

ANÁLISIS DE CUESTIONARIOS DE AUTOEVALUACIÓN EN LA ASIGNATURA DE ESTADÍSTICA EN PSICOLOGÍA. UNA APROXIMACIÓN AL ESTUDIO DE LAS CURVAS CARACTERÍSTICAS DE LOS ÍTEMS.^{1,2}

Joan Guàrdia Olmos^{3,4}, Maribel Però Cebollero⁴, Montserrat Freixa Blanxart⁵, Jaume Turbany Oset⁵ y Amàlia Gordóvil Merino⁵.
Universitat de Barcelona, España

Resumen

El objetivo del presente trabajo consiste en presentar el estudio del comportamiento psicométrico de un cuestionario de autoevaluación para estudiantes universitarios. Para ello se ha trabajado con una muestra de alumnos de la asignatura de *Análisis de Datos en Psicología* (Facultad de Psicología, Universidad de Barcelona) del curso académico 2008-2009. Los alumnos contestaron tres cuestionarios de respuesta cerrada, el número de alumnos que contestó el primer cuestionario fue de 413, el segundo de 390 y el tercero de 372. El estudio de la métrica de los ítems se realizó según la teoría de respuesta a los ítems. A partir de los análisis se puede concluir que el tercer cuestionario es el que presenta mejores índices de discriminación, en tanto que el segundo cuestionario es el que presenta una distribución más equilibrada de los ítems en cuanto a sus índices de dificultad en el continuo del constructo a medir.

Palabras clave: Rendimiento académico, estadística, teoría de respuesta al ítem.

Abstract

The main of this paper is to present a psychometric approach to item's analysis in a self-examination questionnaire in university students. We have worked with a sample of students in the course of Data Analysis in Psychology (School of Psychology, University of Barcelona) in the 2008-2009 academic year. Students answered three closed questionnaires; the number of students who answered the first questionnaire was 413, the second 390 and the third 372. The study of the metric of the items was made according to the item response theory. From the analysis we can conclude that the third questionnaire is the best in discrimination rates, while the second questionnaire is the one that presents a more balanced distribution of items in terms of their rates of difficulty in continuum of construct measured.

Key words: Academic performance, statistics, item response theory.

Introducción

El profesorado de estadística en titulaciones del ámbito de las ciencias sociales sabe del bajo rendimiento académico que obtienen los alumnos en esta materia, entendiendo por rendimiento académico el hecho de superar o no la materia. En concreto, es común encontrar en las diferentes facultades de psicología del territorio español que el porcentaje de alumnos que superan la asignatura de estadística en

¹ Este estudio fue subvencionado por el *Programa de Recerca en Docència Universitària (REDICE-08)* de la Universitat de Barcelona (Nº proyecto: A0801-11), y realizado por miembros del *Grup Consolidat d'Innovació Docent: Grup de desenvolupament per a l'aprenentatge de l'estadística en salut (Código: 2008GIDC-UB/24)* y el *Grup de Recerca Consolidat en Tècniques Estadístiques Avançades Aplicades a la Psicologia* (Código: 2009 SGR 388).

² Los autores agradecen a la Sra. Laia Farras Permanyer y a la Sra. Georgina España Abad por su colaboración en el procedimiento.

³ Correspondencia. Departament de Metodologia de les Ciències del Comportament. Facultat de Psicologia, Universitat de Barcelona. Passeig de la Vall d'Hebron, 171, 08035 Barcelona (Spain). Tel.: 933125090; Fax: 934021359; email: jguardia@ub.edu

⁴ Facultat de Psicologia. Departament de Metodologia de les Ciències del Comportament. Universitat de Barcelona. Institut de Recerca en Cervell, Cognició i Conducta (IR3C).

⁵ Facultat de Psicologia. Departament de Metodologia de les Ciències del Comportament. Universitat de Barcelona

primera convocatoria se sitúa alrededor del 60% en el mejor de los casos. Son diversos los trabajos que intentan determinar los factores que inciden en el rendimiento académico de los alumnos en asignaturas técnico-metodológicas, y particularmente en estadística. Por ejemplo, algunos trabajos estudian la relación del rendimiento académico con variables de tipo motivacional, de personalidad o factores psicosociales (Castejón, Navas & Sampascual, 1996; García & Fumero, 1998; Garrido & Rojo, 1996; Pérez-Sánchez & Castejón, 1996; Roces, González & Touron, 1997). Otros autores apuntan como factores relacionados con el rendimiento académico de los universitarios en asignaturas de tipo metodológico los antecedentes académicos de los alumnos, como por ejemplo, la nota de selectividad, el rendimiento en el bachillerato y la opción cursada en el mismo (Alvarado & García, 1999-2000; García, Alvarado & Jiménez, 2000), así como la asistencia y la participación en las clases presenciales (Alvarado & García, 1999-2000; García et al., 2000; Huberty, 2000). Gal y Ginsburg (1994) ó Garfield (1994) destacan la conveniencia de tener en cuenta las actitudes y creencias de los alumnos hacia la estadística, y Boyle (1999), Gardner y Hudson (1999) y Smith (1998) remarcan el interés de la realización de trabajos de tipo práctico en los que deben aplicarse diferentes técnicas estadísticas en el aprendizaje de la estadística.

Dada nuestra preocupación por el bajo rendimiento de los alumnos en la asignatura de *Análisis de Datos en Psicología*, son diversas las actuaciones que nuestro grupo de innovación docente ha realizado a lo largo de los años. Entre ellas cabe destacar la creación de una asignatura de libre elección desde el curso académico 1999-2000, *Fonaments matemàtics* (Fundamentos Matemáticos), con el fin de suplir el déficit que tienen los alumnos que proceden de bachilleratos de corte más social o humanístico. Algunas actuaciones se han enfocado en el estudio de los factores que influyen en el rendimiento académico de los alumnos en esta asignatura (Guàrdia, et al., 2002; Guàrdia, et al., 2006), aunque en este caso, los resultados de estos trabajos muestran el conocido efecto de dispersión que se acostumbra a encontrar en el estudio del rendimiento académico. Otras actuaciones más activas han implicado la elaboración de material para el alumnado en formato electrónico como puede ser la elaboración del CD-Rom de estadística descriptiva (Guillén et al., 2001), cuya eficacia en el aprendizaje de la estadística descriptiva queda recogida en el trabajo de Però et al. (2004). En los últimos años la acción que se está realizando en este ámbito se ha centrado en una actuación más directa sobre la docencia realizada con el alumnado, en concreto, se ha generado en el aula una dinámica de trabajo colaborativo en grupos de cuatro alumnos a la hora de resolver diferentes problemas que los profesores proponen a lo largo del curso; esta estrategia ha implicado un aumento en el rendimiento académico de los alumnos, siendo el porcentaje de alumnos que superan la asignatura en primera convocatoria de aproximadamente un 70% (Guàrdia, Freixa, Turbany & Però, 2008 y Guàrdia, Però, Freixa, Turbany & Gordóvil, en prensa 2008). Ahora bien, un aspecto que no se ha resuelto aún consiste en la promoción del trabajo autónomo del alumno, aspecto muy importante a tener en cuenta en la implementación que se deriva de la lógica del Espacio Europeo de Educación Superior. Es por este motivo que durante el curso académico 2008-2009 se han puesto en funcionamiento diferentes estrategias para fomentar el trabajo autónomo del alumno, entre ellas cabe destacar la generación de un glosario en el campus virtual de la asignatura, el enlace al observatorio *Pythia*, repositorio de páginas WEB estáticas y dinámicas de diferentes conceptos desarrollados en la asignatura (<http://www1.ub.edu/gid-estadistica/enllacos.htm>) a las que los alumnos podían entrar a lo largo del curso, con el fin de profundizar en los contenidos explicados en la asignatura de *Análisis de Datos en Psicología*, y una serie de pruebas tipo test para la preparación de la asignatura (Guàrdia et al., 2009).

En este punto, cabe comentar que los procesos de autoevaluación en el aprendizaje autónomo de los alumnos es una cuestión aún pendiente; en el presente trabajo se pretende mostrar un estudio psicométrico de un cuestionario de autoevaluación generado a partir de la métrica psicológica necesaria para poder asegurar la adecuación de sus ítems, mediante la aplicación de la teoría de respuesta a los ítems (TRI) (Muñiz, 1997). La elaboración de cuestionarios de autoevaluación con una métrica adecuada puede implicar una gran ayuda en el fomento del trabajo autónomo del alumno, ya que a partir de estos cuestionarios los alumnos pueden determinar su nivel de conocimiento de la materia. La elección de la

TRI como vehículo propicio viene justificada no sólo en las características propias de la técnica de análisis, sino también por la escasa utilización que en este tipo de pruebas de rendimiento se ha hecho de la TRI. Existen antecedentes en el ámbito académico (Guàrdia et al, 2009), pero no dirigidos a pruebas o sistemas de autoevaluación en el ámbito de la estadística en las titulaciones universitarias.

Método

Participantes

La muestra estudiada estuvo formada por 447 estudiantes matriculados en la asignatura *Análisis de Datos en Psicología* de la Facultad de Psicología de la Universidad de Barcelona. El total de grupos de matrícula de la asignatura durante el curso académico 2008-2009 fue de 9, pero en el presente trabajo sólo se incluyeron los alumnos matriculados en los 5 grupos en que se implementó la estrategia para fomentar el trabajo autónomo del alumno. Del total de alumnos estudiados un 18,2% eran varones, la media de edad de los participantes fue de 19.8 años con una desviación típica de 3.4 años, y un 30.2% de los alumnos matriculados en estos cinco grupos habían cursado la asignatura en cursos anteriores.

Instrumentos

Los instrumentos utilizados fueron tres cuestionarios tipo test de selección múltiple con única respuesta. En concreto, las opciones de respuesta de cada ítem eran cuatro, de las que sólo una era correcta. El primer cuestionario estaba formado por un total de 30 ítems, el segundo cuestionario por 28 ítems, y el tercer cuestionario estaba compuesto de 15 ítems. Cada cuestionario hacía referencia a cada uno de los bloques en que se había estructurado el temario de la asignatura, así pues, el primer cuestionario hacía referencia a aspectos vinculados básicamente con estadística descriptiva, el segundo cuestionario hacía referencia a aspectos relacionados con la distribución normal, la distribución muestral y la estimación por intervalo, y finalmente, el tercer cuestionario valoraba aspectos vinculados con las pruebas de hipótesis bivariadas.

Procedimiento

La docencia de la asignatura *Análisis de Datos en Psicología* durante el curso académico 2008-2009 en los cinco grupos de matrícula estudiados se estructuró en una clase de docencia magistral a la semana y dos clases de sesiones de trabajo colaborativo a la semana. El temario de la asignatura estaba estructurado en tres bloques, tal como se ha comentado en el apartado de instrumentos; para cada bloque del temario se realizaba una prueba de evaluación presencial. La del primer bloque se realizó el 13 o 14 de octubre dependiendo del grupo de matrícula, la prueba del segundo bloque se realizó el 3 o 4 de noviembre dependiendo del grupo de matrícula, y la prueba del tercer bloque se realizó el 19 de diciembre para los cinco grupos de matrícula. Con el fin de facilitar al alumnado su trabajo autónomo, para cada uno de los bloques del temario se activó en el campus virtual de la asignatura el cuestionario de autoevaluación explicado en el apartado de instrumentos. Estos cuestionarios estaban a disposición de los alumnos una semana antes de la prueba de evaluación presencial, y los podían entregar de forma voluntaria al profesorado el día de la prueba de evaluación, con el fin de que formaran parte de los criterios de evaluación final de la asignatura. El primer cuestionario fue contestado por un total de 413 alumnos, el segundo por un total de 390 alumnos, y el tercer cuestionario fue contestado por un total de 372 alumnos. La recogida de la información siguió la normativa ética necesaria.

Análisis de datos

El análisis de datos para cada uno de los cuestionarios consistió en el estudio de las curvas características de los ítems a partir de la teoría de respuesta al ítem (TRI). Tal como señalan García-Cueto (1993), García-Cueto & Fidalgo (2005) y Muñiz (1997), la gran contribución de la TRI consiste en la posibilidad de obtener mediciones de los constructos invariantes tanto por lo que respecta a los instrumentos utilizados como de los sujetos implicados. Los modelos de TRI asumen que existe una

relación funcional entre los valores de la variable que miden los ítems y la probabilidad de acertar los ítems, esta función recibe el nombre de curva característica de los ítems (CCI). La CCI queda definida por tres parámetros, el parámetro a que representa el índice de discriminación, el parámetro b que representa el índice de dificultad, y el parámetro c que representa la probabilidad de acertar el ítem al azar. La función matemática que se suele utilizar para modelar las CCI suele ser la logística. A continuación se muestran las expresiones matemáticas de la función logística para los modelos de un parámetro (b) o modelo de Rasch, de dos parámetros (a y b), y de tres parámetros (a , b y c):

$$\text{Modelo de un parámetro: } p_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

$$\text{Modelo de dos parámetros: } p_i(\theta) = \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}}$$

$$\text{Modelo de tres parámetros: } p_i(\theta) = c_i + (1 - c_i) \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}}$$

Donde: $p_i(\theta)$ es la probabilidad de acertar el ítem para un valor determinado de competencia (θ)

θ son los valores de la variable medida

e es la base de los logaritmos neperianos

D es una constante

a_i es el índice de discriminación del ítem i

b_i es el índice de dificultad del ítem i

c_i es el valor de $p_i(\theta)$ para un valor de θ igual a $-\infty$

En el presente trabajo se decidió ajustar modelos de tres parámetros en los tres cuestionarios, dado que al tratarse de pruebas de rendimiento es importante tener en cuenta la posibilidad de acertar por azar el ítem. Sin embargo, en el análisis del segundo cuestionario no se llegó a la convergencia en la solución propuesta inicialmente, y por tanto se ajustó el modelo de dos parámetros. El método de estimación de parámetros en los tres casos ha sido la estimación máximo verosímil condicional. Todos los análisis fueron llevados a cabo con el software Bilog-MG 3 para Windows. En todos los casos, se definió la estimación de parámetros a partir de un máximo de treinta ciclos bajo el algoritmo de Newton. Se utilizaron distribuciones probabilísticas a priori basadas, en todos los casos, en la corrección del error por azar y se fijó una primera solución para el parámetro b derivada de los resultados que, en pruebas análogas, habían obtenido otras cohortes de estudiantes en cursos anteriores. No fue necesaria en ningún caso la utilización de sistemas de reestimación en los parámetros, puesto que en el único caso de no convergencia, el modelo de dos parámetros resultó suficiente y eficiente para el objetivo estadístico previsto.

Resultados

Inicialmente se analizaron las condiciones de aplicación de la TRI para cada uno de los cuestionarios, obteniéndose resultados favorables al cumplimiento de tales condiciones. En la tabla 1 se muestra la estimación de los parámetros del modelo de tres parámetros para los ítems del primer cuestionario, en tanto que en la figura 1 se muestra la curva característica de los 30 ítems. Como se muestra en la tabla 1, el modelo de tres parámetros estimado en los 30 ítems ajusta a los datos tal como se aprecia en el test de χ^2 , a excepción de los ítems 15 y 18 ($\chi^2=25$; g.l.=9; $p=.003$ y $\chi^2=17.9$; g.l.=9; $p=.036$ respectivamente). Por otra parte, los tres parámetros estimados son estadísticamente significativos en los 30 ítems (a excepción del parámetro b en los ítems 6, 12, 16, 21, 23, 24, 26, 27 y 29), así como las saturaciones factoriales de cada ítem para un modelo unifactorial. La convergencia se ha conseguido tras 16 ciclos o iteraciones. Los valores de las saturaciones factoriales oscilan entre .236 y .860.

Tabla 1.
Resumen de los parámetros del modelo de tres parámetros para los 30 ítems del primer cuestionario

Ítem	Constante (E.S.)	Carga factorial (E.S.)	Parámetro a (E.S.)	Parámetro b (E.S.)	Parámetro c (E.S.)	χ^2 (g.l.) <i>p</i>
1	1.078*	0.621*	0.792*	-1.361*	0.217*	4.3 (8)
	0.227	0.136	0.174	0.402	0.094	.831
2	-0.934	0.807*	1.367*	0.683*	0.265*	10.2 (9)
	0.529	0.239	0.404	0.243	0.072	.335
3	-1.275*	0.516*	0.602*	2.119*	0.262*	11.0 (9)
	0.537	0.175	0.204	0.704	0.079	.275
4	2.932*	0.864*	1.716*	-1.709*	0.187*	7.7 (5)
	0.364	0.170	0.338	0.237	0.084	.177
5	1.492*	0.595*	0.740*	-2.018*	0.205*	2.7 (8)
	0.216	0.138	0.172	0.504	0.091	.951
6	-0.115	0.731*	1.071*	0.107	0.221*	3.8 (9)
	0.318	0.158	0.231	0.284	0.083	.925
7	-1.590*	0.762*	1.178*	1.350*	0.117*	5.3 (8)
	0.458	0.212	0.328	0.208	0.047	.726
8	1.242*	0.822*	1.443*	-0.861*	0.213*	8.2 (7)
	0.257	0.167	0.294	0.222	0.090	.315
9	2.614*	0.742*	1.108*	-2.359*	0.195*	2.7 (6)
	0.281	0.156	0.233	0.424	0.088	.848
10	3.514*	0.771*	1.213*	-2.898*	0.199*	1.1 (4)
	0.454	0.248	0.390	0.690	0.089	.900
11	2.547*	0.847*	1.595*	-1.596*	0.206*	3.1 (5)
	0.339	0.182	0.343	0.254	0.089	.683
12	0.355	0.660*	0.879*	-0.404	0.172*	9.4 (9)
	0.224	0.120	0.160	0.285	0.077	.404
13	1.812*	0.817*	1.417*	-1.279*	0.194*	3.7 (7)
	0.255	0.150	0.260	0.229	0.085	.816
14	3.419*	0.860*	1.688*	-2.026*	0.205*	7.8 (4)
	0.605	0.278	0.545	0.374	0.091	.098
15	0.837*	0.525*	0.616*	-1.358*	0.217*	25.0 (9)
	0.226	0.119	0.139	0.493	0.094	.003
16	-0.411	0.721*	1.040*	0.395	0.191*	13.6 (9)
	0.329	0.164	0.236	0.267	0.077	.136
17	-2.477*	0.728*	1.061*	2.334*	0.124*	10.1 (9)
	0.692	0.297	0.433	0.501	0.039	.339
18	-1.133*	0.814*	1.402*	0.808*	0.141*	17.9 (9)
	0.407	0.195	0.336	0.171	0.052	.036
19	0.860*	0.769*	1.204*	-0.714*	0.180*	11.8 (8)
	0.222	0.141	0.221	0.230	0.080	.159
20	1.069*	0.573*	0.699*	-1.528*	0.209*	10.6 (9)
	0.217	0.118	0.144	0.443	0.091	.306
21	0.086	0.236*	0.684*	-0.126	0.564*	15.7 (9)
	0.294	0.095	0.161	0.442	0.133	.075
22	1.637*	0.849*	1.606*	-1.020*	0.190*	6.0 (7)
	0.259	0.158	0.299	0.195	0.083	.535
23	-0.109	0.625*	0.801*	0.136	0.192*	10.8 (9)
	0.280	0.131	0.168	0.336	0.082	.288
24	0.227	0.717*	1.027*	-0.221	0.225*	10.1 (9)
	0.287	0.150	0.215	0.302	0.090	.339
25	0.623*	0.765*	1.189*	-0.524*	0.204*	2.8 (8)
	0.248	0.146	0.226	0.250	0.086	.947
26	0.359	0.773*	1.218*	-0.295	0.176*	3.7 (7)
	0.242	0.143	0.225	0.222	0.077	.809
27	0.281	0.512*	0.597*	-0.471	0.205*	10.7 (9)
	0.245	0.113	0.131	0.451	0.089	.300
28	-1.288*	0.722*	1.042*	1.236*	0.215*	3.1 (9)
	0.533	0.231	0.334	0.279	0.068	.962
29	0.404	0.683*	0.936*	-0.432	0.204*	10.0 (8)
	0.248	0.136	0.187	0.303	0.086	.266
30	0.984*	0.715*	1.023*	-0.962*	0.201*	7.1 (9)
	0.227	0.135	0.194	0.288	0.088	.623

Con * se han marcado todos los parámetros estadísticamente significativos con un nivel de significación inferior a .05.

E.S.: error estándar del parámetro estimado; χ^2 : valor del estadístico chi cuadrado como prueba de bondad de ajuste, g.l.: grados de libertad, y *p*: nivel de significación.

Por lo que respecta al parámetro c (posibilidad de acertar por azar el ítem), en general su valor se sitúa alrededor de .2, inferior al valor teórico esperado de .33. Esta estimación se deriva de la teoría axiomática de la probabilidad a partir de la equiprobabilidad de las tres opciones de respuesta en las que se puede dar respuesta al azar, sus valores oscilan entre .117 y .564. Tal como se observa en la tabla 1, en general, los ítems no presentan una gran dificultad (valores entre -2.898 y 2.334), de hecho se puede afirmar que son bastante fáciles, es decir, requieren de pocas habilidades y/o conocimientos estadísticos (coeficiente b con valores negativos), sólo en 15 de los 30 ítems se podría decir que presentan una dificultad intermedia (ítems 2, 6, 8, 12, 16, 18, 19, 21, 23, 24, 25, 26, 27, 29 y 30), y que 3 de los 30 ítems presentan una dificultad alta (ítems 3, 7 y 17), especialmente el ítem 17, cuyo índice de dificultad es de 2.334. La valoración cualitativa que se hace en este apartado se basa tanto en la significación de los parámetros como a partir del signo de los mismos, debiéndose entender los calificativos alto o bajo como mera descripción general de los parámetros, en ningún caso como criterio estadístico estricto. De hecho, a nivel gráfico se puede apreciar rápidamente este hecho, dado que los ítems de dificultad intermedia presentan una curva característica en la que el punto de inflexión se produce aproximadamente en el punto intermedio de la escala del constructo medido (θ), en tanto que para los ítems de dificultad alta este punto de inflexión se produce en los valores positivos del constructo medido (θ). El resto de ítems presentan una dificultad baja, cabe destacar los ítems 5, 9, 10 y 14 por su extrema facilidad, aspecto que se puede observar en las CCI que se muestran muy próximas al cuadrante superior izquierdo, ya que prácticamente todos los alumnos aciertan estos ítems (figura 1). Finalmente, para el índice de discriminación (parámetro a) los valores estimados oscilan entre .602 y 1.716, en 20 de los 30 ítems el valor del parámetro es superior a 1, lo que indicaría que la pendiente de la CCI es elevada y por tanto los ítems presentan una adecuada discriminación (ítems 2, 4, 6, 7, 8, 9, 10, 11, 13, 14, 16, 17, 18, 19, 22, 24, 25, 26, 28 y 30).

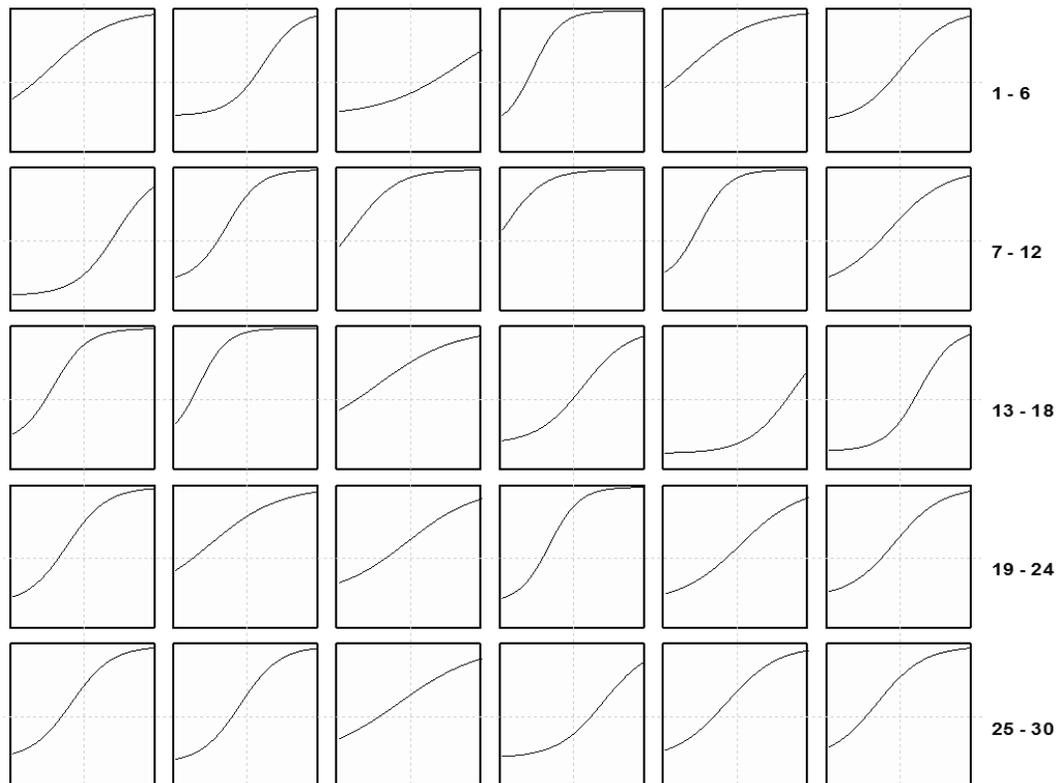


Figura 1. Curvas características de los 30 ítems del primer cuestionario (modelo de 3 parámetros).

Tabla 2.
Resumen de los parámetros del modelo de dos parámetros para los 28 ítems del segundo cuestionario

Ítem	Constante (E.S.)	Carga factorial (E.S.)	Parámetro a (E.S.)	Parámetro b (E.S.)	χ^2 (g.l.) p
1	-0.234* 0.103	0.319* 0.080	0.337* 0.084	0.693* 0.341	17.1 (9) .047
2	-1.679* 0.137	0.296* 0.086	0.310* 0.090	5.416* 1.626	13.4 (8) .099
3	-1.328* 0.135	0.712* 0.117	1.015* 0.166	1.309* 0.226	17.0 (7) .017
4	1.531* 0.179	0.846* 0.141	1.584* 0.264	-0.966* 0.128	4.8 (5) .438
5	3.112* 0.335	0.886* 0.150	1.914* 0.324	-1.626* 0.174	3.2 (4) .525
6	0.074 0.115	0.731* 0.121	1.070* 0.178	-0.069 0.106	12.1 (6) .060
7	3.418* 0.350	0.835* 0.168	1.520* 0.305	-2.249* 0.319	5.8 (6) .446
8	3.346* 0.329	0.821* 0.160	1.438* 0.280	-2.328* 0.329	4.0 (6) .676
9	-0.725* 0.119	0.701* 0.118	0.984* 0.166	0.737* 0.159	7.7 (6) .260
10	1.392* 0.140	0.684* 0.122	0.937* 0.168	-1.487* 0.255	4.5 (6) .613
11	0.395* 0.121	0.790* 0.128	1.287* 0.209	-0.307* 0.098	4.4 (6) .627
12	-0.446* 0.106	0.392* 0.091	0.427* 0.098	1.045* 0.329	0.7 (8) .999
13	-0.426* 0.106	0.420* 0.093	0.463* 0.103	0.921* 0.301	7.3 (8) .505
14	-0.171 0.108	0.574* 0.107	0.701* 0.131	0.244* 0.163	17.9 (8) .022
15	0.777* 0.136	0.802* 0.128	1.342* 0.214	-0.579* 0.104	12.5 (6) .052
16	0.719* 0.130	0.817* 0.135	1.418* 0.233	-0.507* 0.106	5.3 (6) .505
17	-0.247* 0.103	0.370* 0.087	0.399* 0.094	0.620* 0.289	15.2 (9) .086
18	-1.124* 0.134	0.709* 0.126	1.006* 0.179	1.117* 0.194	3.5 (7) .831
19	-0.996* 0.114	0.303* 0.084	0.319* 0.088	3.126* 0.907	20.9 (8) .007
20	-1.052* 0.121	0.486* 0.112	0.556* 0.128	1.891* 0.434	6.0 (7) .541
21	0.976* 0.129	0.735* 0.116	1.085* 0.172	-0.900* 0.156	8.6 (6) .199
22	-1.489* 0.144	0.687* 0.117	0.946* 0.161	1.573* 0.256	4.9 (7) .675
23	-0.100 0.105	0.473* 0.093	0.536* 0.105	0.187 0.196	25.6 (9) .002
24	1.506* 0.161	0.828* 0.120	1.476* 0.215	-1.020* 0.135	12.1 (6) .059
25	-0.606* 0.110	0.530* 0.107	0.625* 0.126	0.969* 0.252	21.5 (7) .003
26	1.488* 0.151	0.728* 0.128	1.061* 0.187	-1.402* 0.222	3.8 (7) .808
27	1.396* 0.168	0.843* 0.135	1.567* 0.251	-0.891* 0.121	5.5 (5) .356
28	-1.383* 0.149	0.747* 0.127	1.123* 0.192	1.232* 0.191	10.8 (6) .094

Con * se han marcado todos los parámetros estadísticamente significativos con un nivel de significación inferior a .05.

E.S.: error estándar del parámetro estimado; χ^2 : valor del estadístico chi cuadrado como prueba de bondad de ajuste, g.l.: grados de libertad, y p: nivel de significación.

En la tabla 2 se muestran el valor de los parámetros estimados para los ítems del segundo cuestionario, y en la figura 2 se muestran las respectivas curvas características para modelos de 2 parámetros. En 6 de los 28 ítems el modelo estimado no ajusta a los datos, se trata de los modelos estimados en los ítems 1 ($\chi^2=17.1$; $g.l.=9$; $p=.047$), 3 ($\chi^2=17.0$; $g.l.=7$; $p=.017$), 14 ($\chi^2=17.9$; $g.l.=8$; $p=.022$), 19 ($\chi^2=20.9$; $g.l.=8$; $p=.007$), 23 ($\chi^2=25.6$; $g.l.=9$; $p=.002$) y 25 ($\chi^2=21.5$; $g.l.=7$; $p=.003$). Para los 28 modelos parciales ajustados, todos los parámetros estimados han resultado estadísticamente significativos a excepción del parámetro b en los ítems 6 y 23. En el caso de las saturaciones factoriales de los ítems a un modelo unifactorial los valores oscilan entre .296 y .886.

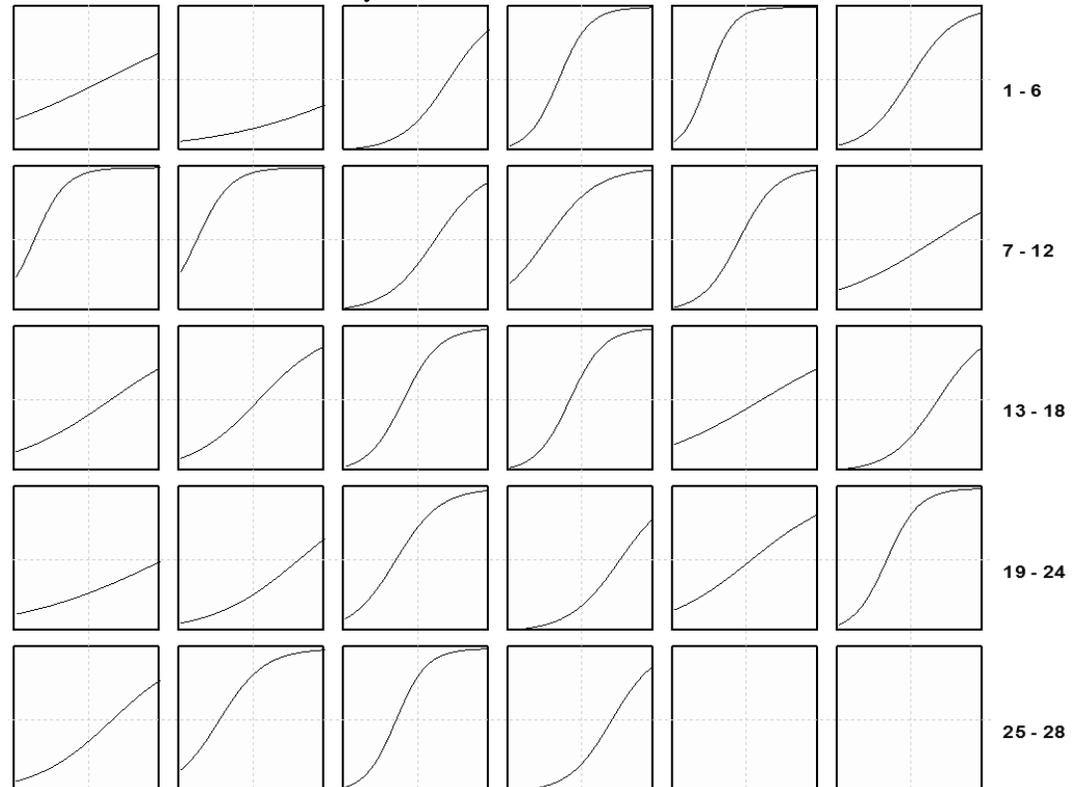


Figura 2. Curvas características de los 28 ítems del segundo cuestionario (modelo logístico de 2 parámetros).

Por lo que respecta al índice de dificultad, tal como se puede observar en la tabla 2, sus valores oscilan entre -1.626 y 5.416 , además, 8 de los 28 ítems presentan una dificultad alta (ítems 2, 3, 12, 18, 19, 20, 22 y 28), especialmente por lo que respecta a los ítems 2 ($b= 5.416$), y 19 ($b= 3.126$); 6 ítems son muy fáciles (5, 7, 8, 10, 24 y 26), especialmente los ítems 7 ($b= -2.249$), y 8 ($b= -2.328$). Finalmente, en relación al índice de discriminación, sus valores oscilan entre $.310$ y 1.914 . Quince de los 28 ítems presentan un índice de discriminación superior a 1, lo que indica el adecuado poder discriminativo de estos ítems (3, 4, 5, 6, 7, 8, 11, 15, 16, 18, 21, 24, 26, 27 y 28). Los ítems más fáciles son aquellos que presentan una curva característica más próxima al cuadrante superior izquierdo, en tanto que los más difíciles están más próximos al lado derecho de la curva, por otra parte, los ítems con adecuado poder discriminativo son los que presentan un mayor cambio de pendiente.

En la tabla 3 se presenta el valor de los parámetros estimados para el modelo de tres parámetros y en la figura 3 se muestran las curvas características de los 15 ítems del tercer cuestionario. En este caso no se ha podido valorar el ajuste de los modelos estimados dados el bajo número de ítems del cuestionario. De

todos modos, todos los parámetros estimados han resultado estadísticamente significativos (a excepción del parámetro b en los ítems 5, 9, 10 y 11), siendo 36 el número de ciclos necesarios para llegar a la convergencia en la estimación. Los valores de las saturaciones factoriales para el modelo unifactorial oscilan entre .055 y .961, para 12 de los 15 ítems el valor de la saturación factorial es superior a .9. Los valores del parámetro c oscilan entre .052 y .878, aunque cabe destacar que en 13 ítems el valor de este parámetro estimado es inferior a .1, valor inferior al valor teórico esperado (.33).

Tabla 3.
Resumen de los parámetros del modelo de tres parámetros para los 15 ítems del tercer cuestionario

Ítem	Constante (E.S.)	Carga factorial (E.S.)	Parámetro a (E.S.)	Parámetro b (E.S.)	Parámetro c (E.S.)
1	1.839*	0.955*	3.236*	-0.568*	0.071*
	0.276	0.285	0.965	0.219	0.033
2	-0.569	0.916*	2.279*	0.250*	0.068*
	0.328	0.307	0.764	0.081	0.032
3	-1.422*	0.956*	3.260*	0.436*	0.060*
	0.565	0.384	1.308	0.048	0.029
4	1.658*	0.900*	2.060*	-0.805*	0.054*
	0.255	0.307	0.704	0.365	0.027
5	0.842*	0.918*	2.307*	-0.365	0.065*
	0.271	0.281	0.705	0.219	0.032
6	1.970*	0.055*	1.837*	-1.072*	0.878*
	0.241	0.027	0.564	0.401	0.270
7	-1.110*	0.073*	1.152*	0.964*	0.755*
	0.267	0.035	0.430	0.238	0.282
8	-0.915*	0.866*	1.731*	0.529*	0.085*
	0.308	0.296	0.592	0.094	0.036
9	1.731*	0.941*	2.791*	-0.620	0.052*
	0.396	0.402	1.192	0.390	0.026
10	1.027*	0.911*	2.206*	-0.465	0.075*
	0.279	0.289	0.701	0.257	0.035
11	0.975*	0.907*	2.149*	-0.454	0.065*
	0.258	0.299	0.708	0.254	0.032
12	-0.677*	0.901*	2.075*	0.326*	0.062*
	0.310	0.302	0.696	0.074	0.030
13	-2.204*	0.900*	2.062*	1.069*	0.064*
	0.516	0.376	0.863	0.243	0.030
14	1.773*	0.922*	2.381*	-0.744*	0.062*
	0.279	0.312	0.806	0.330	0.030
15	1.475*	0.961*	3.456*	-0.427*	0.089*
	0.285	0.269	0.967	0.178	0.035

Con * se han marcado todos los parámetros estadísticamente significativos con un nivel de significación inferior a .05.

E.S.: error estándar del parámetro estimado; χ^2 : valor del estadístico chi cuadrado como prueba de bondad de ajuste, $g.l.$: grados de libertad y p : nivel de significación.

En relación al índice de dificultad, los valores oscilan entre -1.702 y 1.069 , ninguno de los ítems ha resultado extremadamente fácil o difícil, de hecho el ítem más fácil es el 6 ($b=-1.072$) y el más difícil el 13 ($b=1.069$). Finalmente, en relación al índice de discriminación los valores estimados oscilan entre 1.152 y 3.456 , por lo que los 15 ítems presentan un adecuado poder discriminativo. A las mismas conclusiones se puede llegar a partir de la figura 3, dado que la pendiente de la tangente a la curva es bastante pronunciada para los 15 ítems (índice de discriminación) y todos ellos se sitúan en torno al valor 0 (índice de dificultad) del constructo medido (θ).

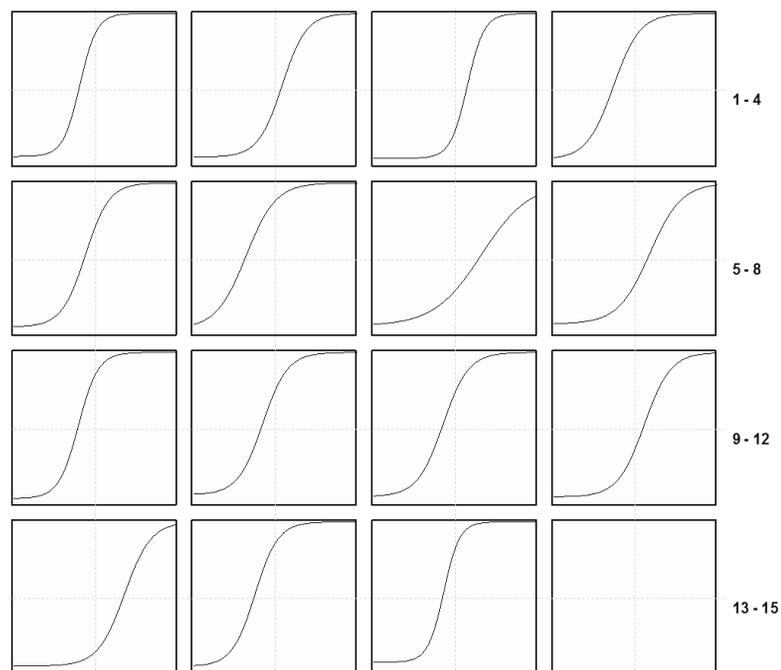


Figura 3. Curvas características de los 15 ítems del tercer cuestionario (modelo logístico de 3 parámetros).

Conclusiones

El objetivo del presente trabajo consistía en la elaboración de instrumentos de autoevaluación con una métrica adecuada, a fin de disponer de diferentes herramientas que ayuden y guíen al alumnado en la realización del trabajo autónomo que deben realizar para superar con éxito la materia de estadística. Como ya se ha comentado, se han realizado tres cuestionarios de respuesta cerrada llevándose a cabo el estudio de las propiedades métricas de los ítems a partir de la teoría de respuesta a los ítems (TRI), y en consecuencia, estimando los parámetros de las curvas características de los ítems.

A partir del análisis realizado de los ítems de los tres cuestionarios se puede concluir que el cuestionario que presenta mejores curvas características de los ítems es el tercero, dado que para los 15 ítems de este cuestionario los índices de discriminación presentan un valor muy adecuado, es decir, la pendiente de la tangente a la curva característica del ítem es muy pronunciada. Ahora bien, por lo que respecta a la dificultad de los ítems, es el segundo cuestionario el más adecuado, dado que los 28 ítems se distribuyen por igual entre ítems de dificultad baja, intermedia y alta (Martínez, Moreno & Muñiz, 2005; Muñiz, 1996), es decir, cubren de forma bastante amplia el continuo de los valores del constructo a medir. Debemos recordar que este segundo cuestionario no mostró ajuste aceptable al modelo de tres parámetros y se ajustó finalmente a un modelo de dos parámetros. Ello lo interpretamos en base a los propios contenidos de lo que se evalúa, puesto que el temario de la asignatura se centra en aspectos muy matemáticos y en muchos casos, en conocimientos que se suponen ya adquiridos en fases académicas previas a la que nos interesa en este trabajo. Como sea que el perfil de estudiante que accede a nuestra titulación carece en general de formación matemática, este cuestionario resulta especialmente difícil y por tanto, con los parámetros de distribución y de discriminabilidad resulta un modelo estadístico en TRI suficientemente explicativo. De todos modos cabe comentar que para los tres cuestionarios elaborados se puede asumir el supuesto de unidimensionalidad o independencia local de los ítems (García-Cueto & Fidalgo, 2005; Muñiz, 1997), dado que las saturaciones factoriales de los ítems de los tres cuestionarios han resultado estadísticamente significativas, por lo que constituyen una evidencia a favor del modelo

unifactorial. Además, para los cuestionarios, 1 y 3, en que se ha podido estimar el parámetro c , en general el valor estimado ha sido inferior al valor teórico esperado.

Además, tal como se mostró en un trabajo anterior (Guàrdia, et al., 2009), existe relación entre el hecho de haber superado la asignatura y el hecho de haber entregado los tres cuestionarios. Por otra parte en este trabajo también se muestra que los alumnos con mejor nota en la asignatura son también los que presentan mayor puntuación en los cuestionarios analizados, lo que iría a favor de una adecuada validez de criterio predictiva de los tres cuestionarios analizados en el presente trabajo.

Debemos resaltar alguna de las limitaciones que se dan en este trabajo, y fundamentalmente se centran en la utilización de distribuciones a priori de carácter teórico (siempre susceptibles de discusión), y el hecho de que la estimación de los parámetros propios de la TRI se ha efectuado con muestras distintas a la estudiada. En el primer caso hemos recurrido al álgebra de probabilidad para la definición de esas distribuciones, y en el segundo debimos asumir la imposibilidad de estar evaluando reiteradas veces al alumnado. Así pues, en las siguientes fases de esta línea de trabajo se abordará la depuración de estos cuestionarios y el análisis de los distintos formatos en los que se pueden presentar a los estudiantes este tipo de sistema de autoevaluación.

Referencias

- Alvarado, J.M., & García, M.V. (1999/2000). *Incidencia de la asistencia a clase, del trabajo efectivo y de factores individuales sobre el rendimiento académico*. Actas del V Congreso de Metodología de las C.C. Humanas y Sociales. Coordinadores, A. M. López, J.L. López y R. Moreno (pp. 683 – 686). Sevilla, septiembre 1997.
- Boyle, C.R. (1999). A problem-based learning approach to teaching biostatistics. *Journal of Statistics Education*, 7 (1). Disponible en <http://www.amstat.org/publications/jse/secure/v7n1/boyle.cfm>
- Castejón, J.L., Navas, L., & Sampascual, G. (1996). Un modelo estructural del rendimiento académico en matemáticas en la educación secundaria. *Revista de Psicología General y Aplicada*, 49(1), 27 – 43.
- Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: towards an assessment framework. *Journal of Statistics Education*, 2 (2). Disponible en <http://www.amstat.org/publications/jse/v2n2/gal.html>
- García, M.V., Alvarado, J.M., & Jiménez, A. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema*, 12(Supl. 2), 248 – 252.
- García, L., & Fumero, A. (1998). Personalidad y rendimiento académico en estudiantes universitarios: Un estudio predictivo en tres cursos académicos. *Análisis y Modificación de Conducta*, 24(93), 65 – 77.
- García-Cueto, E. (1993). *Introducción a la psicometría*. Madrid: Siglo XXI de España Editores, S.A.
- García-Cueto, E. & Fidalgo, A. M. (2005). Análisis de los ítems. En J. Muñiz, A.M. Fidalgo, E. García-Cueto, R. Martínez & R. Moreno (Eds.). *Análisis de los ítems*, (pp. 53-131). Madrid: La Muralla, S.A.
- Gardner, P., & Hudson, I. (1999). University students' ability to apply statistical procedures. *Journal of Statistics Education*, 7 (1) . Disponible en <http://www.amstat.org/publications/jse/secure/v7n1/gardner.cfm>
- Garfield, J.B. (1994). Beyond testing and grading: using assessment to improve student learning. *Journal of Statistics Education*, 2 (1). Disponible en <http://www.amstat.org/publications/jse/v2n1/garfield.html>
- Garrido, I., & Rojo, C. (1996). Motivación, cognición y rendimiento. *Revista de Psicología General y Aplicada*, 49 (1), 5 – 12.
- Guàrdia, J., Freixa, M., Però, M., Turbany, J., Coscolluela, A., Barrios, M., et al. (2006). Factors related to the academic performance of students in the statistics course in Psychology. *Quality & Quantity*, 40(4), 461-474.
- Guàrdia, J., Freixa, M., Turbany, J., Coscolluela, A., Però, M., Barrios, M., et al. (2002). Estudio sobre los factores que inciden en el rendimiento académico en la asignatura análisis de datos en Psicología. *Metodología de las Ciencias del Comportamiento, Volumen Especial*, 275 - 278.

- Guàrdia-Olmos, J., Freixa-Blanxart, M., Turbany-Oset, J. & Però-Cebollero, M. (2008). Collaborative learning in the teaching of statistics in psychology: an alternative to traditional teaching. *Revista de Formació e Innovació Educativa Universitaria*, 1(4), 96-106.
- Guàrdia, J., Però, M., Freixa, M., Turbany, J., & Gordóvil, A. (en prensa, 2008). Analysis of the implications of collaborative work in terms of performance and satisfaction among students registered in the *Data Analysis in Psychology Course*. *Quality & Quantity*.
- Guàrdia, J., Però, M., Freixa, M., Turbany, J., Barrios, M., & Gordóvil, A. (2009). *Elaboración de material para el aprendizaje autónomo del alumno en Análisis de Datos en Psicología impartida en la Universidad de Barcelona*. Comunicación presentada en el I Congreso de Docència Universitaria celebrado en Vigo (España) del 2 al 4 de julio.
- Guillén, M., Alea, V., Muñoz, C., Soldevilla, C., Torrelles, E., Viladomiu, N. et al. (2001). *Estadística Descriptiva Básica*. Barcelona: Edicions UB Colección: Team, 8 (CD-Rom). (ISBN: 84-8338-292-X).
- Huberty, C.J. (2000). Assessment of student performance in statistics. *Teaching Statistics*, 22(2), 44 – 48.
- Martínez, R., Moreno, R., & Muñoz, J. (2005). Construcción de los ítems. En J. Muñoz, A.M. Fidalgo, E. García-Cueto, R. Martínez & R. Moreno (Eds.). *Análisis de los ítems*, (pp. 9-52). Madrid: La Muralla, S.A.
- Muñoz, J. (1996). *Teoría clásica de los tests*. Madrid: Ediciones Pirámide.
- Muñoz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Ediciones Pirámide.
- Pérez-Sánchez, A.M., & Castejón, J.L. (1996). Factores psicosociales en la predicción del rendimiento académico. *Revista de Psicología de la Educación*, 19, 31 – 51.
- Peró, M., Turbany, J., Freixa, M., Guàrdia, J., Barrios, M., Ferrer, R. et al. (2004). Evaluación de la implementación de un CD-Rom de estadística descriptiva en la asignatura de Análisis de Datos en Psicología. *Metodología de las Ciencias del Comportamiento, Volumen Especial*, 485 – 490.
- Roces, C., González, M.C., & Touron, J. (1997). Expectativas de aprendizaje y de rendimiento de los alumnos universitarios. *Revista de Psicología de la Educación*, 22, 99 – 123.
- Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education*, 6 (3).. Disponible en <http://www.amstat.org/publications/jse/v6n3/smith.html>