

Identificación de tipologías de Patrones Atípicos de Respuesta en pruebas tipo test

Eduardo Doval y M. Dolors Riba

*Departament de Psicobiologia i de Metodologia de les Ciències de la Salut.
Facultat de Psicologia, Universitat Autònoma de Barcelona.*

eduardo.doval@uab.cat

dolors.riba@uab.cat

INTRODUCCIÓN

Las pruebas tipo **test** constituyen una forma **objetiva** e **imparcial** de evaluar **conocimientos** (Lane, Haladyna y Raymond, 2016, Haladyna y Rodríguez, 2013; Moreno, Martínez y Muñiz, 2015).

La **facilidad en la administración y corrección**, han hecho de las pruebas tipo test uno de los formatos elegidos en la mayoría de las evaluaciones a **gran escala** (PISA, TIMMS y PIRLS, a nivel internacional, o las pruebas de evaluación de las **competencias básicas** del Instituto Nacional de Evaluación Educativa o las diversas comunidades autónomas, a nivel nacional).

INTRODUCCIÓN

La forma más simple de obtener un **indicador del conocimiento** consiste en otorgar un punto a cada respuesta correcta (y cero puntos a las incorrectas), y **sumar** el número de puntos obtenidos.

Con el mismo criterio, la **proporción de respuestas correctas** de un grupo de alumnos evaluados a una determinada pregunta es un **indicador inverso** de su nivel de **dificultad** (a menor proporción de respuestas correctas más dificultad).

Cuando se infiere su nivel de conocimientos en la materia evaluada a partir de su puntuación, se está dando por hecho que ese alumno ha seguido una **pauta de respuestas coherente** con la dificultad de las preguntas.

INTRODUCCIÓN

Esta puntuación, sin embargo, puede obtenerse a partir de respuestas muy variadas.

	+ FACIL					+DIFICIL					Puntuación
	1	1	1	1	1	0	0	0	0	0	5
	1	1	1	1	0	1	0	0	0	0	5
	1	1	1	1	0	0	1	0	0	0	5
					.						
					.						
patrón atípico de respuesta → (PAR)	0	1	1	1	1	0	0	0	0	1	5
					.						
					.						
	0	0	0	0	0	1	1	1	1	1	5

INTRODUCCIÓN

Para la identificación de PAR se han propuesto numerosos índices (Karabatsos, 2003; Meijer y Sitjsma, 2001). La mayoría de ellos puede calcularse fácilmente con el paquete PerFit de R (Tendeiro, 2015).

Uno de los índices no paramétricos que ofrecen un buen rendimiento en la detección de PAR es el **Índice de Precaución Modificado (MCI)**; Harnisch y Linn (1981).

INTRODUCCIÓN

El MCI proporciona valores entre 0 (patrón de respuestas esperado) y 1 (patrón de respuesta completamente contrario al esperado), y puede considerarse que los valores superiores a 0.35 corresponden a presencia de PAR.

	+ FACIL					+DIFICIL					MCI
	1	1	1	1	1	0	0	0	0	0	0
	1	1	1	1	0	1	0	0	0	0	.
	1	1	1	1	0	0	1	0	0	0	.
											.
	-----										0.35
											.
patrón atípico de respuesta (PAR) →	0	1	1	1	1	0	0	0	0	1	.
											.
											.
	0	0	0	0	0	1	1	1	1	1	1

Igual que ocurre con los otros índices, el MCI permite identificar la presencia de PAR, pero no ofrece información acerca del tipo concreto de respuestas que ha dado lugar a ese patrón atípico.

OBJETIVO

Presentar una metodología de análisis que permita, no solo identificar la presencia de PAR, sino también su tipología.

MÉTODO

El método propuesto consiste en

- 1) Ordenar las preguntas de la prueba en función de su índice de dificultad (proporción de respuestas correctas),
- 2) Calcular los centiles 33 y 66 de los valores de dificultad de las preguntas y utilizarlos como puntos de corte para definir tres categorías de preguntas en función de su nivel de dificultad

baja, B			media, M				alta, A			
1	1	1	1	1	0	0	0	0	0	0
1	1	1	1	0	1	0	0	0	0	0
1	1	1	1	0	0	1	0	0	0	0
				.						
				.						
0	1	1	1	1	0	0	0	0	1	1
				.						
				.						
0	0	0	0	0	0	1	1	1	1	1

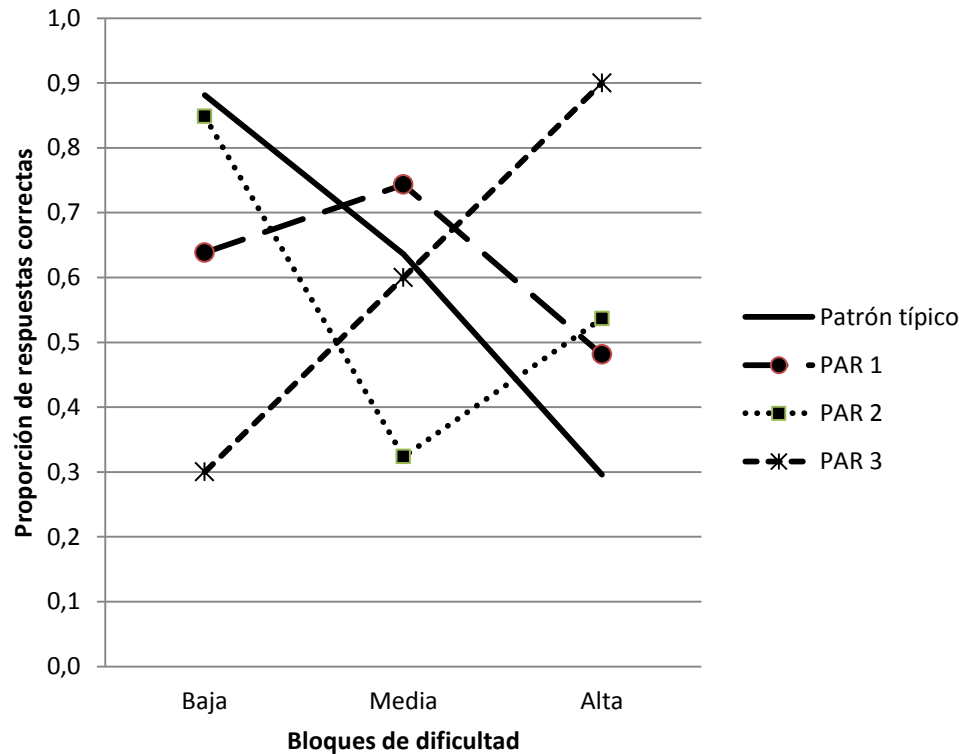
MÉTODO

El método propuesto consiste en

- 1) Ordenar las preguntas de la prueba en función de su índice de dificultad (proporción de respuestas correctas),
- 2) Calcular los centiles 33 y 66 de los valores de dificultad de las preguntas y utilizarlos como puntos de corte para definir tres categorías de preguntas en función de su nivel de dificultad
baja, B media, M alta, A
- 3) A nivel individual de cada estudiante, identificar los PAR con el índice MCI.
- 4) Calcular, para cada alumno, el porcentaje de respuestas correctas obtenido en cada bloque de preguntas (B-M-A, por ese orden), con el objeto de identificar su perfil de respuestas.

RESULTADOS

Patrones de respuesta



Un patrón no atípico debería tener un perfil descendente o plano, pero nunca creciente en ninguna de sus partes.

Los PAR estarán asociados a una tendencia creciente entre los bloques B-M (patrón 1) o entre los bloques M-A (patrón 2) o a una tendencia de respuestas correctas totalmente ascendente (patrón 3).

APLICACIÓN

Se analizan respuestas a una prueba realizadas por el Consell Superior d'Avaluació de Catalunya para comprobar las competencias lingüísticas básicas en lengua catalana.

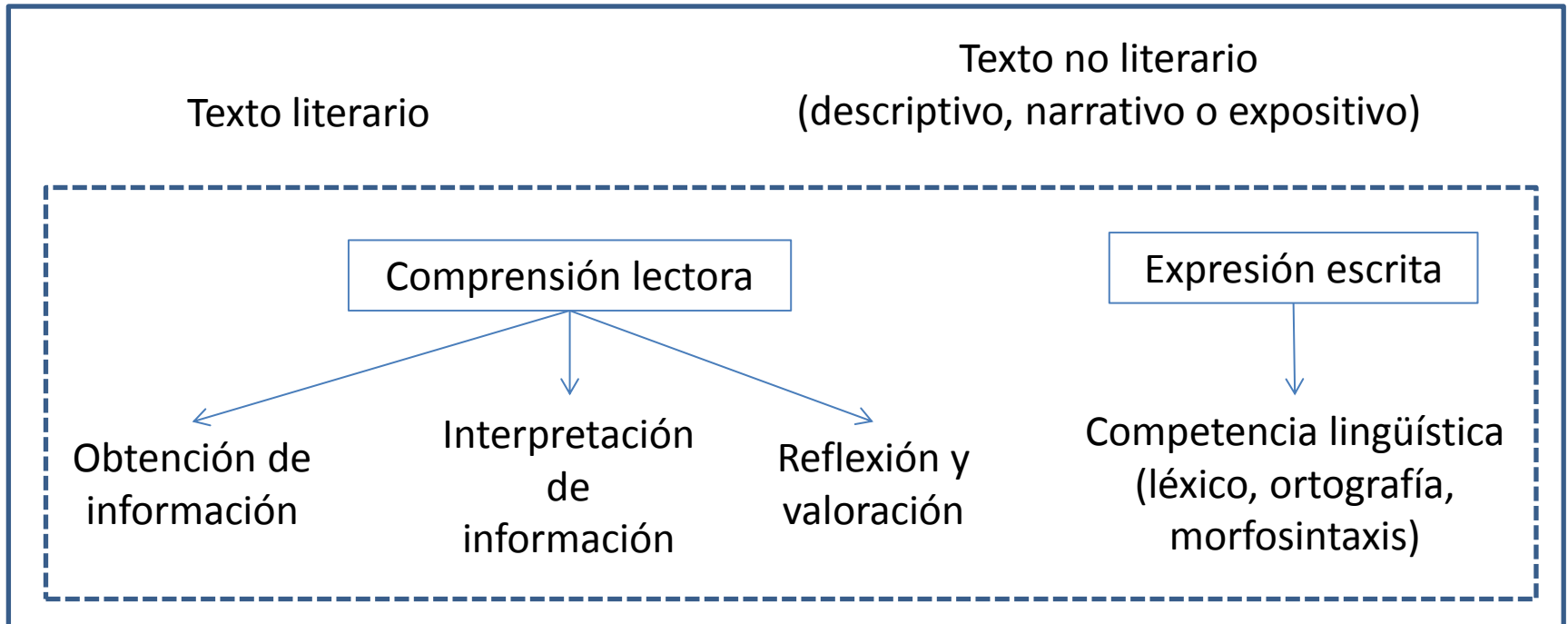
Población

Alumnado de 4º de ESO de Cataluña matriculado en el curso 2013-14.
64.505 alumnos de colegios públicos, concertados o privados.
55133 alumnos (85.5%) realizaron la prueba de castellano.

MÉTODO DE INVESTIGACIÓN

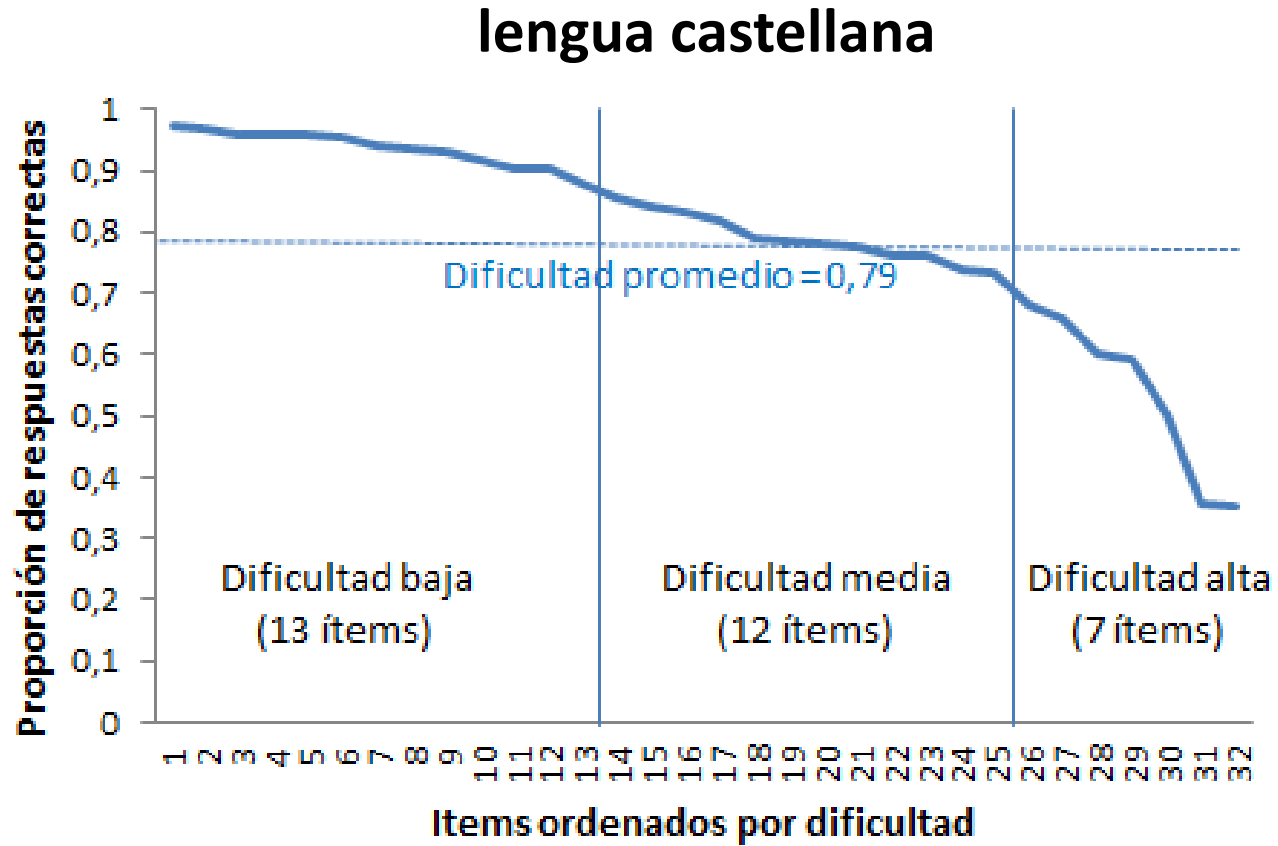
Instrumentos

Pruebas de competencias básicas en lengua catalana y castellana.



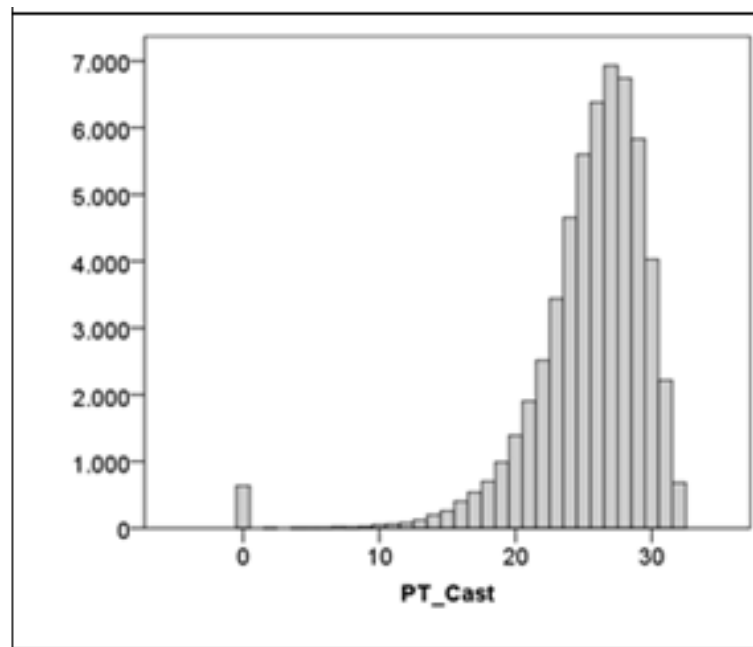
Ambas pruebas tienen 32 preguntas tipo test (4 opciones o V/F).

RESULTADOS



RESULTADOS

Puntuación total lengua castellana

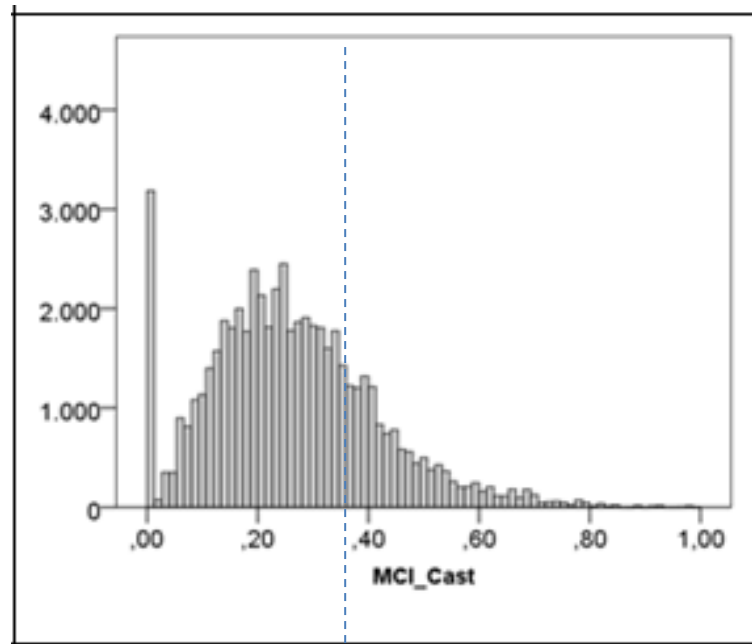


Media = 25,41

D.E. = 4,54

RESULTADOS

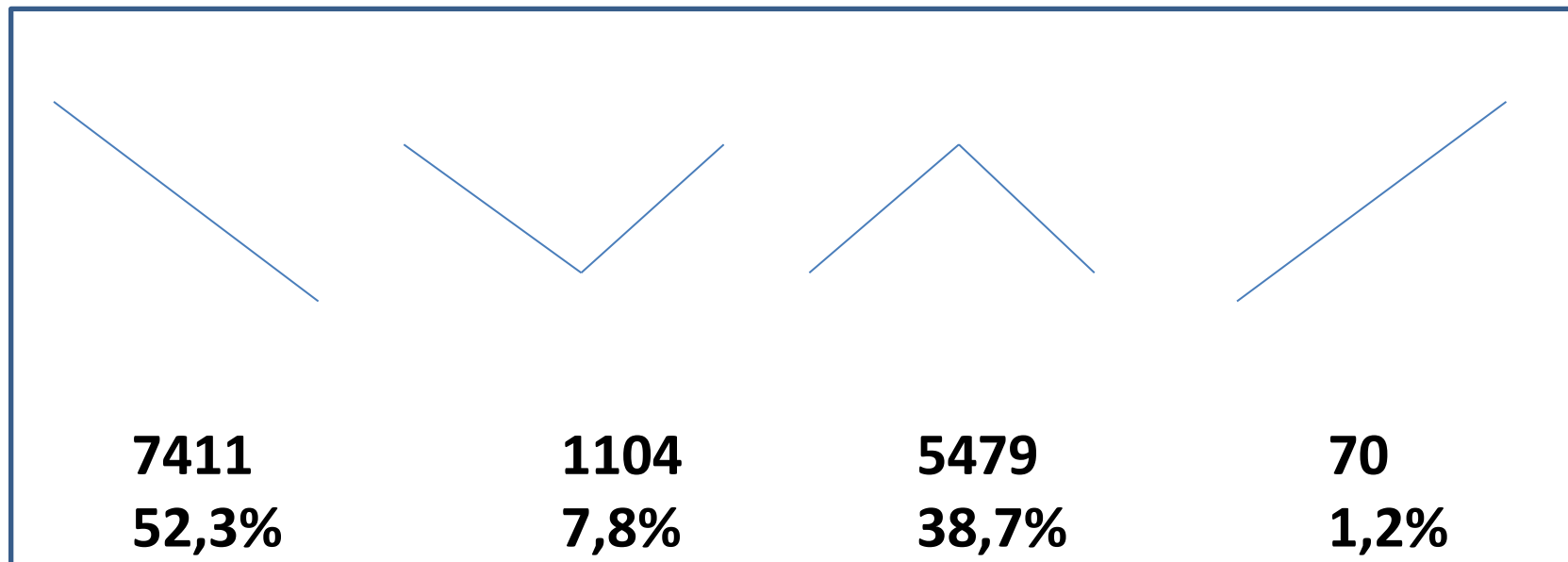
MCI lengua castellana



No PAR
40.969
94.3%

PAR
14.164
25.7%

RESULTADOS



SUS	9%	27,3%	51,3%	12,4%
APR	43,2%	12,6%	42,8%	1,5%
NOT	64 %	7%	28,5%	0,4%
EXC	38,7 %	6,6%	53%	1,7%

CONTRIBUCIÓN DEL TRABAJO

La identificación de los PAR es necesaria para realizar evaluaciones de los conocimientos más válidas (AERA, APA y NCME, 2014).

En caso de presencia de PAR, la identificación de las causas de atipicidad en las respuestas es imprescindible para ir más allá de una evaluación puramente sumativa y aprovechar la información que ofrece la manera en que se ha respondido a la prueba para realizar evaluaciones formativas.

CONTRIBUCIÓN DEL TRABAJO

Los patrones atípicos pueden estar relacionados con **estilos de aprendizaje**:

¿Por qué un alumno puede contestar correctamente las preguntas más difíciles y sin embargo responde incorrectamente a las más fáciles?

Los alumnos con un estilo de aprendizaje **reflexivo** pueden interpretar que las preguntas más fáciles son “demasiado fáciles” y buscar explicaciones alternativas que acaben siendo incorrectas.

¿Por qué un alumno puede contestar a las preguntas de dificultad baja y moderada de forma decreciente y sin embargo contestar bien a un elevado número de preguntas más difíciles?

Los estudiantes **pragmáticos** podrían haber dedicado mucho esfuerzo a estudiar aspectos difíciles y desatender los aspectos más fáciles

CONTRIBUCIÓN DEL TRABAJO

Por su sencillez conceptual y de aplicación, pretendemos que la metodología propuesta resulte de utilidad a los docentes que utilizan pruebas tipo test como método de evaluación.

REFERENCIAS

AERA, APA y NCME (2014). *The Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Haladyna, T.M., y Rodriguez, M.C. (2013). *Developing and validating test items*. New York, NY: Routledge

Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement*, 18, 133–46.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education* 16, 277–298.

Lane, S., Haladyna, T.M. y Raymond, M. (2016). *Handbook of test development* (2nd ed.). New York, NY: Routledge.

Meijer, R.R. and Sitjsma, K. (2001). Methodology Review: Evaluating Person Fit. *Applied Psychological Measurement*, 25(2), pp. 107–135.

Moreno, R., Martínez, R.J. y Muñiz, J. (2015). Guidelines based on validity criteria for the development of multiple choice items. *Psicothema*, 27(4), 388-394.

Tendeiro, J.N. (2015). Package ‘PerFit’ [Programa informático]. University of Groningen. Disponible en <http://cran.r-project.org/web/packages/PerFit>

MUCHAS GRACIAS POR LA ATENCIÓN

eduardo.doval@uab.cat

dolors.riba@uab.cat

Esta investigación se ha realizado gracias a la ayuda de la Dirección General de Investigación y Gestión del Plan Nacional de I+D+I, del Ministerio de Economía y Competitividad, a través del proyecto EDU2013-41399-P.

Agradecemos al Consell Superior d'Avaluació de Catalunya su colaboración en este proyecto.