4 de febrero de 2015 en <http://www.fyde-cajacanarias.es/inic/download.php?idfichero=647>

- Oreja-Rodríguez, J. R.; Y. Armas-Cruz y F. J. García-Rodríguez (2012): “Caracterización del posicionamiento competitivo de la empresa hotelera en Tenerife: Una aproximación a partir de la metodología de Rasch” en Oreja-Rodríguez, J. R. y J. Febles-Acosta. [Coord] (2012) **Modelos de Rasch en Administración de Empresas: Técnicas de Diagnóstico y Posicionamiento**. Santa Cruz de Tenerife: FYDE-CajaCanarias -Colección E-BOOKS nº 6, pp. 73/93. Disponible el 4 de febrero de 2015 en <http://www.fyde-cajacanarias.es/inic/download.php?idfichero=647>
- Parra-López, E. y J. R. Oreja-Rodríguez (2014): ““Evaluation of the competitiveness of tourist zones of an island destination: An Application of a Many- Facet Rasch Model (MFRM)” **Journal of Destination Marketing and Management**. Vol 3(2), pp 114/121.
- 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).
- 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
- 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. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/memo1960.pdf>
- Rasch, G. (1972): Objektiviteten i samfundsvidenskaberne et metodeproblem (Objectivity in the social sciences: A methodological problem). *Nationalekonomisk Tidsskrift* 110, 161-196 (citado por Masters y Wright, 1984)

- Rasch, G. (1980): **Probabilistic Models for Some Intelligence and Attainment Tests** (Expanded Ed.) Chicago: University of Chicago Press, en originalmente publicado en 1960, Copenhagen: Denmarks paesdagogiske Institut.
- Reese, T.W (1943): Norman Campbell's Theory of Fundamental Measurement, as described in "The application of the Theory of Physical Measurement to the Measurement of psychological Magnitudes, with Three Experimental Examples", **Psychological Monographs**, Vol. 55:3, 251, pp 6-20. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rn8.htm>
- 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
- Salas- i - Martin, X.; B. Bilbao-Orsorio, A. Di Battista, M. Drzeniek Hanouz, T. Geiger y C. Galván (2014): "The Global Competitiveness Index 2014 -15: Accelerating a Robust Recovery to Create Productive Jobs and Support Inclusive Growth", en Schwab, K. [Editor] (2014): **The Global Competitiveness Report 2014-15**. World Economic Forum. Geneva, pp. 3/52. Disponible el 4 de febrero de 2015 en http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2014-15.pdf
- Samejima, F. 1997. Graded Response Model. In W. J. Van der Linden, & R. K. Hambleton (Eds.), **Handbook of Modern Item Response Theory**. NY: Springer.
- Samejima, F. (1969): "Estimation of Latent Ability Using a Response Pattern of Graded Scores. **Psychometrika Monographs**, 34 (17)
- Sarabia, F. J. (1999): **Metodología para la investigación en Marketing y Dirección de Empresas**. Madrid: Pirámide
- Schwab, K. [Editor] (2014): **The Global Competitiveness Report 2014-15**. World Economic Forum. Geneva. Disponible el 4 de febrero de 2015 en http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2014-15.pdf
- Senders, V.L. (1958): **Measurement and Statistics**. Oxford University Press. New York

- Siegel, S. (1956): **Nonparametric Statistics**. McGraw Hill. New York
- Smith R.M. (1996): “Polytomous Mean-Square Fit Statistics” **Rasch Measurement Transactions** 10:3 p. 516/517. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt103a.htm>
- Smith, R. M. (2004): “Detecting Item Bias with the Rasch Model” en Smith, E.V. y R.M. Smith (2004): **Introduction to Rasch Measurement. Theory, Models and Applications**. Maple Grove, Minnesota: JAM Press, pp. 391/ 418
- 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).
- Stenner, A. J. (1990): “Objectivity: Specific and General”. **Rasch Measurement Transactions**, 4 (3), p. 111 Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt43a.htm>
- Stenner, A. J. (1994): “Specify objectivity – local and general” **Rasch Measurement Transactions**, 8 (3), p. 374. Disponible el 4 de febrero de 2015 en <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
- 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. Disponible el 4 de febrero de 2015 en https://www.brocku.ca/MeadProject/Thurstone/Thurstone_1925b.html

- Thurstone (1927): “A Law of Comparative Judgment” **Psychological Review** 34, pp. 273/286. Disponible el 4 de febrero de 2015 en <http://mpg.ndlab.net/wp-content/uploads/2009/07/thurstone94law.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, 2003)
- 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
- 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. (1977): “Fundamental Measurement for Psychology” en Embretson , S.E. y Hershberger, S.L (1977): **The New Rules of Measurement. What Every Psychologist and Educator Should Know**. Capítulo 4, pp. 65/104. NJ: Lawrence Erlbaum Ass. Inc.
- Wright, B. D. (1992a): “Rasch vs. Birnbaum,” **Rach Measurement Transactions**, vol. 5 (4), p. 178. Disponible el 4 de febrero de 2015 en <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. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt61a.htm>
- Wright, B. D. (1996): “Reliability and separation” **Rasch Measurement Transactions**, Vol. 9(4), p.472. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt94.htm>
- Wright, B. D. (1997a): S.S. Stevens Revisited. **Rasch Measurement Transactions** vol. 11 (1), pp. 552 – 553. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt111n.htm>
- Wright, B. D. (1997b): “Fundamental Measurement for Outcome Evaluation” **IOM Research Memoranda Memo66**. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/memo66.htm>
- Wright, B. D. (2003): “Rack and Stack: Time 1 vs Time 2 or Pre-Test vs Post-Test en **Rasch Measurement Transactions** vol 17(1), pp 905-906. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt171a.htm>
- Wright, B. D. y G. N. Maters (1982): **Rating Scale Analysis**. Chicago: MESA Press
- Wright, B. D. y G. N. Maters (2002): “Number of Person o Item Strata” **Rasch Measurement Transactions**, 16(3), p. 888. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/rmt/rmt163f.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, BD; RJ Mead y LH Ludlow (1980): “KIDMAP: Person-by-item interaction mapping”. **Research Memorandum N° 29 MESA Psychometric Laboratory**. Department of Education, University of Chicago. Chicago: MESA Press. Disponible el 4 de febrero de 2015 en <http://www.rasch.org/memo29.pdf>

Yanes-Estévez, V. y J. R. Oreja-Rodríguez (2008): “Los análisis Rack y Stack del dinamismo del entorno” en Febles, J. (Coord.) (2008): **Los Modelos de Rasch en Administración de Empresas. Aplicaciones avanzadas**. Santa Cruz de Tenerife: FYDE-CajaCanarias e IUDE-ULL, pp. 253/273. Disponible el 4 de febrero de 2015 en <http://www.fyde-cajacanarias.es/inic/download.php?idfichero=429>

Zwick, R.; D.T. Thayer y C. Lewis (1999): “An Empirical Bayes Approach to Mantel – Haenszel DIF Analysis” **Journal of Educational Measurement**, 36 (1): 1/28