

# Contribución de los ítems a los patrones atípicos de respuesta: cálculo del índice de precaución modificado sin el ítem

Ester Boixadera<sup>1</sup>, Rebeca García-Rueda<sup>2</sup>, Eduardo Doval<sup>3</sup>, M. Dolors Riba<sup>3</sup>, Jordi Renom<sup>4</sup>, Marta Fuentes<sup>5</sup>

<sup>1</sup> Servei d'Estadística Aplicada, UAB, <sup>2</sup> Servei de Llengües, UAB

<sup>3</sup> Departament de Psicobiologia i de Metodologia de les Ciències de la Salut. UAB

<sup>4</sup> Departament de Metodologia de les Ciències del Comportament. UB.

<sup>5</sup> Departament de Psicologia Bàsica, Evolutiva i de l'Educació. UAB.

## 1. INTRODUCCIÓN

Ante una prueba de evaluación de conocimientos, es de esperar que las personas evaluadas contesten correctamente a las preguntas fáciles e incorrectamente a las difíciles. Ante esta pauta de respuestas plausible, el nivel de capacidad de la persona evaluada debería estar determinado por la dificultad de los ítems que marcan la zona de transición entre los aciertos y los errores. En ocasiones las respuestas siguen patrones atípicos (PAR). En estos casos, se puede poner en riesgo la validez de las puntuaciones de la prueba (Meijer y Sitjsma, 2001).

Conocer qué ítems contribuyen a que una pauta sea atípica resulta necesario para interpretar la puntuación total resultante y valorar si a partir de ella pueden extraerse conclusiones válidas acerca del nivel de conocimientos evaluados con la prueba.

## 2. OBJETIVO

Programar una rutina de R, basada en el paquete PerFit (Tendeiro, 2015), que permita identificar los ítems de la prueba que contribuyan a la presencia de PAR.

## 3. ÍNDICE DE PRECAUCIÓN MODIFICADO (IPM)

La rutina se basa en el cálculo del índice de Precaución Modificado (IPM), uno de los indicadores no paramétricos que mejor detecta la presencia de PAR (Karabatsos, 2003).

Propuesto por Harnish y Linn (1985), el índice se basa en la covariación del patrón de respuestas, codificado como 1 (respuesta correcta) y 0 (respuesta incorrecta) con el vector de dificultades de los ítems. Los ítems más difíciles contestados correctamente o los más fáciles respondidos incorrectamente, tienen un gran peso en la identificación de una pauta de respuesta como atípica.

Toma valores entre 0 y 1. Los valores altos (superiores a 0.30) indican PAR.

## 4. RUTINA de R

```
# Llamada a la librería PerFit (Tendeiro, 2015)
library(PerFit)
```

```
#Lectura de la matriz de datos
Datos<- read.csv2("Datos.csv", header=TRUE)
```

```
#Cálculo del índice de Precaución Modificado
IPM<-Cstar(Datos)$PFscores
```

```
resultados<-matrix(IPM,ncol=1)
```

```
# Bucle: Cálculo IPM sin el ítem
for(i in 1:ncol(Datos))
{
  item<-Cstar(Datos[,i])$PFscores
  vector<-matrix(c(item),ncol=1)
  resultados<-cbind(resultados,vector)
}
```

```
TablaRes<-data.frame(resultados)
names(TablaRes)<-paste("IPM-i",0:ncol(Datos))
names(TablaRes)[1]<-c("IPM")
```

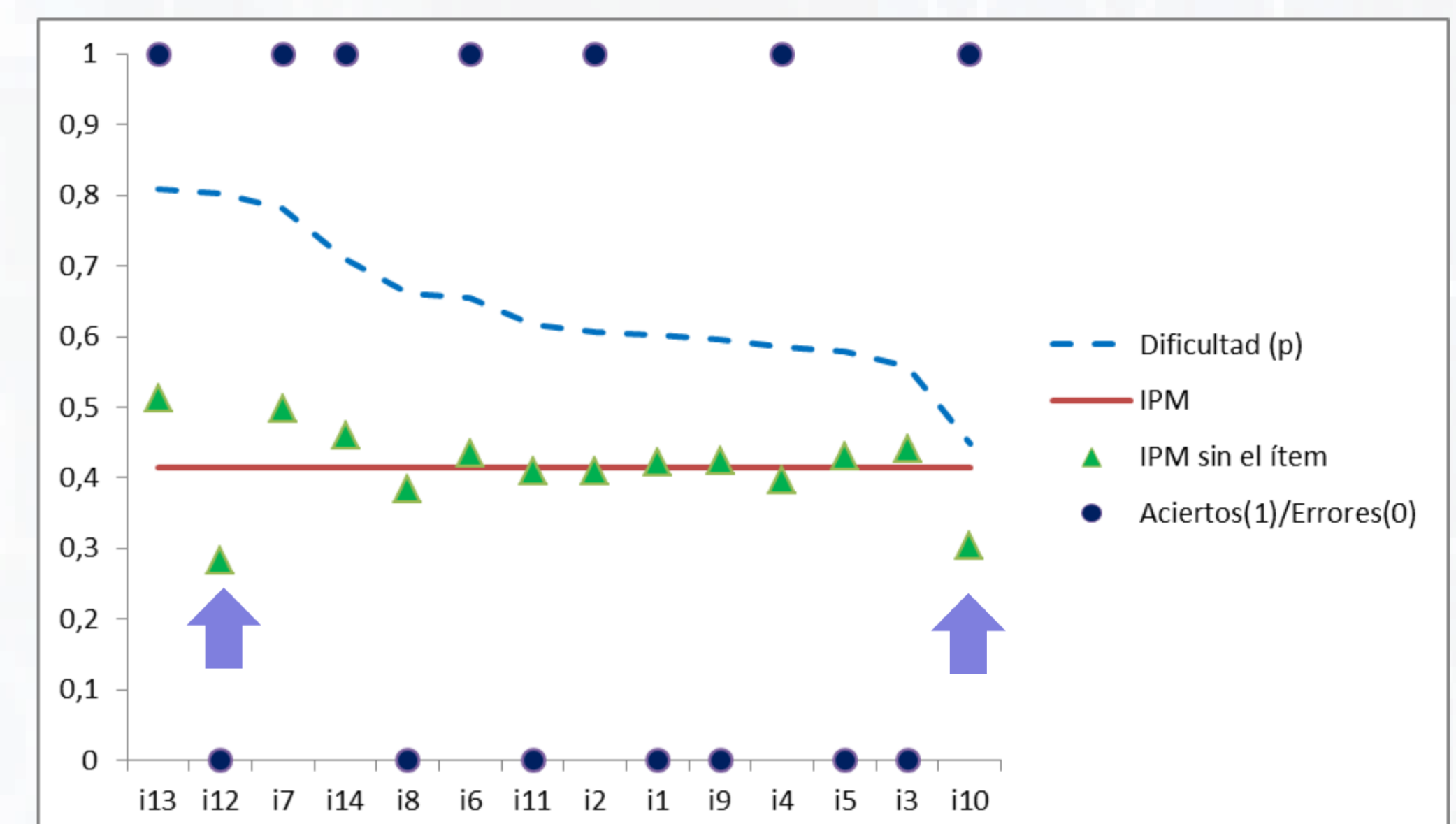
El archivo 'Datos.csv' contiene 290 respuestas (correctas=1, incorrectas=0) de una muestra aleatoria de 290 alumnos de 4º de la ESO que durante el curso 2014-2015 realizaron la prueba de competencias básicas en inglés. Los 14 ítems de este ejemplo evalúan la comprensión oral.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1	1	0	0	0	0	0	0	1	1	1	0	1	0
2	0	0	0	0	0	0	0	1	1	1	0	1	0	0
3	0	1	0	0	0	1	1	0	1	1	0	1	0	0
4	1	1	0	1	0	0	0	0	0	1	0	0	0	1
5	0	0	0	1	0	0	0	0	0	1	0	0	1	0
6	0	1	0	0	0	0	0	0	0	0	0	0	0	1
7	0	0	0	1	0	0	0	0	1	1	0	0	0	1
8	1	0	0	0	1	1	0	0	1	1	0	0	0	0
9	1	1	0	0	0	1	0	0	1	0	1	0	1	0
10	0	0	0	0	0	1	0	0	1	1	0	0	0	0

Resultados

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	IPM	IPM-1	IPM-2	IPM-3	IPM-4	IPM-5	IPM-6	IPM-7	IPM-8	IPM-9	IPM-10	IPM-11	IPM-12	IPM-13	IPM-14
2	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,11	0,08	0,08	0,08	0,06	0,08
3	0,23	0,23	0,23	0,24	0,23	0,23	0,23	0,23	0,00	0,23	0,30	0,23	0,21	0,40	0,23
4	0,23	0,23	0,23	0,24	0,23	0,23	0,23	0,31	0,23	0,23	0,28	0,04	0,13	0,32	0,23
5	0,69	0,69	0,69	0,71	0,69	0,59	0,69	0,66	0,72	0,61	0,83	0,69	0,65	0,65	0,69
6	0,50	0,37	0,38	0,55	0,38	0,37	0,53	0,48	0,53	0,50	0,60	0,50	0,44	0,46	0,53
7	0,53	0,53	0,53	0,55	0,38	0,37	0,53	0,48	0,53	0,53	0,63	0,53	0,73	0,46	0,53
8	0,28	0,28	0,28	0,29	0,28	0,05	0,28	0,38	0,28	0,28	0,34	0,28	0,39	0,18	0,28

Figura 1: Contribución de los ítems al IPM en una pauta de respuestas con 7 aciertos



## 5. RESULTADOS

La Figura 1 contiene información sobre la pauta de respuesta de una persona que ha contestado bien la mitad de los ítems de la prueba. Los ítems (eje de abscisas) están ordenados en función de su dificultad (representada por la línea discontinua azul) desde el más fácil al más difícil. Los puntos muestran las respuestas a cada ítem (1=respuesta correcta y 0=respuesta incorrecta). La línea roja señala el valor del Índice de Precaución Modificado de la pauta, que en este caso es 0.42). Los triángulos verdes muestran los valores que toma el índice IPM sin cada uno de los ítems.

Los ítems que más contribuyen a valores altos del IPM global son aquellos que tienen valores más bajos del IPM-sin el ítem. En este caso son la respuesta incorrecta al ítem 12 (un ítem fácil) y la respuesta correcta al ítem 10 (el ítem más difícil) las que contribuyen más al valor del IPM de esta pauta de respuestas.

## 6. CONCLUSIONES

En este caso se ha calculado el Índice de Precaución Modificado utilizando la función *Cstar* del paquete PerFit (Tendeiro, 2015). La rutina puede adaptarse a cualquier otro indicador de PAR implementado en el paquete PerFit con un simple cambio en la llamada de la función.

Este tipo de análisis resulta especialmente útil en pruebas con pocos ítems y bien diferenciados por su nivel de dificultad. Si el número de ítems es elevado, el cambio en el índice al no tener en cuenta un ítem será irrelevante.

El conocimiento de los ítems que están implicados en una pauta atípica tiene gran interés en las evaluaciones retroalimentativas, esto es, cuando se aprovecha la información que ofrece dicha pauta para proporcionar retroalimentación del proceso de aprendizaje y proporcionar elementos de mejora del mismo.

## 7. REFERENCIAS

- Harnisch, D. L. y Linn, R. L. (1981) Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement*, 18(3), 133-146.
- Karabatsos, G. (2003) Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics. *Applied Measurement In Education*, 16(4), 277-298.
- Meijer, R. R. y Sitjsma, K. (2001) Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25(2), 107-135.
- Tendeiro, J. (2015). Package "PersonFit"., v. 1.3.1. <http://cran.r-project.org/web/packages/PerFit/PerFit.pdf>

Este trabajo ha recibido el apoyo económico de la Dirección General de Investigación y Gestión del Plan Nacional de I+D+I, del Ministerio de Economía y Competitividad, con el proyecto EDU2013-41399-P.

Agradecemos al Consell Superior d'Avaluació de la Generalitat de Catalunya habernos facilitado los datos de este estudio.