

ASOCIACIÓN DE QUÍMICOS DEL PRINCIPADO DE ASTURIAS

SECCIÓN TÉCNICA DE ENSEÑANZA

**XXIV OLIMPIADA DE QUÍMICA**

ASTURIAS – 2010

*ANÁLISIS DE LA PRUEBA*

La prueba de la XXIV Olimpiada de química constó de dos partes cuyas características se resumen en la tabla 1

**Tabla 1**

	<b>Tipo</b>	<b>Preguntas</b>	<b>Observaciones</b>	<b>Valor</b>
<b>Prueba 1</b>	Test opción múltiple	30	Cuatro respuestas	40 %
<b>Prueba 2</b>	Problemas	3	Varios apartados	60 %

#### **Test de opción múltiple:**

Se seleccionaron 30 preguntas de opción múltiple con cuatro respuestas posibles de las que sólo una era la más cierta, cada respuesta errónea se penalizaba con la tercera parte de su valor de modo que la nota de cada alumno se hallaba según **[Bien]** –  $\frac{1}{3}$  **[Mal]** reduciéndolo a 10 puntos.

La corrección se llevó a cabo con una plantilla y la realizaron profesores de secundaria anónimamente ya que cada examen estaba encabezado por una clave desconociendo la identidad del alumno. No se suscitó ninguna incidencia.

La ubicación de las respuestas correctas entre las diferentes alternativas fue aleatoria y equitativa [7 para A, 8 para B, 6 para C y 9 para D], para que fuese totalmente equitativo debería haber una D menos y una C más como respuestas correctas, en todo caso es un sesgo mínimo.

#### **Problemas**

Se propusieron tres, el primero y tercero tenían dos apartados cada uno y el segundo, cuatro apartados. Todos se calificaron sobre dos puntos con lo que se obtenía un valor total de 6 puntos que se redujeron a 10 para tratar homogéneamente los resultados.

Fueron corregidos por profesores de universidad, anónimamente, sin incidencia alguna; no se detectó ningún error en el planteamiento de cada problema o apartado.

#### **Resultados**

Para los test los resultados se recogen en la tabla 2, en ella se registra la frecuencia absoluta y porcentaje de aciertos para cada posible respuesta (A, B, C, D) y para las respuestas en blanco. La última fila recoge los resultados globales.

En la misma tabla se anota en [azul negrita] la respuesta correcta y con fondo [amarillo] el máximo de respuestas. Se observa que coinciden estos valores en 25 de los ítems, es decir eligen la respuesta correcta el mayor número de alumnos.

Tabla 2

Item	Respuesta correcta	A		B		C		D		BLANCO	
1	D	18	16%	31	28%	8	7%	25	23%	28	25%
2	A	69	63%	12	11%	4	4%	7	6%	18	16%
3	B	19	17%	34	31%	14	13%	25	23%	18	16%
4	A	34	31%	6	5%	21	19%	28	25%	21	19%
5	C	11	10%	20	18%	41	37%	17	15%	21	19%
6	D	18	16%	43	39%	1	1%	38	35%	10	9%
7	D	9	8%	3	3%	13	12%	73	66%	12	11%
8	A	32	29%	8	7%	19	17%	27	25%	24	22%
9	B	31	28%	47	43%	8	7%	9	8%	15	14%
10	C	10	9%	21	19%	38	35%	12	11%	29	26%
11	C	0	0%	7	6%	72	65%	4	4%	27	25%
12	B	11	10%	31	28%	11	10%	16	15%	41	37%
13	C	7	6%	24	22%	64	58%	0	0%	15	14%
14	B	7	6%	85	77%	11	10%	4	4%	3	3%
15	D	7	6%	11	10%	14	13%	67	61%	11	10%
16	A	53	48%	4	4%	14	13%	10	9%	29	26%
17	D	20	18%	11	10%	17	15%	49	45%	13	12%
18	D	11	10%	10	9%	28	25%	55	50%	6	5%
19	B	8	7%	61	55%	11	10%	13	12%	17	15%
20	D	6	5%	8	7%	15	14%	74	67%	7	6%
21	A	38	35%	40	36%	16	15%	4	4%	12	11%
22	B	13	12%	52	47%	12	11%	5	5%	28	25%
23	B	7	6%	41	37%	13	12%	27	25%	22	20%
24	B	37	34%	29	26%	13	12%	8	7%	23	21%
25	A	23	21%	10	9%	11	10%	15	14%	51	46%
26	A	69	63%	9	8%	8	7%	7	6%	17	15%
27	C	17	15%	1	1%	78	71%	9	8%	5	5%
28	D	13	12%	15	14%	37	34%	36	33%	9	8%
29	D	31	28%	21	19%	7	6%	38	35%	13	12%
30	C	5	5%	22	20%	49	45%	13	12%	21	19%
Total		634	19%	717	22%	668	20%	715	22%	566	17%

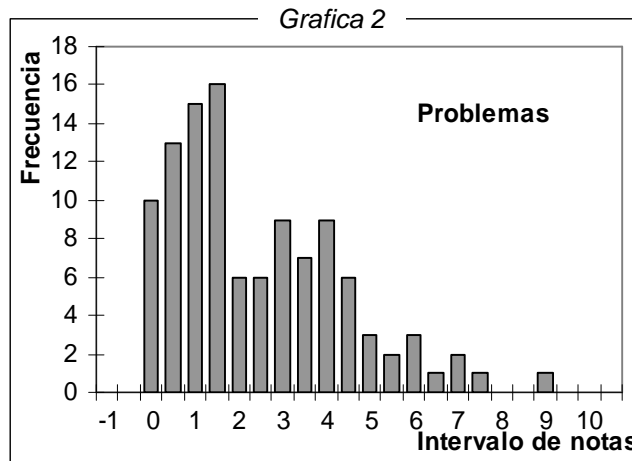
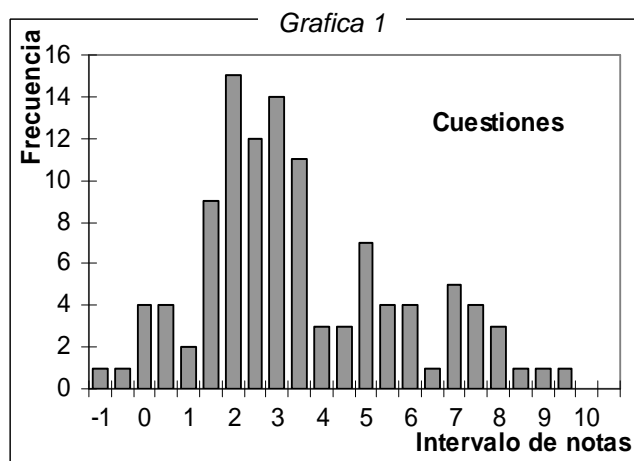
La distribución de frecuencias entre diferentes intervalos de notas se recoge en la tabla 3

Tabla 3

		Frecuencia de notas en los distintos intervalos																						
Intervalos		-1	-0,5	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10
Cuestiones		1	1	4	4	2	9	15	12	14	11	3	3	7	4	4	1	5	4	3	1	1	1	0
Problemas		0	0	10	13	15	16	6	6	9	7	9	6	3	2	3	1	2	1	0	0	1	0	0

Se aprecia una desviación hacia valores menores de cinco puntos. Este sesgo se aprecia mejor en las gráficas 1 y 2 correspondientes a las cuestiones y a los problemas.

Se aprecia claramente que hay un grupo de alumnos con un bajo nivel de instrucción que realmente no compitieron con sus compañeros por obtener un buen resultado.



En la tabla 4 se recogen diferentes parámetros estadísticos tanto de las cuestiones como de los problemas.

**Tabla 4**

<i>Estadística</i>	<i>cuestiones</i>	<i>problemas</i>
Media	3,28	2,24
Error típico	0,22	0,19
Mediana	2,73	1,7
Moda	2,45	0
Desviación estándar	2,27	1,94
Varianza de la muestra	5,14	3,77
Curtosis	-0,10	0,45
Coefficiente de asimetría	0,67	0,95
Mínimo	-1	0
Máximo	9,23	8,67
Total	110	110

### **Índice de dificultad de los ítems del test**

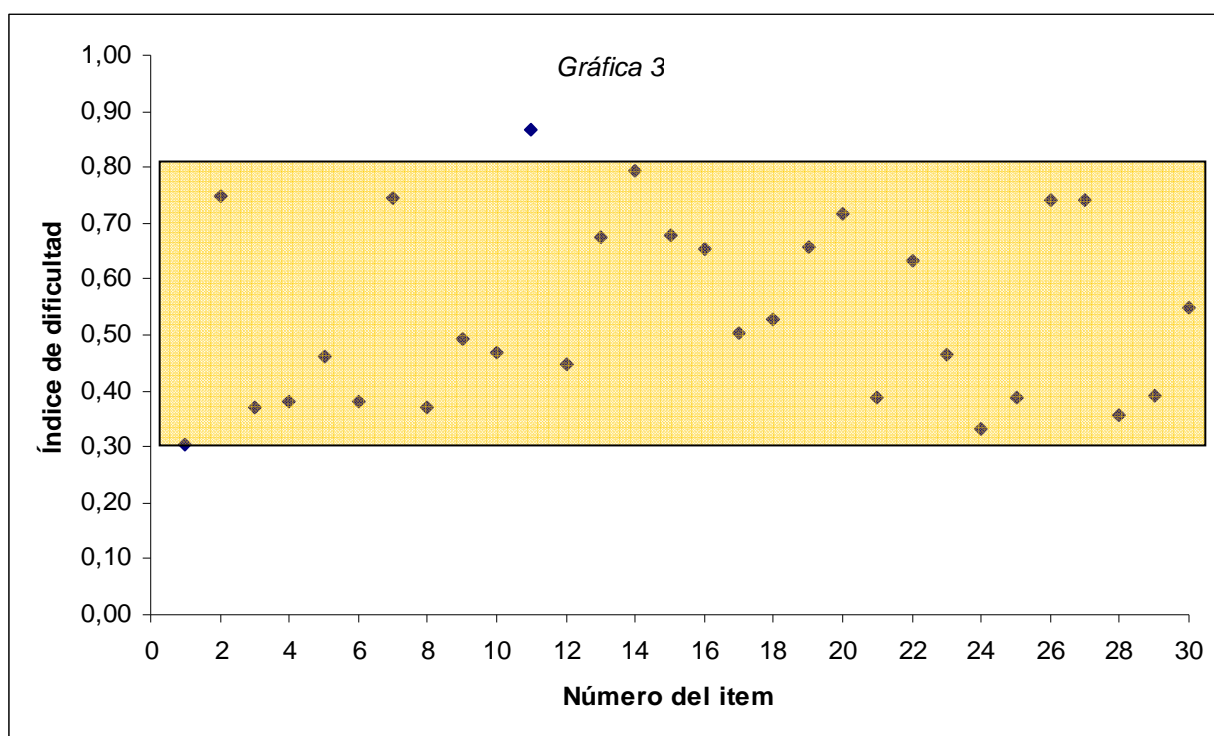
Aunque la dificultad de un ítem podríamos verla por observación directa de los alumnos que la contestan bien o mal, se suele usar el Índice de dificultad que se define como la proporción de alumnos que aciertan el ítem. Oscila desde 0 hasta 1.

Un índice de dificultad bajo indica que el ítem ha sido contestado por un pocos alumnos (si está bien planteado es difícil) y un índice de dificultad alto indica que el ítem es correctamente contestado por muchos alumnos (si está bien planteado es fácil). No hay que olvidar que este parámetro estadístico no es un valor estadístico absoluto, está condicionado no sólo por la naturaleza del ítem sino también por la población a la que se somete y su variabilidad puede ser muy diferente si la población a la que se destina la prueba fuera otra. Lo ideal es que los ítems no sean ni demasiado fáciles ni demasiado difíciles. Si la mayoría de los

estudiantes aciertan o la mayoría fallan el ítem no puede discriminar bien puesto que no hay variabilidad en las respuestas.

En la gráfica 3 se aprecia la distribución de los índices de dificultad a lo largo de las diferentes preguntas indicándose qué preguntas fueron las que resultaron más fáciles y las que fueron más difíciles.

Se observa que gran parte de las preguntas están en una franja que va de 0,30 a 0,80 de índice de dificultad, predominando más las fáciles (índice alto) que las difíciles (índice bajo). Destacan la 11 y la 14 como muy fáciles seguidas de la 2, 7, 26 y 27.



### **Índice de discriminación de los ítems del test**

Uno de los propósitos de todo ítem es discriminar académicamente entre alumnos de buen rendimiento y alumnos de mal rendimiento y así para definirlo se siguen las clásicas recomendaciones de establecer en la muestra dos grupos extremos, el alto con el 27% con las puntuaciones más altas y el bajo formado por el 27 % con las puntuaciones más bajas.

El Índice de discriminación se define como la diferencia entre la proporción de acertantes al ítem que pertenecen al grupo superior menos la proporción de acertantes al ítem que pertenecen al grupo inferior.  

$$\text{Ind.disc} = pA - pB.$$

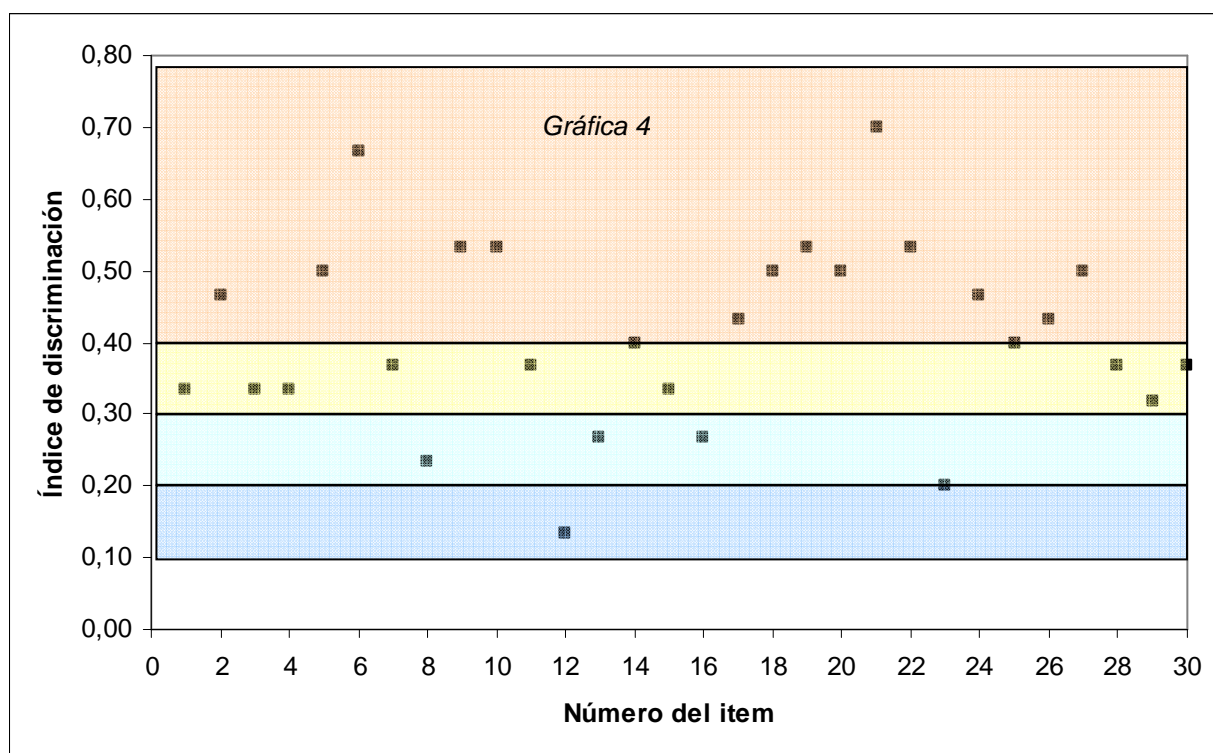
Este índice varía desde  $-1$  a  $+1$  y si todas las personas del grupo alto acertaran el ítem y ninguna del grupo bajo valdría  $1$  y si todas las personas del grupo bajo acertaran el ítem y ninguna del grupo alto tomaría un valor de  $-1$ . El índice de discriminación de un ítem mide la calidad de un ítem en el sentido que un buen ítem tiene que ser acertado por una proporción de alumnos que obtengan mejor puntuación que por aquellos que obtengan menos puntuación.

Si este índice es alto indica que el ítem discrimina bien entre buenos y malos. El índice de discriminación de un ítem mide la correspondencia entre la puntuación de los alumnos en el ítem y la puntuación en el test

Valores habituales de este índice son los que se muestran en la tabla 5

<b>Tabla 5</b>	
<i>Valores</i>	<i>Discriminación</i>
Igual o mayor que 0,40	Muy buena
Entre 0,30 y 0,39	Buena
Entre 0,20 y 0,29	Regular
Entre 0,10 y 0,19	Debe mejorarse.
Menor de 0,10	Sin utilidad

Realizando el cálculo para todos los ítems y representándolo se obtiene la gráfica 4, en ella se aprecia que hay dos ítems (números 12 y 23) que necesitan mejorarse y tres (números 8, 11 y 16) regulares, el resto tiene una buena o muy buena discriminación

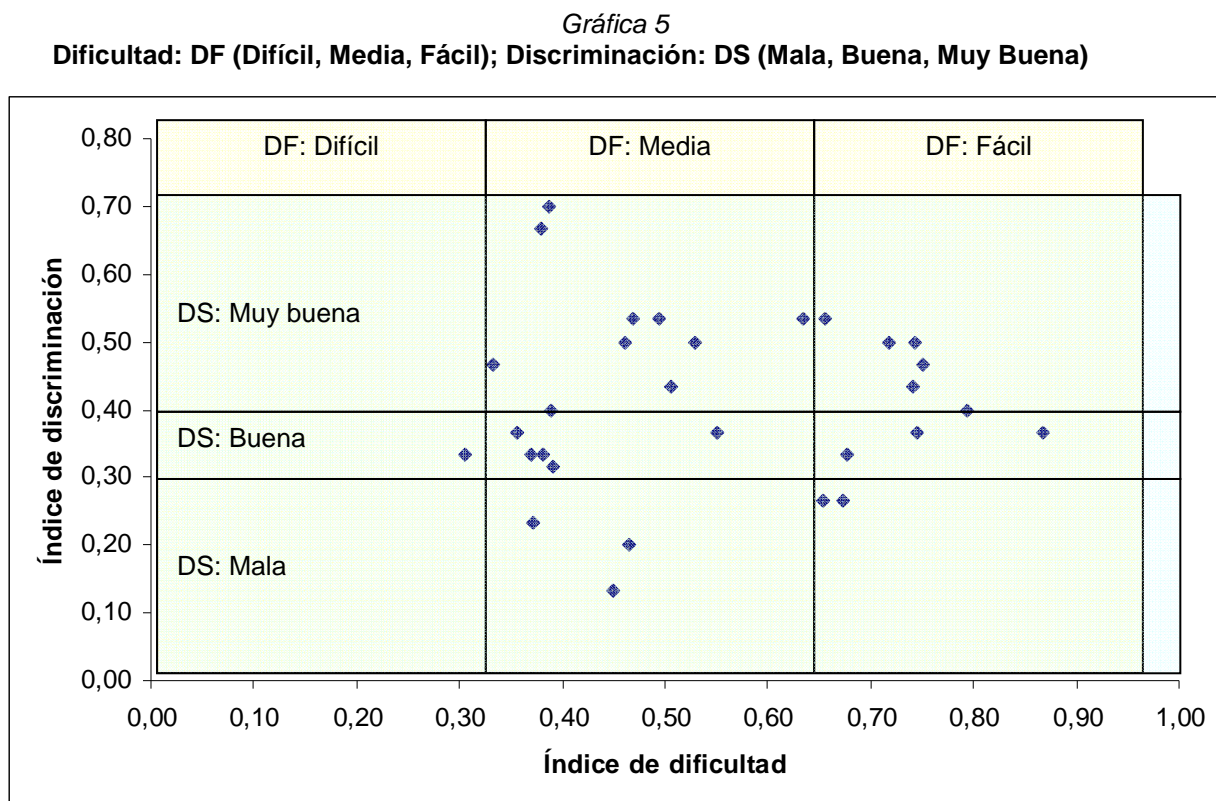


Estos datos son coherentes si aplicamos a todo el test la fórmula 20 de Kuder y Richardson que nos permite estimar la confiabilidad de consistencia interna de una prueba. El máximo valor es 1 y valores por encima de 0,7 se consideran adecuados. Este índice analiza el test como instrumento de medida y de alguna manera mide la capacidad del test para obtener siempre resultados semejantes cuando se aplica en condiciones semejantes.

El test se manifiesta como fiable ya que aplicándole a este test da un valor de 0,804. Se podría aumentar la fiabilidad aumentando la longitud del mismo o suprimiendo alguno de los ítems problemáticos.

### Discriminación frente a la dificultad

Cruzando para cada ítem la dificultad y la discriminación se obtiene la gráfica 5, los resultados avalan los análisis anteriores.



Como se aprecia en la gráfica, la mayoría de los ítems son de una dificultad razonable (media o fácil) y sin embargo discriminan correctamente.

### Conclusiones

A falta de un análisis más preciso, parece que la prueba fue razonable desde el punto de vista de la dificultad de los diferentes ítems y existe una buena correlación entre la calificación de los problemas y la de las cuestiones de opción múltiple, hallado el coeficiente de correlación entre las calificaciones en los dos apartados dio un valor de 0,68.

Por otra parte debe reflexionarse sobre el número de presentados que realmente no compiten, hecho observable en el número de ellos que obtienen una baja calificación tanto global como en cada parte.

La discusión puede establecerse en dos extremos:

- Fomentar la participación exclusivamente de alumnos de elevado rendimiento con un espíritu fuertemente competitivo, haciendo una selección previa por calificaciones obtenidas. Existen algunas agrupaciones que optan por esta solución recomendándolo expresamente en sus convocatorias.
- Fomentar la participación de todos los alumnos interesados, aun los que tienen un conocimiento precario de conceptos químicos, para intentar dinamizar el interés por el estudio de la Química entre estudiantes de secundaria. En esta opción, los estudiantes que realmente tienen conocimientos e interés van a participar de igual modo por lo que realmente no se hace daño alguno, salvo las complicaciones organizativas y la obtención de recursos.

Por ahora nos declaramos partidarios de la segunda opción, aunque es una cuestión abierta.