

SOFTWARE, INSTRUMENTACIÓN Y METODOLOGÍA

Comparación de estrategias de ajuste de los datos al modelo de crédito parcial

Antonio J. Rojas Tejada, Andrés González Gómez**, José L. Padilla García** y Cristino Pérez Meléndez**
Universidad de Almería y ** Universidad de Granada

En el presente trabajo se comparan dos estrategias de ajuste de los datos al modelo. La primera de ellas, llamada estrategia Items-Sujetos-Total (IST), se presenta estructurada en tres etapas donde se evalúa: primero, el ajuste de los ítems; segundo el de los sujetos, y por último, el ajuste global de los datos al modelo (ítems y sujetos). La segunda estrategia, llamada Sujetos-Items-Total (SIT), invierte el orden de una de las etapas: primero se evalúa el ajuste de los sujetos; segundo el de los ítems; y, por último, el ajuste global de los datos al modelo. Se han comprobado los resultados que producen ambas estrategias cuando administramos un conjunto de ítems (30), elaborados para medir actitud religiosa, a una muestra de sujetos (821). El modelo de Teoría de Rasgo Latente utilizado ha sido el Modelo de Crédito Parcial. Los resultados muestran diferencias entre ambas estrategias: la IST maximiza el número de sujetos ajustados frente a la SIT donde se maximiza el número de ítems ajustados. Además se propone una forma de controlar el problema de la sensibilidad de los estadísticos de ajuste al tamaño muestral.

Comparison of strategies of the data fit to the partial credit model. A comparative study of the results provided by two strategies for fitting data to Latent Trait Theory Models has been performed. The first, called Total-Subjects-Items (TSI), is structured in three phases: 1) assessment of item fit, 2) assessment of subject fit; and finally, 3) overall fit of data to the models (items and persons). The second strategy, the Total-Items-Subjects (TIS), changes the order of the phases: 1) assessment of subject fit, 2) assessment of item fit and, 3) overall fit of data to the models. To verify the results of these two strategies, a set of 30 items, designed to measure religious attitude, was administered to a sample of 821 persons. The Latent Trait Theory Model used were the Partial Credit Model. The results underline an important difference between the two procedures: the TSI maximizes the number of persons with good fit and the TIS maximizes the number of items with good fit. Moreover, a procedure for controlling the sensitivity of fit to sample size is proposed.

El estudio del ajuste de los datos al modelo es una actividad obligada en la utilización de los modelos de Teoría de Rasgo Latente, ya que a partir de sus resultados se derivarán las características diferenciales de esta teoría sobre el test, el conjunto de ítems y la medición de los sujetos (López Pina e Hidalgo, 1996). Es decir, si los datos no se ajustan al modelo, entonces los datos no pueden utilizarse para calibrar los ítems ni para medir a los sujetos (Wright, 1980). De hecho, todas las ventajas y posibilidades que proporciona la utilización de la Teoría de Rasgo

Latente en la construcción de tests sólo se obtienen cuando el ajuste entre los datos y el modelo es adecuado (p.e. Barbero, 1999; Hambleton, 1990; Hambleton y Swaminathan, 1985; Wright y Stone, 1979).

A partir del resultado del ajuste podemos detectar ítems de 'mala calidad' (Wright y Stone, 1979) y sujetos que no pueden ser medidos con el conjunto de ítems utilizado (O'Brien, 1992). En general, las pruebas de ajuste están basadas en el principio de semejanza entre los datos empíricos procedentes de la administración de un conjunto de ítems a una muestra de sujetos y los datos pronosticados por el modelo que se está utilizando (p.e. Gustafsson, 1980): a mayor semejanza entre ellos mayor ajuste y a mayores diferencias mayor desajuste.

El estudio del ajuste se puede considerar desde una triple perspectiva: ajuste de los ítems, de los sujetos y ajuste global de los da-

tos al modelo. En cierta forma, las dos primeras cumplen la función de depurar los datos mientras que la última sirve como indicador de lo adecuado que resulta el modelo de medida utilizado, en el sentido de posibilitar un escalamiento conjunto de sujetos y estímulos en una misma variable psicológica (p.e. Hoijtink, Molenaar y Post, 1994). Desde este punto de vista, es lógico que el ajuste global de los datos al modelo sea la última de las tres aproximaciones para el análisis del ajuste. Sin embargo, no hay ninguna razón a priori para determinar cuál de las dos aproximaciones restantes debe realizarse en primer lugar. Esta indefinición está presente en los trabajos psicométricos dedicados a la construcción de tests con modelos basados en la Teoría del Rasgo Latente, de forma tal que el profesional que quiera aplicar estos modelos se encuentra desamparado al no contar con una secuencia de pasos bien definidos.

En este estudio planteamos dos posibilidades para realizar el análisis del ajuste:

1. Estrategia Items-Sujetos-Total (IST). Consistente en: 1) comprobar el ajuste de los ítems, 2) eliminar los ítems desajustados, 3) realizar de nuevo el proceso de calibración y, sólo cuando resulten los ítems ajustados pasar a, 4) comprobar el ajuste de los sujetos, y posteriormente, 5) evaluar el ajuste del conjunto global de datos al modelo.

2. Estrategia Sujetos-Items-Total (SIT). La secuencia de acciones es: 1) comprobar el ajuste de los sujetos, 2) eliminar los sujetos desajustados, 3) realizar de nuevo el proceso de calibración y, sólo cuando los sujetos resulten ajustados pasar a, 4) comprobar el ajuste de los ítems y, por último, 5) evaluar el ajuste del conjunto global de datos al modelo.

De forma general, podemos considerar que una de las razones para encontrar sujetos desajustados es que entre los ítems utilizados para estimar la habilidad/actitud de estos sujetos estén presentes algunos de baja calidad. Si eliminamos estos ítems, es razonable que se consiga una mejor estimación de los niveles de habilidad/actitud de los sujetos. Por tanto, cabe esperar que la estrategia IST maximice el número de sujetos con los que se alcanza un ajuste adecuado.

A la inversa, una de las razones para detectar ítems desajustados es que entre los sujetos utilizados se encuentren algunos con características especiales que distorsionan su proceso de respuesta a los ítems (O'Brien, 1992; Fed Li y Olejnik, 1997). En definitiva, si utilizamos la estrategia SIT y eliminamos del análisis los sujetos desajustados, se puede esperar que aumente el número de ítems que se ajustan al modelo.

Vemos, pues, cómo la estrategia utilizada para el estudio del ajuste podría llevar a resultados distintos, en función del orden en que se realicen las mismas acciones.

Por otro lado, existe el problema habitual en el estudio del ajuste que es el de la sensibilidad de los estadísticos al tamaño muestral. Este aspecto está muy bien recogido en los trabajos de Hambleton (1989), Hambleton y Murray (1983) y López Pina e Hidalgo (1996). Hambleton (1989) y Hambleton y Murray (1983) muestran, para el modelo logístico de 3 parámetros, con datos simulados (por tanto con un ajuste teórico perfecto) cómo al variar el tamaño muestral, para determinar el grado de ajuste de los ítems, se obtienen diferencias en cuanto al porcentaje de ítems desajustados. Así con muestras de sujetos grandes (N=2400), para un total de 50 ítems, se obtienen porcentajes de desajustes de los ítems entre el 76% y 84%, mientras que con muestras pequeñas (N=150) estos porcentajes oscilan entre el

10% y el 40%. Similares resultados obtienen López Pina e Hidalgo (1996) para el modelo logístico de 2 parámetros (también con datos simulados) con 40 ítems: con muestras pequeñas (N=50) se desajustaba un 15% de los ítems, mientras que con muestras grandes (N=2000) este porcentaje se elevaba hasta el 67,5%. En definitiva, con tamaños muestrales grandes es muy fácil que pequeños desajustes de los estadísticos de ajuste resulten valores estadísticamente significativos, llevando a la conclusión de que los ítems no se ajustan al modelo; por contra, al utilizar tamaños muestrales pequeños corremos el peligro de que las estimaciones sean inconsistentes y se obtengan errores de estimación muy elevados. Para solucionar en parte este problema seguimos un procedimiento, llevado a cabo por Rojas (1998), consistente en realizar diferentes análisis de ajuste utilizando distintos tamaños muestrales.

El modelo utilizado en este trabajo para analizar el funcionamiento de las estrategias de análisis del ajuste ha sido uno perteneciente a la familia de modelos politómicos de Rasch (Rasch, 1960; Rasch, 1980). Concretamente, nos centraremos en el Modelo de Crédito Parcial (Masters, 1982, 1988a, 1988b; Masters y Wright, 1984, 1997; Wright y Masters, 1982). Para este modelo se han desarrollado distintos tipos de índices de ajuste al modelo. Masters y Wright (1997) los dividen en los tres tipos anteriormente comentados: 1) índices de ajuste de los ítems, donde se identifican los ítems que tienen 'problemas' con el modelo, 2) índices de ajuste de los sujetos, donde se indican los sujetos que no siguen el patrón de respuestas general que propone el modelo; y, 3) índices de ajuste global, que señalan el grado de ajuste general del conjunto de datos al modelo (ítems y sujetos).

El objetivo fundamental de este trabajo es realizar un estudio con datos reales donde se muestre si las distintas estrategias de ajuste de los datos al modelo producen resultados diferentes, concretamente respecto al número de ítems y de sujetos que presentan ajuste. Un objetivo complementario es mostrar un procedimiento para controlar el problema de la sensibilidad al tamaño muestral.

Método

Sujetos

La muestra utilizada está formada por un total de 821 sujetos. Del total, el 22.61% está compuesto por hombres y el 77.4% son mujeres; la edad media de los sujetos es de 21.30 años (mediana=21 y moda=20), con un rango de 28 (máxima=46 y mínima=18), y desviación típica de 3.39; todos son alumnos pertenecientes a las universidades de Almería (80.4%) o Sevilla (19.6%); mayoritariamente de la carrera de Psicología (59.7%), seguido de Psicopedagogía (8.8%), Magisterio (7.3%), Ciencias Ambientales (7.1%), Pedagogía (6.8%) y otras (10.3%).

Materiales

Se utilizó un conjunto de 30 ítems elaborados para medir actitudes religiosas. Estos ítems han sido utilizados por Morales (1988) para la elaboración de la 'Escala de Actitudes Religiosas (R-1)'. En estos ítems se pretende medir la 'tónica religiosa o la cercanía a la fe' (Morales, 1988, p. 488). La definición del cons-

tructo es muy genérica, y no pretende aludir a ninguna religión o práctica religiosa concreta, tal y como lo conciben Hood (1970) y Pargament et al. (1988) (*cit. en* Gorsuch, 1988). Los items vienen expresados en una escala tipo Likert de 5 puntos. La elección de este test ha estado motivada fundamentalmente por dos cuestiones: 1) el test permite someterse al control del Modelo de Crédito Parcial dadas sus características: utilización de items politómicos con formato de escalas de clasificación para recoger en una única dimensión la actitud o tónica religiosa general; y, 2) es un test que no está finalizado; es más, Morales (1988) Morales (1988) al elaborarlo con fines didácticos, presenta desde una versión inicial de la escala con 30 items, hasta llegar a una tercera versión o test final de 18 items, describiendo el proceso de análisis de items clásico.

Análisis

Dentro del Modelo de Crédito Parcial se han desarrollado dos estadísticos que nos proporcionarán el grado de ajuste de los datos (tanto de los items como de los sujetos) al modelo. Estos estadísticos son: 'Outfit' o Residual Cuadrático Medio No Ponderado estandarizado, e 'Infit' o Residual Cuadrático Medio Ponderado estandarizado (p.e. Wright y Linacre, 1992a).

La puntuación residual estandarizada (z_{ij}) o diferencia estandarizada entre la puntuación observada (x_{ij}) y la esperada (E_{ij}) de un sujeto j a un item i viene definida como:

$$z_{ij} = \frac{(x_{ij} - E_{ij})}{\sqrt{W_{ij}}}$$

Donde W_{ij} es la varianza de x_{ij} . Para cada *item* i , se puede calcular el Residual Cuadrático Medio No Ponderado estandarizado - outfit- (u_i), cuya formulación sería:

$$u_i = \frac{\sum_j z_{ij}^2}{N}$$

Donde N es el número de sujetos de la muestra. Igualmente para un sujeto j se puede calcular el valor de outfit (u_j), donde lo único que cambia en la formulación anterior es que N hace referencia al número de items utilizado (n).

Estos Residuales Cuadráticos Medios No Ponderados ('outfit') tanto de items como de sujetos (u_i y u_j) pueden ser muy sensibles a valores extremos o 'outliers' (Wright y Linacre, 1992b). Para reducir la influencia de estos valores extremos podemos calcular un nuevo estadístico llamado Residual Cuadrático Medio Ponderado estandarizado ('infit'). Para cada item i este estadístico (v_i) viene definido por:

$$v_i = \frac{\sum_j W_{ij} \cdot z_{ij}^2}{\sum_j W_{ij}}$$

De forma similar, también podemos calcular este Residual Cuadrático Medio Ponderado estandarizado o 'infit' para un sujeto j (v_j), donde debemos sustituir en la formulación anterior el número de sujetos (N) por el número de items utilizado (n).

Considerando ambos estadísticos, el criterio para considerar si un item o un sujeto tienen un mal ajuste es que obtengan valores de 'infit' o 'outfit' iguales o mayores de 2 (o una correlación item-total o sujeto-total negativa). Para comprobar si los datos (items o sujetos) se ajustan al modelo se examinará la distribución de los valores de 'infit' y de 'outfit' de dos formas: a) calculando los valores de los estadísticos de 'outfit' e 'infit' de items y de sujetos, sus valores medios y desviaciones típicas, conociendo que cuando existe un ajuste adecuado de los datos al modelo sus medias deben estar cercanas a 0 y sus desviaciones típicas ser similares a 1; y, b) observando la distribución de los valores de los estadísticos 'outfit' e 'infit', ya que deben distribuirse de forma aproximadamente normal.

Para llevar a cabo la aplicación del Modelo de Crédito Parcial se ha empleado el programa desarrollado por Wright y Linacre (1992b): BIGSTEPS ver. 2.29. Los procedimientos de estimación utilizados han sido PROX y UCON, descritos por Wright y Masters (1982).

Procedimiento

El cuestionario fue administrado de forma colectiva por un investigador en el horario y aula donde los alumnos recibían las clases. Todos los sujetos debían responder a todas las preguntas que se les hacía en el cuadernillo. La recogida de datos tuvo lugar en las instalaciones de las universidades de Almería y Sevilla.

Resultados

Estrategia IST

Etapas I: Ajuste y selección de los items con diferentes tamaños muestrales

La finalidad de esta etapa es considerar cuántos items del conjunto de los 30 utilizados se ajustan al Modelo de Crédito Parcial y, por tanto, se seleccionarán como adecuados para medir la variable propuesta. Pero, como hemos comentado anteriormente, la evaluación de la bondad del ajuste de los datos al modelo presenta un problema importante: la *sensibilidad de los estadísticos al tamaño muestral* del estudio. Esto es especialmente importante cuando tratamos del ajuste de los items, ya que es en esta fase donde decidiremos qué items se ajustan, y por tanto se seleccionan para medir la variable, y qué items son descartados por su desajuste. Por ello, en esta etapa, se ha intentado controlar el efecto de sensibilidad de los estadísticos al tamaño muestral, de forma que se han considerado cuatro tamaños muestrales. El primer tamaño, el mayor, ha estado compuesto por la muestra total de los sujetos ($n=821$). Las otras tres muestras son submuestras aleatorias con el 75%, 50% y 25% de los sujetos.

Los resultados indican (ver tabla 1) que existen 16 items que se consideran desajustados para el tamaño muestral 100% según sus valores de 'infit' y 'outfit'; con tamaño muestral 75%, se desajustan 15 items; 14 son los items desajustados con el tamaño muestral del 50%, y tan sólo 7 items se consideran desajustados si tenemos un tamaño muestral del 25%. Vemos claramente cómo el

tamaño muestral está influyendo en los estadísticos ‘infit’ y ‘outfit’. Todos los items muestran correlaciones item-total (CIT) positivos, salvo el item 4 en el tamaño muestral 25%, que tiene un valor de -0.04.

Los items 1, 3, 4, 11, 21, 24 y 28 son considerados desajustados por las cuatro submuestras. Los items 2, 12, 23, 25, 27 y 30 muestran desajustes para los tamaños muestrales del 50%, 75% y 100%. Los items 10 y 26 sólo aparecen desajustados para tamaños muestrales del 75% y 100%. Los items 19 y 20 sólo se desajustan con un determinado tamaño muestral, 50% y 100% respectivamente.

Esto nos debe servir para tomar una decisión respecto a qué items debemos rechazar o aceptar debido a su ajuste al modelo. En nuestro caso se van a considerar no ajustados los items que hayan mostrado desajuste en tres o más tamaños muestrales. De acuerdo con este criterio se considerarán desajustados los 13 items siguientes: 1, 2, 3, 4, 11, 12, 21, 23, 24, 25, 27, 28 y 30. Items que serán eliminados todos (a la vez) para la siguiente etapa.

Etapa II: Ajuste y selección de los sujetos

Una vez seleccionados los 17 items que se consideran ajustados al modelo, se procede a analizar el ajuste de los sujetos con estos 17 items. De los 821 sujetos se observa que 188 de ellos muestran desajuste, bien por valores inadecuados de ‘infit’ y ‘outfit’, bien por valores negativos de la correlación sujeto-total (correlación idéntica a la de item-total salvo que en vez de tomar la puntuación de los items se toman las de los sujetos y para la suma total se utilizan las suma de puntuaciones de los items en todos los sujetos). Este dato supone un 22.9% del total de la muestra. Estos sujetos muestran patrones de respuestas muy diferentes de los esperados y serán eliminados de la muestra (tal y como se hizo con los items).

Etapa III: Ajuste final de los items seleccionados con los sujetos elegidos

Ajuste global de los items

Para el conjunto de nuestros items de actitudes, los resultados muestran (ver tabla 2) evidencias para concluir un ajuste aceptable de los items. Tan sólo en un item (19) se obtuvieron valores de ‘infit’ y ‘outfit’ excesivos (2.31 y 2.72, respectivamente). Por otro lado, los valores medios de los estadísticos ‘infit’ y ‘outfit’ están cercanos a 0 (0.10 y 0.11, respectivamente); con desviaciones típicas de 1.20 y 1.29, respectivamente. En cuanto a la forma de la distribución, y a pesar de contar sólo con 17 items se ha efectuado el contraste con la prueba de Kolmogorov-Smirnov, resultando una

$p \leq 0.769$ para los valores del estadístico ‘infit’ y una $p \leq 0.909$ para los valores del estadístico ‘outfit’, por tanto, no se puede rechazar la hipótesis nula de normalidad.

Ajuste global de los sujetos

Para los sujetos seleccionados en nuestra muestra final los resultados también muestran evidencia para concluir un ajuste adecuado, al ser los valores medios de los estadísticos ‘infit’ y ‘outfit’ cercanos a 0 (0.02 y 0.01, respectivamente) y sus desviaciones típicas cercanas a 1 (1.00 y 0.95, respectivamente), tal y como cabe esperar cuando existe un buen ajuste. Los resultados para la prueba de Kolmogorov-Smirnov para comprobar si la distribución de los valores de ‘infit’ y ‘outfit’ siguen una distribución normal, arrojaron respectivamente unos valores de $p \leq 0.054$ y $p \leq 0.068$, con lo que se mantiene la hipótesis de normalidad.

Estos resultados apoyan la interpretación de un ajuste global adecuado tanto para los items como para los sujetos.

Estrategia SIT

Etapa I: Ajuste y selección de los sujetos

En esta primera etapa se ha realizado el proceso de ajuste de los 30 items con el total de sujetos que componen la muestra (821). Buscamos examinar el ajuste de los sujetos, con el fin de descartar a aquellos que no tengan un patrón de respuestas acorde con lo propuesto por el modelo.

De los 821 sujetos se observan que 243 de ellos muestran patrones de respuestas incoherentes con lo pronosticado por el modelo (bien por valores inadecuados de ‘infit’ y ‘outfit’, bien por valores negativos de la correlación sujeto-total). Este dato supone un 28.5% de la muestra total. Estos sujetos serán eliminados (todos a la vez) de la muestra para la segunda etapa.

Etapa II: Ajuste y selección de los items con diferentes tamaños muestrales

En esta etapa, al igual que con la estrategia IST, se ha intentado controlar el efecto de sensibilidad de los estadísticos utilizando los 578 sujetos que superan la etapa I (nueva muestra 100%) y tres submuestras aleatorias (muestras 75%, 50% y 25%). En este caso, observamos (ver tabla 3) que existen 16 items que se consideran desajustados para el tamaño muestral 100%; con tamaño muestral 75%, se desajustan 12 items; los items desajustados con el tamaño muestral del 50% son 10, y también 10 items se consideran desajustados con un tamaño muestral del 25%. Todos los items mues-

Tabla 1
Resumen del ajuste de los items para las distintas muestras: Estrategia IST
(Nota: M. indica Muestras; X indica desajuste)

	ITEMS																													
M.	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
25%	X		X	X							X									X		X					X			7
50%	X	X	X	X							X	X							X		X	X	X		X	X	X	X	X	14
75%	X	X	X	X						X	X	X								X		X	X	X	X	X	X	X	X	15
100%	X	X	X	X						X	X	X								X	X		X	X	X	X	X	X	X	16
	4	3	4	4	0	0	0	0	0	2	4	3	0	0	0	0	0	0	1	1	4	0	3	4	3	2	3	4	0	3

tran CIT positivos, salvo el ítem 4 en el tamaño muestral 25%, que tiene un valor de -0.04 y el ítem 3 en el tamaño muestral 50%, que tiene un valor de -0.01.

Los ítems 1, 3, 4, 11, 21, 24, 25, 27 y 28 se consideran desajustados en las cuatro submuestras. El ítem 2 muestra desajuste para los tamaños muestrales del 50%, 75% y 100%. Los ítems 10, 20 y 30 aparecen desajustados para dos tamaños muestrales. Los ítems 12, 23 y 29 sólo se desajustan con un determinado tamaño muestral (100%).

Para contrarrestar el efecto de la sensibilidad de los estadísticos de ajuste al tamaño muestral, al igual que se hizo con la estrategia IST, se van a considerar no ajustados los ítems que hayan mostrado desajuste en tres o más tamaños muestrales. De esta forma se considerarán desajustados, y se eliminarán, los 10 ítems siguientes: 1, 2, 3, 4, 11, 21, 24, 25, 27 y 28.

Etapas III: Ajuste final de los ítems seleccionados con los sujetos elegidos

Ajuste global de los ítems

Resultó adecuado en todos los casos, a excepción del 2 (para el valor de 'infit') y del 20 (para el valor de 'infit' y 'outfit'). Todas

las correlaciones son positivas y oscilan entre el valor mínimo de 0.31 y el máximo de 0.50 (ver tabla 4).

Al estudiar la distribución de los valores de ajuste, sus medias y sus desviaciones típicas observamos lo siguiente: 1) la distribución de los estadísticos indica un ajuste aceptable de los ítems, ya que al efectuar el contraste con la prueba de Kolmogorov-Smirnov para comprobar el ajuste de la distribución a la distribución normal, no podemos afirmar que esos datos no se distribuyan según una normal ($p \leq 0.802$ para los valores 'infit' y $p \leq 0.759$ para los valores 'outfit'); y, 2) los valores medios de los estadísticos 'infit' y 'outfit' están cercanos a 0 (0.09 y 0.07, respectivamente); con desviaciones típicas de 1.22.

Ajuste global de los sujetos

En este caso, el resultado de la prueba de Kolmogorov-Smirnov para comprobar si las distribuciones de los valores de 'infit' y 'outfit' siguen una distribución normal, fue de $p \leq 0.627$ y $p \leq 0.250$, respectivamente, significando que no podemos decir que sus distribuciones no sean normales. En cuanto a las medias y desviaciones típicas fueron 0.00 y 0.01 para el 'infit' y 1.09 y 1.05 para el 'outfit'.

Resultados que también apoyan la interpretación de un ajuste global adecuado con esta estrategia.

Conclusiones

En este trabajo se exponen dos estrategias de ajuste de los datos al Modelo de Crédito Parcial. En la comparación que se hace de dichas estrategias los resultados muestran diferencias en cuanto al número de ítems y sujetos que se seleccionan por su ajuste adecuado, si bien con ambas aproximaciones se consigue un ajuste de los datos al modelo adecuado. Partiendo de que las dos estrategias son equivalentes en cuanto a las operaciones que se hacen y que tan solo difieren en el orden de sus dos primeras etapas, podemos decir que, para el caso que nos ocupa, según la estrategia seguida se maximiza el número de ítems seleccionados por su ajuste o el número de sujetos. En este trabajo la estrategia IST maximiza el número de sujetos seleccionados por mostrar patrones de respuestas coherentes con el modelo (87.1%) frente a la estrategia SIT (71.5%). Por el contrario, y tal y como apuntábamos al principio, la estrategia SIT maximiza el número de ítems seleccionados por mostrar ajuste al modelo (66.67%) frente a la estrategia IST (56.67%).

No nos cabe la menor duda de que este estudio debería complementarse con un estudio donde se utilicen datos simulados, donde se hayan determinado de antemano las propiedades de los ítems. De esta forma se podrían corroborar y generalizar los re-

Tabla 2
Ajuste final con la Estrategia IST: 17 ítems y 633 sujetos
(Nota: CIT indica Correlación Ítem-Total)

Nº Ítem	Infit	Outfit	CIT
5	0,38	0,44	0,39
6	1,57	1,94	0,35
7	1,04	0,93	0,36
8	0,64	0,64	0,39
9	0,46	0,63	0,38
10	-1,41	-1,90	0,47
13	-1,38	-1,25	0,43
14	-1,04	-1,13	0,42
15	-0,90	-0,73	0,44
16	1,26	1,41	0,29
17	1,93	1,45	0,29
18	-0,40	-0,39	0,42
19	2,31	2,72	0,35
20	-0,93	-0,84	0,43
22	-0,13	-0,58	0,40
26	-1,01	-1,22	0,42
29	-0,71	-0,24	0,42
Media	0,10	0,11	0,39
Desv. T.	1,20	1,29	0,50

Tabla 3
Resumen del ajuste de los ítems para las distintas muestras: Estrategia SIT
(Nota: M. indica Muestras; X indica desajuste)

M.	ITEMS																														
	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	
25%	X		X	X						X	X									X			X	X		X	X				10
50%	X	X	X	X							X									X		X	X		X	X		X	X		10
75%	X	X	X	X							X								X	X		X	X		X	X		X	X		12
100%	X	X	X	X						X	X	X								X	X		X	X	X		X	X	X	X	16
	4	3	4	4	0	0	0	0	0	2	4	1	0	0	0	0	0	0	0	2	4	0	1	4	4	0	4	4	1	2	

Tabla 4
Ajuste final con la Estrategia SIT: 20 ítems y 578 sujetos
(Nota: CIT indica Correlación Item-Total)

Nº Ítem	Infit	Outfit	CIT
5	-0,04	-0,02	0,42
6	2,02	1,85	0,35
7	1,47	1,16	0,36
8	0,89	0,89	0,40
9	1,10	1,09	0,38
10	-0,81	-1,27	0,47
12	-1,37	-1,24	0,48
13	-0,41	-0,27	0,39
14	-0,63	-0,68	0,42
15	0,00	0,25	0,43
16	0,93	1,05	0,31
17	1,35	1,39	0,32
18	0,31	0,09	0,42
19	1,98	1,99	0,39
20	-1,30	-1,27	0,46
22	1,14	1,32	0,37
23	-1,38	-1,21	0,49
26	-0,38	-0,55	0,43
29	-0,93	-1,08	0,46
30	-2,07	-2,11	0,50
Media	0.09	0.07	0.41
Desv. T.	1.22	1.22	0.50

sultados aquí expuestos. De hecho, de repetirse estos resultados podríamos sacar algunas conclusiones importantes en función del objetivo para el que van a utilizarse las calibraciones de los ítems y las mediciones de los sujetos. Por ejemplo, si elaboramos un conjunto de ítems para medir una variable psicológica y el interés está centrado en conseguir el mayor número de ítems que midan dicha variable para, por ejemplo, crear un banco de ítems, y donde el interés no radica en obtener las medidas de los sujetos, utilizaríamos la estrategia SIT. Si la finalidad de los ítems elaborados no es tanto incluirlos en un banco de ítems como obtener una medida de los sujetos en la variable psicológica que pretenden medir, la estrategia que conseguiría un mayor número de medidas de los sujetos coherente con el modelo aplicado sería la IST.

Queremos apuntar también que los resultados obtenidos con el Modelo de Crédito Parcial deben extenderse a la aplicación de

cualquier modelo basado en la Teoría de Rasgo Latente, y por tanto, en este sentido se hace necesario investigar estas estrategias desde distintos modelos.

No queremos dejar de señalar que tanto el desajuste de los ítems como el de los sujetos nos deben hacer reflexionar sobre las posibles causas que originan tal desajuste, causas que ya indicaba O'Brien (1992) para los ítems, al apuntar bien problemas con la teoría psicológica en la que se basan los ítems o bien en el proceso en el que la teoría ha sido operativizada en los ítems o quizás en ambas. Fed Li y Olejnik (1997) reflexionan sobre estas causas para los sujetos desajustados, donde debería considerarse si dichos sujetos: 1) entienden los ítems erróneamente (Tatsuoka, 1984, 1985); 2) tienen una excepcional creatividad (Levine y Drasgow, 1982); 3) manifiestan diferencias culturales (Van Der Flier, 1982); 4) están haciendo trampa o defraudando (Mandsen, 1987); 5) muestran diferencias instruccionales (Padilla, 1994; Padilla, Pérez y González, 1998, 1999; Harnish, 1983; Miller, 1982); 6) están siendo afectados por ítems sesgados (Frary, 1982); 7) muestran deseabilidad social (Schmitt, Cortina y Whitney, 1993); o, 8) muestran ausencia de la variable medida (Reise y Waller, 1993).

Una novedad que se presenta en este estudio, y que ya se recogía en el trabajo de Rojas (1998), es una forma de controlar el efecto de sensibilidad de los estadísticos de ajuste al tamaño muestral con las dos estrategias seguidas para el ajuste. Así, se ha controlado este efecto dividiendo la muestra total en cuatro submuestras, donde se ha calculado por separado los estadísticos de ajuste a los ítems, siendo el criterio de selección de ítems para considerarlos ajustados que aparezcan los valores de dichos estadísticos adecuados en, al menos, dos submuestras. No obstante, esta novedosa estrategia es una forma, y no la única forma para intentar disminuir este efecto. Queremos indicar con ello que lo expuesto aquí no es más que un apunte para intentar solucionar el tan indeseable efecto de sensibilidad de los estadísticos de ajuste al tamaño muestral, cuestión olvidada en la literatura sobre el tema.

Agradecimientos

Queremos hacer un agradecimiento explícito al Profesor Vicente Ponsoda (Universidad Autónoma de Madrid) por algunas de las ideas y reflexiones sugeridas para la realización del presente trabajo.

Referencias

- Barbero, M.I. (1999). Desarrollos recientes de los modelos psicométricos de la teoría de respuesta a los ítems. *Psicothema*, 11, 195-210.
- Frary, R.B. (1982). A comparison of person fit measures. Paper presented at the *Annual Meeting of the American Educational Research Association*. New York.
- Fred Li, M. y Olejnik, S. (1997). The power of Rasch person-fit statistics in detecting unusual response patterns. *Applied Psychological Measurement*, 21, 215-231.
- Gorsuch, R.L. (1988). Psychology of religion. *Annual Review of Psychology*, 39, 201-221.
- Gustafsson, J. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33, 205-233.
- Hambleton, R.K. (1989). Principles and Selected applications of item response theory. En R.L. Linn (Ed.). *Educational measurement*. (pp. 147-200). New York: McMillan.
- Hambleton, R.K. (1990). Item response theory: introduction and bibliography. *Psicothema*, 2, 97- 107.
- Hambleton, R.K. y Murray, L. (1983). Some goodness of fit investigations for item response models. En R.K. Hambleton (Ed.). *Applications of item response theory*. (pp.71-94). Vancouver, B.C.: Educational Research Institute of British Columbia.
- Hambleton, R.K. y Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Kluwer Academic Publishers.
- Harnish, D.L. (1983). Item response patterns: applications for educational practice. *Journal of Educational Measurement*, 20, 191-206.

- Hojitink, H., Molenaar, I. y Post, W. (1994). *PARELLA. User's manual*. The Netherlands, Groningen: IEC ProGAMMMMA.
- Hood, R.W. Jr. (1970). Religious orientation and the reported religious experience. *Journal for the Scientific Study of Religion*, 9, 285-291.
- Levine, M.V. y Drasgow, F. (1982). Appropriateness measurement: review, critique and validating studies. *British Journal of Mathematical and Statistical Psychology*, 35, 42-56.
- López Pina, J.A. e Hidalgo, M.D. (1996). Bondad de ajuste y teoría de respuesta a los ítems. En J. Muñiz (Coor.). *Psicometría*. (pp. 643-703). Madrid: Universitat.
- Mandsen, H.S. (1987). *Utilizing Rasch analysis to detect cheating on language examinations*. ERIC Document Reproduction Service N° ED 287 284.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Masters, G.N. (1988a). The analysis of partial credit scoring. *Applied Measurement in Education*, 1, 279-297.
- Masters, G.N. (1988b). Partial credit model. En J.P. Keeves, (Ed.). *Educational research, methodology and measurement: an international handbook*. (pp. 292-297). Elmsford, N.Y.: Pergamon Press.
- Masters, G.N. y Wright, B.D. (1984). The essential process in a family of measurement models. *Psychometrika*, 49, 529-544.
- Masters, G.N. y Wright, B.D. (1997). The partial credit model. En W. J. van der Linden y R.K. Hambleton (Eds.). *Handbook of modern item response theory*. (pp. 101-121). New York: Springer-Verlag.
- Morales, P. (1988). *Medición de actitudes en psicología y educación*. San Sebastián: Ttattalo.
- O'Brien, M.L. (1992). Using Rasch procedures to understand psychometric structure in measures of personality. En M. Wilson (Ed.). *Objective measurement: theory into practice*. (pp. 61- 76). Norwood, NJ: Ablex Publishing Corporation.
- Padilla, J.L.; Pérez, C. y Gómez, A. (1998). La explicación del sesgo en los ítems de rendimiento. *Psicothema*, 10, 481-490.
- Padilla, J.L.; Pérez, C. y Gómez, A. (1999). La dimensionalidad del test y las diferencias instruccionales. *Psicothema*, 11, 183-193.
- Pargament, K.I., Kennell, J., Hathaway, W., Grevengoed, N. ; Newman, J. y Jones, W. (1988). Religion and the problem-solving process: three styles of coping. *Journal for the Scientific Study of Religion*, 27, 90-104.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. (Reprinted by the Chicago University Press, 1980).
- Reise, S.P. y Waller, N.G. (1993). Traitness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology*, 65, 143-151.
- Rojas, A.J. (1998). *Aplicación del Modelo de Crédito Parcial y Modelo de Escalas de Clasificación a la medición de actitudes*. Almería: Servicio de Publicaciones de la Universidad de Almería. [Edición en CD-ROM].
- Schmitt, N. Cortina, J.M. y Whitney, D.J. (1993). Appropriateness fit and criterion-related validity. *Applied Psychological Measurement*, 17, 143-150.
- Tatsuoka, K.K. (1984). Caution indices based on item response theory. *Psychometrika*, 49, 95-110.
- Tatsuoka, K.K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, 10, 55-73.
- Van Der Flier, H. (1982). Deviant response patterns and comparability of tests scores. *Journal of Cross-Cultural Psychology*, 13, 267-298.
- Wright, B.D. (1980). Afterword. En Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. (pp. ix-xxiii). Chicago: The University of Chicago Press.
- Wright, B.D. y Linacre J.M. (1992a). *A user's guide to BIGSTEPS*. Chicago: Mesa Press.
- Wright, B.D. y Linacre, J.M. (1992b). *BIGSTEPS ver. 2.2. Computer Program*. Chicago: Mesa Press.
- Wright, B.D. y Masters, G.N. (1982). *Rating scale analysis*. Chicago: Mesa Press.
- Wright, B.D. y Stone, M.H. (1979). *Best test design: Rasch measurement*. Chicago: Mesa Press.

Aceptado el 19 de julio de 1999