

El Análisis de Reactivos con el Modelo Rasch

Manual Técnico A

Serie: Medición y Metodología

Manuel Jorge González Montesinos



UNIVERSIDAD DE SONORA

**INSTITUTO NACIONAL PARA
LA EVALUACION DE LA EDUCACION**

Marzo del 2008

México, D. F.

CONTENIDO

Presentación 3

Parte 1 La Base Teórica

Introducción 6

Dirección en el modelamiento Rasch 8

Descripción inicial del Modelo 9

Bases de estimación del Modelo 13

Método de estimación de parámetros 15

Análisis de ajuste de los datos al Modelo 20

Implementación computacional 21

Criterios de bondad de ajuste 23

Parte 2 Práctica Elemental

Bases de operación de los programas Winsteps y Ministep 29

El archivo de especificaciones 30

El archivo de datos 33

Tratamiento de casos con datos faltantes 41

Ejercicio inicial de programación 47

Corrida inicial en Ministep 49

Tablas con estadísticas por reactivo y respondente 58

Opciones básicas de programación 67

Opciones especiales de programación 60

¿Se ajustan los datos observados al Modelo Rasch? 70

Gráficas en Winsteps y Ministep 72

La grafica CCR empírica 76

Ejercicio de integración 78

Interpretación práctica 80

Casos especiales: Análisis Rasch con ítems politómicos 92

Caso 1 Estructura RSM 93

Caso 2 Estructura PCM 94

Reactivos en sentido inverso 106

Post Scriptum 108

Bibliografía 110

“Porque con la misma medida que midan serán medidos”.
San Lucas 6, 38

Presentación

El Modelo de Rasch es tal vez el principal recurso técnico propuesto desde mediados del siglo XX para analizar las propiedades psicométricas de reactivos que forman parte de exámenes de uso académico. En forma más general el modelo y sus procedimientos son aplicables a reactivos que forman parte de instrumentos de medición utilizados en disciplinas como la psicología, la sociología, ciencias de la salud y en una amplia gama de disciplinas con las que se estudia el comportamiento humano.

En la aplicación de este recurso y sus procedimientos asociados, el centro de la atención es el encuentro entre los sustentantes y los reactivos en un examen o instrumento. Si bien el sustento y los procedimientos bajo el modelo puedan parecer al principio complicados, el fundamento y proceso lógico del Modelo Rasch es sumamente simple, elegante y parsimonioso.

El modelo postula *que un reactivo eficiente sólo debe ser contestado correctamente por aquellos sustentantes que posean la habilidad requerida por las demandas cognitivas del reactivo. A la inversa, un reactivo eficiente no debe ser contestado correctamente por aquellos sustentantes que no posean la habilidad requerida para responderlo.*

Bajo estas dos premisas el procedimiento de análisis procede primero a determinar las dificultades de los reactivos y las habilidades de los sustentantes

con una unidad de medida especial que se aplica a una muestra de respuestas emitidas por un grupo de n sustentantes a un conjunto de k reactivos.

Después de esta calibración inicial de dificultades de los reactivos y habilidades de sustentantes el análisis Rasch procede a establecer una expectativa probabilística acorde a la dificultad de cada reactivo y a la habilidad de cada respondente.

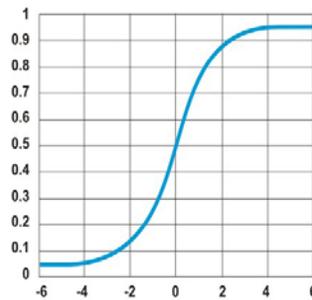
Una vez establecidas las expectativas probabilísticas para k reactivos y n sustentantes, el procedimiento avanza comparando cada patrón de respuesta reactivo por reactivo y sustentante por sustentante. En esta serie de comparaciones anidadas el procedimiento identifica con toda precisión aquellos reactivos y sustentantes que se comportaron de manera distinta a las expectativas probabilísticas del Modelo Rasch. Es decir el procedimiento identifica aquellos reactivos y sustentantes que presentan respuestas anómalas. El proceso identifica aquellos reactivos que no fueron contestados correctamente por sustentantes que poseen la habilidad para hacerlo. Por extensión se identifican también aquellos reactivos que fueron contestados correctamente por sustentantes que no poseen la habilidad requerida para hacerlo. De igual manera el procedimiento identifica aquellos sustentantes que no contestaron correctamente a reactivos dentro de su nivel de habilidad; o que contestaron correctamente a reactivos fuera de su nivel de habilidad.

En resumen, el procedimiento identifica como anomalías todos los casos que se apartan de la expectativa del Modelo Rasch.

Los conjuntos de anomalías se suman y al rebasarse un límite definido de

tolerancia, los casos se remiten a posibles características resultantes de la construcción de los reactivos o de las idiosincrasias particulares de algunos sustentantes. Una vez separados esos casos los reactivos que se comportan dentro de la expectativa del Modelo se conservan como reactivos probados empíricamente y productivos en la información que aportan sobre los conocimientos y habilidades de los sustentantes que los respondieron correctamente y de los sustentantes que no los respondieron correctamente.

Al ingresar en esta rama de la psicometría contemporánea el lector debe tener siempre presente la figura:



Esta es la llamada ojiva logística o curva logística. La curva de forma de S o sigmoidea describe la trayectoria que bajo el Modelo Rasch debe seguir un reactivo eficiente. La trayectoria que un reactivo debe seguir es que la probabilidad de respuesta correcta (eje Y, de 0 a 1) aumente conforme aumenta la habilidad del respondente (eje X, de -6 a +6).

En las siguientes secciones se describen a detalle los supuestos teóricos, los procedimientos estadísticos y los criterios de ajuste de reactivos al modelo logístico que han emergido de la aproximación de Georg Rasch (1901-1980) para el Análisis Individualizado de Reactivos.

Este manual técnico tiene como objetivo principal el servir como material de apoyo teórico y práctico para el entrenamiento de académicos e investigadores interesados en conocer y aplicar los fundamentos de la medición Rasch en las ciencias sociales. El material ha sido seleccionado y preparado para aplicarse en el contexto de un curso presencial con taller de cómputo. Sin embargo –con un esfuerzo adicional por parte de lector- esta exposición de principios y procedimientos de la medición Rasch puede aplicarse a manera de un curso autónomo acompañado de prácticas de cómputo realizadas de manera individual.

En la primera parte se condensan las bases teóricas mínimas para aplicar el sistema de análisis Rasch a reactivos que forman parte de escalas o subescalas en exámenes de uso académico o bien en otra clase de instrumentos conformados por reactivos de respuesta binaria (correcto – incorrecto, si – no, de acuerdo- no de acuerdo). En la segunda parte se detalla la implementación computacional de la medición Rasch mediante ejercicios de programación e interpretación con un programa de distribución libre. En las secciones finales se describe e implementa el análisis Rasch en reactivos de respuesta graduada (tipo Likert) que requieren un tratamiento distinto con extensiones del modelo Rasch desarrolladas especialmente para estos casos.

El usuario de este material debe advertir que por tratarse de una síntesis especialmente compacta, es necesario que este material se combine con la lectura de una introducción más completa y detallada a los razonamientos en

que se sustenta la medición Rasch. (Se sugiere: Tristán, L. A. 2002).

Como ya se expresó el material que se presenta se orienta a combinar la teoría con la práctica, en la segunda parte se detallan los procedimientos elementales pero completos para realizar el análisis Rasch en casos generales mediante el uso de un programa de cómputo especialmente diseñado (Ministep/Winsteps, Linacre, J.M.1991, 2006). El proceso formativo se completa con una descripción y ejemplificación de casos especiales donde los reactivos tienen estructura de escala de rating como los reactivos de tipo Likert.

En concreto la intención es ofrecer al lector un recurso técnico de nivel intermedio para iniciarse y avanzar en la especialidad del análisis y construcción de medidas utilizando el potencial de la metodología Rasch. Es importante aclarar que los contenidos de ambas secciones presuponen una familiarización de nivel intermedio con los métodos estadísticos comunes.

Parte I

La Base Teórica

Introducción

El análisis Rasch está basado en un modelo matemático propuesto por Georg Rasch (1953, 1960), en el que se describe la relación entre la probabilidad de una respuesta correcta a un reactivo y la diferencia entre la habilidad del respondente y la dificultad del reactivo. El modelo es dicotómico ya que únicamente toma en cuenta las diferencias en la dificultad del reactivo para producir una medida de la habilidad del respondente. El procedimiento

compara cada respondente y cada reactivo en términos de su ajuste a un modelo unidimensional en el que un solo constructo o variable latente subyace y se manifiesta en la respuesta correcta al reactivo.

El Modelo Rasch tiene dos supuestos fundamentales:

Unidimensionalidad e Independencia local. Si bien estos supuestos se elaboran con detalle más adelante, por ahora es necesario concebir a un conjunto de reactivos como elementos que funcionan en una sola dirección para definir un constructo discreto y unidimensional. La escala de reactivos al funcionar en conjunto, forma una estructura jerárquica con continuidad de menor a mayor grado de dificultad. Sin embargo la unidimensionalidad no significa que el desempeño del respondente se deba a un solo proceso cognitivo. Por el contrario, cuando los reactivos funcionan en conjunto, las medidas se comportan como componentes que definen el constructo de interés (D' Agostino, 2002). Si bien los aspectos cognitivos individuales pueden funcionar independientemente, cuando se aplican a una escala de reactivos que funcionan en conjunto, los componentes del proceso aportan evidencia de unidimensionalidad. Una de las principales ventajas del proceso de modelamiento Rasch es su capacidad para detectar la contribución de áreas de contenido en los conjuntos de reactivos para definir el peso relativo de los subconjuntos en la composición de un examen determinado ya sea de logros o de habilidades.

Dirección del Modelamiento Rasch.

El análisis con el modelo Rasch es actualmente el método de elección

para determinar las propiedades psicométricas de grupos de reactivos en procedimientos de medición a gran escala (Wright, B.D. & Stone, M. H. 1979). Antes de entrar en mayor detalle, es de crucial importancia establecer desde el principio la direccionalidad del procedimiento de modelación.

En los análisis estadísticos convencionales se procede a partir de los datos observados y en seguida se propone el modelo de relaciones que mejor explique las regularidades detectadas en los datos. Esta direccionalidad en el análisis pone al conjunto de datos observados en primer plano.

En el análisis Rasch el procedimiento toma una dirección inversa. El modelo matemático es derivado de la función logística que relaciona una probabilidad creciente de respuesta a los reactivos a medida que aumenta la habilidad de los respondientes. En este caso el modelo pasa a primer plano y los razonamientos de ajuste proceden desde los datos observados hacia la expectativa del modelo. Aquellos datos que no se ajusten a la expectativa del modelo se consideran como “anomalías” en los patrones de respuesta a los reactivos. Los criterios de bondad de ajuste de los datos al modelo serán descritos en secciones posteriores. Lo importante por ahora es dejar en claro que la determinación de ajuste de los datos al modelo desde cada reactivo y desde cada respondiente permitirá tomar decisiones sobre las propiedades reales de los reactivos.

Descripción Inicial del Modelo

El centro de atención del procedimiento es el encuentro del respondiente con cada reactivo en un examen. El modelo es monoparamétrico ya que

describe la relación entre la probabilidad de respuesta correcta a un reactivo $P(x = 1 | \beta_s)$ y la diferencia entre la habilidad del respondente β y la dificultad del reactivo δ . Esta probabilidad se expresa mediante la fórmula:

$$P_{t(x=1|\beta)} = \frac{e^{(\beta-\delta)}}{1 + e^{(\beta-\delta)}}$$

$$t = 1, 2, \dots, k$$

El análisis Rasch es a su vez uno de los fundamentos de la Teoría de Respuesta al Ítem (TRI) que abarca una serie de técnicas de medición que se originaron a partir del Modelo Rasch (Rasch, G. 1960). Los principios en que se sustenta la medición Rasch son:

- El desempeño de los sustentantes en un examen puede predecirse mediante la especificación de un conjunto de factores llamados rasgos latentes o simplemente habilidades.
- La relación entre el desempeño de los sustentantes ante un reactivo y el conjunto de habilidades que subyace al desempeño se describe mediante una función monótona incremental denominada la curva característica del reactivo.
- La curva característica del reactivo (CCR) traza la relación entre el nivel de habilidades de los sustentantes y la probabilidad de respuesta correcta a un reactivo (Figura 1). Esto implica que aquellos sustentantes

de mayor nivel de habilidad tendrán una mayor probabilidad de emitir la respuesta correcta a un reactivo determinado.

- Los únicos parámetros estimados en el modelo Rasch son la dificultad del reactivo δ y la habilidad del sustentante β . La teoría matemática que sustenta al modelo postula que estos dos parámetros son suficientes para explicar y predecir el desempeño de los sustentantes ya que los parámetros contienen toda la información necesaria. Estos parámetros se expresan en una unidad de medida especial denominada lógito que se obtiene multiplicando los momios de respuesta a cada reactivo por el logaritmo natural **Ln**. Con esta transformación de los momios se logra expresar el parámetro de dificultad de reactivo en una escala uniforme.
- Las unidades de medida en lógitos se aplican para caracterizar tanto los grados de dificultad de los reactivos como los grados de habilidad de los sustentantes a partir de las siguientes ecuaciones (Tristán, L. A. 2002, pp.11-15):

$M_r = \frac{q}{p}$ Momio o cociente de errores sobre aciertos observados para el reactivo

$M_s = \frac{P}{q}$ Momio o cociente de aciertos sobre errores observado para el sustentante

$\delta_r = \text{Ln } M_r$ Medida de dificultad en lógitos para el reactivo

$\beta_s = \text{Ln } M_s$ Medida de habilidad en lógitos para el sustentante

$P(\delta_r) = 1 / (1 + e^{\beta})$ Probabilidad de respuesta correcta a reactivo dado

su grado de dificultad

$P(\beta_s) = e^{\beta} / (1 + e^{\beta})$ Probabilidad de respuesta correcta del sustentante dado su grado de habilidad que es la expresión propia del modelo Rasch.

Ahora bien en concreto, las transformaciones logarítmicas de los momios del reactivo $\ln M_r$ y de los momios del respondente $\ln M_s$ convierten los cocientes de errores sobre aciertos (en el caso del reactivo) y los cocientes de aciertos sobre errores (en el caso del respondente) en una medida lineal de dificultad y de habilidad expresada en una métrica común a los dos casos. Así las dificultades de los reactivos δ_r y las habilidades de los respondentes β_s se expresan en la medida denominada *Lógito* que se deriva del término “Logit” que a su vez proviene de “Log odd unit” que corresponde al logaritmo de una unidad de momios (o proporción de errores sobre aciertos para un reactivo dado o de aciertos sobre errores para un respondente dado).

En la práctica la obtención de las medidas de dificultad de reactivos y de habilidad de respondentes se llama calibración de reactivos y de respondentes. A partir de estas calibraciones se pueden expresar las probabilidades de respuesta correcta a un reactivo dada su dificultad calibrada, o de que un respondente emita la respuesta correcta a un reactivo dada su habilidad calibrada. Para ello es necesario convertir la calibración de habilidad β_s re-expresándola como el exponente de la constante e que es igual a 2.711828. Esta transformación permite a su vez representar las calibraciones en una métrica estándar de la que se obtienen las expresiones de probabilidad de

acuerdo a cada reactivos individual dada su dificultad y la habilidad de cada respondente.

Una vez obtenidas las probabilidades asociadas a cada reactivo y a cada respondente es posible expresar de forma grafica la relación entre habilidad de respondentes calibrada en lógitos y probabilidad de respuesta correcta de lo que resulta la llamada curva característica del reactivo o CCR.

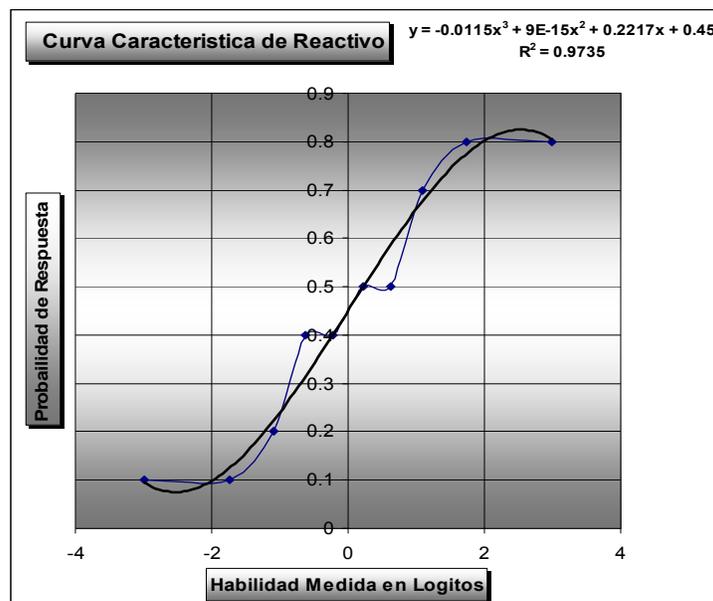


Figura 1:
Ejemplo de CCR
 (Tristán, L. A.2002, pp. 136-141)

Esta curva representa a su vez la expectativa del Modelo Rasch para cada reactivo. Como ya se expresó arriba la CCR resulta de trazar las medidas de habilidad en lógitos y las probabilidades de respuesta correcta a cada reactivo después de aplicar las ecuaciones anteriores a un conjunto de patrones

de respuesta observados en n sustentantes.

La aplicación del modelo requiere que se adopten los siguientes supuestos:

- Unidimensionalidad: este supuesto especifica que una sola habilidad se mide mediante un conjunto dado de reactivos. Este supuesto se satisface cuando se detecta la presencia de un factor dominante que influye sobre el desempeño en el examen.
- Independencia Local: este supuesto especifica que cuando las habilidades que influyen sobre el desempeño se mantienen constantes las respuestas de los sustentantes a cualquier par de reactivos dado, son estadísticamente independientes.

Bases de Estimación en el Modelo

Como ya se expresó en la sección anterior el modelo se describe mediante la ecuación:

$$P_i(x=1|\beta_s) = e^{(\beta_s - \delta_r)} / (1 + e^{(\beta_s - \delta_r)}) \quad i = 1, 2, \dots, k$$

Donde:

- $P_i(x=1|\beta)$ es la probabilidad de que un sustentante seleccionado aleatoriamente responda correctamente al reactivo i .
- β_s es la habilidad del sustentante
- δ_r es la dificultad del reactivo
- k es el número de reactivos en el examen
- e es la constante trascendental 2.71828

- La función $P_i(\beta_s)$ traza la curva logística en forma de S con valores de 0 y 1 sobre la escala de habilidad y que conforma la curva característica de un reactivo dado.
- El punto de inflexión **0** es el aquel en la escala de habilidad donde la probabilidad de respuesta correcta es **.50** y consecuentemente a mayor dificultad del reactivo se requerirá un mayor nivel de habilidad para emitir la respuesta correcta con grado de probabilidad mayor al **50%**.

En la estimación de los parámetros de habilidad de sustentantes y dificultad de reactivos, la característica fundamental del modelo es la propiedad de invariabilidad. El modelo postula que bajo esta propiedad los parámetros que caracterizan a un reactivo no dependen de la distribución de habilidades de los sustentantes. De igual forma, los parámetros que caracterizan a los sustentantes no dependen del conjunto de reactivos en un examen determinado (Hambleton, R.K. Swaminathan, H., Rogers, H. J. 1991, p.15). Esta condición se sostiene cuando las predicciones del modelo y los patrones de respuesta observados ajustan entre si ya que las curvas características de cada reactivo se obtienen independientemente de la distribución de habilidades de los sustentantes con cuyos patrones de respuesta se estiman los parámetros.

La propiedad de invariabilidad descrita se deriva de la regresión lineal. En el contexto de regresión lineal, la línea trazada por los valores conjuntos de las variables X y Y, predice los valores de Y a partir de cualquier valor posible de X. Cuando el modelo de regresión se sostiene -es decir el modelo y los datos

observados ajustan entre sí- la misma línea de regresión se obtendrá para cualquier rango restringido de los valores de X en la ecuación de regresión. Esto es que, para cualquier subpoblación de valores de X la pendiente e intersección permanecerán iguales siempre y cuando el modelo y lo datos observados ajusten entre sí. En el caso de TRI los modelos de respuesta a reactivos retienen esta propiedad ya que pueden considerarse como modelos de regresión no lineal (Hambleton, R.K. Swaminathan, H., Rogers, H. J. 1991, p. 19).

Si bien la propiedad de invariabilidad no se puede observar directamente en sentido estricto, si puede inferirse. Para esta inferencia basta determinar cuando un modelo de respuesta a reactivo y los datos observados ajustan entre sí al evaluar la congruencia entre dos o más conjuntos de parámetros estimados. Si se toman dos muestras de sustentantes de diferente habilidad de una misma población, la comparación de parámetros característicos de los reactivos en un examen se postula como indicación del grado en que se mantiene la propiedad de invariabilidad para el caso particular bajo análisis. El grado de congruencia puede observarse indirectamente calculando la correlación entre los dos conjuntos de parámetros estimados y examinando el gráfico de dispersión resultante.

Método de Estimación de Parámetros

Ya que el modelo se aplica a partir de datos observados en patrones de respuesta reales la estimación probabilística a priori no es aplicable. El método apropiado y matemáticamente justificable es el de Máxima Verosimilitud, o ML

(Maximum Likelihood). Este es un método a posteriori que da por resultado estimaciones de los parámetros que con mayor probabilidad habrían producido los patrones de respuesta observados en los datos. Bajo el modelo Rasch la estimación de parámetros de dificultad y habilidad se denomina calibración de propiedades de reactivos y habilidades de sustentantes.

Como se describió previamente una curva característica de reactivo CCR dada traza las probabilidades de respuesta correcta al reactivo combinando los parámetros β_s y δ_r . En la construcción del trazo estos parámetros se combinan a través de su diferencia $(\beta_s - \delta_r)$.

En teoría esta diferencia determina el resultado del encuentro entre el sustentante y el reactivo. Sin embargo, se requiere una restricción para el tratamiento apropiado de esta diferencia. Dado que la diferencia $(\beta_s - \delta_r)$ puede variar de $-\infty$ a $+\infty$ y la probabilidad de que el sustentante emita la respuesta correcta solo puede representarse mediante valores entre 0 y 1. Esta restricción se logra especificando la diferencia $(\beta_s - \delta_r)$ como un exponente de la constante natural $e = 2.71828$. La expresión resultante se representa

$e^{(\beta_s - \delta_r)} = \exp^{(\beta_s - \delta_r)}$ que varía entre 0 y $+\infty$ y se convierte al intervalo entre 0 y 1 formando la razón de:

$$e^{(\beta_s - \delta_r)} / (1 + e^{(\beta_s - \delta_r)})$$

Esta razón se aplica para especificar la probabilidad de respuesta correcta al reactivo en la ya conocida ecuación característica del modelo:

$$P_i(x=1|\beta_s) = e^{(\beta_s - \delta_r)} / (1 + e^{(\beta_s - \delta_r)}) \quad i = 1, 2, \dots, n$$

La ecuación ahora descrita en detalle es el modelo Rasch que especifica las CCR de cada reactivo en un examen o subescala. El modelo permite la estimación de los parámetros β_s y δ_r de manera tal que los estimados son independientes entre sí de tal manera que los estimados de β_s quedan libres de los efectos de δ_r y los estimados de δ_r quedan libres de los efectos de β_s . Esta propiedad del modelo fue definida por G. Rasch como la objetividad específica de la estimación. En la demostración de esta propiedad se obtuvo el modelo en una escala lineal con generalización de propiedades métricas. (Rasch, G. 1960).

Siguiendo la ecuación característica se establece que cuando se da el encuentro de un sustentante s de habilidad más alta que la dificultad del reactivo r la diferencia $(\beta_s - \delta_r)$ es positiva y la probabilidad de respuesta correcta es mayor a .50. En la medida que la habilidad del sustentante exceda la dificultad del reactivo se incrementa la diferencia positiva. A la inversa, cuando la dificultad del reactivo es mayor a la habilidad del sustentante la diferencia $(\beta_s - \delta_r)$ es negativa y la probabilidad de respuesta correcta es menor a .50. Esta probabilidad decrece en la medida en que se incrementa la diferencia negativa hasta que aproxima a 0.

Las unidades de la medida con que se representan las estimaciones de β_s y δ_r se denominan lógitos porque los parámetros para sustentantes y reactivos, se calibran como el logaritmo natural de los momios de respuesta correcta a los reactivos que se eligen para definir el punto 0 de la escala. La dificultad del reactivo en logitos es el logaritmo natural de momio donde se

emitiría una respuesta incorrecta por sustentante situado en el punto 0 de habilidad en la escala.

Bajo el modelo Rasch los estimados de habilidad y dificultad se obtienen mediante cálculo manual empleando el algoritmo PROX (Wright & Stone, 1979, pp.20-27) que se describe a continuación.

El algoritmo PROX asume que las habilidades de los sustentantes β_s se distribuyen de una manera aproximadamente normal con una media M y una desviación estándar σ . Las dificultades de los reactivos δ_r también se distribuyen de una forma aproximadamente normal con una dificultad promedio H una desviación estándar ω .

Así cuando $\beta_v \sim N(M, \sigma^2)$ y $\delta_r \sim N(H, \omega^2)$ para cualquier sustentante s con puntaje r_v en un examen de L reactivos la estimación de habilidad se obtiene con la ecuación:

$$\beta_s = H + X \ln [r_v / (L - r_v)] .$$

Para cualquier reactivo r con puntaje s_r en una muestral de N sustentantes la estimación de dificultad se obtiene con la ecuación:

$$\delta_r = M + Y \ln [(N - s_r) / s_r] .$$

Los coeficientes X y Y son factores de expansión que se requieren para seguir la dispersión de la dificultad de reactivos en el caso de X y la dispersión de la habilidad de los sustentantes en el caso de Y .

Específicamente, $X = (1 + \omega^2 / 2.89)^{1/2}$ y $Y = (1 + \sigma^2 / 2.89)^{1/2}$ y el valor $2.89 = (1.7)^2$ resulta del factor de escala 1.7 que limita la curva logística S para que coincida aproximadamente con la curva normal. Esto sucede porque en la curva logística S los valores $1.7z$ no exceden más a una diferencia del 1% a los valores de la curva normal z .

Los estimados de β_s y δ_r tienen su error estándar SE característico:

$$SE(\beta_s) = X [L/r_v(L - r_v)]^{1/2}, \quad SE(\delta_r) = Y [N/s_i(N - s_i)]^{1/2}.$$

El algoritmo PROX puede aplicarse directamente a puntajes observados en reactivos calculándose así el valor en lógitos del reactivo r con:

$$\delta_r = \ln [(N - s_i) / s_i].$$

El valor en lógitos para el sustentante frente al reactivo se calcula:

$$y_v = \ln [r_v / (L - r_v)].$$

Los factores de expansión X y Y se estiman a continuación con las ecuaciones:

$X = [(1 + U/2.89)/(1 - UV/8.35)]^{1/2}$ para factor de expansión en logitos para sustentantes, y:

$Y = [(1 + V/2.89)/(1 - UV/8.35)]^{1/2}$ para el factor de expansión en logitos para reactivos.

En las ecuaciones anteriores $2.89 = (1.7)^2$ y $8.35 = (2.89)^2 = (1.7)^4$, U es la varianza de las unidades lógitos para reactivos, y V es la varianza de

unidades logit para personas. Estas varianzas se obtienen a partir de las fórmulas:

$$U = (\sum x_i^2 - Lx.^2)/(L-1) \text{ y } V = (\sum y_v^2 - Ny.^2) (N-1).$$

La estimación de parámetros procede al definir el centro la prueba a 0 de tal manera que $H=0$.

Luego se obtienen los parámetros:

$$\delta_r = M + Yx_i = Y(x_i - x.) \text{ para cada dificultad de reactivo } b_v = H + Xy_v =$$

Xy_v para cada habilidad de sustentante.

Los errores estándar correspondientes se obtienen mediante:

$$SE (\delta_r) = Y [N/s_i (N - s_i)]^{1/2} \simeq 2.5/N^{1/2}$$

$$SE (\beta_s) = X [L/r_v (L - r_v)]^{1/2} \simeq 2.5/N^{1/2}$$

Finalmente la media y desviación estándar de la muestra de sustentantes se obtiene con:

$$M \approx -Yx \text{ y } \sigma \approx 1.7 (Y^2 - 1)^{1/2}$$

Análisis de Ajuste de los Datos al Modelo

Una vez que se han calculado las estimaciones de parámetros para los reactivos y los sustentantes el paso siguiente es evaluar el ajuste de los datos al modelo obtenido en el procedimiento. Con las estimaciones de β_s y δ_r obtenidas de los patrones de respuesta de la muestra observada, se obtiene la diferencia entre lo que el modelo predice y los datos observados. Los residuales entre modelo y datos se calculan a partir de la expectativa del modelo para cada

x_{sr} y en cada β_s δ_r , sustrayendo la diferencia entre la expectativa y el x_{sr} observado.

La expectativa del modelo para x_{vi} se obtiene de: $E \{ x_{sr} \} = \pi_{sr}$ y la varianza del modelo es $V\{ x_{sr} \} = \pi_{sr} (1 - \pi_{sr})$ donde π_{sr} se obtiene de : $\pi_{sr} = \exp(\beta_s - \delta_r) / [1 + \exp(\beta_s - \delta_r)]$. El residual estandarizado se obtiene de: $z_{sr} = (x_{sr} - \pi_{sr}) / [\pi_{sr} (1 - \pi_{sr})]^{1/2}$.

Cuando los datos se ajustan al modelo, los residuales estandarizados toman una distribución que aproxima a la normal con una media de 0 y una varianza de 1.

Es fundamental tener en cuenta que los residuales capturan las desviaciones de las expectativas del Modelo Rasch de cada reactivo o de cada respondente. Es decir, los residuales son una medida de las anomalías observadas para cada reactivo o cada respondente con respecto al Modelo.

Estas “anomalías” observadas a su vez se transformarán en una serie de medida de bondad de ajuste de los datos al modelo. Si bien existen criterios técnicos de tolerancia o flexibilidad en el ajuste de los datos al modelo también existen límites claramente definidos para determinar cuando las anomalías observadas para cada reactivo o respondente rebasan los límites admisibles de tolerancia. Estos y otros aspectos se describen a detalle en las siguientes secciones sobre interpretación de la bondad de ajuste.

Implementación Computacional

Debe notarse que el algoritmo PROX descrito es apropiado para lograr aproximaciones a la estimación de parámetros por cálculo manual. El procedimiento equivalente para cálculo automatizado se denomina UCON y se implementa en el software Rasch WINSTEPS desarrollado por J.M. Linacre, (1991,2006). Este programa está diseñado para construir medición Rasch a partir de un conjunto de sustentantes y sus patrones de respuesta a un conjunto de reactivos. Las bases elementales de operación del programa se describen en la Parte II de este trabajo.

WINSTEPS comienza con un estimado central para cada calibración de sustentante y reactivo. Estos estimados iniciales se producen empleando una versión iterativa del procedimiento PROX para alcanzar una convergencia aproximada al patrón de datos observado. Enseguida se aplica el procedimiento UCON de forma iterada para obtener mayor exactitud en las estimaciones de los parámetros para sustentantes y reactivos. Adicionalmente, con UCON se obtienen errores estándar y estadígrafos de ajuste (FIT). La implementación de UCON se basa en los métodos de estimación de máxima verosimilitud incondicional y máxima verosimilitud conjunta para llegar a los estimados finales de los parámetros de sustentantes y reactivos. Las medidas resultantes se reportan en lógitos y los estadígrafos de ajuste se reportan como medias cuadráticas de residuales (MSQ) y como residuales estandarizados $N(0,1)$.

La estimación de ajuste entre datos y modelos inicia con un cálculo de residuales entre cada sustentante s al responder a cada reactivo r . Este cálculo produce una estimación de qué tanto se apartan los patrones de respuesta de las expectativas del modelo, bajo las calibraciones de habilidad de sustentantes y dificultad de reactivos. Estos residuales se calculan con la fórmula:

$$Y_{sr} = x_{sr} - E_{sr}$$

Los residuales resultantes se suman en un estadígrafo de ajuste expresado como una media cuadrática o como un estadígrafo estandarizado con una distribución z o t . Esta media cuadrática se obtiene de elevar al cuadrado las diferencias entre los valores observados y las expectativas del modelo para los reactivos o respondientes. En concreto al elevarse al cuadrado las diferencias entre X_{sr} y E_{sr} se eliminan los números negativos y la media cuadrática resultante captura una medida de las anomalías o discrepancias entre los datos observados y el Modelo Rasch. En el output del programa estos estadígrafos se reportan como valores **INFIT** y **OUTFIT** tanto para los reactivos como para los respondientes y su aplicación para decidir sobre la bondad de ajuste entre los datos y el modelo se describe en la siguiente sección.

Criterios de Bondad de Ajuste

El valor de **INFIT** se interpreta como una media cuadrática ponderada de residuales que es sensible a patrones de respuesta irregulares. Este estadígrafo captura comportamientos de respuestas no esperadas (es decir “anomalías”

según el modelo) a reactivos calibrados cerca del nivel de habilidad del sustentante. El **INFIT** se calcula:

$$\mathbf{INFIT} = \Sigma Z_{sr}^2 W_{sr}/N.$$

Donde los residuales están ponderados por sus varianzas individuales (W_{sr}) para reducir la influencia de comportamiento de respuestas lejos del nivel de habilidad calibrado para el sustentante o el nivel de dificultad del reactivo.

El estadígrafo **OUTFIT** es el promedio de los residuales estandarizados derivados tanto de sustentantes como de reactivos. Este promedio no es ponderado para producir estimaciones más sensibles a respuestas no esperadas lejos de las habilidades calibradas para los respondentes o para las dificultades de reactivos. El **OUTFIT** se calcula:

$$\mathbf{OUTFIT} = \Sigma Z_{sr}^2/N.$$

Su valor se interpreta como una media cuadrática no ponderada sensible a los comportamientos extremos no esperados en los patrones de respuesta. Este estadígrafo de ajuste es sensible a valores extremos y aun más sensible a comportamientos no esperados que afectan respuestas a reactivos que se encuentran lejos del nivel de habilidad del sustentante.

Debe notarse que ambos INFIT y OUTFIT se obtienen de la suma de cuadrados de la diferencia entre la expectativa del modelo y los residuales (o diferencias observadas) para cada reactivo y para cada respondente.

Para la interpretación de los valores de los estadígrafos de ajuste se aplican los siguientes criterios:

Para **INFIT**: el estadígrafo de media cuadrática (MNSQ) tiene una expectativa de 1. Valores sustancialmente menores a 1 indican dependencia (determinismo) en los datos observados. Valores sustancialmente mayores a 1 indican ruido (aleatoriedad) en los datos. Un **INFIT MNSQ** de 1.0 es indicativo de un ajuste perfecto entre los datos y el modelo. Valores superiores a 1.3 son indicativos de falta de ajuste. Valores superiores al umbral de 1.3 indican ruido o alta variabilidad aleatoria en los datos. Los valores menores a 1.00 son también indicativos de falta de ajuste ya que aparentemente ajustan demasiado bien.

Bajo **INFIT** aparecen también valores estandarizados **ZSTD**. Este es el estadígrafo de media cuadrática de los residuales representado en logaritmos naturales. Cuando este estadígrafo toma valores de rango -2 a +2, los valores están en el intervalo de lógitos aceptable para determinar ajuste razonable tanto en sustentantes como en reactivos. Los valores superiores a +2 o inferiores a -2 son indicativos de falta de ajuste entre los datos y el modelo.

Para **OUTFIT**: la media cuadrática **MNSQ** tiene también una expectativa de 1. Valores sustancialmente menores a 1 indican dependencia (determinismo) en los datos observados mientras que valores sustancialmente superiores a 1 indican la presencia de valores extremos. Un **OUTFIT MNSQ** de 1.0 indica ajuste perfecto. Un valor de 1.3 es indicio de falta de ajuste mientras que valores

superiores a 1.3 definitivamente indican falta de ajuste. Los valores menores a 1.00 son también indicativos de falta de ajuste ya que aparentemente ajustan demasiado bien.

Bajo **OUTFIT** aparecen también valores estandarizados **ZSTD**. Para este estadígrafo los valores entre -2 y +2 están dentro del intervalo de logitos aceptable para determinar un ajuste razonable tanto para sustentantes como para reactivos. Los valores superiores a +2 o inferiores a -2 son indicativos de falta de ajuste.

Tratándose de valores no estandarizado **INFIT MNSQ** y **OUTFIT MNSQ**, el razonamiento que origina los umbrales de 1. y 1.3 es el siguiente:

Un valor **MNSQ** de $1 + x$ explica $100x\%$ más variación entre los datos observados y los patrones de respuesta esperados si el modelo y los datos observados fueran compatibles. Por lo tanto, valores superiores a 1.30 indican 30% más de variación entre lo que el modelo predice y los patrones de respuesta observados de facto. De la misma forma valores menores a 1.00, por ejemplo .80 indican 20% menos de variación que la esperada bajo el modelo. (Bond & Fox, 2001).

La interpretación de estadígrafos **INFIT** y **OUTFIT** se fortalece al especificar que una medida se considera “muda” o atenuada cuando se observa dependencia (determinismo) no modelados en los datos o bien cuando hay extrema redundancia o sesgos significativos de error aleatorio. Esto

ocurre cuando los estadígrafos presentan valores **MNSQ < .80** o bien valores **ZSTD < -2**. En el otro extremo una medida se considera “ruido” (aleatorio) cuando los estadígrafos de **MNSQ >1.3** o bien **ZSTD >+2**.

Ejemplo de Aplicación

A continuación se presenta un ejemplo con la información de salida obtenida de subescala de un examen aplicado en un ejercicio de medición en el nivel de educación media superior así como los lineamientos para la interpretación de la tabla de calibración de reactivos y estadígrafos de ajuste. La información que sigue tiene solo un propósito ilustrativo ya que obtención de las tablas y sus bases de interpretación se tratan en la segunda parte de este manual.

Tabla 1

TABLE 13.1 2DO. SEMESTRE FORMA A											
											Apr 10 13:01 2007
INPUT: 13062 PERSONS, 20 ITEMS MEASURED: 13061 PERSONS, 20 ITEMS, 2 CATS											3.37

PERSON: REAL SEP.: .92 REL.: .46 ... ITEM: REAL SEP.: 50.38 REL.: 1.00											
ITEMS STATISTICS: MEASURE ORDER											

ENTRY	RAW				INFIT		OUTFIT		PTBIS		
NUMBER	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZEMP	MNSQ	ZEMP	CORR.	DISCR	ITEMS

20	2337	12954	1.58	.02	1.06	.6	1.25	1.4	.00	.89	R20
2	2533	13017	1.48	.02	1.04	.5	1.13	.8	.04	.92	R2
5	3048	13001	1.22	.02	.99	-.1	1.04	.3	.13	.99	R5
3	3322	13014	1.10	.02	1.08	1.0	1.17	1.4	.01	.84	R3
16	3352	13014	1.09	.02	1.04	.6	1.08	.6	.07	.91	R16
1	3380	13016	1.07	.02	1.10	1.4	1.20	1.6	-.02	.79	R1
14	3847	12956	.87	.02	1.07	1.1	1.11	1.1	.03	.82	R14
15	6214	12970	.00	.02	.96	-1.1	.95	-.8	.21	1.21	R15
19	6353	13009	-.04	.02	1.01	.3	1.02	.4	.13	.93	R19
12	6465	12964	-.08	.02	1.04	.9	1.04	.7	.10	.81	R12
9	6598	13013	-.12	.02	.92	-2.1	.90	-1.7	.28	1.41	R9
18	6755	13008	-.18	.02	.98	-.6	.97	-.5	.19	1.12	R18
8	6821	12977	-.21	.02	.99	-.4	.98	-.3	.18	1.07	R8
17	7168	13012	-.32	.02	.96	-.9	.95	-.8	.21	1.18	R17
10	7347	13009	-.38	.02	.90	-2.5	.87	-2.1	.31	1.47	R10
7	8138	13026	-.66	.02	.98	-.5	.96	-.5	.19	1.08	R7
6	9404	13038	-1.15	.02	.98	-.3	.96	-.4	.18	1.04	R6
4	9455	13031	-1.17	.02	.98	-.3	.95	-.4	.18	1.05	R4
13	10984	13017	-1.95	.03	.97	-.2	.94	-.3	.16	1.02	R13
11	11282	13029	-2.15	.03	.92	-.6	.79	-1.1	.25	1.08	R11

MEAN	6240.	13004.	.00	.02	1.00	-.2	1.01	.0			

S.D.	2690.	25.	1.06	.00	.06	1.0	.11	1.0			
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+											

Para la interpretación de los valores en esta tabla de calibración de reactivos se aplican los criterios siguientes:

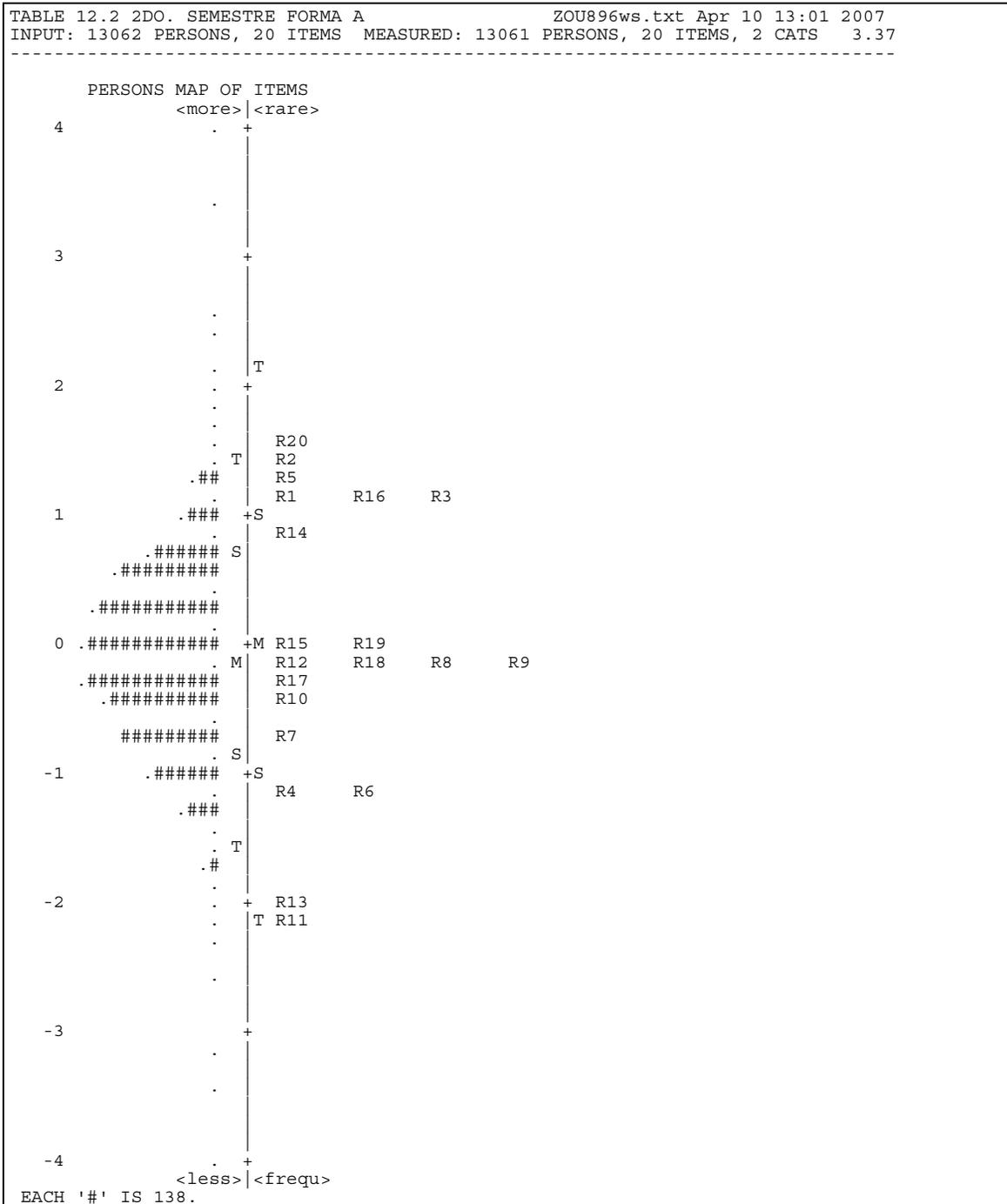
- **MEASURE** es la calibración de dificultad del reactivo expresada en lógitos.
- **ERROR** es el error estándar de la medida en lógitos. Cuanto más estos valores aproximen 0, la precisión de la medida es mayor.
- Para ajuste interno **INFIT MSQ** y ajuste externo **OUTFIT MSQ** la expectativa del modelo es de valores entre **1.00 y 1.30**.
- Para ajuste interno estandarizado **INFIT ZEMP** y ajuste externo **OUTFIT ZEMP** la expectativa del modelo es de valores entre **-2 y +2**. En este análisis se utiliza ZEMP como estandarización empírica de los valores relativa a la muestra observada.
- Los valores que no satisfacen estos criterios indican reactivos que no se ajustan al modelo
- **PTBIS** indica la correlación punto biserial entre cada reactivo calificado dicotómicamente y la puntuación total observada para el reactivo. Entre más se incremente el valor de **CORR** mayor la indicación de unidimensionalidad en la escala.
- **DISCR** indica el potencial de cada reactivo a distinguir sustentantes de baja y alta habilidad. El valor expresa la pendiente de la **CCR** donde la expectativa del modelo es **1.00**. Valores superiores a 1.00 indican mayor poder de discriminación.

Por ahora conviene notar que solo dos reactivos R 9 y R10 presentaba valores de INFIT y OUTFIT inferiores a -2.0 lo que indica desajuste estos a la expectativa del modelo. Estos criterios y su manejo se ampliarán en la sección de prácticas e interpretación.

La tabla que se presenta a continuación muestra la alineación de los reactivos y la habilidad de los respondientes organizados por grupos de habilidad. En ambos casos las medidas están expresadas en la escala común de lógitos tanto para reactivos como para respondientes. La tabla 12.2 presenta la alineación empíricamente observada y como puede verse los reactivos se alinean con los sustentantes y existen dos secciones vacías de ítems entre

R14 y R15 y entre R6 y R13. Los respondentes se distribuyen en cuanto a su habilidad de manera aproximadamente normal alrededor de la media de 0 lógito

Tabla 2



Parte II: Práctica Elemental

Bases de Operación del Programa Winsteps / Ministep

Winsteps, (Linacre, J.M. 1991-2006) es un recurso de cómputo que implementa el análisis de tipo Rasch en patrones de respuesta a ítems emitidos por sustentantes en exámenes de conocimientos y habilidades así como en otros instrumentos de medición. El programa contiene un sistema de análisis de ítems exhaustivo que puede emplearse con ítems calificados en forma binaria (0. 1) y en cadenas respuesta de opción múltiple (A B C D). También es posible analizar respuestas a ítems tipo escala Likert. Si bien el uso Winsteps requiere precisión en la preparación previa; el programa es práctico y poderoso ya que genera información abundante y exacta sobre los ítems y los sustentantes en un examen.

Existe una versión de entrenamiento denominada **Ministep** disponible de manera gratuita en el sitio: <http://www.winsteps.com/ministep.htm> . Con esta versión se pueden realizar todos los procedimientos de análisis de la versión completa aunque limitados a 25 ítems y 75 respondentes o casos. El contenido de estas notas se refiere a la versión completa y a la de entrenamiento.

La operación de Winsteps / Ministep (en adelante WS / MS) se efectúa a través de archivos de control (ACs) que son grupos de instrucciones en los que el usuario especifica las modalidades de su análisis a través de una secuencia de pasos ordenados por variables de control (VCs). Estas variables de control, también llamadas “especificaciones”, regulan la aplicación del procedimiento a una base de datos previamente preparada en un formato de archivo de datos

(AD). Por medio de las especificaciones o VCs, el usuario puede seleccionar opciones básicas y especiales para el análisis de su interés siguiendo las reglas y convenciones que se describen en este trabajo.

Esta sección práctica tiene dos propósitos:

- 1) Describir la preparación de archivos de control y de datos así como las principales reglas de operación y opciones de WS / MS.
- 2) Describir las características de las principales tablas de información de salida de WS / MS y sus bases de interpretación.

Con lo anterior no se busca substituir el uso del manual del programa. El objetivo es lograr una base de práctica suficiente para que el usuario se inicie en los procedimientos y pueda avanzar desde esta base hacia análisis más completos para propósitos especiales. Se destaca también que estas notas para práctica elemental no substituyen la formación teórica mínima necesaria sobre las bases conceptuales del modelo Rasch y su lógica de aplicación. Estos requisitos mínimos previos se presentan en las secciones iniciales de este trabajo.

El Archivo de Especificaciones de Control (AC).

Un AC es un archivo de texto (con extensión .txt) preparado en un editor de textos común (Notepad u otro similar) que contiene las especificaciones o variables de control (VC's) para un análisis particular y comentarios del usuario. Los componentes son las VC's que se incluyen en el archivo seguidas de un signo de = tal como se muestra:

De manera completa un archivo de control (AC) básico consta de los

elementos que se muestra en el siguiente ejemplo:

Ejemplo 1

```

; WINSTEPS archivo de control para demostración
; Con este elemento se inician variables de control
&INST
; Titulo para imprimir en la información de salida
TITLE = "Ejemplo Básico de Operación 1"
NI = 20 ; Numero de ítems 20 = longitud del examen
ITEM1 = 1 ; Columna del primer dato a nivel de ítems
NAME1 = 21 ; Columna del primer dato a nivel de personas
NAMLEN = 3 ; Longitud máxima de datos de identificación personas
XWIDE = 1 ; Ancho de columnas por campo de datos
PERSON = Alumno ; Las personas se llaman... Alumno
ITEM = Reactivo; Los ítems se llaman... Reactivo
; Calificación de respuestas observadas (datos)
CODES = ABCD ; Códigos validos de respuesta
; Clave de respuestas correctas 2do Semestre Español FORMA A
KEY1 = BAAABBCBCACCBADDBCCC
; Lectura de Datos en archivo externo desde ruta hasta archivo con
extensión txt
DATA = "C:\Documents and Settings\Usuario\My Documents\Ejemplo1.txt"
&END
; Las etiquetas para los reactivos / ítems son
R1 ; Item 1
R2 ; Item 2
R3 ; Item 3
R4 ; Item 4
R5 ; Item 5
R6 ; Item 6
R7 ; Item 7
R8 ; Item 8
R9 ; Item 9
R10 ; Item 10
R11 ; Item 11
R12 ; Item 12
R13 ; Item 13
R14 ; Item 14
R15 ; Item 15
R16 ; Item 16
R17 ; Item 17
R18 ; Item 18
R19 ; Item 19
R20 ; Item 20
END NAMES

```

Desglosando cada línea se tiene:

Para insertar comentarios que no operan como instrucción basta incluir un punto y coma (;) antes del comentario. Como sigue:

; El comentario es un recordatorio ilustrativo para el programador / usuario.

; El programa ignorará lo anterior al operar.

; Cualquier otro caso no precedido por ; contará como parte de las VC's de un AC

; Las instrucciones inician con la expresión:

&INST

TITLE = "*nombre de análisis*"

DATA = "*ruta y nombre del archivo de datos*"

Para estos casos se utiliza la doble comilla (" ") sólo cuando los nombres o las rutas tienen espacios intermedios.

NI = *num* Indica numero de ítems a analizarse

ITEM1= *num* señala el numero de la columna en el archivo de datos donde inicia la información sobre los ítems (patrones de respuesta)

NAME1 = *num* señala el numero de la columna en el archivo de datos donde inicia la información (identificador) de los respondentes.

NAMLEN = *num* indica el ancho del identificador (numero de caracteres)

XWIDE = *num* indica el ancho de las columnas de datos (campo) usualmente 1

PERSON = *nombre* designación de los respondentes en la información de salida

ITEM = *nombre* designación de los reactivos en la información de salida

CODES = *tipos de código* en archivo de datos (A B C D.... 0 1....o 1 2 3 4 5 etc.)

KEY = *clave de respuestas correctas* (BACDAC etc.....)

DATA = "C:\Documents and Settings\Usuario\My Documents\Datos1.txt"

; Las instrucciones terminan con la expresión

&END

; Resulta conveniente asignar etiquetas a los ítems

R1 ; Reactivo 1

R2 ; Reactivo 2

.

.

.

.

Etc.

END NAMES ; Finaliza el etiquetado

Con lo anterior se describe la preparación en secuencia de un análisis Rasch básico. Las VCs han sido resaltadas solo como ilustración de su forma y posición.

Este análisis correrá sobre un archivo de datos y su conformación se muestra a continuación.

El Archivo de Datos (AD)

La lectura de los datos hacia el archivo de control se realiza a partir de un archivo de datos. Por lo regular los datos originales (patrones de respuesta a ítems) estarán capturados en archivos de algún programa para manejo de bases de datos como EXCEL, FOXPRO, o DBASE. En principio es necesario guardar estos archivos en formato de texto, es decir con extensión .txt o .dat. Es importante que estos archivos solo contengan los patrones de respuesta y los identificadores de respondente. Esto es importante ya que WS/MS requieren que el archivo de datos esté libre de elementos que no correspondan a datos de ítems y respondentes. También es posible que los datos estén capturados en

archivos de algún paquete estadístico como SPSS o similar. Para el caso de SPSS, existe una opción en WS para leer archivos de este tipo que se describirá en la sección de opciones.

Hay dos formas iniciales para la lectura de los datos de los patrones de respuesta a los ítems. La primera consiste en ingresar los datos en formato .txt en el mismo archivo de control al final de la expresión END NAMES. Si bien esta forma es relativamente práctica, no es muy recomendable cuando el archivo de datos es grande. Para estos casos es preferible crear un archivo de datos por separado como se describirá más adelante. Sin embargo para análisis de patrones de respuesta con un menor número de casos (por ejemplo 40 casos) es posible utilizar la porción final el archivo de control que quedaría de la siguiente forma:

Ejemplo 2

```
R15      ; Item 15
R16      ; Item 16
R17      ; Item 17
R18      ; Item 18
R19      ; Item 19
R20      ; Item 20
END NAMES
BDCAABCABBBBCADBCBBC001
CCBABCDBBACCBCDCCCDD002
CDDADCBBBCBCCBCBCCCAC003
ABCAABCBBBCDBDDDBCDA004
BCDBDBCBCACDBBDCCDBB005
ACBADCAACBDACBCADBAC006
ADDAABCABACCBBDCBABA007
ACCBDCCBABCCBBCCABCC008
DDABABCBAABCBCADCCAC009
ACDACBCBBBCBCACCCABCC010
ADDDBCBBAACCBADDDDBCC011
BAADBADBBACCBCACABDC012
DDAADCBBAACCBDBBCCBB013
BAAADCBBABCBBCCCCADC014
CDBBBBCCCCCBCABCCCD015
BAAAABCBBABCBCDDBCCB016
CDAADBCDCAC . CBDCBCBD017
CCCADBCBCBCABABCCCB018
BDDBACBBDBCCBBDCCACCD019
CDDABBCBAACCCAACBCCB020
```

```

BAADCBCAAACBBADBDCAA021
BCBADBCACACDBCDDBCCB022
ADDACBCBCBCCBACDCCCC023
ADBABBBCBDDCABAADDCDA024
BADADBCDBBCDBBADBBAB025
ADBAABCBCACDBAACBACA026
ABDACCCABDCBDCDCDBD027
CDAADBABDBCCBDDCBBCD028
BADDADBCACBDBCDBCCAD029
CBDACBABCBCBCACDACBA030
CDDACBABCBCBDDCBCCD031
BADDADBCBCBDBCDBCAAD032
DCDAABABAACBBBACBACA033
CDAACBABDBCCBDDCBBCD034
BACBACAAAACBBDACBDDA035
BCDCCCDACACCBBDDCCDD036
ADDAACBBCBCCBBDDBDAA037
ADBAABDBCACBBCDABCCA038
CDAADBABCACCBACDBACD039
CDDBACADBCBBBCCADCAB040

```

Como puede verse en este caso el ingreso y la lectura de datos se realizan directamente en el archivo de control. Lo anterior simplemente le indica al programa que se tienen 20 ítems (en las columnas) y 40 sustentantes (en las filas).

A manera de ejemplo, el arreglo anterior es útil para describir las características de un archivo de datos en WS / MS. La posición de las columnas indica por sí sola la designación de los reactivos. Los sustentantes están numerados del 001 al 040. Ampliando este ejemplo, si los datos están capturados en forma binaria (1, 0) el final del archivo de control y datos combinados se vería de la siguiente forma:

Ejemplo 3

```

R15      ; Item 15
R16      ; Item 16
R17      ; Item 17
R18      ; Item 18
R19      ; Item 19
R20      ; Item 20
END NAMES
000110010000010000000001
000100000100000000000002
000100000000000000010003

```

```

100110000000000000000000000000001004
00000000001000000000000000000005
1001001100010001000100100006
100110010100000000101007
10000000100000001000007
00101000100000100010008
1001000000000100100009
100000001100010000000010
01100100010000101000011
001100001100000000000012
011100001000000000100013
0000000000000010000014
011100001000000000000015
00110000010.00000000016
00010000000101000001017
0000100000000000100018
00010000110001100000019
01100001110001000011020
000100010100000000000021
10010000000001000000022
10010000000101100001023
01010000000000100010024
10011000010001100101025
10010001000000000000025
00110010000000000000027
01001000100000000010028
00010010000001001001029
00010010000000000000030
010010000000000000110031
00011010110000100101032
00110010000000000000033
01001011110000100001034
00000001010000000000035
10011000000000000011036
10011000010000010001037
00110010010001000100038
00001010000000010010039
00001010000000010010040

```

En este caso la lectura de los datos asigna automáticamente “1” como acierto y “0” como error por lo que no se utilizarían las instrucciones CODES y KEY ya que estos valores quedan definidos con la forma binaria. Los identificadores de respondente se encuentran en las últimas tres columnas.

En suma, cuando el número de casos lo permite el análisis puede manejarse de la forma descrita. Si se opta por esta combinación de especificaciones y datos en el mismo archivo de control se omitirá la especificación de DATA = ya que el programa no la requiere al encontrar los

datos después de END NAMES.

La descripción anterior ilustra los aspectos básicos de la lectura de datos pero no es muy práctica cuando el análisis incluyen grandes números de respondientes, por ejemplo: 100, 1000, o aún más casos. La versión completa de WINSTEPS admite hasta 30,000 reactivos y 10,000,000 de respondientes. Para estos casos de análisis de gran escala los archivos de datos se preparan por separado ya que con ello se logra mayor control sobre las propiedades del procedimiento. Bajo esta modalidad será necesario incluir una instrucción con la forma:

```
DATA ="C:\Documents and Settings\Usuario\My Documents\Ejemplo1.txt"  
&END
```

Aquí la ruta C:\ y los nombres de los directorios y subdirectorios dirigen el programa hacia el archivo de datos que ha sido preparado y guardado con un nombre específico, en este caso: Ejemplo1.txt

Un archivo de datos aparte del archivo de control es simplemente un arreglo rectangular con k reactivos en las columnas y n respondientes en las filas. Los identificadores de caso (por lo general algún número de folio) pueden colocarse al principio o al final de cada patrón individual de respuestas. En este punto es importante recordar que las instrucciones **ITEM1= num** y **NAME1 = num**, le indicarán al programa las columnas donde se encuentran los datos de los reactivos y los de los identificadores respectivamente.

Aunque existen varias opciones para el manejo de archivos de datos, en

su forma más elemental el arreglo de los datos queda de la siguiente forma:

Ejemplo 4

```

BDCAABCABBBBCADBCBBC105
CCBABCDBBACCBCDCCDD109
CDDADCB . . BCCBCBCCCAC111
ABCAABCBBBCBDDDBCD113
B . DBDBCBCACDBBDCCDBB115
ACBADCAACBDACBCADBAC117
ADDAABCABACCBBDCBABA119
ACCBDCCBABCCBBCCABCC121
DDABABCBAACBBACDCCAC123
ACDACBCBBBCB . . CCABCC125
ADDDBCBBAACCBADDDDBCC129
BAADBADBBACCBCACABDC131
DDAADCBBA . CCBBDBBCBB133
BAAADCBBABCBBCCCCADC135
CDBBBBCCCC . CBCABCCCD137
BAAABCBBABCBCDDBCCB139
CDAADBCDCAC . CBDCBCBD141
CCCADBCBCBCABABCCCB145
BDDBACBBDBCCBBDCCACD147
CDDABBCBAACCCAACBCCB151
BAADCBCAAACBBADBDCAA153
BCBAD . . . CACDBCDDBCCB155
ADDACBCBCBCCBACDCCCC157
ADBABBCCBBDCAADDCDA161
BADADBCDBBCDBBADBBAB163
ADBAABCBCACDBAACBACA165
etc.

```

En este caso el archivo no tiene espacios intermedios entre los campos de datos y los datos faltantes (ítems no contestados) han sido substituidos por un punto. El tratamiento de casos con datos faltantes es de suma importancia y se detalla en la siguiente sección. Los identificadores de caso se encuentran en las ultimas 3 columnas.

De inicio el manual de WINSTEPS recomienda esta forma básica de arreglo aunque existen también otras opciones que se comentarán más adelante. Por ahora la idea básica es que el programa por default viene diseñado para leer los archivos de datos de esta forma elemental.

Ahora bien, si el archivo de datos está en un formato diferente, por

ejemplo cuando los campos de datos están delimitados por coma (formato .csv) o bien por espacios de de tabulador (TAB) es necesario incluir una especificación para que el archivo de control reconozca el formato del archivo de datos. Esto se logra incluyendo la especificación **DELIMITER = “,”** o bien **DELIMITER= TAB** según sea el caso. La sección de la AC que dirige a la lectura de datos toma la siguiente forma:

```
DATA ="C:\Documents and Settings\Usuario\My Documents\Ejemplo1.csv"
DELIMITER = “,”
&END
```

Es posible utilizar esta opción para diferentes tipos de delimitador:

```
DELIMITER = “,” ; los datos están delimitados por coma
DELIMITER = BLANK ; los datos están delimitados por un espacio
DELIMITER = SPACE ; los datos están delimitados por un espacio
DELIMITER = TAB ; los datos esta delimitados por TAB
DELIMITER = “,” ; los datos están delimitados por punto y coma
```

La elección anterior dependerá de las preferencias del usuario en lo relativo al manejo de bases de datos. Sin embargo es recomendable que siempre que sea posible, el archivo de datos quede compactado sin delimitadores tal como se ha mostrado en los ejemplos previos.

Tratamiento de Casos con Datos Faltantes

Con mucha frecuencia los archivos de datos contendrán datos faltantes o valores perdidos. Esto puede deberse a que los respondientes omitieron respuestas o a problemas de captura o codificación de los datos. Es importante que el usuario determine por inspección el origen de los casos con datos faltantes y tome una decisión sobre cómo tratarlos.

De inicio MS y WS calculan los estimados sobre la base de los datos

efectivamente existentes en el archivo de datos. Es crucial tener en cuenta que el procedimiento simplemente “ignora” cualquier símbolo que no esté expresamente declarado en la instrucción CODES. Así por ejemplo en las líneas:

```
CDDADCB . . BCCBCBCCCAC111
ABCAABCBBBCBDDDBCDA113
B . DBDBCBCACDBBDCDBB115
ACBADCAACBDACBCADBAC117
ADDAABCABACCBBDCBABA119
ACCB CCBABCCBBCCABCC121
DDABABCABC BCADCCAC123
ACDACBCBBBCB . . CCABCC125
ADDDBCBBAACCBADDDBC129
BAADBADBBACCBACABDC131
DDAADCBBA . CCBBDBBCBB133
```

Los puntos “.” serán ignorados y MS/WS calcularán los estimados de dificultad de reactivos y de habilidad de los respondientes en base a los patrones de respuesta efectivamente presentes y los espacios en blanco “ ” o con “.” no serán tomados en cuenta para la estimación de parámetros. Esto es una ventaja ya que en los casos de observaciones con datos faltantes no será necesario imputar valores perdidos. Sin embargo esta ventaja debe utilizarse con cuidado ya que su uso puede tener implicaciones importantes para la precisión de los estimados y del análisis en conjunto.

Repasando material anterior se tiene que cuando MS/WS procesa cada observación individual. Para cada caso observado X_{ni} de la persona n en el ítem i , el procedimiento calcula una expectativa E_{ni} basada en la medida actual de B_n de habilidad estimada de la persona n y la calibración actual D_i de dificultad del reactivo.

Pero debe tenerse en cuenta que estos cálculos se realizan “saltándose” los

datos faltantes. Y de esta manera los estimados no estarán reflejando la influencia de los datos faltantes en el análisis de cada caso individual. Esto es, las medidas actuales de habilidad y de dificultad reflejan solamente los patrones de respuesta (aciertos o errores) efectivamente registrados y presentes en el archivo de datos.

Por lo anterior es necesario aplicar una instrucción especial: MISSCORE, cuando los datos faltantes se deben a respuestas expresamente omitidas por los respondientes. Esto es particularmente necesario en el análisis de exámenes de logro. En estos casos las respuestas omitidas por los respondientes usualmente se consideran como errores, ya que esto aumenta la precisión de la estimación de habilidad real de los respondientes.

Ante esta situación es importante representar las omisiones con un símbolo específico al caso. Por lo general es recomendable utilizar “9” en lugar de “.” o el espacio en blanco “ ”.

Así por ejemplo en el siguiente archivo de datos :

```
00011001000001000000001
00010000010000000000002
000100000000.0000010003
10011000000000000001004
000000000100000000010005
10010011000100010010006
100110010..000000101007
100000001000000001000007
00101000100000100010008
10010000000001001000009
10000000110001000000010
01100100010000101000011
0011000..1000000000012
011100001000000000100013
00000000000000100000014
01110000100000000000015
00110000010.00000000016
00010000000101000001017
000010000000000001000018
00010000110001100000019
01100001110001000011020
00010001010000000000021
```

```

10010000000001000000022
1001000....101100001023
0101000000000100010024
10011000010001100101025
10010001000000000000025
00110010000000000000027
01001000100000000010028
00010010000....01001029
00010010000000000000030
01001000000000000110031
00011010110000100101032
00..0010000000000000033
01001011110000100001034
00000001010000000000035
10011000000000000011036
10011000010000010001037
00110010010001000100038
00001010000000010010039
....1010000000010010040

```

Si se trata de un examen de logro, la presencia de los valores faltantes en “.” producirían estimaciones de habilidad sesgadas en aquellos casos en que no se reflejan las omisiones de los respondentes. Para este caso el tratamiento apropiado consiste en substituir los datos faltantes con un símbolo 9 (por ejemplo) y declarar la instrucción MISSCORE = 0

De esta forma el archivo:

```

00011001000001000000001
00010000010000000000002
00010000000090000010003
10011000000000000001004
00000000010000000010005
10010011000100010010006
10011001099000000101007
10000000100000001000007
00101000100000100010008
10010000000001001000009
10000000110001000000010
01100100010000101000011
00110009910000000000012
01110000100000000100013
00000000000000100000014
01110000100000000000015
00110000010900000000016
00010000000101000001017
000010000000000001000018
00010000110001100000019
01100001110001000011020

```

```

000100010100000000000021
100100000000010000000022
10010009999101100001023
01010000000000100010024
10011000010001100101025
10010001000000000000025
00110010000000000000027
010010001000000000010028
00010010000999901001029
00010010000000000000030
010010000000000000110031
00011010110000100101032
00990010000000000000033
01001011110000100001034
00000001010000000000035
10011000000000000011036
10011000010000010001037
00110010010001000100038
00001010000000010010039
99991010000000010010040

```

Se trata con la siguiente instrucción:

CODES = 01

MISSCORE = 0

Lo cual especifica que todo símbolo que no haya sido expresamente declarado en CODES será tratado como error de tal manera que los “9” serán calificados a 0 = error, lo que produce una estimación más precisa de las habilidades.

En el caso de un archivo de datos con códigos ABCD y una clave de respuestas

Correctas KEY=ABCCADDAD etc. el archivo de datos queda:

```

DDADCB99BCCBCBCCCAC111
ABCAABCBBBCBDDDBCD113
B9DBDBCBCACDBBDCDBB115
ACBADCAACBDACBCADBAC117
ADDAABCABACCBBDCBABA119
ACCB CCBABCCBCCABCC121
DDABABCABC BCADCCAC123
ACDACBCBBBCB99CCABCC125
ADDDBCBBAACCBADDDDBCC129
BAADBADBBACCBCACABDC131
DDAADCBBA9CCBBDBBCBB133

```

Y la instrucción apropiada es:

CODES = ABCD

KEY= ABCCCADDAD etc.

MISSCORE = 0

Lo que en efecto hace que los “9” sean calificados con 0 (error).

Es necesario notar que las previsiones sobre datos faltantes deben ejecutarse de manera obligatoria en el análisis de patrones de respuesta en exámenes de logro. Esto debido a las implicaciones que la presencia de datos faltantes tiene en las estimaciones de habilidad de respondentes y de dificultad de los reactivos.

En el caso de patrones de respuesta a reactivos de tipo Likert –cuyo análisis se trata en secciones posteriores- también es posible declarar MISSCORE=0 para que en el análisis de la dificultad de emitir acuerdos sobre afirmaciones se tome en cuenta la presencia de respuestas omitidas. Por ejemplo en reactivos de opinión o percepción con una escala Likerts de 1 a 7 el caso es:

CODES=123457

MISSCORE=0

MISSCORE=0 trata como respuesta negativa a todo símbolo que no haya sido expresamente declarado en la instrucción CODES.

Bajo estos casos existe una variante que permite imputar valores faltantes hacia un punto medio (o neutral) si ese tratamiento se puede justificar en teoría.

En uno de estos casos

CODES=12345678

MISSCORE=4

Calificaría imputando como postura intermedia o neutral a 4 los datos faltantes.

Pero la decisión de imputar de esta forma debe ser tomada a la luz de teoría sustantiva y en el contexto particular de estudio que se realice. En las secciones finales se describen los procedimientos especiales para analizar reactivos de actitud, afinidad, opinión o percepción.

Ejercicio Inicial de Programación

Antes de proceder hacia la descripción de la interpretación de la información de salida es conveniente que el lector realice un breve ejercicio de programación y corrida del programa. Aunque es posible copiar y pegar en Notepad el mini-programa que se presenta más abajo, se recomienda escribirlo en el editor Notepad u otro y guardarlo con extensión .txt para así lograr una familiarización inicial con las reglas de construcción de un archivo de control (AC).

El siguiente ejercicio consiste en analizar los patrones de respuesta a 20 ítems emitidos por 40 respondentes. Se asume que para esta práctica se empleará la versión de entrenamiento Ministep, por lo tanto el mini-programa es:

Ejemplo 5

```

; Ejemplo para entrenamiento en Ministep
&INST
TITLE = "Análisis de Examen Basico"
NI = 20           ; 20 reactivos
ITEM1 = 1        ; los datos de reactivos inician en la columna 1
NAME1=21        ; el identificador inicia en la columna 21
NAMELEN = 2     ; el identificador de respondentes es de 2 dígitos
XWIDE = 1       ; cada campo de datos es de 1 columna
ITEM = Reactivo ; los ítems se llaman Reactivo
PERSON = Alumno ; las personas se llaman Alumno
&END
R1      ;Item 1
R2      ;Item 2

```

```

R3      ;Item 3
R4      ;Item 4
R5      ;Item 5
R6      ;Item 6
R7      ;Item 7
R8      ;Item 8
R9      ;Item 9
R10     ;Item 10
R11     ;Item 11
R12     ;Item 12
R13     ;Item 13
R14     ;Item 14
R15     ;Item 15
R16     ;Item 16
R17     ;Item 17
R18     ;Item 18
R19     ;Item 19
R20     ;Item 20
END NAMES

```

El archivo de datos que en este caso puede copiarse y pegarse bajo END NAMES, (después de escribir el mini-programa) es:

```

100100011000100100101
100110100111001000002
101000010001111100103
100010011111111100104
111111001111001110005
010011111111001100106
110101010111011101007
111111001110011110008
110110100111111100008
000011111011111011110
101110100011111111011
111001010111111011012
101001111111111001113
111101010101111101114
111101011111111110015
111010001111111111116
110011111111111010117
111011010111111011118
011101111111111001119
11101001111111101120
111001110011111111121
101111111111111001022
110111011111101101123
11110111111111101024
111111011011111101125
101111111111110111.26
101101111111111111027
01101111111111101128
11111101111111011029
110011111111111111130
10111111111111101131
10111111111111101132
11101111111101111133
11100111111111111134
10111111111111111135
1111111111111101136

```

```

1111111111111111101137
1111111111111111111138
1111111111111111111139
1111111111111111111140

```

Para esta práctica se recomienda guardar el archivo de control con los datos combinados con algún nombre corto como “**Examen1.txt**” en una carpeta de documentos llamada “**Practicas Rasch 1**” (o algún otro nombre similar) ya que para correr este mini-programa el usuario dirigirá Ministep hacia esa carpeta.

Corrida Inicial en Ministep

Los pasos para correr el mini-programa en Ministep se ilustran a continuación;

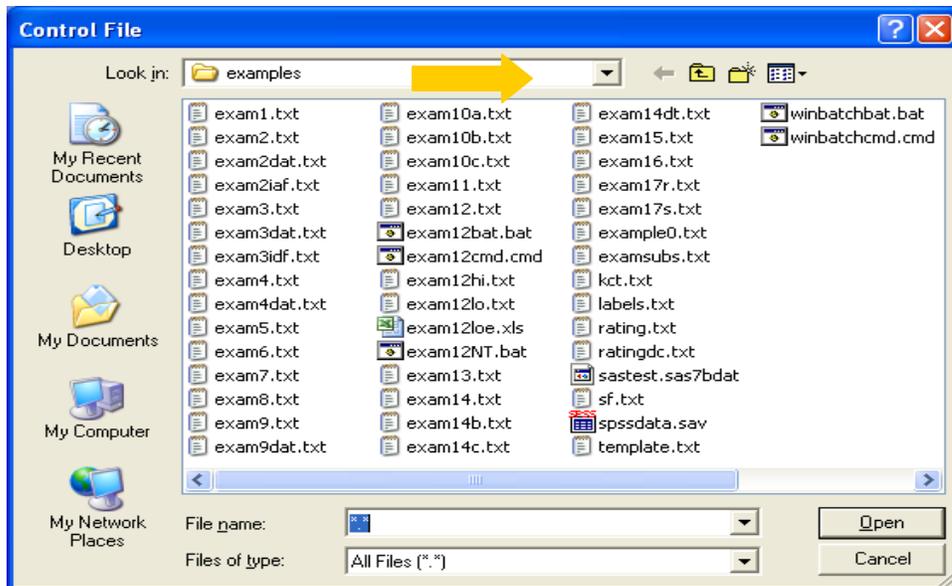
- (1) Abrir Ministep desde su icono de acceso directo en el escritorio (o carpeta)
- (2) Se abrirá la siguiente pantalla:

```

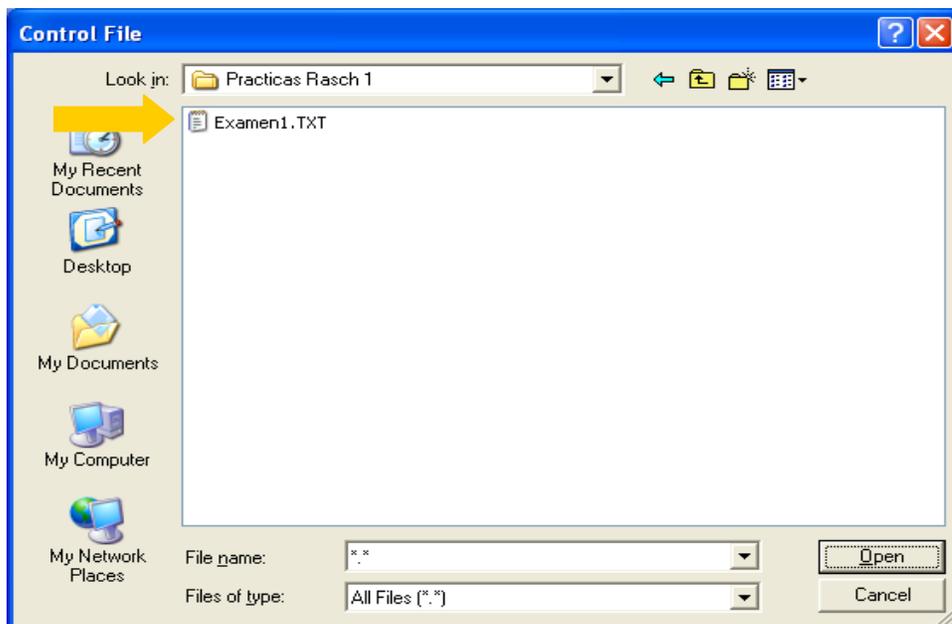
MINISTEP Version 3.63.0 Apr 11 9:19 2007
Current Directory: C:\WINSTEPS\examples\
Control file name? (e.g., exam1.txt). Press Enter for Dialog Box:

```

- (3) Al pulsar **ENTER** se abrirá la siguiente ventana de dialogo:



(4) Pulsar la pestaña de selección para ir a la carpeta que contiene el archivo de control (AC) y el archivo de datos (en su caso) y pulsar **OPEN**.



(5) Seleccionar el AC que desea correr (en este caso Examen1.TXT) y pulsar **OPEN** y se abrirá la siguiente pantalla:

```

MINISTEP Version 3.63.0 Apr 11 9:19 2007
Current Directory: C:\WINSTEPS\examples\

Control file name? (e.g., exam1.txt). Press Enter for Dialog Box:

Previous Directory: C:\WINSTEPS\examples
Current Directory: C:\Documents and Settings\Becario\My Documents\Act1-UEE\Presentaciones MGM\Practicas Rasch 1
C:\Documents and Settings\Becario\My Documents\Act1-UEE\Presentaciones MGM\Practicas Rasch 1\Examen1.TXT

Extra specifications (or press Enter):

```

(6) Pulsar **ENTER** y Ministep comenzará a calcular con iteraciones hasta llegar la siguiente pantalla de información de salida (OUTPUT):

```

Reading Control Variables ..
Input in process..
Input Data Record:
100100011000100100101
^I
^P
40 Alumno Records Input.

CONVERGENCE TABLE
*Control: \Practicas Rasch 1\Examen1.TXT Output: \ZOU136WS.TXT
| PROX ACTIVE COUNT EXTREME 5 RANGE MAX LOGIT CHANGE |
| ITERATION Alumnos Reactiv CATS Alumnos Reactiv MEASURES STRUCTURE |
|-----|-----|-----|-----|-----|
| 1 40 20 2 2.67 2.10 2.9444 |
|-----|-----|-----|-----|-----|
| 2 37 20 2 2.90 2.27 .6859 |
|-----|-----|-----|-----|-----|
| 3 37 20 2 3.02 2.31 .1258 |
Checking connectivity ...
*Control: \Practicas Rasch 1\Examen1.TXT Output: \ZOU136WS.TXT
| JMLE MAX SCORE MAX LOGIT LEAST CONVERGED CATEGORY STRUCTURE |
| ITERATION RESIDUAL* CHANGE Alumno Reacti CAT RESIDUAL CHANGE |
|-----|-----|-----|-----|-----|
| 1 .27 .2273 5 19* |
|-----|-----|-----|-----|-----|
| 2 .06 .0618 5 19* |
Calculating Fit Statistics
Standardized Residuals N(0,1) Mean: .01 S.D.: .98
Análisis de Examen Básico
-----+-----+-----+-----+-----+
| Alumnos 40 INPUT 40 MEASURED INFIT OUTFIT |
| SCORE COUNT MEASURE ERROR IHNSQ ZSTD OHNSQ ZSTD |
| MEAN 14.7 19.5 1.49 .65 1.00 .1 .96 .1 |
| S.D. 2.6 .6 .92 .15 .15 .6 .52 .7 |
| REAL RMSE .67 ADJ.SD .64 SEPARATION .96 Alumno RELIABILITY .48 |
-----+-----+-----+-----+-----+
| Reactivos 20 INPUT 20 MEASURED INFIT OUTFIT |
| MEAN 27.3 36.1 .00 .48 .99 .1 .95 .0 |
| S.D. 6.0 3.9 .91 .11 .19 1.1 .34 1.1 |
| REAL RMSE .49 ADJ.SD .76 SEPARATION 1.54 Reacti RELIABILITY .70 |
-----+-----+-----+-----+-----+
Output written to C:\Documents and Settings\Becario\My Documents\Act1-UEE\Presentaciones MGM\Practicas Rasch 1\ZOU136WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus

```

(7) Esta pantalla describe la información general sobre la corrida del

programa y las iteraciones (repeticiones de los algoritmos) fueron necesarias para converger a la solución. A partir de este paso pueden seleccionarse desde el menú principal OUTPUT TABLES las tablas principales de salida para interpretar los estadígrafos de ajuste al modelo de los ítems y sustentantes.

(8) Por ejemplo, al seleccionarse la **TABLA 3.1 SUMMARY OF PERSONS AND ITEMS**, Ministep presenta la estadística descriptiva sumaria de sustentantes y reactivos de la siguiente forma (se resaltan los encabezados de columna para describirlos a detalle):

(9) Descripción Parcial de la Información de Salida

Tabla 2

RAW SCORE		COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
SCORE					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	14.7	19.5	1.49	.63	1.00	.1	.96	.1
S.D.	2.6	.6	.92	.15	.15	.6	.52	.7
MAX.	18.0	20.0	3.18	1.05	1.48	1.9	3.61	2.1
MIN.	7.0	18.0	-.71	.49	.75	-1.5	.35	-1.3
REAL RMSE	.67	ADJ.SD	.64	SEPARATION	.96	Alumno	RELIABILITY	.48
MODEL RMSE	.65	ADJ.SD	.65	SEPARATION	1.00	Alumno	RELIABILITY	.50
S.E. OF Alumno	MEAN = .15							
MAXIMUM EXTREME SCORE:				3 Alumnos				
VALID RESPONSES:				97.4%				

Tabla 4

SUMMARY OF 40 MEASURED (EXTREME AND NON-EXTREME) Alumnos								
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	15.1	19.5	1.71	.72				
S.D.	2.7	.5	1.18	.35				
MAX.	19.0	20.0	4.41	1.83				
MIN.	7.0	18.0	-.71	.49				
REAL RMSE	.81	ADJ.SD	.85	SEPARATION	1.05	Alumno RELIABILITY	.52	
MODEL RMSE	.80	ADJ.SD	.86	SEPARATION	1.07	Alumno RELIABILITY	.53	
S.E. OF Alumno MEAN = .19								

Alumno RAW SCORE-TO-MEASURE CORRELATION = .93 (approximate due to missing data)
 CRONBACH ALPHA (KR-20) Alumno RAW SCORE RELIABILITY = .60 (approximate due to missing data)

Tabla 5

SUMMARY OF 20 MEASURED (NON-EXTREME) Reactivos								
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	27.3	36.0	.00	.47	.99	.1	.95	.0
S.D.	6.0	3.9	.91	.12	.19	1.1	.34	1.1
MAX.	35.0	37.0	1.44	.75	1.38	2.1	1.70	2.0
MIN.	10.0	19.0	-1.76	.36	.64	-2.8	.46	-2.6
REAL RMSE	.49	ADJ.SD	.76	SEPARATION	1.54	Reacti RELIABILITY	.70	
MODEL RMSE	.48	ADJ.SD	.77	SEPARATION	1.59	Reacti RELIABILITY	.72	
S.E. OF Reactivo MEAN = .21								

U=MEAN=.000 USCALE=1.000
 Reactivo RAW SCORE-TO-MEASURE CORRELATION = -.80 (approximate due to missing data)
 721 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 646.68

Seguendo el Manual de WS/MS (Linacre, 2006, pp. 161-163) se desglosan a continuación los rubros resaltados de la Tabla 1:

- A) El título (encabezado) resulta de la especificación TITLE en el AC.
- B) Los **datos sumarios** para Personas (Alumno) se presentan primero omitiendo los casos extremos. Es decir, presentan los casos NON-EXTREME, y se descartan aquellos respondentes que contestaron correctamente a todos los ítems o no contestaron ningún ítem

correctamente. En esta tabla se reportan 3 casos extremos que fueron omitidos del análisis. (La siguiente Tabla 2 incluye todos los casos).

- C) **RAW SCORE** es el número de respuestas correctas observadas y en este caso como se trata de estadísticas sumarias se presentan la media (MEAN) la Desviación Estándar (S. D.), el puntaje de aciertos máximo MAX y el mínimo MIN.
- D) **COUNT** es simplemente el número total de respuestas registradas
- E) **MEASURE** es la medida de habilidad en lógitos estimada para los respondentes y de calibración de dificultad para reactivos.
- F) **MODEL ERROR** el error estándar asociado a cada estimación de medida.
- G) **INFIT** es el estadígrafo de ajuste interno ponderado por la información y que es más sensible a comportamientos inesperados (según la expectativa del Modelo) que afectan respuestas a ítems cercanos al nivel de habilidad medida de las personas.
- H) **MSNQ** es el estadígrafo de media cuadrática de ajuste interno que según el modelo tiene una expectativa de 1. Los valores sustancialmente menores a 1 indican dependencia (determinismo) en los datos observados. Los valores sustancialmente mayores a 1 indican ruido (o demasiada estocasticidad) en los datos observados.

- I) **ZSTD** es el mismo estadígrafo de media cuadrática de ajuste interno pero estandarizado para aproximar una distribución con media teórica de 0 y una varianza de 1. Existen opciones para modificar esta estandarización. LOCAL = Y, y LOCAL =L cuyo uso e interpretación se describirán en la sección dedicada a opciones especiales.
- J) **OUTFIT** es el estadígrafo de ajuste externo sensible a comportamientos inesperados (según el Modelo) que afecta respuestas a ítems lejanos al nivel de habilidad medida de las personas.
- K) **ZSTD** es el mismo estadígrafo de media cuadrática de ajuste externo pero estandarizado para aproximar una distribución con media teórica de 0 y una varianza de 1. Existen opciones para modificar esta estandarización.

LOCAL = Y, y LOCAL =L cuyo uso e interpretación se describirán en la sección dedicada a opciones especiales.

- L) **RMSE** es la raíz del error cuadrático medio calculado sobre todas las personas o sobre todos los ítems presentes en el archivo de datos. El error cuadrático medio (Mean Square Error o MSE) es un indicador del error total (variación aleatoria) en una muestra. Incluye las fuentes de error totales, es decir variación aleatoria atribuible al muestreo y variación aleatoria no atribuible al muestreo. Se interpreta como la **desviación estándar del estimador** con respecto al parámetro que

se trata de aproximar.

M) **REAL RMSE** es la raíz del error cuadrático medio, pero calculado sobre la base de que los desajustes en los datos se deben desviaciones de los datos hacia el modelo. Este valor representa “el peor caso” de la confiabilidad y se reporta como el límite inferior de los estimados de confiabilidad obtenidos a partir de una muestra individual de datos observados.

N) **MODEL RMSE** es también la raíz del error cuadrático medio calculado sobre la base de los datos que se ajustan al modelo. Este cálculo asume que todo desajuste en los datos es un reflejo de la naturaleza estocástica del modelo. Este valor representa “el mejor caso” de la confiabilidad y se reporta como el límite superior de los estimados de confiabilidad obtenidos a partir de una muestra individual de datos observados.

O) **ADJ. S.D.** es la desviación estándar de los estimados ajustada después de sustraer su varianza de error de su varianza observada. La varianza de error de los estimados es atribuible al error estándar de medición.

Así:

$$(\text{ADJ. SD})^2 = (\text{S.D. Medida})^2 - (\text{RMSE})^2$$

El valor de **ADJ. S.D.** de un estimado de la desviación estándar

“verdadera” al que se le ha removido el sesgo derivado de error de medición.

P) **SEPARATION** (para PERSONAS o ITEMS) es la razón de **ADJ. S.D. a RMSE**. Proporciona una razón medida de separación (entre personas o ítems) en unidades de RMSE y es más fácil de interpretar que la correlación de confiabilidad.

Q) **PERSON (ITEM) SEP REL** es una forma de separación (confiabilidad) de las medidas, equivalente a KR-20, el Alfa de Cronbach y el Coeficiente de Generalizabilidad. La relación entre separación SEP y confiabilidad REL es:

$$REL = SEP^2 / (1 + SEP)^2$$

O bien

$$SEP = (REL/(1-REL))^{1/2}$$

Donde: SEPARACION = SD. Verdadera / Error RMSE

$$CONFIABILIDAD = Varianza Verdadera / Varianza Observada$$

R) **S.E. OF MEAN** es el error estándar de la media de la persona o del ítem

S) **RELIABILITY** es el índice convencional de confiabilidad ALFA, KR-20 y representa una aproximación a la confiabilidad del examen / escala basado en los puntajes crudos de la muestra observada.

T) RAW SCORE TO MEASURE CORRELATION (de Personas o Ítems)

es el coeficiente de correlación Pearson entre los puntajes crudos y las medidas (en lógitos) que incluye los puntajes extremos. Cuando los datos están completos (es decir cuando no hay valores faltantes) estos coeficientes deben estar cerca de 1.0 para personas y -1.0 para ítems

U) APPROXIMATE LOG LIKELIHOOD CHI SQUARE es el logaritmo de la

función de verosimilitud expresada como χ^2 . Representa un criterio global de bondad de ajuste entre los datos y el modelo en la muestra observada. Se busca que el valor de este criterio sea el menor posible ya que LL χ^2 menor indica un mejor ajuste global. Al comparar 2 muestras sobre el mismo conjunto de ítems (escala) se preferirá aquella muestra con una LL χ^2 menor.

Tablas con Estadísticas por Ítem y por Sustentante

Una vez examinadas e interpretadas las estadísticas sumarias de la muestra, el paso siguiente consiste en interpretar la información que el programa produce sobre los ítems y las personas. La tabla **13.1 ITEM MEASURE ORDER** (Reactivos en Orden de Medida) contiene esta información:

Tabla 6

TABLE 13.1 Análisis de Examen Básico 1											ZOU724WS.TXT Apr 18 9:29 2007		
INPUT: 40 Alumnos			20 Reactivos			MEASURED: 40 Alumnos			20 Reactivos			2 CATS 3.63.0	

Alumno: REAL SEP.: .96			REL.: .48			... Reactivo: REAL SEP.: 1.54			REL.: .70				
Reactivo STATISTICS: MEASURE ORDER													

ENTRY	RAW		MODEL	INFIT	OUTFIT	PTMEA	EXACT	MATCH					
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	OBS%	EXP%	Reactivo	

17	19	37	1.44	.36	1.03	.3	1.02	.2	.43	62.2	66.0	R17
4	22	37	1.05	.36	1.31	2.1	1.37	1.8	.20	54.1	68.5	R4
7	22	37	1.05	.36	1.00	.0	.94	-.2	.44	64.9	68.5	R7
2	24	37	.78	.37	1.15	1.0	1.35	1.5	.28	62.2	70.9	R2
20	10	19	.73	.48	.64	-2.8	.61	-2.6	.81	89.5	63.5	R20
19	24	36	.66	.38	1.10	.6	1.03	.2	.35	63.9	71.7	R19
5	25	37	.64	.38	1.16	1.0	1.12	.6	.30	64.9	72.3	R5
16	26	37	.49	.39	1.38	1.9	1.60	2.0	.10	67.6	73.8	R16
3	27	37	.33	.40	.87	-.6	.84	-.5	.48	75.7	76.0	R3
18	27	37	.33	.40	.71	-1.5	.60	-1.5	.60	81.1	76.0	R18
9	28	37	.17	.41	.94	-.2	.78	-.6	.44	78.4	78.1	R9
6	29	37	-.01	.43	.78	-.9	.65	-1.0	.53	86.5	80.1	R6
8	31	37	-.42	.48	1.00	.1	.75	-.4	.36	83.8	84.2	R8
10	31	37	-.42	.48	.98	.0	.89	-.1	.34	83.8	84.2	R10
13	31	37	-.42	.48	.99	.1	.95	.1	.35	83.8	84.2	R13
14	32	37	-.66	.51	.82	-.5	.61	-.7	.45	89.2	86.6	R14
1	33	37	-.94	.56	1.21	.6	1.24	.6	.13	89.2	89.2	R1
11	34	37	-1.29	.63	.83	-.2	.56	-.4	.38	91.9	91.9	R11
12	35	37	-1.76	.75	.88	.0	.46	-.4	.33	94.6	94.6	R12
15	35	37	-1.76	.75	.94	.1	1.70	.9	.18	94.6	94.6	R15
MEAN	27.3	36.1	.00	.47	.99	.1	.95	.0		78.1	78.7	
S.D.	6.0	3.9	.91	.12	.19	1.1	.34	1.1		12.3	9.3	

De inicio se han resaltado los reactivos R4 con INFIT ZSTD de 2.1 y el R2 con INFIT ZSTD de -2.8 de OUTFIT ZSTD. Claramente estos reactivos en esta muestra no se ajustan a las expectativas del modelo. Estos deberán separarse para su escrutinio en lo relativo a contenido, redacción y congruencia con el constructo unidimensional que trata de captar con la escala. El reactivo R16 está exactamente en el límite de 2.0 de OUTFIT ZSTD.

Sin embargo, lo adecuado para efectos de entrenamiento es interpretar la tabla en todos sus rubros:

REAL SEPARATION y RELIABILTY para alumnos e ítems es solamente un recordatorio de los valores ya reportados en estos rubros en la tabla de estadísticas sumarias (3.1 Summary Statistics).

ENTRY NUMBER es sencillamente el orden que toman los reactivos según su medida individual de mayor a menor dificultad.

RAW SCORE es el número de aciertos registrados para ese ítem.

COUNT es el número de respondientes (alumnos) que emitieron respuestas a cada ítem. En este caso, 37 indica que para el cálculo se tomaron en cuenta los patrones de respuesta de 37 alumnos de los 40 originales en la muestra. Esto se debe a que 3 alumnos respondieron correctamente a los 20 reactivos del examen y esa información fue descartada.

MEASURE es la medida en lógitos de la dificultad de cada reactivo (es decir la calibración del reactivo). En este caso estos valores van desde +1.44 lógitos a -1.77 lógitos en dificultad. Se observan 11 reactivos (R17 a R9) por arriba de la dificultad media de 0 lógitos. Se observan 9 reactivos (R6 a R15) por debajo de la dificultad media de 0 lógitos.

MODEL SE es el error estándar de cada medida individual expresado en lógitos. Es decir, la variación del estimador observado con respecto al parámetro que trata de estimar. En este caso el error estándar fluctúa entre .36 lógitos para los reactivos de mayor dificultad, hasta .75 lógitos para los reactivos de menor dificultad. Es de notar que en esta muestra el error estándar de los estimados decrece a medida que aumenta la dificultad de los reactivos. Inversamente, el error estándar aumenta hasta .75 lógitos en reactivos de menor dificultad.

INFIT MNSQ es el estadígrafo (ya conocido) de ajuste interno que indica ajuste razonable del reactivo al modelo cuando los valores están en el intervalo de

.80 y 1.30. Los valores sustancialmente menores a 1 indican dependencia (o determinismo) en los datos observados. Los valores mayores a 1.30 indican ruido (o demasiada aleatoriedad) en los datos observados.

El valor de 1.0 indica ajuste perfecto del ítem (o persona) a la expectativa del modelo. Los reactivos R7 y R8 cumplen este criterio. Las medias cuadráticas sin estandarizar (MNSQ) representan desviaciones entre el total de los datos observados (del reactivo) y el modelo. Sin embargo al no estar estandarizados estos valores retienen las características particulares de la distribución muestral observada para el reactivo (o persona).

INFIT ZSTD es el estadígrafo estandarizado (ya conocido) de ajuste interno que indica ajuste razonable cuando los valores están en el intervalo de -2 a +2. El valor 0 indicaría un ajuste estandarizado perfecto. Debe recordarse que INFIT es sensible a “anomalías” en los patrones de respuesta a reactivos cercanos al nivel de habilidad medida de los respondentes. La estandarización a una media de 0 y una varianza de 1, permite una interpretación más directa y generalizable. Esto se debe a que al estandarizarse estos valores se liberan de las características de la distribución muestral observada para el reactivo o persona. En la tabla los reactivos R7, R8 y R12 presentan esta característica.

En una descripción detallada de los **umbrales críticos de INFIT** tenemos:

- **INFIT > 2.0** el ruido aleatorio (fuera de la variable de interés) es mayor que la información útil que aporta el reactivo. Este efecto indeseable

(debido a características particulares del reactivo o del respondente)
degrada el proceso de medición.

- **INFIT > 1.5** el ruido aleatorio (fuera de la variable de interés) es notable. Esta situación no aporta al proceso de medición pero tampoco lo degrada.
- **INFIT de 0.5 a 1.5** situación **PRODUCTIVA** (óptima) para el proceso de medición.
- **INFIT < 0.5** esta situación ajusta demasiado bien (es decir es demasiado predecible). Es confundente en el sentido de que aparenta que se está midiendo muy bien el constructo de interés cuando de hecho no es el caso.

Por regla general en la toma de decisiones sobre reactivos deben corregirse primero los desajustes (MISFITS) grandes. Los desajustes menores a 1.0 son de considerarse solo cuando se va a acortar la longitud de un examen. (Linacre, J. M. 2006, p.169).

Para la decisión de si deben interpretarse los estadígrafos no estandarizados (MNSQ) o los estandarizados (ZSTD) se debe tener en cuenta que los valores de ambos deben coincidir entre si dentro de sus umbrales críticos y dentro del rubro del que se trate INFIT o OUTFIT). Sin embargo, es importante notar que se recomienda depender de un criterio único el ZSTD solamente en aquellos casos en que se trate de salvar una situación de un MNSQ > 1.5, y el tamaño

de la muestra sea pequeño o el examen sea de corta longitud. (Wright, B. citado en Linacre, 2006, P.169-170).

OUTFIT MSNQ y ZSTD son los ya conocidos estadígrafos de ajuste externo. Estos criterios son sensibles a valores extremos (outliers) en los datos. Es decir captan las “anomalías” en patrones de respuesta a reactivos lejanos al nivel de habilidad medida de los sujetos respondientes. Se interpretan con los mismos umbrales críticos que bajo INFIT. Debe recordarse que el ajuste externo se refiere a las anomalías detectadas en los extremos de la distribución muestral.

Así tenemos que:

- **OUTFIT > 2.0** el ruido aleatorio (fuera de la variable de interés) es mayor que la información útil que aporta el reactivo. Este efecto indeseable (debido a características particulares del reactivo o del respondiente) degrada el proceso de medición.
- **OUTFIT > 1.5** el ruido aleatorio (fuera de la variable de interés) es notable. Esta situación no aporta al proceso de medición pero tampoco lo degrada.
- **OUTFIT de 0.5 a 1.5** situación **PRODUCTIVA** (optima) para el proceso de medición.
- **OUTFIT < 0.5** esta situación ajusta demasiado bien (es decir es demasiado predecible). Es confundente en el sentido de que aparenta

que se está midiendo muy bien el constructo de interés cuando de hecho no es el caso.

Es muy importante destacar que los valores de INFIT y OUTFIT se comportan de manera independiente entre sí. Es decir un reactivo puede presentar un buen ajuste interno INFIT pero también puede presentar un mal ajuste externo en OUTFIT. La situación también puede darse a la inversa donde un reactivo presente buen ajuste en OUTFIT y un mal ajuste en INFIT. Esto se debe a que ambos estadígrafos detectan anomalías en partes diferentes del continuo de dificultades de reactivos y habilidades de los respondentes en un examen y muestra.

En lo anterior se aplica también la recomendación de que el criterio único ZSTD solo debe aplicarse (en el caso de OUTFIT) cuando aquellos casos en que se trate de salvar una situación de un $MNSQ > 1.5$, y el tamaño de la muestra sea pequeño o el examen sea de corta longitud. (Wright, B. citado en Linacre, 2006, P.169-170).

PTMEA CORR es la correlación punto – medida. Este estadígrafo es una modificación de la correlación punto biserial (PTBIS), y se interpreta como una correlación que mide el grado de asociación entre el puntaje particular observado para el reactivo (o respondente) individual y el puntaje total observado en el examen para el reactivo (o respondente). Los valores altos de la correlación biserial puntual PTBIS o de PTMEA CORR son un indicativo de que

el reactivo “trabaja en conjunto” con la escala a la que pertenece. Los valores negativos en PTBIS o PTMEA indican que el reactivo probablemente ha sido calificado mal debido a un error en la clave de respuestas (KEY). La diferencia específica entre PTBIS y PTMEA es que PTBIS pierde significado cuando hay datos faltantes en los patrones de respuesta. PTMEA conserva su significado en casos donde existen datos faltantes ya que las medidas Rasch que se obtienen se utilizan para reemplazar los valores totales en el cálculo de este tipo modificado de correlación punto-biserial.

EXACT OBS % representa el porcentaje de aciertos efectivamente observado

MATCH EXPECT % representa el porcentaje en el que cada reactivo (o cada respondente) cumplió con la expectativa (predicción) del modelo. Se observa que los porcentajes efectivamente observados están ligeramente por debajo de las expectativas del modelo en los reactivos de mayor dificultad. El porcentaje observado y esperado se acerca considerablemente en los reactivos de menor dificultad.

Finalmente la última columna de la tabla 13.1, indica el nombre de los reactivos en su orden de dificultad de acuerdo al etiquetado que se dio al final del archivo de control (AC).

Las secciones finales de la tabla 13.1 contienen estadísticas sumarias (Media y Desviación Estándar, ya comentadas) para cada rubro.

A continuación se presenta la **TABLA 17.1 PERSON MEASURE ORDER** (Personas en Orden de Medida) que contiene los estadígrafos de ajuste para cada respondente individual. La interpretación de INFIT y OUTFIT sigue los mismos criterios que cuando se trata de reactivos.

Debe notarse que los primeros 3 casos P38,.39 y 40 no tienen calibración completa ya que son casos que respondieron correctamente a los 20 reactivos y por no presentar variabilidad no aportan información útil al proceso.

Tabla 7

TABLE 17.1 Análisis de Examen Básico 1												ZOU724WS.TXT Apr 18 9:29 2007			
INPUT: 40 Alumnos 20 Reactivos												MEASURED: 40 Alumnos 20 Reactivos		2 CATS 3.63.0	
Alumno: REAL SEP.: .96 REL.: .48 ...												Reactivo: REAL SEP.: 1.54 REL.: .70			
Alumno STATISTICS: MEASURE ORDER															
ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT OBS%	MATCH EXP%	Alumno			
38	19	19	4.41	1.83	MAXIMUM ESTIMATED MEASURE								8		
39	19	19	4.41	1.83	MAXIMUM ESTIMATED MEASURE								9		
40	19	19	4.41	1.83	MAXIMUM ESTIMATED MEASURE								0		
35	18	19	3.18	1.05	1.00	.3	.63	.2	.21	94.7	94.8		5		
36	18	19	3.18	1.05	1.04	.3	.83	.3	.14	94.7	94.8		6		
37	18	19	3.18	1.05	.86	.1	.35	-.2	.38	94.7	94.8		7		
30	17	19	2.40	.77	.99	.2	.72	.0	.27	89.5	89.5		0		
31	17	19	2.40	.77	.84	-.1	.50	-.4	.43	89.5	89.5		1		
32	17	19	2.40	.77	1.02	.2	.73	.0	.25	89.5	89.5		2		
33	17	19	2.40	.77	1.05	.3	1.19	.5	.13	89.5	89.5		3		
34	17	19	2.40	.77	.95	.1	.61	-.2	.33	89.5	89.5		4		
26	16	18	2.30	.77	1.16	.5	3.61	2.1	-.16	88.9	88.9		6		
24	16	19	1.89	.66	.86	-.2	.59	-.5	.45	84.2	84.2		4		
25	16	19	1.89	.66	.90	-.1	.88	.1	.35	84.2	84.2		5		
27	16	19	1.89	.66	.98	.1	.68	-.3	.35	84.2	84.2		7		
28	16	19	1.89	.66	.93	.0	1.25	.6	.26	84.2	84.2		8		
29	16	19	1.89	.66	.95	.0	.67	-.3	.37	84.2	84.2		9		
15	16	20	1.61	.59	.96	.0	.73	-.4	.37	80.0	80.0		5		
16	16	20	1.61	.59	1.05	.3	1.02	.2	.23	80.0	80.0		6		
17	16	20	1.61	.59	1.02	.2	.81	-.2	.30	80.0	80.0		7		
18	16	20	1.61	.59	.94	-.1	.74	-.3	.38	80.0	80.0		8		
19	16	20	1.61	.59	1.02	.2	1.19	.5	.22	80.0	80.0		9		
20	15	19	1.51	.59	.79	-.6	.64	-.6	.52	78.9	78.9		0		
21	15	19	1.51	.59	1.09	.4	1.00	.2	.22	78.9	78.9		1		
22	15	19	1.51	.59	.84	-.4	.62	-.6	.50	78.9	78.9		2		
23	15	19	1.51	.59	.96	.0	.97	.1	.33	78.9	78.9		3		
13	15	20	1.29	.55	.78	-.8	.61	-.9	.56	80.0	75.3		3		
14	15	20	1.29	.55	1.00	.1	1.26	.7	.26	80.0	75.3		4		
10	14	20	1.01	.52	1.17	.8	1.20	.6	.16	65.0	71.2		0		
11	14	20	1.01	.52	1.26	1.2	1.18	.6	.09	55.0	71.2		1		
12	14	20	1.01	.52	.88	-.5	.73	-.7	.49	75.0	71.2		2		
5	13	20	.75	.50	1.22	1.2	1.17	.6	.15	55.0	68.2		5		
8	13	20	.75	.50	1.28	1.5	1.59	1.7	.02	55.0	68.2		8		
4	12	20	.50	.49	.75	-1.5	.66	-1.3	.64	75.0	66.8		4		
6	12	20	.50	.49	1.17	1.0	1.15	.6	.21	55.0	66.8		6		
7	12	20	.50	.49	.87	-.7	.79	-.8	.52	75.0	66.8		7		
9	12	20	.50	.49	1.04	.3	.93	-.2	.36	55.0	66.8		8		

3	9	20	-.21	.49	.92	-.4	.88	-.5	.50	75.0	68.3	3
2	8	20	-.46	.50	1.00	.1	1.07	.4	.39	75.0	69.4	2
1	7	20	-.71	.51	1.48	1.9	1.52	1.7	-.07	55.0	71.8	1

MEAN	15.1	19.5	1.71	.72	1.00	.1	.96	.1		78.0	79.3	
S.D.	2.7	.5	1.18	.35	.15	.6	.52	.7		11.9	8.7	

Los niveles de habilidad calibrados para este grupo de respondentes van desde 3.18 lógitos hasta -.71 lógitos. Predominan los casos por encima de la habilidad media de 0 lógitos, y hay solo 3 casos por debajo de 0 lógitos. En este caso debe notarse que los errores estándar de los estimados son más altos que si se tratara de reactivos.

Opciones Básicas de Programación

Un análisis inicial puede repetirse con las opciones de eliminar de las corridas subsecuentes aquellos reactivos o personas que presenten desajustes considerables. Se anotan aparte aquellos reactivos o personas que presenten desajustes considerables respecto a las expectativas del modelo. Para eliminar reactivos o personas existen dos opciones que modificarán el inicio de la corrida del programa.

IDELQ = YES

IDELQ (Ítem Deletion Question) el programa preguntará al usuario si desea eliminar reactivos en la siguiente corrida. En este caso el usuario responderá con los números de reactivo que desea eliminar seguidos de ENTER y finalizará la serie de preguntas con un 0 y ENTER.

PDELQ = YES

PDELQ (Person Deletion Question) el programa preguntará al usuario si desea eliminar personas en la siguiente corrida. En este caso el usuario responderá YES a las preguntas en pantalla de eliminar reactivos y /o personas y en seguida ingresará uno por uno los números identificadores de reactivo y/o persona que desea eliminar seguidos de ENTER y finalizara la serie de reactivos o personas con un 0 y ENTER.

```

MINISTEP Version 3.63.0 Apr 23 11:18 2007
Current Directory: C:\WINSTEPS\examples\
Control file name? (e.g., exam1.txt). Press Enter for Dialog Box:
Previous Directory: C:\WINSTEPS\examples
Current Directory: C:\Documents and Settings\Becario\My Documents\Act1-UEE\Presentaciones MGM\Practicas Rasch 1
C:\Documents and Settings\Becario\My Documents\Act1-UEE\Presentaciones MGM\Practicas Rasch 1\Examen1.TXT
Extra specifications (or press Enter):
Temporary Workfile Directory: C:\TEMP\
Reading Control Variables ..
DO YOU WANT TO READ SOME Reactiv TO DELETE FROM A FILE?
DO YOU WANT TO TYPE IN SOME Reactiv TO DELETE ?
yes
Processing Reactiv Deletions from: Keyboard
INPUT Reacti TO DELETE (0 TO END):

```

Estas opciones pueden añadirse al archivo de control (AC) antes de &END de la siguiente manera:

Ejemplo 6

```

&INST
TITLE = "Análisis de Examen Básico"
NI = 20           ; 20 reactivos
ITEM1 = 1        ; los datos de reactivos inician en la columna 1
NAME1=21         ; el identificador inicia en la columna 21
NAMELEN = 2      ; el identificador de respondentes es de 2 dígitos
XWIDE = 1        ; cada campo de datos es de 1 columna
ITEM = Reactivo  ; los ítems se llaman Reactivo
PERSON = Alumno  ; las personas se llaman Alumno
IDELQ = YES      ; pregunta si se eliminan reactivos de la corrida
PDELQ = YES      ; pregunta si se eliminan personas de la corrida
&END

```

El usuario advertirá que al repetir los análisis eliminando reactivos y/o personas que presenten desajustes considerables, los índices de ajuste de los reactivos y personas restantes en el análisis. Para estos casos, lo recomendable es eliminar primero aquellos ítems o personas que presenten los mayores desajustes. Es decir aquellos desajustes que claramente se salgan los umbrales críticos (ver pp. 46-48 arriba). Los desajustes menores a 1 son de considerarse solo cuando se está tratando de reducir la longitud de un examen o instrumento. (Linacre, 2006. P.169)

Opciones Especiales de Programación

LOCAL = YES

LOCAL efectúa una estandarización empírica, es decir basada en la distribución realmente observada localmente en la muestra sobre la que se calcula. En estos casos los rubros de INFIT y OUTFIT estandarizados aparecerán en la información de salida con el encabezado ZEMP (Z Empírica).

Esta opción se aplica en los siguientes casos

Los estadígrafos estandarizados someten a prueba la siguiente hipótesis:

¿Se ajustan los datos observados al MODELO Rasch?

Bajo condiciones de tamaños de muestra grandes se aumenta la potencia estadística (power) de la prueba y la hipótesis NUNCA puede ser confirmada ya que los datos empíricos siempre tienen algún grado de desajuste al modelo. Esto puede hacer que los estadígrafos de ajuste aumenten hasta ser no interpretables. Por ello los estadígrafos de INFIT ZEMP y OUTFIT ZEMP se ajustan a la muestra local.

Con esta opción activada, la escala de los estadígrafos de ajuste se modifica en el contexto (local) de la cantidad de “distorsión” en las base de datos bajo análisis. En los valores ZEMP los estadígrafos INFIT y OUTFIT se dividen entre su desviación estándar, así con el reescalamiento la Desviación Estándar es $SD = 1.0$

La escala ZEMP (Z Empírica) produce estadígrafos que se interpretan como sigue:

¿Qué tan no- factible es esta cantidad de desajuste en el contexto del patrón general de desajuste en los datos empíricamente observados?

DISCRIMINATION = YES

El modelo Rasch afirma que los ítems (ideal) tienen una expectativa de discriminación de 1.0 (perfecta distinción entre respondentes de alta y baja habilidad en el rasgo medido). Pero los índices de discriminación empíricos tienden a variar Por ello se recomienda DISCRIM = YES.

DISCR con valor = 1.0 indica que el ítem cumple la expectativa del modelo monoparamétrico.

Valores superiores a 1.0 indican sobre-discriminación e inferiores a 1.0 sub-discriminación. En el análisis TRR de ítems, la sobre-discriminación es una característica deseable.

PTBIS = Y

Calcula y reporta el coeficiente (de Ítems y Personas) biserial puntual r_{pbis} que en el análisis Rasch es un indicador útil para diagnosticar errores en la codificación de ítems o de claves. Valores negativos o en 0 indican ítems o personas con patrones de respuesta que contradicen la variable. Esta opción no debe emplearse si la base de datos tiene numerosos valores faltantes ya que la correlación biserial puntual pierde su significado cuando hay datos faltantes.

En caso de que la base de datos tenga datos faltantes puede usarse la alternativa:

PTBIS = RPM

Calcula y reporta coeficientes punto medida r_{pm} (Ítems o Personas). Estos indican la asociación de las respuestas a ítems (o de personas) con las respectivas medidas en lógitos. r_{pm} retiene mejor su significado en los casos en donde hay datos faltantes numerosos. Valores negativos o en 0 indican ítems o personas que contradicen la variable. Este estadígrafo es igual al **PTMEA CORR** comentado arriba (p.48).

Si se emplean las opciones especiales descritas el archivo de control queda de la siguiente forma:

Ejemplo 7

```

&INST
TITLE = "Análisis de Examen Basico"
NI = 20           ; 20 reactivos
ITEM1 = 1        ; los datos de reactivos inician en la columna 1
NAME1=21        ; el identificador inicia en la columna 21
NAMELEN = 2     ; el identificador de respondentes es de 2 dígitos
XWIDE = 1       ; cada campo de datos es de 1 columna
ITEM = Reactivo ; los ítems se llaman Reactivo
PERSON = Alumno ; las personas se llaman Alumno
IDELQ = YES     ; pregunta si se eliminan reactivos de la corrida
PDELQ = YES     ; pregunta si se eliminan personas de la corrida
LOCAL = YES     ; estandarización empírica ZEMP en estad. de ajuste
DISCR = YES     ; reportar índice de discriminación
PTBIS = YES     ; reportar la correlación biserial puntual
& END

```

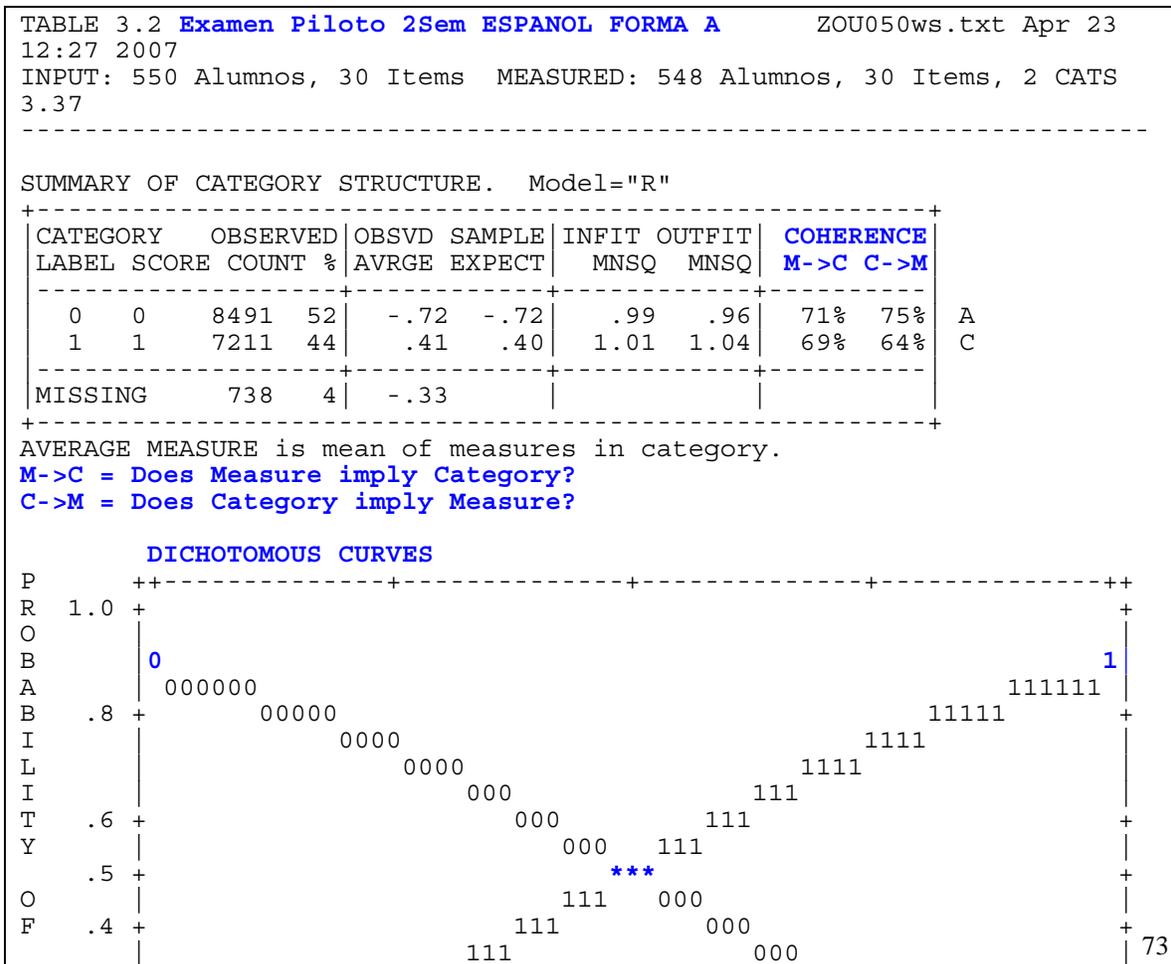
Graficas en Winsteps / Ministep

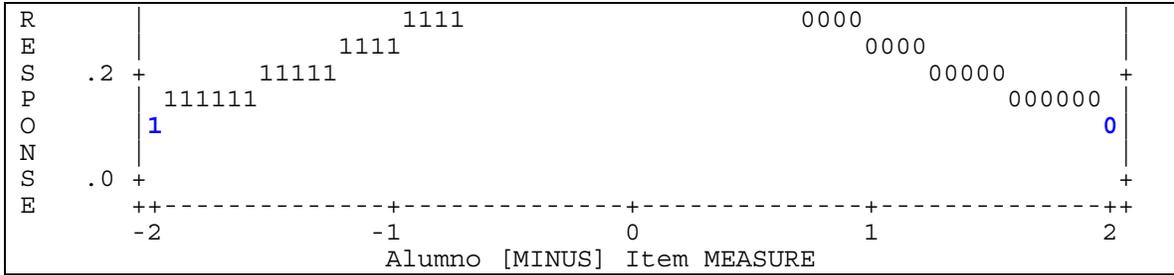
El usuario podrá constatar que la información en forma tabular que el programa ofrece es abundante. En la sección anterior se ha detallado la interpretación de las principales tablas. 3.1, (Sumarias) 13.1(Ítems), y 17.1 (Personas). En esta sección se describen las principales graficas que el programa produce así como su interpretación.

Tabla 3.2

Esta grafica aparece bajo el menú de OUTPUT TABLES. Sumariza el comportamiento de las categorías de respuesta. En este caso hay dos categorías reportadas 0 y 1. Con los porcentajes empíricamente observados para cada una así como los porcentajes esperados conforme al modelo. Presenta también los promedios de INFIT y OUTFIT de cada categoría. Es importante notar que toda esta tabla indica como trabajó la escala en esta muestra en particular.

Tabla 8





La columna M -> C presenta la coherencia de la escala y la muestra expresada por el porcentaje real de las medidas que se esperaba que produjeran observaciones en esta categoría y que de hecho lo hicieron. Es decir, la coherencia como porcentaje real de medidas que efectivamente implican la categoría en esta muestra.

La columna C -> M muestra el porcentaje de las observaciones que fueron producidas por medidas que corresponden a la categoría. Es el grado de coherencia entre la categoría que implica a la medida.

Ambas columnas deben tomarse como un indicador global de la adecuación entre la muestra y las expectativas del modelo.

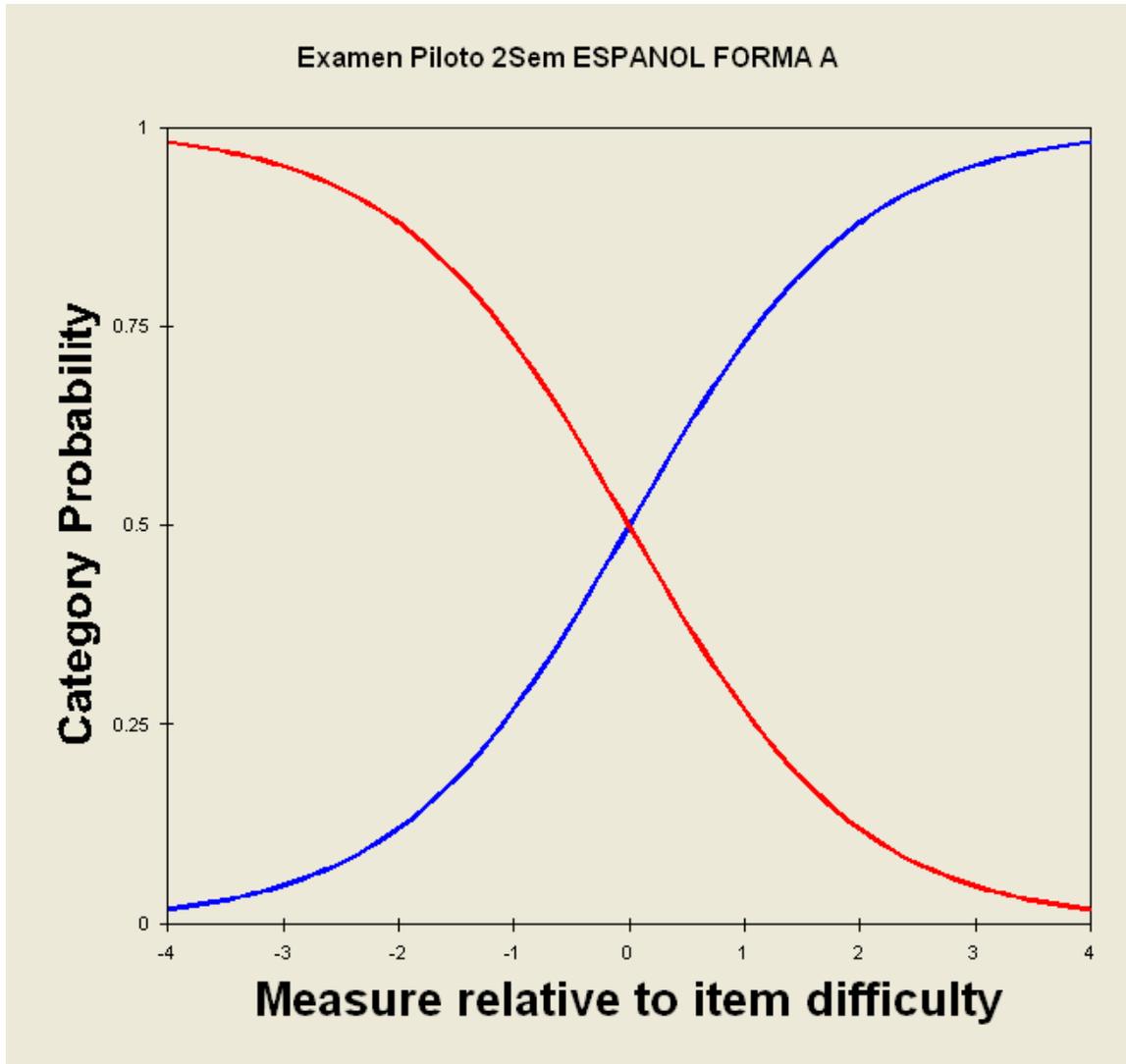
Las curvas dicotómicas de probabilidad presentan una predicción de cómo se espera que la escala trabaje en muestras similares en el futuro. Para este caso, se espera que en muestras equivalentes tomadas de la misma población de respondentes la categoría 0 (respuesta incorrecta) tendrá una probabilidad de .90 a -2 lógitos de habilidad calibrada. En consecuencia la categoría 1 (respuesta correcta) tendrá una probabilidad de .90 a +2 lógitos de habilidad calibrada.

De manera inversa a -2 lógitos de habilidad la respuesta correcta (1) tiene .10 de probabilidad. La respuesta incorrecta tiene .10 de probabilidad a +2 lógitos de habilidad.

En 0 lógitos (la habilidad media) las categorías 0 y 1 tienen una probabilidad de .50 como lo predicen los comportamientos característicos de los reactivos de acuerdo al modelo.

Esta grafica también se obtiene como la primera opción bajo el sub-menú de GRAPHS y toma la siguiente forma:

Gráfica 1

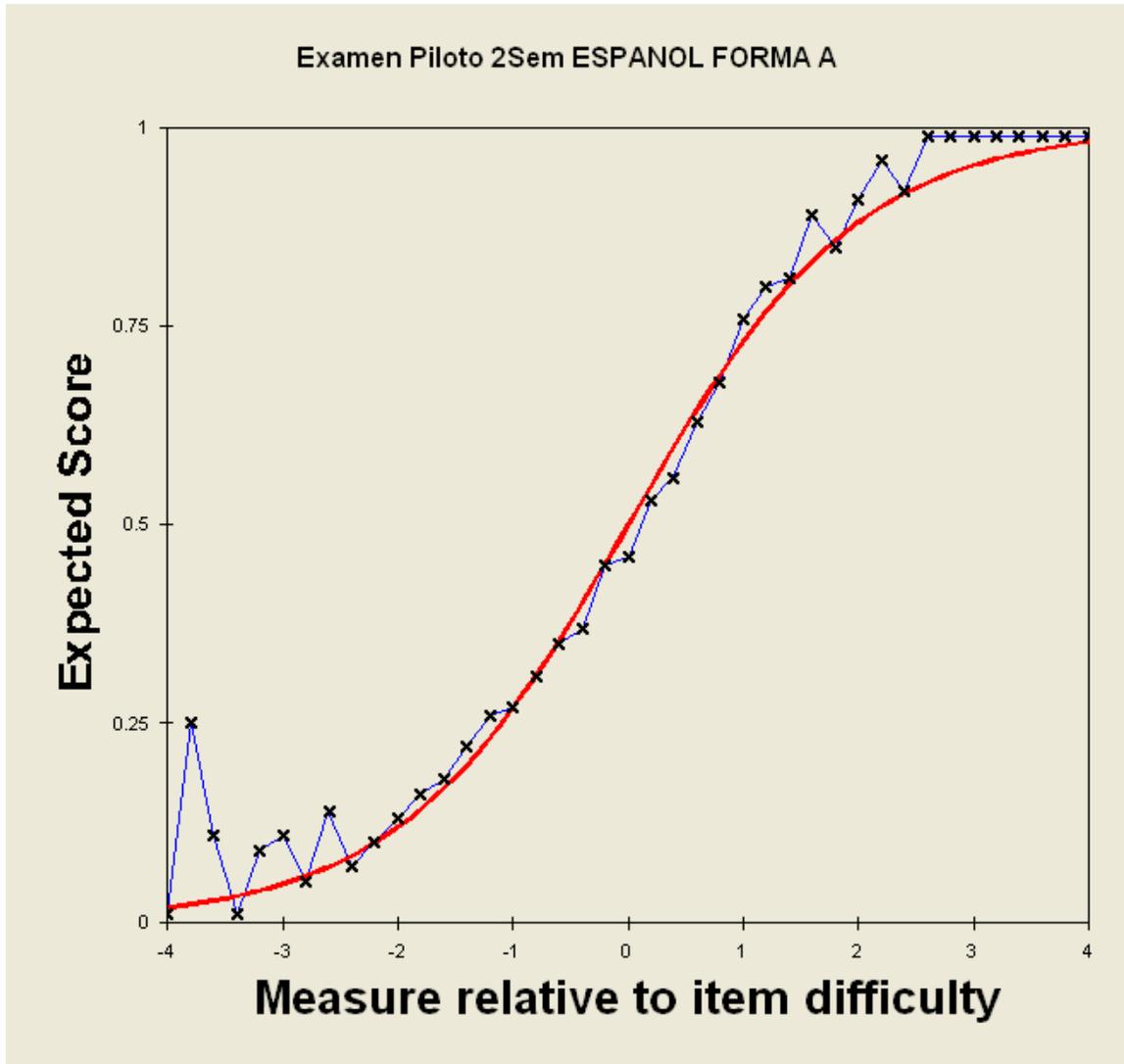


En esta se puede apreciar la misma categorización donde la línea roja representa la probabilidad de respuesta incorrecta decreciente conforme aumenta la habilidad de los respondentes MEASURE relativa a la dificultad de los reactivos ambas expresadas en lógitos. La línea azul representa la probabilidad de respuesta correcta creciente conforme aumenta la habilidad en lógitos En esta grafica se amplía la perspectiva para considerar habilidades de -4 hasta + lógitos aun cuando estas calibraciones no hayan sido empíricamente

observadas en la muestra particular de calibración.

La Grafica CCR Empírica

Grafica 2



Esta gráfica se obtiene bajo la segunda opción del sub-menú GRAPH5. Representa la congruencia entre los patrones de respuesta observados en esta muestra y las probabilidades de respuesta esperadas según el modelo. En

este caso, se observa que los patrones de respuestas coinciden con las expectativas del modelo (expresadas como probabilidades definidas) entre -2.1 y +2.1 lógitos de habilidad efectivamente calibrada. Esta grafica es una representación de cómo efectivamente trabajo la escala en esta muestra y permite anticipar como trabajará en aplicaciones futuras en situaciones donde se presenten las medidas de habilidad calibradas y relativas a la dificultad de reactivos.

Ejercicio de Integración

Para lograr el propósito práctico de esta sección se recomienda al lector realizar el siguiente ejercicio:

- 1) Preparar un archivo de control (AC) en **Ministep** para analizar el siguiente archivo de datos (AD):

Ejemplo 8

```

10000111000100AA
00100011100011AB
111111110100000AC
00010000100000AD
00000001010001AE
00000010000010AF
11011010000001AG
00101000000100AH
00000111101100AI
10101101100100AJ
01111001011000AK
00000111000000AL
11111001101011AM
01000011011100AN
01000001100101AO
01011011100000AP
11101000000000AQ
10000111000101AR
00011101000000AS
00111000000100AT
10100111000100AU
10000000110111AV
11001010000001AW
00101000001101AX
00100001101011AY
00000000100001AZ
10001110100001BA

```

```

00100101001000BB
11101110101100BC
01101001001100BD
10000000110111BE
00111010000000BF
01001000000100BG
01001111011111BH
11011001010001BI
11101001000010BJ
01000100100110BK
01101010111001BL
01011111100011BM
11101001100100BN
00100101011100BO
00101110011101BP
00000011000011BQ
11101111110111BR
01000010001000BS
00001000100000BT
00100000000010BU
00101110000000BV
00001000000001BW
11011101110111BX

```

- 2) Como puede observarse se trata de un arreglo de 14 reactivos y 50 respondentes. Las columnas 1 a 14 contienen los patrones de respuesta en binario (1= correcto. 0 = incorrecto) por lo que no es necesario incluir una clave de respuestas en el AC. Los respondentes se identifican por un código de dos caracteres AA....etc. En este caso es posible combinar en un solo archivo AC las especificaciones de control con el archivo de datos al final como se describió en el Ejemplo 2 (Pág. 27). Se sugiere guardar el archivo de control con un nombre como "Practica2.txt" o similar.
- 3) Para verificar su práctica de programación el lector debe correr el análisis en **Ministep** y comparar sus resultados con la siguiente información de salida:

Tabla 9

TABLE 3.1 Análisis de Reactivos Practica 2								
INPUT: 50 Alumnos, 14 Reactivos					MEASURED: 50 Alumnos, 14 Reactivos, 2 CATS 3.37			
SUMMARY OF 50 MEASURED Alumnos								
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	5.5	14.0	-.50	.61	1.00	.0	1.00	.0
S.D.	2.4	.0	.85	.07	.10	.6	.18	.7
MAX.	12.0	14.0	1.88	.78	1.26	1.7	1.61	1.6
MIN.	2.0	14.0	-1.87	.55	.79	-1.7	.60	-1.7
REAL RMSE	.63	ADJ.SD	.57	SEPARATION	.91	Alumno	RELIABILITY	.45
MODEL RMSE	.61	ADJ.SD	.59	SEPARATION	.95	Alumno	RELIABILITY	.48
S.E. OF Alumno MEAN = .12								
Alumno RAW SCORE-TO-MEASURE CORRELATION = 1.00								
CRONBACH ALPHA (KR-20) Alumno RAW SCORE RELIABILITY = .46								
SUMMARY OF 14 MEASURED Reactivos								
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	19.8	50.0	.00	.32	1.00	.0	1.00	-.1
S.D.	5.0	.0	.50	.02	.07	.6	.13	.7
MAX.	29.0	50.0	.81	.35	1.10	1.0	1.22	.9
MIN.	12.0	50.0	-.88	.30	.87	-1.3	.82	-1.4
REAL RMSE	.32	ADJ.SD	.38	SEPARATION	1.19	Reacti	RELIABILITY	.59
MODEL RMSE	.32	ADJ.SD	.39	SEPARATION	1.22	Reacti	RELIABILITY	.60
S.E. OF Reactivo MEAN = .14								

Interpretación Práctica

Una vez obtenida la Tabla de Estadísticas Sumarias 3.1, el lector puede ensayar la interpretación con los siguientes lineamientos:

- 1) Los **datos sumarios** para Personas (Alumno) se presentan primero omitiendo los casos extremos. Es decir, presentan los casos NON-EXTREME, y se descartan aquellos respondentes que contestaron correctamente a todos los ítems o no contestaron ningún ítem correctamente. **En este análisis no se detectaron casos extremos.**
- 2) **RAW SCORE** es el numero de respuesta correctas observadas y en

este caso como se trata de estadísticas sumarias se presentan la media (MEAN) que es de **5.5** para el promedio de calificación de alumnos. la Desviación Estándar (S. D.), que es de **2.4** el puntaje de aciertos máximo MAX que es **12** y el mínimo MIN que es **2**.

- 3) COUNT** es simplemente el número total de respuestas registradas **14** ítems de 14 posibles fueron intentados por los 50 respondentes.
- 4) MEASURE** es la medida de habilidad en lógitos estimada para los respondentes y de calibración de dificultad para reactivos. **-.50** lógitos indican que en promedio los reactivos están $\frac{1}{2}$ lógito por debajo de la media de 0 lógitos la desviación estándar (SD) en lógitos es de .85. LE máximo de habilidad de los respondentes es de **1.88** lógitos y el mínimo es de **-1.87**, lo que muestra que este grupo de respondentes **no excede** del rango de **-2 a +2** lógitos.
- 5) MODEL ERROR** el error estándar asociado a cada estimación de medida. Para el caso del promedio de habilidad en lógitos el error es de **.61**. Debe notarse que esta fluctuación es mayor que el propio promedio de medida **-.50**. Esto se debe a que en este análisis la estimación está más propensa a la fluctuación con respecto al MODELO debido al tamaño de muestra que es de $n=50$.

- 6) **INFIT** es el estadígrafo de ajuste interno ponderado por la información y que es más sensible a comportamientos inesperados (según la expectativa del Modelo) que afectan respuestas a ítems cercanos al nivel de habilidad medida de las personas.
- 7) **MSNQ** es el estadígrafo de media cuadrática de ajuste interno que según el modelo tiene una expectativa de 1. Los valores sustancialmente menores a 1 indican dependencia (determinismo) en los datos observados. Los valores sustancialmente mayores a 1 indican ruido (o demasiada estocasticidad) en los datos observados. En este análisis el valor promedio de **INFIT es de 1.0**. Lo que indica que **en promedio** la muestra tiene buen ajuste al modelo. Hay un máximo en promedio de **1.26** con un mínimo de **.79**.
- 8) **ZSTD** es el mismo estadígrafo de media cuadrática de ajuste interno pero estandarizado para aproximar una distribución con media teórica de 0 y una varianza de 1. Existen opciones para modificar esta estandarización. LOCAL = Y, y LOCAL =L cuyo uso e interpretación se describirán en la sección dedicada a opciones especiales. En este análisis el **promedio de INFIT** estandarizado es de **0**. Lo cual es consistente con el **MSNQ de 1**. Nótese que en este análisis no se empleó la estandarización empírica ZEMP debido al tamaño de la muestra.

- 9) OUTFIT** es el estadígrafo de ajuste externo sensible a comportamientos inesperados (según el Modelo) que afecta respuestas a ítems lejanos al nivel de habilidad medida de las personas. El MSNQ de OUTFIT es en promedio de **1.0** con un máximo de **1.61** y un mínimo de **.60**. El valor de 1.61 es tal que rebasa el límite máximo de 1.50. Este resultado indica que hay algunos respondentes que exceden la expectativa del modelo.
- 10)ZSTD** es el mismo estadígrafo de media cuadrática de ajuste externo pero estandarizado para aproximar una distribución con media teórica de 0 y una varianza de 1. En este caso el valor promedio de OUTFIT es **de1.00**
- 11)LOCAL = Y**, y **LOCAL =L** cuyo uso e interpretación se aplica cuando se tienen tamaños de muestra grandes. En este análisis no se utilizó.
- 12)RMSE** es la raíz del error cuadrático medio calculado sobre todas las personas o sobre todos los ítems presentes en el archivo de datos. El error cuadrático medio (Mean Square Error o MSE) es un indicador del error total (variación aleatoria) en una muestra. Incluye las fuentes de error totales, es decir variación aleatoria atribuible al muestreo y variación aleatoria no atribuible al muestreo. Se interpreta como la **desviación estándar del estimador** con respecto al parámetro que se trata de aproximar. De nuevo se advierte que este valor es alto, .63 debido al tamaño de muestra reducido.

13) REAL RMSE es la raíz del error cuadrático medio, pero calculado sobre la base de que los desajustes en los datos se deben desviaciones de los datos hacia el modelo. Este valor representa “el peor caso” de la confiabilidad y se reporta como el límite inferior de los estimados de confiabilidad obtenidos a partir de una muestra individual de datos observados. De nuevo se advierte que este valor es alto, .63 debido al tamaño de muestra reducido.

14) MODEL RMSE es también la raíz del error cuadrático medio calculado sobre la base de los datos que se ajustan al modelo. Este cálculo asume que todo desajuste en los datos es un reflejo de la naturaleza estocástica del modelo. Este valor representa “el mejor caso” de la confiabilidad y se reporta como el límite superior de los estimados de confiabilidad obtenidos a partir de una muestra individual de datos observados. En este caso es de .61, lo cual es un valor alto que se debe al reducido tamaño de muestra.

15) ADJ. S.D. es la desviación estándar de los estimados ajustada después de sustraer su varianza de error de su varianza observada. La varianza de error de los estimados es atribuible al error estándar de medición. Así:

$$a. (\text{ADJ. SD})^2 = (\text{S.D. Medida})^2 - (\text{RMSE})^2$$

16) El valor de **ADJ. S.D.** de un estimado de la desviación estándar “verdadera” al que se le ha removido el sesgo derivado de error de

medición. Para este análisis es de **.57** en promedio.

17)SEPARATION (para PERSONAS o ITEMS) es la razón de **ADJ. S.D. a RMSE**. Proporciona una razón medida de separación (entre personas o ítems) en unidades de RMSE y es más fácil de interpretar que la correlación de confiabilidad. Este valor es de .91 para respondentes en este análisis.

18)PERSON (ITEM) SEP REL es una forma de separación (confiabilidad) de las medidas, equivalente a KR-20, el Alfa de Cronbach y el Coeficiente de Generalizabilidad. En este caso el índice es de .91. La relación entre separación SEP y confiabilidad REL es:

$$\text{a. } \text{REL} = \text{SEP}^2 / (1 + \text{SEP})^2$$

19) O bien

$$\text{a. } \text{SEP} = (\text{REL}/(1-\text{REL}))^{1/2}$$

i. Donde: SEPARACION = SD. Verdadera / Error RMSE

ii. CONFIABILIDAD = Varianza Verdadera / Varianza
Observada

20)S.E. OF MEAN es el error estándar de la media de la persona o del ítem, en este caso .12 en promedio.

21)RELIABILITY es el índice convencional de confiabilidad ALFA, KR-20 y representa una aproximación a la confiabilidad del examen / escala basado en los puntajes crudos de la muestra observada. Este índice es muy bajo para el conjunto de respondentes .45 y .60 para reactivos en este análisis debido al reducido tamaño de muestra.

22)RAW SCORE TO MEASURE CORRELATION (de Personas o Ítems) es el coeficiente de correlación Pearson entre los puntajes crudos y las medidas (en lógitos) que incluye los puntajes extremos. Cuando los datos están completos (es decir cuando no hay valores faltantes) estos coeficientes deben estar cerca de 1.0 para personas y -1.0 para ítems. Este supuesto se cumplió en este caso, ya que el valor de este índice es de 1.0

Tabla 10

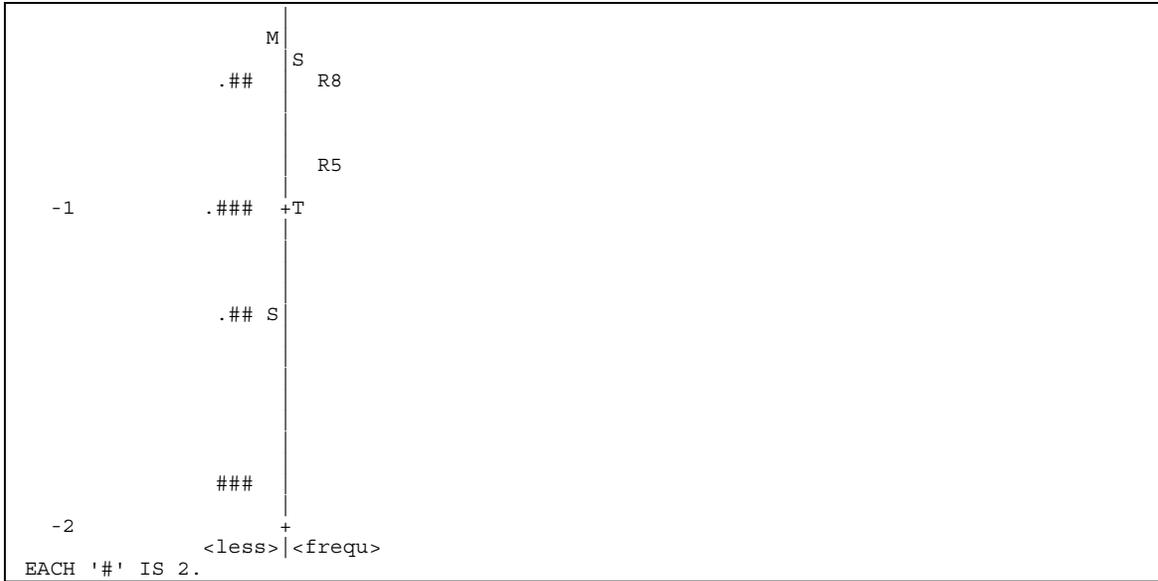
TABLE 13.1 Análisis de Reactivos Practica 2												May 18 10:37 2007	
INPUT: 50 Alumnos, 14 Reactivos												MEASURED: 50 Alumnos, 14 Reactivos, 2 CATS 3.37	
Alumno: REAL SEP.: .91 REL.: .45 ... Reactivo: REAL SEP.: 1.19 REL.: .59													
Reactivos STATISTICS: MEASURE ORDER													
ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT		OUTFIT		PTBIS	DISCR	React		
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.				
4	12	50	.81	.35	1.09	.5	1.22	.8	.05	.83	R4		
10	12	50	.81	.35	.90	-.5	.83	-.7	.30	1.15	R10		
11	14	50	.57	.34	1.06	.4	1.01	.0	.12	.91	R11		
13	14	50	.57	.34	1.00	.0	1.18	.8	.14	.93	R13		
1	18	50	.15	.32	.91	-.7	.82	-1.1	.30	1.28	R1		
6	19	50	.05	.31	.97	-.3	.92	-.5	.22	1.13	R6		
9	21	50	-.14	.31	.99	-.1	1.06	.4	.19	.97	R9		
2	22	50	-.23	.31	.87	-1.3	.82	-1.4	.35	1.50	R2		
12	22	50	-.23	.31	1.02	.2	.98	-.2	.14	.96	R12		
14	22	50	-.23	.31	1.02	.2	1.03	.2	.16	.93	R14		
3	23	50	-.33	.30	1.10	1.0	1.13	.9	.04	.59	R3		
7	23	50	-.33	.30	1.10	1.0	1.13	.9	.04	.58	R7		
8	26	50	-.60	.30	.96	-.4	.92	-.6	.20	1.19	R8		
5	29	50	-.88	.31	1.01	.1	.94	-.4	.18	1.03	R5		
MEAN	20.	50.	.00	.32	1.00	.0	1.00	-.1					
S.D.	5.	0.	.50	.02	.07	.6	.13	.7					

En la tabla anterior referida a ítems es notable el hecho de que no aparecen casos que excedan los valores críticos de .80 a 1.30 en MNSQ INFIT y MNSQ OUTFIT. Además, se advierte en los criterios los valores para todos los ítems se comportan dentro de la banda de .50 a 1.50 lo cual los sitúa en una condición óptima y productiva para el proceso de medición. Es importante notar que si bien en este conjunto de 14 reactivos predomina la facilidad con un rango de -.88 a +.81 lógitos, todos los reactivos se comportan dentro de la expectativa del Modelo Rasch.

Por lo que respecta a la siguiente Tabla 11, el lector podrá advertir si bien las habilidades de los respondentes son bajas ya que no exceden el rango -2 a +2 lógitos, los respondentes se comportan dentro de la expectativa del modelo. El caso del **respondente 4** es la única excepción ya que presenta un valor de **MNSQ OUTFIT de 1.61**.

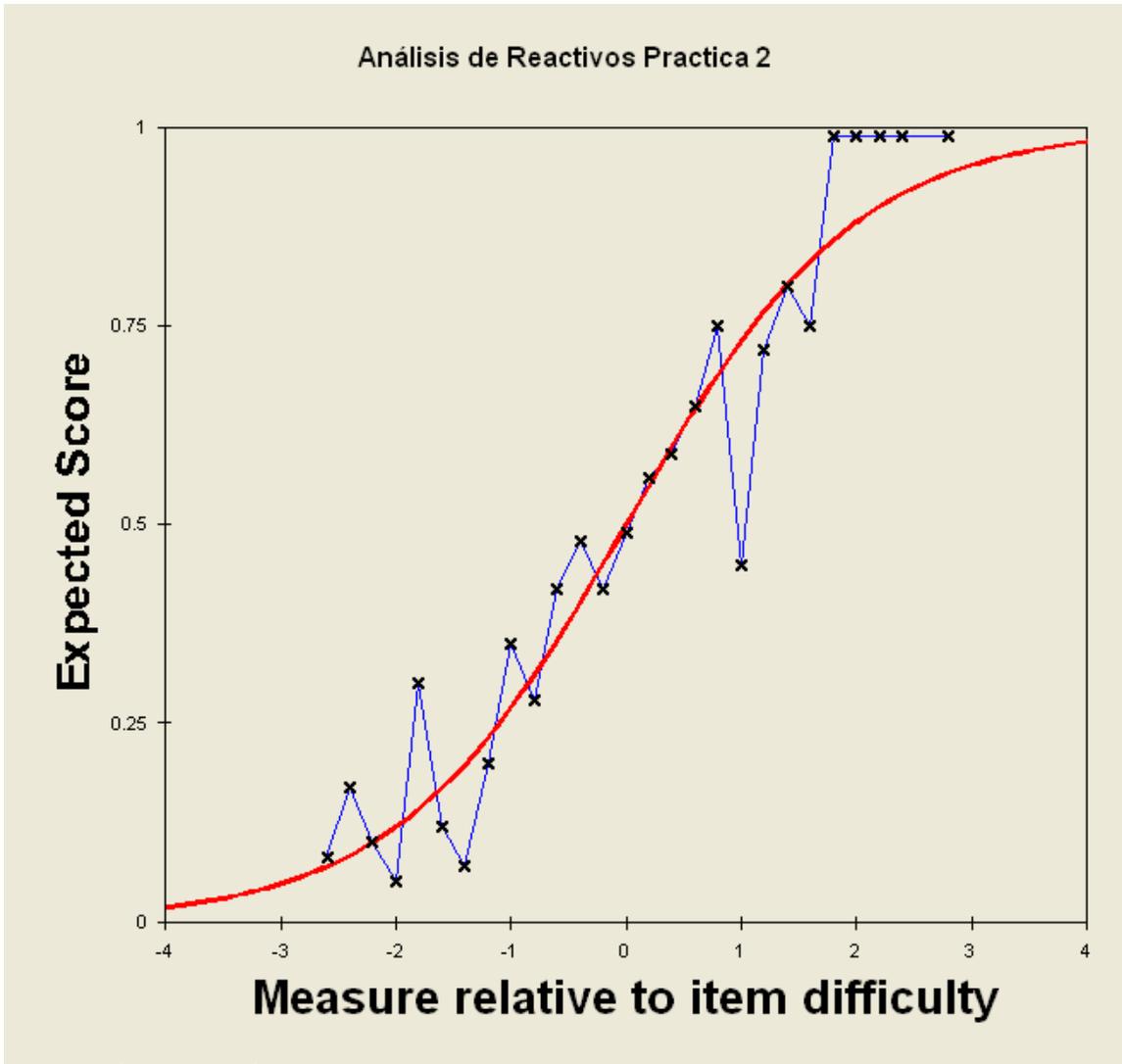
Tabla 11

TABLE 17.1 Análisis de Reactivos Practica 2											May 18 10:37 2007
INPUT: 50 Alumnos, 14 Reactivos											MEASURED: 50 Alumnos, 14 Reactivos, 2 CATS 3.37
Alumno: REAL SEP.: .91 REL.: .45 ... Reactivo: REAL SEP.: 1.19 REL.: .59											
Alumno STATISTICS: MEASURE ORDER											
ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT		OUTFIT		PTBIS CORR.	AI	
					MNSQ	ZSTD	MNSQ	ZSTD			
44	12	14	1.88	.78	.84	-.3	.60	-.7	.50	BR	
50	11	14	1.37	.67	1.11	.3	1.16	.4	-.10	BX	
13	10	14	.97	.61	1.07	.3	1.06	.2	.01	AM	
34	10	14	.97	.61	1.04	.2	1.02	.1	.07	BH	
29	9	14	.62	.57	.90	-.5	.92	-.4	.30	BC	
39	9	14	.62	.57	.98	-.1	.95	-.2	.20	BM	
3	8	14	.30	.56	.99	-.1	.99	.0	.16	AC	
38	8	14	.30	.56	1.01	.1	1.01	.1	.13	BL	
42	8	14	.30	.56	1.04	.3	1.04	.2	.07	BP	
10	7	14	.00	.55	.84	-1.3	.82	-1.4	.48	AJ	
11	7	14	.00	.55	1.15	1.1	1.16	1.1	-.11	AK	
35	7	14	.00	.55	1.10	.7	1.10	.7	-.03	BI	
40	7	14	.00	.55	.79	-1.7	.78	-1.7	.58	BN	
2	6	14	-.31	.56	.97	-.2	.96	-.3	.20	AB	
7	6	14	-.31	.56	1.01	.1	1.02	.1	.11	AG	
9	6	14	-.31	.56	1.03	.2	1.01	.1	.08	AI	
14	6	14	-.31	.56	1.11	.8	1.14	.8	-.09	AN	



En la Tabla 12 anterior se confirma que la alineación de respondentes y reactivos es muy dispersa y escasa prevaleciendo un número importante de respondentes con habilidad por debajo de las dificultades de los reactivos.

Figura 3



Para concluir este ensayo de interpretación, la Figura 3 nos indica claramente que el comportamiento de las habilidades medidas con respecto a la dificultad de los reactivos en esta muestra de respondentes traza algunas anomalías en casos fuera del rango de -1 a +1 lógitos. Este resultado se debe en particular a las imprecisiones de estimación debido al reducido tamaño de muestra y los escasos grados de habilidad detectados en el análisis con un rango efectivo de -1.87 a +1.88.

Con base en los lineamientos descritos y con la información de salida a la mano, en las Tablas 10 y 11, el lector puede avanzar a la interpretación individual de las calibraciones de dificultad de los reactivos y las medidas de habilidad de los respondentes. En seguida pueden interpretarse los valores de INFIT y OUTFIT obtenidos tanto para los reactivos como para los respondentes. Aplicando los criterios descritos en las páginas 47 y 48, es posible analizar el comportamiento de cada reactivo y de cada respondente respecto a las expectativas del modelo Rasch.

En esta Práctica 2 puede concluirse que la distribución de reactivos y respondente muestra algunas áreas de dificultad y de habilidad que no coinciden y que en general la medida de habilidad de los respondentes tiende a la baja en la escala de lógitos. También puede observarse a partir de la figura 3 que la probabilidad de respuesta correcta decrece considerablemente en varios niveles de la escala de lógitos. Se invita al lector para que con base en su experiencia, ensaye posibles explicaciones para dar cuenta de estas características de los

reactivos y respondentes según los datos observados.

Casos Especiales Análisis Rasch con Ítems Politómicos

Hasta la sección anterior se procedió a las aplicaciones del modelo Rasch a reactivos dicotómicos. En este apartado se destaca que esta metodología no se limita solo a reactivos dicotómicos sino que su aplicación se ha extendido a reactivos politómicos como es el caso de las escalas de respuesta graduada tipo Likert.

Se han desarrollado extensiones del modelo Rasch para analizar reactivos de tipo rating bajo el Rating Scale Model (RSM) de Andrich. También se ha desarrollado un modelo denominado Partial Credit Model (PCM) de Masters. La diferencia entre estos radica en que bajo RSM se asume que las distancias entre las categorías de los reactivos rating son uniformes de reactivo a reactivo. Bajo PCM las distancias entre las categorías al interior de los reactivos no son uniformes de reactivo a reactivo. En este caso cada reactivo tiene su propia gradación única e independiente de los demás reactivos en el instrumento. En ambos casos es posible utilizar la técnica Rasch para analizar el ajuste de grupos de reactivos a un modelo unidimensional que captura una variable latente medida por la escala bajo análisis.

Es de suma importancia notar que para todos los casos de reactivos de respuesta graduada o tipo Likert, los conceptos de facilidad o dificultad de reactivo se convierte en la disposición de los respondentes a estar de acuerdo o no de acuerdo con una afirmación (aceptabilidad o no-aceptabilidad). O bien estos conceptos se pueden trasladar a la disposición de los respondentes

para emitir una opinión favorable (facilidad) o desfavorable (dificultad) ante una afirmación o situación dada.

Para aquellos casos en que se pueda sostener y justificar la uniformidad de distancias de las categorías graduadas entre reactivos el modelo RSM es:

Caso 1 Estructura RSM “Rating Scale Model” (Andrich)

Se parte de supuesto: de todos los ítems comparten la misma estructura de escala rating en cuanto a distancias entre categorías.

El Modelo Politémico es:

$$\text{Log}(P_{nij}/P_{ni(j-1)}) = B_n - D_i - F_j$$

Donde:

P_{nij} = Probabilidad de que la persona n en su encuentro con el ítem i emita su respuesta en la categoría j .

B_n = La medida de “habilidad” (afinidad) de la persona n

D_j = La medida de “dificultad” (aceptabilidad) del ítem i (MEASURE)

F_j = La medida de calibración de la categoría j relativa a la categoría $j-1$

Trasladando a estos casos los expuestos en secciones previas tenemos que

para evaluar la bondad de ajuste de los datos al modelo umbrales críticos de

INFIT son:

- $\text{INFIT} > 2.0$ el ruido aleatorio (fuera de la variable de interés) es mayor que la información útil que aporta el reactivo. Este efecto indeseable (debido a características particulares del reactivo o del respondente) degrada el proceso de medición.
- $\text{INFIT} > 1.5$ el ruido aleatorio (fuera de la variable de interés) es notable. Esta situación no aporta al proceso de medición pero tampoco lo degrada.

- INFIT de 0.5 a 1.5 situación PRODUCTIVA (optima) para el proceso de medición.
- INFIT < 0.5 esta situación ajusta demasiado bien (es decir es demasiado predecible). Es confundente en el sentido de que aparenta que se esta midiendo muy bien el constructo de interés cuando de hecho no es el caso.

Para evaluar la bondad de ajuste de los datos al modelo umbrales críticos de OUTFIT son:

- OUTFIT > 2.0 el ruido aleatorio (fuera de la variable de interés) es mayor que la información útil que aporta el reactivo. Este efecto indeseable (debido a características particulares del reactivo o del respondente) degrada el proceso de medición.
- OUTFIT > 1.5 el ruido aleatorio (fuera de la variable de interés) es notable. Esta situación no aporta al proceso de medición pero tampoco lo degrada.
- OUTFIT de 0.5 a 1.5 situación PRODUCTIVA (optima) para el proceso de medición.
- OUTFIT < 0.5 esta situación ajusta demasiado bien (es decir es demasiado predecible). Es confundente en el sentido de que aparenta que se está midiendo muy bien el constructo de interés cuando de hecho no es el caso.

Como podrá observarse la aplicación puntual de los criterios de bondad de ajuste de los reactivos a las expectativas del modelo, no difieren en el caso de ítems politómicos.

Ahora bien, para aquellos casos en que NO se pueda sostener ni justificar la uniformidad de distancias de las categorías graduadas entre reactivos el modelo PCM es:

Caso 2 Estructura PCM “Partial Credit Model” (Masters)

Supuesto: Cada ítem tiene su propia estructura de escala rating

Modelo Politómico

$$\text{Log}(P_{nij}/P_{ni(j-1)}) = B_n - D_i - F_j = B_n - D_{ij}$$

Donde:

P_{nij} = Probabilidad de que la persona n en su encuentro con el ítem i emita su respuesta en la categoría j .

B_n = La medida de “habilidad” (afinidad) de la persona n

D_j = La medida de “dificultad” (aceptabilidad) del ítem i (MEASURE)

F_j = La medida de calibración de la categoría j relativa a la categoría $j-1$

De nueva cuenta los criterios de bondad de ajuste son como ya se han definido.

Ahora bien, la implementación en WS o MS de análisis de ítems politómicos se describe como sigue:

- 1) La programación del archivo de control sigue la forma general acostumbrada pero con variantes específicas al caso, como se presenta en el ejemplo 8
- 2) Cabe aclarar que en este tipo de análisis es preferible de inicio asumir que los ítems NO COMPARTEN estructura de categorías uniformes entre ellos. Es decir adoptamos el caso PCM de Rasch-Masters de tal manera que cada reactivo tenga su propia (y única) estructura de categorías y distancias. Esto es razonable en instrumentos en proceso de desarrollo para los que no se tiene evidencia de uniformidad compartida entre ítems.
- 3) Debe estar claro que como es de esperarse los respondientes expresan sus grados de percepción de frecuencia con respecto a 40 afirmaciones tales como: “En este programa académico todas las personas reciben un trato respetuoso”.

- 4) La dinámica de presentación de las afirmaciones es la típica en cualquier caso de aplicación de instrumentos con reactivos tipo Likert.

El archivo de control (AC) para un caso como este es.

Ejemplo 8

```

Title = "Instrumento de 40 Ítems Políticos Adaptado"
DATA = "C:\My Documents\.....\Rasch Model\CFK1.DAT"
NI = 40 ; 40 reactivos
Item1 = 1 ; Los datos reactivo columna 1
Xwide = 1 ; Ancho de columnas
Codes = 12345 ; códigos validos son 1-5.
DELIM = TAB ; Campos delimitados por TAB
IGROUPS = 0 ; CASO PCM cada item tiene su propia estructura
cfile=* ; Etiquetar las categorías de respuesta
1 Casi Nunca
2 A Veces
3 Frecuentemente
4 Casi Siempre
*
Total = Yes ; Mostar totales por reactivo
Chart = Yes ; Producir tabla cruzada
Mnsq = No ; Utilizar solo estadígrafos de ensute ESTANDARIZADOS
Uimean= 0 ; Centrar la media de dificultad en 0.
Stbias= Yes ; Ajustar el sesgo de estimación
&End
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

```

```

32
33
34
35
36
37
38
39
40
END NAMES

```

Es de notarse que la principal diferencia en este archivo de control radica en las secciones intermedias donde la estructura de los ítems esta especificada por las líneas:

```

IGROUPS = 0      ; CASO PCM cada ítem tiene su propia estructura
cfile=*         ; Etiquetar las categorías de respuesta
1 Casi Nunca
2 A Veces
3 Frecuentemente
4 Casi Siempre
*

```

Aquí IGROUPS=0 especifica que NO existen grupos de ítems que compartan estructura de categorías y distancias. Es decir cada ítem tiene su propia estructura única.

Cfile=* señala el inicio de los valores de respuesta al interior de cada ítem. Y es crucial notar que * señala el fin de los valores de respuesta. Un segmento de la base de datos donde corre este análisis es:

```

4444444434311131111444432333131323323233
3334424242422442212443434434134324.43244
4444444333444..3..34444444433333.4.33334
4444444443333322333332333333333322233334
4344332243333423112343323343423424433223
4444424444423441112221224321143413213124
433334233333433422433333333433343433333
4444343333343332122422424434444324344444
34443232333212111232333333343324434344
3343342422424114113144334442343211222134

```

```

34344433332322233333233233333344433233
444444444444442333344433443144434433134
3344444344341332114343433444144434344344
34444323333312211243343332333223333334
44442323332241241111324324422.3212223144
44444333322413311232232322334313222223
4444444444344443323433343343444324344444
4434322343144313112324434323343312331123
44344424444443244343444344444234344334
444344444434444214.443444434434433444444
4444444444444332223433433334443324343333
4444444443444433223443434443334334344344
3333344434433443444344322313234313312223
4433332333334322223333432322233334333434
444444334343443411233343332323232344333
44343333323412221233332323233223333233
4444344344334323313444344443444321424134
4444434343333322244333343333323434334
433333432233433211234333433323332333331
4444434334432332333433322433343433433334
444444343323221222133434323444434423334
3344434122122121111332221432243213123134
443444434443434444443344434444344434443
4444443343334234112422333433343323444344
44444444434343332244333443443333343344
444444233234434224444444433434443433444
444444344334433234444334333444343344444
44434433343334324333334343444443433333
4434443333344222111344343322342214313243
443431333222422112242332322133323343123
    
```

Ahora bien al correr el archivo de control de este ejemplo (8) obtenemos la información de salida convencional con algunas adiciones importantes propias de los casos politómicos como se ejemplifica en:

TABLE 13.1 Instrumento Items politocmicos ZOU714WS.TXT Mar
 INPUT: 345 PERSONS 40 ITEMS MEASURED: 345 PERSONS 40 ITEMS 4 CATS 3.64.2

 PERSON: REAL SEP.: 3.51 REL.: .92 ... ITEM: REAL SEP.: 8.39 REL.: .99

 ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEA CORR.	EXACT OBS%	MATCH EXP%	ITEM
17	719	340	2.01	.07	1.26	3.4	1.27	3.3	.59	45.9	47.6	17
18	816	340	1.53	.07	1.31	4.1	1.32	4.0	.56	41.4	46.4	18
16	884	344	1.23	.07	1.62	7.5	1.90	9.7	.37	39.8	46.7	16
33	874	340	1.23	.07	1.07	.9	1.06	.8	.62	46.3	46.7	33
19	912	340	1.04	.07	1.05	.8	1.07	.9	.60	48.5	47.6	19
15	950	340	.83	.07	.84	-2.3	.87	-1.7	.66	54.1	49.2	15
38	997	343	.62	.07	1.00	.0	1.05	.7	.61	52.2	50.6	38
14	994	341	.60	.07	1.06	.9	1.05	.6	.55	49.3	50.6	14
29	1040	343	.38	.08	.96	-.5	1.00	.0	.57	58.4	52.5	29
25	1053	342	.29	.08	.90	-1.3	.94	-.6	.53	57.6	53.2	25
27	1064	342	.22	.08	1.06	.9	1.07	.8	.54	51.8	53.8	27
28	1068	343	.22	.08	.70	-4.3	.72	-3.4	.60	59.2	53.7	28
10	1070	342	.20	.08	.97	-.3	1.12	1.3	.50	55.9	54.3	10
11	1083	342	.10	.08	.79	-2.8	.74	-3.1	.60	61.2	55.0	11

35	1079	338	.06	.08	.92	-1.1	.96	-.4	.55	61.6	55.6	35
32	1076	336	.04	.08	.76	-3.2	.81	-2.1	.57	65.6	55.8	32
34	1099	343	.03	.08	1.04	.6	.99	-.1	.53	56.3	55.7	34
8	1098	343	.03	.08	.90	-1.3	.98	-.2	.50	57.2	55.7	08
24	1095	341	.01	.08	.76	-3.3	.75	-2.9	.58	61.1	55.7	24
31	1093	340	.00	.08	1.03	.4	.98	-.1	.52	57.7	55.9	31
36	1104	343	.00	.08	.98	-.2	.90	-1.0	.58	56.0	55.9	36
12	1097	338	-.07	.08	.76	-3.3	.74	-2.9	.59	64.0	56.5	12
39	1116	344	-.07	.08	.76	-3.2	.72	-3.2	.62	62.0	56.5	39
20	1120	343	-.11	.08	.95	-.6	.86	-1.5	.60	61.3	57.2	20
22	1126	343	-.15	.08	.94	-.7	.91	-.8	.52	60.1	57.8	22
7	1132	344	-.17	.08	1.09	1.1	1.07	.7	.48	52.3	57.9	07
21	1154	345	-.31	.09	.81	-2.4	.80	-2.0	.58	66.2	58.9	21
26	1158	345	-.34	.09	1.09	1.1	1.15	1.4	.41	58.9	59.2	26
23	1155	342	-.39	.09	.85	-1.9	.84	-1.5	.51	62.9	59.6	23
6	1166	344	-.42	.09	.82	-2.2	.77	-2.3	.56	64.6	59.8	06
9	1171	344	-.46	.09	.79	-2.7	.77	-2.2	.54	64.3	60.4	09
37	1162	341	-.47	.09	.87	-1.5	.76	-2.3	.58	63.7	60.5	37
30	1167	341	-.50	.09	.90	-1.2	.83	-1.6	.54	64.0	60.9	30
3	1177	340	-.61	.09	.97	-.3	1.16	1.3	.41	65.4	61.7	03
13	1191	342	-.68	.09	1.67	6.4	1.73	5.1	.38	55.6	62.4	13
5	1197	341	-.75	.09	.97	-.3	.91	-.7	.43	63.7	63.2	05
40	1228	344	-.96	.10	1.27	2.8	1.19	1.4	.42	66.4	66.1	40
1	1248	343	-1.20	.11	1.36	3.4	1.45	2.7	.32	65.7	70.0	01
4	1269	344	-1.40	.11	1.12	1.2	1.12	.8	.34	71.3	72.9	04
2	1274	341	-1.62	.12	1.06	.6	1.04	.3	.34	74.6	76.2	02

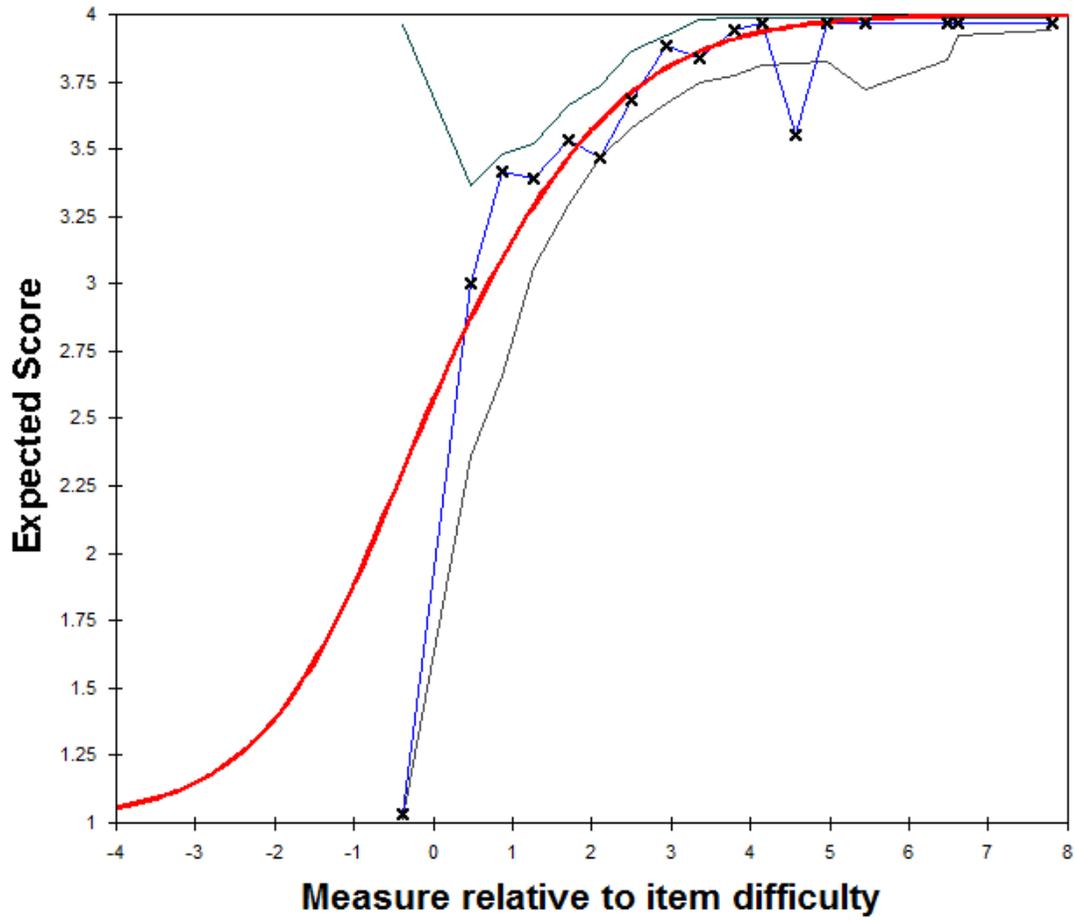
MEAN	1079.0	339.9	.00	.08	1.00	-.1	1.01	.0		58.6	56.9	
S.D.	117.8	2.0	.74	.01	.21	2.5	.25	2.5		7.5	6.6	

Como puede observarse se han señalado con rojo algunos (no todos) de los reactivos que presentan desajuste a la dimensión principal. Lo indicado en este caso es inspeccionar la redacción y la carga semántica de los reactivos señalados para determinar si en uno o en ambos aspectos se están produciendo las condiciones de desajuste. Eventualmente de ser así estos reactivos deberán ser reformulados para que se ajusten mejor al modelo como lo hacen los otros o bien en su caso deberán ser substituidos.

Por otra parte es posible extraer de la información de salida bajo GRAPHS ICC la curva característica de cada uno de los reactivos (CCR). A continuación se presenta el reactivo 1 que cuyo ajuste al modelo es razonable:

Figura 4

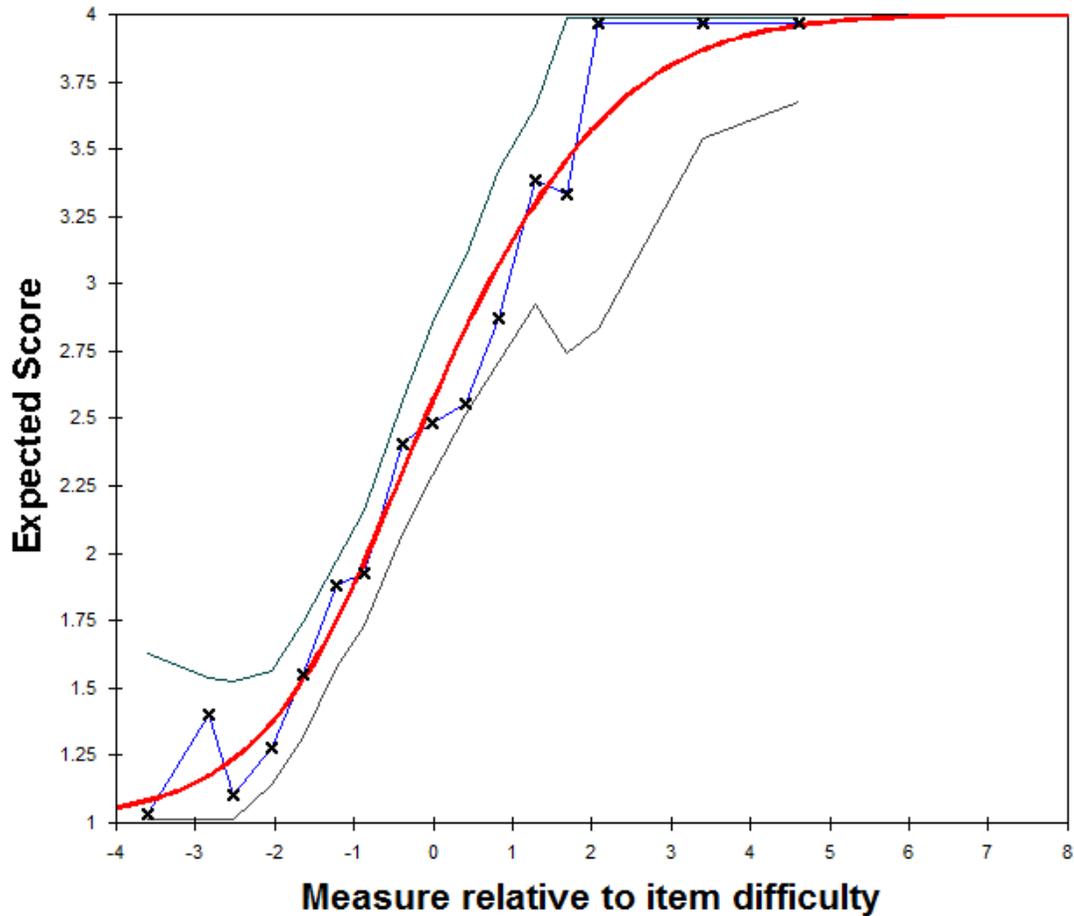
1.01 Respeto Alumnos



Ahora comparamos lo anterior con la CCR del reactivo 17 cuyo ajuste a la dimensión es malo:

Figura 5

17.17 Enterado Asuntos



Donde se observa que este reactivo en este caso produjo ratings muy altos (límite superior) en calibraciones de entre 2 y 4 lógitos de dificultad. Esto es indicativo de que el reactivo se está contestando por la mayoría de los respondentes más por su alta deseabilidad social que por patrones de respuesta efectivos y productivos. También es muy importante señalar que este desajuste por sí solo no garantizaría la eliminación del reactivo. Más bien es una indicación de que la redacción y sentido del mismo deben revisarse.

En la información de salida hay una gran cantidad de información gráfica que se puede explorar reactivo por reactivo. A manera de ejemplo considérense las dos figuras siguientes:

Figura 6

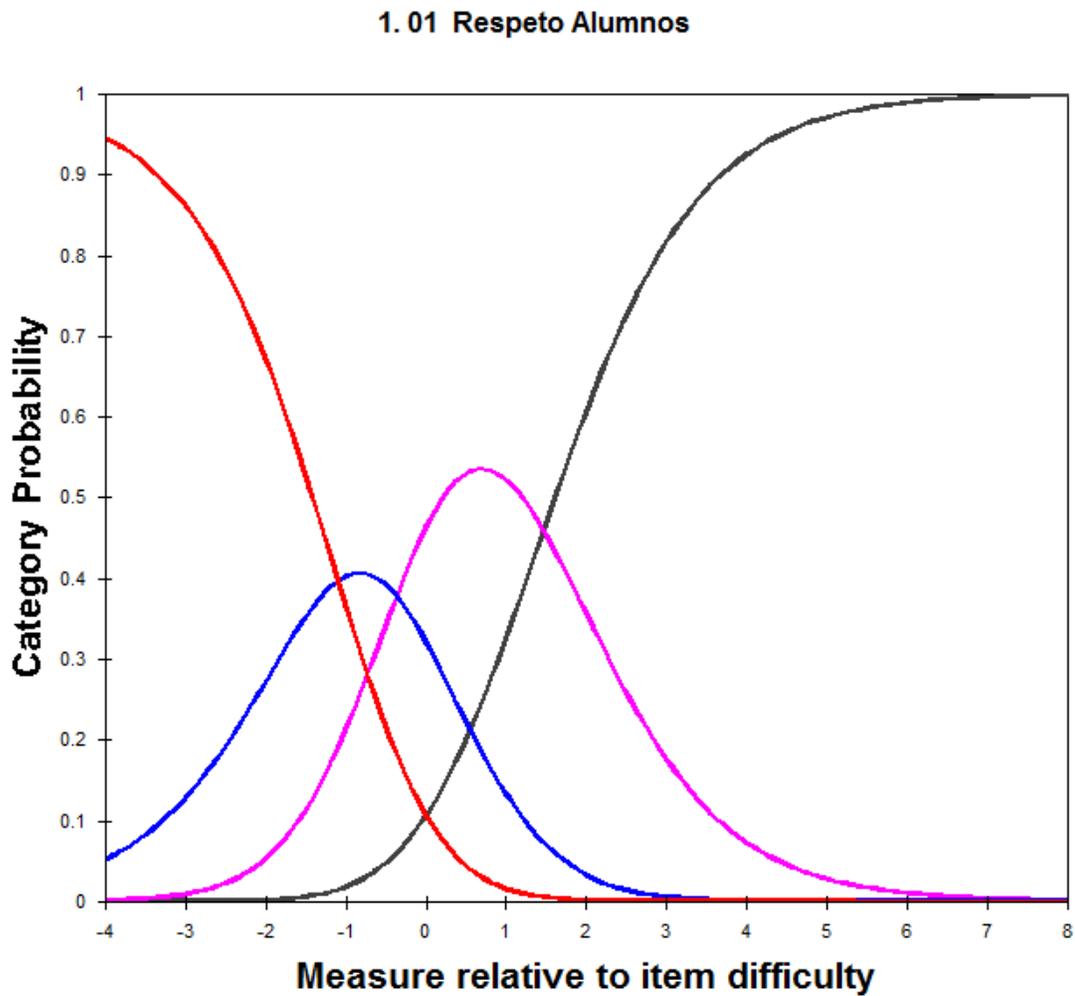
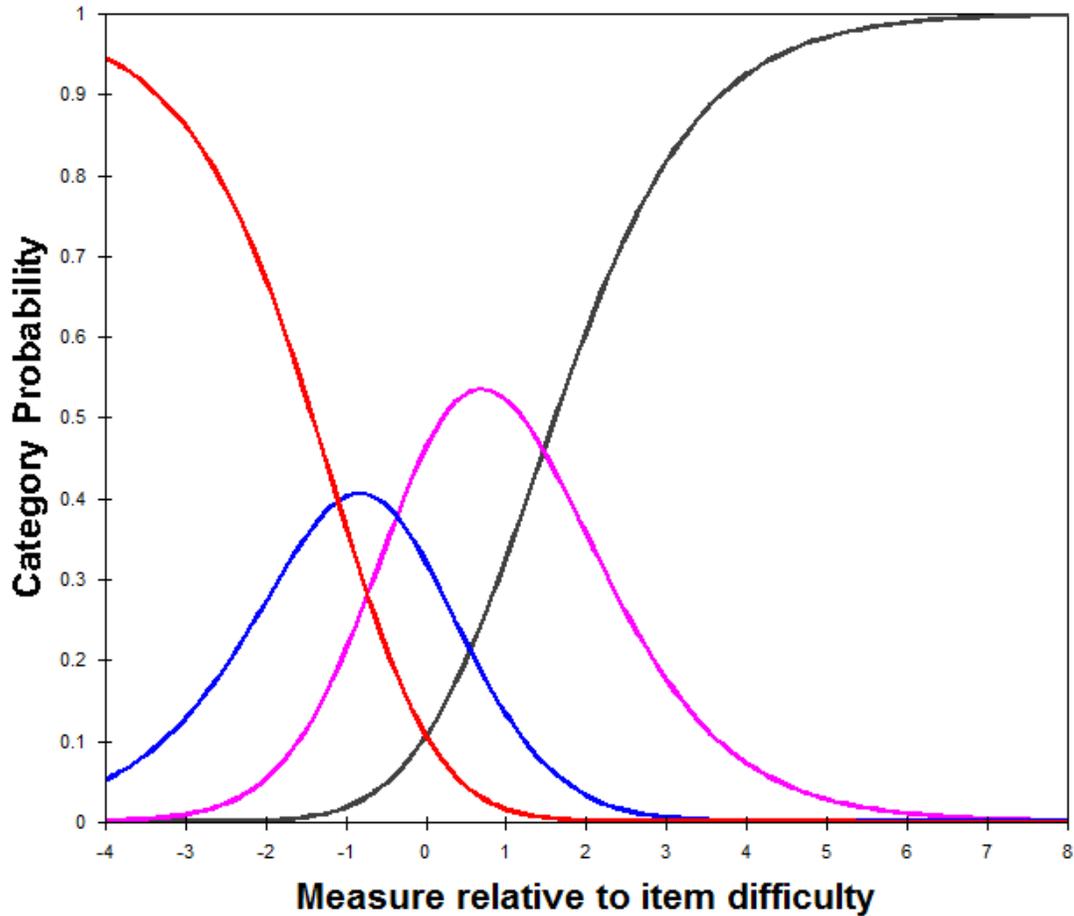


Figura 7

17. 17 Enterado Asuntos



Ambas representan para los ítems bajo ejemplo (1 y 17) la probabilidad acumulada para cada una de las categorías de respuesta que en este caso son:

- | | | |
|---|----------------|---------|
| 1 | Casi Nunca | ROJO |
| 2 | A Veces | AZUL |
| 3 | Frecuentemente | VIOLETA |
| 4 | Casi Siempre | NEGRO |

Se observa con claridad que en los extremos de las escalas graduadas de estos ítems las opciones 1 y 4 tiene un comportamiento distinto a las categorías

intermedias. En la curva del extremo “Casi Nunca” o 1 se delimita una función decreciente en probabilidad de respuestas a medida que baja la habilidad en lógitos de los respondentes, esto es a medida que decrece la capacidad de los respondentes para expresar acuerdo baja la probabilidad de respuesta hasta aproximar 0. En el otro extremo la situación es precisamente a la inversa donde la curva del extremo “Casi Siempre” o 4 delimita una función ascendente en la probabilidad de respuestas que aumenta a medida que aumenta la capacidad (o disposición) de los respondentes para expresar acuerdo con las afirmaciones. Esta función aproximaría a la unidad “1” si se observan valores de calibración de más de 4 lógitos. Lo cual es solo una proyección de la tendencia generalizada ya que efectivamente no se observaron empíricamente esas calibraciones tan altas.

Por otra parte puede verse también con claridad que las categorías intermedias de respuesta 2 y 3 alcanzan un pico de altura moderada en probabilidad de respuesta en calibraciones de lógitos -1 y +1 respectivamente. Puede además notarse que la opción intermedia describe una campana no regular entre calibraciones de -4 a +3 lógitos de disponibilidad para expresar acuerdo (Opción 2 “A Veces”). El caso para la opción 3 “Frecuentemente” sigue la misma forma entre -3 y + lógitos de “afinidad” de los respondentes con las afirmaciones presentadas en los ítems 1 y 17.

Es necesario notar que las curvas de probabilidad acumulada no difieren entre los ítems 1 y 17 lo cual a su vez indica que hay consistencia de estructura entre estos dos ítems y este caso pudiera darse si no para todos, si para un

buen numero de ítems en el instrumento. Esto solo es determinable bajo una inspección ítem por ítem que el lector puede emprender a manera de ejercicio individual.

Otro aspecto importante es que el output nos permite seleccionar otros recursos como por ejemplo el llamado Mapa de Wright:

TABLE 12.2 CFK Adapatdo		ZOU010WS.TXT Mar 4 15:28 2008			
INPUT: 345 PERSONS 40 ITEMS		MEASURED: 345 PERSONS 40 ITEMS		4 CATS	3.64.2

PERSONS - MAP - ITEMS					
<more> <rare>					
6	.	+			
	.#				
	.				
5		+			
	.				
	.				
4	.	+			
	#		T		
	#				
	##				
	.#				
3	#	+			
	.##				
	.###				
	.###		S		
	###				
	.###				
2	#####	+	17	Enterado Asuntos	
	.#####				
	.#####				
	#####		T	18 Decisiones Acuerdo	
	.#####		M		
	#####			16 Asistencia Alumnos	
				33 Ayuda Estudiantes	
1	.#####	+	19	Participacion Acuerdos	
	#####			15 Siento en Cuenta	
	.#####		S	14 Ideas en Cuenta	
				38 Consideran Persona	
	#####			25 Actividades Responden	
	#####			29 Diferencias Contribuyen	
	.####	S		10 Docentes Confian	
				11 Entusiasmo Aprender	
				27 Union Participantes	
				28 Maestros Colaboran	
0	#####	+M	08	Atencion Intereses	
				12 Orgullo Carrera	
				24 Maestros Progresando	
				31 Procedimientos Resolucion	
				32 Maestros Apoyo	
				34 Preferencia Progreso	
				35 Cambios Adaptados	
				36 Cuento con Alguien	
				39 Carrera Agradable	
	.###			07 Discusión Abierta	
				20 Cuento Como Persona	
				22 Maestros Actualizados	

.		21	Estusiasmo Maestros
		23	Metodos Adecuados
		26	Preferencia Estudiantes
.##		06	Apoyo Alumnos
		09	Docentes Escuchan
		30	Estudiantes Bienvenidos
		37	Estudiantes Importantes
.	S	03	Colaboradores
		13	Asistencia Maestros
		05	Orgullo Docente
-1	.	T+	40 Personas Amigables
			01 Respeto Alumnos
			04 Respeto Maestros
			T
	.		02 Trato Personas
-2			+
			<less> <frequ>
EACH '#' IS 3.			

En el que se puede observar que una gran cantidad de reactivos quedan situados por debajo de la media 0 de dificultad para emitir acuerdo u opinión favorable, mientras que la gran mayoría de los respondentes quedaron situados por encima de la media 0 de dificultad de emitir acuerdo u opinión favorable.

Es decir aquí se tiene un caso de una mayoría de reactivos fáciles de aceptar como afirmaciones y una mayoría de respondente con alta afinidad o disposición actitudinal favorable a las afirmaciones contenidas en los reactivos.

Reactivos en sentido inverso

Puede darse el caso y de hecho es frecuente en la práctica que algunos de los ítems tengan un sentido negativo. Es decir hay ítems que por razón de su construcción presentan afirmaciones tales que al expresar desacuerdo con ellas el respondente exhibe una afinidad con el constructo. Por ejemplo considérese un reactivo politómico destinado a explorar a percepción del respondente sobre la igualdad de oportunidades para ambos géneros: “Las mujeres con hijos no

deberían tener las mismas oportunidades de desarrollo profesional que los hombres debido a su condición de madres”. El expresar desacuerdo con esta afirmación es de hecho una actitud favorable – positiva hacia la equidad de género. Por ello los patrones de respuesta de esta clase de ítems requieren de ser calificados en sentido inverso.

Lo anterior se logra insertando en el archivo de control las siguientes líneas:

```
Title = "Escala con Items Inversos"
DATA = "C:\Documents and Settings\IRT-Escalas\DATA1.dat"
NI = 20 ; 20 reactivos
Item1 = 1 ; Los datos reactivo columna 1
Xwide = 1 ; Ancho de columnas
Codes = 12345 ; Codigos validos 1-5.
IREFER= FRRFRFRFFRFFRFFRFFRFF ; Dos tipos de item: F = Forward, R = Reverse
IVALUEF=12345 ; Forward (positivos) ítems con sentido positivo
IVALUER=54321 ; Reverse (negativos) ítems con sentido negativo
IGROUPS=0 ; Cada ítem tiene su propia escala PCM
cfile=* ; Etiquetas de las categorías de respuesta
1 Nunca
2 Rara vez
3 Algunas veces
4 Usualmente
5 Siempre
*
LOCAL = Yes
Total = Yes ; total puntajes
Chart = Yes ; produce tabla cruzada
Uimean= 0 ; fijar la dificultad media a 0.
Stbias= Yes ; ajustar para sesgo de estimación
&End
```

Se puede apreciar claramente que IVALUEF =12345 valora las categorías de respuesta observadas –emitidas por el respondente- en forma positiva o (FORWARD) mientras que IVALUER=54321 valora las categorías de respuesta observada en forma negativa (REVERSE) cambiando a sentido positivo las respuestas en desacuerdo a ítems en sentido inverso o negativo.

El usuario declara mediante la instrucción IREFER= FRRFRFRFFRFFRFFRFFRFF; Dos tipos de ítem: F = Forward, R = Reverse, la posición de los ítems que serán valorados (o calificados) en sentido inverso.

Post scriptum

El material presentado cubre en teoría y práctica las fases elementales del Análisis Rasch para la determinación de las propiedades psicométricas de reactivos que forman parte de instrumentos de medición en desarrollo o en su caso de instrumentos consolidados. Como se podrá apreciar, el Modelo Rasch es a la vez simple en sus fundamentos pero exigente en los resultados que admite como válidos. El potencial de la metodología Rasch radica –entre otras cosas- en que esta solo admite como reactivos eficientes solo aquellos que satisfacen los más estrictos criterios de control de calidad establecidos desde un basamento matemático y probabilístico. Este doble basamento es matemático porque se formula como una aplicación derivada de la función logística cuyas propiedades distribucionales están firmemente establecidas. El basamento es también probabilístico ya que trata los encuentros entre sustentantes y reactivos como procesos estocásticos pero que en los cuales es posible diagnosticar la sistematicidad observada en patrones de respuesta –que satisfacen la expectativa matemática de la función logística- y que por lo tanto capturan o reflejan la variable latente de interés.

Se presentan y distribuyen estos materiales con la firme intención de que sean de utilidad para el conjunto de usuarios que posean el muy legítimo interés en depurar y mejorar la eficiencia y productividad métrica de los instrumentos que desarrollen y utilicen en sus procesos de investigación.

La Metodología de Medición Rasch ha sido calificada como la medición fundamental en las ciencias sociales. Sea el lector atento el árbitro final de esta sugerente y polémica aseveración.

El potencial de la metodología descrita radica en que proporciona una especie de filtro matemático para -que en el proceso de desarrollo de instrumentos de medición psicológica- se seleccionen solo aquellos reactivos que efectivamente posean las propiedades necesarias para captar con eficiencia y productividad la variable latente para la que fueron diseñados.

M. G. M.
Marzo de 2008 México, D.F
mgm@caborca.uson.mx

Bibliografía

- Bond, T.G., Fox, Ch. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New Jersey: Erlbaum.
- D'Agostino, J. Et Al. (2002). *Validation of Authentic Performance Assessment: a process suited for Rasch Modeling*. American Journal of Pharmaceutical Education. Vol. 66.
- Hambleton, R.K., Swaminathan, H., Rogers, J. H (1991) . *Fundamentals of Item Response Theory*. Quantitative Methods for the Social Sciences. Newbury Park, CA: Sage Publications.
- Linacre, J. M. (1991- 2006). *Rasch-Model Computer Programs* ©, Chicago, IL: Electronic Publication. www.winsteps.com
- Linacre, J. M. (2006). *A User's Guide to Winsteps Ministeps: Rasch-Model Computer Programs*. Chicago, IL: Electronic Publication. www.winsteps.com
- Rasch, G. (1960). *Probabilistic Models for some Attainment and Intelligence Tests* Copenhagen, Danmarks Pedagogiske Institut
- Tristán L.A. (2002). *Análisis de Rasch para Todos: una guía para evaluadores educativos* CENEVAL, México, D.F.
- Wright, B. D., & Stone, M. H. (1979). *Best Test Design*. Chicago, IL: MESA Press.
- Wright, B. D., & Stone, M. H. (1998). *Diseño de Mejores Pruebas*. CENEVAL, México. D.F.