

**Publicado en texto impreso, en:**

García Aretio, L. (1997). *Criterios de evaluación de la enseñanza a distancia*. En Martínez, C. (Coord.). *Encuentros en la Facultad de Educación sobre evaluación*. Madrid: UNED, pp. 309-318.

## **CRITERIOS DE EVALUACIÓN EN LA ENSEÑANZA A DISTANCIA**

**Lorenzo García Aretio**  
**Titular de la Cátedra UNESCO de Educación a Distancia**  
**Decano Facultad de Educación**  
**UNED**

### **1. LAS PRUEBAS PRESENCIALES EN LA UNED**

Las pruebas presenciales o exámenes *cara a cara* de la UNED consisten en un tipo de *reactivo que proponen los equipos docentes de la Sede Central a todos los estudiantes matriculados en esa materia, con el fin de que éstos, de forma personal, y bajo el control de tribunales expresamente enviados desde la Sede Central, pongan de manifiesto la posesión de determinados conocimientos, aptitudes, habilidades o actitudes que permitan, tras el pertinente juicio de valor, tomar la correspondiente decisión.*

Es muy diversa la forma de evaluar estos aprendizajes presencialmente en esta Universidad. Téngase en cuenta que en la actualidad se imparten 16 carreras universitarias correspondientes a las diversas áreas del saber. La estructura de las pruebas que se proponen a los alumnos son muy diferentes unas de otras debido a la propia estructura del saber que pretende evaluarse, pero también debido a la concepción que el profesor correspondiente tiene del sentido de la evaluación: importancia, validez, fiabilidad, representatividad, número de alumnos a evaluar, etc.

Hemos elaborado un estudio sobre las pruebas presenciales propuestas a los alumnos de los estudios reglados de la UNED, durante la segunda semana de los exámenes de febrero de 1995. En el análisis que hicimos prescindimos de las materias correspondientes a las carreras de la Facultad de Ciencias (Matemáticas, Física y Química), E.T.S. de Ingenieros Industriales y E.U. de Ingeniería Técnica (Informática de Gestión e Informática de Sistemas), por entender que estas

pruebas mayoritariamente consistían en comprobar procesos de aplicación de los conocimientos a realidades propuestas (problemas). Por otra parte, para nosotros tenía mayor dificultad analizar este tipo de pruebas.

Igualmente desechamos de nuestro estudio todos aquellos exámenes de idioma y de lenguas clásicas que proponían traducciones de textos por la similitud de procesos mentales que todos ellos tratan de evaluar.

Consideramos que este trabajo, aunque referido sólo a los exámenes de una semana de una de las convocatorias, tiene suficiente validez habida cuenta de que los profesores, al menos durante una serie consecutiva de cursos, suelen proponer a sus alumnos exámenes muy similares en cuanto a su estructura externa e interna.

Dicho ésto observemos en la siguiente tabla el número de exámenes que fueron analizados por Facultades. Diríamos que, hechas las salvedades precedentes, todos los que se propusieron a los alumnos en las citadas pruebas:

<b>NÚMERO DE EXÁMENES ANALIZADOS</b>		
<b>CARRERA</b>	<b>N. Exám.</b>	<b>%</b>
DERECHO	33	13
G. HISTORIA	22	8,5
FILOLOGÍA	19	7,5
FILOSOFÍA	25	10
EDUCACIÓN	28	11
EC.EMPRESAR.	42	16,5
PSICOLOGÍA	38	14,5
POL. SOCIOLOG.	49	19
<b>TOTAL</b>	<b>256</b>	<b>100</b>

Las facultades de Económicas y Empresariales y de Políticas y Sociología imparten dos carreras cada una de ellas, de ahí que el número de exámenes de las mismas sea superior al de las otras facultades. También ha de observarse que los exámenes se han asignado a las facultades de acuerdo con la información que en el propio examen aparecía (nombre de la Facultad o del Departamento) dado que, como es sabido no todo examen correspondiente a una carrera pertenece a la Facultad a la que está adscrita dicha carrera. Es decir, por ejemplo, existen exámenes de la Facultad de Derecho en las carreras de Económicas y Empresariales o exámenes de la Facultad de Psicología en la carrera de Pedagogía. El escaso número de exámenes recogidos de Filología es debido a que no se consideraron las pruebas de idiomas o lenguas clásicas que proponían traducciones.

En la siguiente tabla ofrecemos una perspectiva general de la tipología externa de los exámenes, considerando básicamente aquellos que contemplaban Pruebas Objetivas. Así existían exámenes que proponían sólo este tipo de pruebas, exámenes que no las contemplaban y otros que pudiéramos considerar mixtos, los cuales consideraban las pruebas objetivas y además otra parte con otro tipo de cuestiones: tradicionales de ensayo de respuesta amplia o breve, comentarios de texto, resolución de casos, problemas, interpretación de gráficos o imágenes, etc. A éstos en esta tabla los hemos denominado como "otro".

TIPOS DE EXÁMENES							
CARRERA	P.OBJ.		P.OBJ+OTRO		OTRO		TOT. N
	N	%	N	%	N	%	
DERECHO	3	9	8	24	22	67	33
G.HISTORIA	0	0	0	0	22	100	22
FILOLOGÍA	0	0	0	0	19	100	19
FILOSOFÍA	0	0	0	0	25	100	25
EDUCACIÓN	0	0	1	3,5	27	96,5	28
EC. EMPRES.	10	24	2	4,5	30	71,5	42
PSICOLOGÍA	15	39,5	3	8	20	52,5	38
POL.SOCIOL.	0	0	1	2	48	98	49
<b>TOTAL</b>	<b>28</b>	<b>11</b>	<b>15</b>	<b>6</b>	<b>213</b>	<b>83</b>	<b>256</b>

En esta tabla podemos destacar que son, sobre todo, los profesores de dos facultades los que más utilizan, para comprobar el rendimiento académico de sus alumnos, sólo pruebas objetivas en sus exámenes. Destaca la Facultad de Psicología con un 39,5% de sus exámenes, seguida de la de Económicas y Empresariales, 24%. Los profesores de la Facultad de Derecho utilizan también bastante el instrumento de la prueba objetiva para comprobar el rendimiento, sin embargo, éstos las utilizan más, combinadas con otro tipo de pruebas, 24%. En los exámenes de las Facultades de Geografía e Historia, Filología y Filosofía ningún profesor utilizó este instrumento y sólo uno lo utilizó en las Facultades de Educación y de Políticas y Sociología.

En lo sucesivo sólo utilizando porcentajes, dado que los "N" de cada Facultad ya los conocemos, vamos a aproximarnos algo más a la tipología de exámenes propuestos a los alumnos en esta convocatoria de febrero de 1995 a la que nos estamos refiriendo.

En este caso vamos a contemplar qué técnica de evaluación han aplicado nuestros profesores para evaluar a los alumnos. Vamos a dar un paso más ofreciendo el porcentaje de exámenes, dentro de cada Facultad, que han incluido en su propuesta una de estas cuatro técnicas: prueba objetiva, problema o casos prácticos, comentario de texto, prueba tradicional de ensayo. Estos porcentajes no

tienen por qué sumar 100 dado que muchos exámenes contemplaban más de una de estas técnicas de evaluación.

<b>TIPOS DE EXÁMENES (porcentajes)</b>				
<b>CARRERA</b>	<b>P.OBJETIVA</b>	<b>PROBLEMA .PRÁCTIC</b>	<b>C.TEXTO</b>	<b>P.ENSAYO</b>
<b>DERECHO</b>	33	18	3	91
<b>G.HISTORIA</b>	0	25	12,5	100
<b>FILOLOGÍA</b>	0	0	79	94,5
<b>FILOSOFÍA</b>	0	12	36	100
<b>EDUCACIÓN</b>	3	21	17	100
<b>EC. EMPRES.</b>	29	29	0	50
<b>PSICOLOGÍA</b>	48	21	0	50
<b>POL.SOCIOL.</b>	2	9	20	83,5
<b>TOTAL</b>	17	18	13	77,5

Algo más de las tres cuartas partes de los exámenes de estas facultades contemplan como técnica preferida la de la prueba de ensayo, composición o prueba abierta en la que el alumno sea de forma más o menos breve ha de construir su respuesta íntegramente. Todos los exámenes analizados de las facultades de Geografía e Historia, Filosofía y Educación contemplan esa modalidad de prueba.

Las siguientes técnicas más utilizadas, globalmente en cuanto al número, de forma muy similar, son las correspondientes a las pruebas objetivas y a los problemas o resolución de casos prácticos. Un 17% y 18% de las pruebas, respectivamente, utilizaron esta fórmula. Casi la mitad de los exámenes de Psicología utilizaron la prueba objetiva. Igual hicieron la tercera parte de los profesores de materias de Derecho y cerca de ese porcentaje se situaron los de Económicas y Empresariales. En realidad son estas tres las facultades que vienen utilizando con más asiduidad esta técnica de evaluación

Obviamente es la Facultad de Económicas y Empresariales la que más utiliza para evaluar a sus alumnos la propuesta para resolver problemas. Muchas de sus materias requieren de la comprobación de estas destrezas para la evaluación. En la Facultad de Geografía e Historia puede extrañar el alto porcentaje de exámenes que han utilizado la resolución de problemas o casos prácticos. Hemos incluido en esta categoría las pruebas que requerían comentar figuras de arte, catalogación de rasgos geográficos, transcripciones, interpretación de diagramas, mapas, etc. Excepto en la Facultad de Filología, en todas las restantes se han incluido pruebas dentro de esta categoría. De todas formas conviene aclarar que si hubiésemos contemplados los exámenes de las distintas lenguas vivas o muertas, probablemente los hubiésemos incluido en esta categoría con lo que esta Facultad quedaría en ella altamente representada.

## 2. TIPOS DE EXÁMENES PROPUESTOS

### 2.1. De las pruebas abiertas

¿De cuántas preguntas constaban las pruebas abiertas? En la siguiente tabla hemos considerado estas cuestiones diferenciándolas por carreras. No presentamos las características de cada caso en los que se presentaron este número de cuestiones. Es decir, no es lo mismo que un examen contemple dos preguntas de prueba de ensayo y además proponga una prueba objetiva, un problema o caso práctico o un comentario de textos o que sencillamente proponga esas dos cuestiones sólo. E incluso pueden ser dos cuestiones que respondan a epígrafes concretos del programa o dos cuestiones aunque se presenten sólo en un examen que requieran poner en juego diversos procesos mentales y relacionar distintos temas motivo de examen. Evidentemente, poca fiabilidad tiene una prueba que sólo contenga una o dos cuestiones en las que se requiera a los alumnos verter por escrito una pequeña parte de lo que han estudiado, con la misma estructura que se contempla en el texto de estudio. En este caso la representatividad de los contenidos estudiados es mínima y muy pobres los procesos mentales o ámbitos cognoscitivos evaluados.

PRUEBAS ABIERTAS/PROBLEMAS/C.PRÁCTICO/C. TEXTO						
CARRERA	Nº DE PREGUNTAS (%)					TOT.
	1-2	3-4	5-6	7-8	9 o +	
DERECHO	9	59	23	4,5	4,5	100
G. HISTORIA	33,5	58,5	4	0	4	100
FILOLOGÍA	47,5	47,5	5	0	0	100
FILOSOFÍA	40	40	12	8	0	100
EDUCACIÓN	14	32	50	0	4	100
EC. EMPR.	10	47	30	10	3	100
PSICOLOGÍA	6	29	41	12	12	100
POL. SOCIOL.	44	37	14	5	0	100
<b>TOTAL</b>	<b>27</b>	<b>44</b>	<b>22</b>	<b>4,5</b>	<b>2,5</b>	<b>100</b>

Como decimos, es muy alto en algunas facultades el número de exámenes que sólo contemplaban una o dos cuestiones. La mayoría de estas pruebas se estructuraron con tres o cuatro preguntas. Tampoco es de despreciar el 22% de exámenes que consideraron cinco o seis preguntas.

Posteriormente hemos procedido a realizar un análisis más sutil de las pruebas presenciales propuestas a los alumnos en la segunda semana de febrero de 1995. Hemos intentado aproximarnos al conocimiento de las capacidades procesos o ámbitos cognoscitivos que trataban de poner en juego, intentaban medir a través de las preguntas hechas a los alumnos.

Por la propia dificultad de este análisis hemos prescindido del estudio de las pruebas objetivas, aunque no porque pensemos que con ellas no pueden evaluarse capacidades distintas a las de conocimientos memorísticos. No somos de esta opinión sino que, por el contrario, entendemos que a través de unas pruebas objetivas adecuadamente elaboradas pueden ponerse en juego procesos mentales complejos y no meramente repetitivos. La razón de prescindir de ellas en este análisis es por la propia dificultad de valorar tantos ítems en el corto espacio de tiempo que tuvimos entre febrero que se realizaron las pruebas y primeros de abril en que hubo que presentar este trabajo. Las áreas del saber ajenas a nuestra especialidad comportan dificultades agregadas para este ensayo.

Por tanto nos hemos limitado a constatar las preguntas abiertas de ensayo, los problemas, casos prácticos y los comentarios de texto. En algunos exámenes de carreras distintas de las de Pedagogía donde encontramos dudas para ubicar determinadas preguntas hicimos la pertinente consulta a profesores de esa especialidad, algunos de la UNED y otros de fuera de esta institución, a los que previamente explicamos lo que entendíamos por cuestiones que pretendían medir una u otra capacidad del dominio cognoscitivo.

En el análisis optamos por acudir a la más conocida de las taxonomías de objetivos, la de Bloom y otros (1979). Entendimos que era adecuada para los propósitos de este trabajo. Así tratamos de contabilizar el número de exámenes que recogía una o más pregunta referente a cada uno de estos seis tipos en orden jerárquico de menor a mayor complejidad: conocimiento, comprensión, aplicación, análisis, síntesis y evaluación. Si ya a los autores de esta reconocida obra les fue muy difícil clasificar objetivos y pruebas según esta determinada taxonomía no habrá de aclararse las dificultades propias de nuestro trabajo, referido en este caso al nivel universitario.

Hay que advertir que las cuestiones incluidas en un determinado tipo se servirán, muy probablemente, de las actividades realizadas en los tipos precedentes fundamentándose en ellos. Nosotros hemos tratado de incluir cada pregunta o cuestión en el tipo más elevado o complejo de los de la serie, y sólo en uno, excepción hecha de los comentarios de texto, a los que después aludimos.

Pues bien, con estos instrumentos procedimos a analizar las pruebas. Tras mucha reflexión decidimos que todos aquellos exámenes que incluyesen problemas o casos prácticos serían clasificados dentro del tipo de *3.0 Aplicación*. Más problemática resultó la decisión de ubicar los comentarios de texto. Dada la complejidad de este tipo de pruebas, finalmente optamos clasificar todos los comentarios de texto dentro de dos de los tipos de la taxonomía: comprensión y análisis. Aunque éramos conscientes de que un tipo superior viene requiriendo, en buena parte, capacidades recogidas en los tipos precedentes, no teníamos claro cuál podría ser el espíritu del evaluador, si el de captar si realmente el texto se comprendía, se interpretaba correctamente, se extraían conclusiones, consecuencias, etc., o por el contrario el de valorar el análisis del texto extrayendo intenciones, argumentos, contrastando, comparando relaciones, analizando

principios, teorías o modelos, etc. De esa manera decidimos, confesamos que de forma poco ortodoxa, incluir estas pruebas dentro de cada uno de los dos tipos. No nos faltaron dudas para incluir este tipo de pruebas dentro de la categoría de *síntesis* e incluso de *evaluación*.

Como los exámenes suelen estar compuestos de más de una cuestión, se da el caso de que en muchos de ellos se han encontrado preguntas que evalúan un tipo de capacidad y otras cuestiones que miden otros procesos mentales. Es decir en un mismo examen podemos encontrar más de un tipo de los que componen la taxonomía utilizada.

<b>PRUEBAS ABIERTAS/PROBLEMAS/C.PRÁCTICO/C. TEXTO</b>						
<b>CARRERA</b>	<b>CAPACIDADES QUE SE EVALÚAN (porcentaje)</b>					
	<b>Conocimiento</b>	<b>Comprensión</b>	<b>Aplicación</b>	<b>Análisis</b>	<b>Síntesis</b>	<b>Evaluación</b>
<b>DERECHO</b>	<b>90</b>	<b>13</b>	<b>20</b>	<b>13</b>	<b>0</b>	<b>0</b>
<b>G. HISTORIA</b>	<b>100</b>	<b>17</b>	<b>25</b>	<b>4</b>	<b>0</b>	<b>0</b>
<b>FILOLOGÍA</b>	<b>89</b>	<b>89</b>	<b>0</b>	<b>89</b>	<b>5</b>	<b>0</b>
<b>FILOSOFÍA</b>	<b>80</b>	<b>88</b>	<b>12</b>	<b>52</b>	<b>4</b>	<b>0</b>
<b>EDUCACIÓN</b>	<b>100</b>	<b>76</b>	<b>21</b>	<b>17</b>	<b>3</b>	<b>0</b>
<b>EC. EMPR.</b>	<b>50</b>	<b>22</b>	<b>38</b>	<b>19</b>	<b>3</b>	<b>0</b>
<b>PSICOLOGÍA</b>	<b>100</b>	<b>24</b>	<b>33</b>	<b>5</b>	<b>19</b>	<b>0</b>
<b>POL. SOCIOL.</b>	<b>80</b>	<b>43</b>	<b>9</b>	<b>43</b>	<b>9</b>	<b>0</b>
<b>TOTAL</b>	<b>84</b>	<b>40</b>	<b>20</b>	<b>29</b>	<b>5</b>	<b>0</b>

De manera que en el 84% de los exámenes que contenían pruebas distintas a las objetivas recogían entre sus preguntas, las de memorización o repetición de definiciones, términos, convenciones, procesos, secuencias, criterios, tendencias, procedimientos, principios, teorías, etc. En realidad buena parte de las cuestiones aquí recogidas son propias de epígrafes o subepígrafes de los distintos temas de estudio. Incluso en muchas de ellas se señala el tema y el número del epígrafe del programa al que corresponde la cuestión planteada. Se vuelve a recordar que sobre los ítems de las pruebas objetivas no se realizó este análisis.

Todos los exámenes de Geografía e Historia, Educación y Psicología recogieron preguntas de esta categoría inferior. El elevado número de comentarios de textos incluidos en los exámenes de Filología y Filosofía hacen que en estas carreras el porcentaje de pruebas que valoran capacidades de comprensión y análisis sea elevado. El buen número de exámenes que en las Facultades de Económicas y Empresariales y Psicología corresponden al tipo de aplicación se debe a que en muchos de ellos se presentan problemas.

No nos decidimos por incluir en la máxima categoría, *Evaluación*, a ninguna de las preguntas de los distintos exámenes aunque, como es lógico, surgieron muchas dudas en el proceso.

## 2.2. De las pruebas objetivas

Veamos ahora la estructura de las distintas pruebas objetivas propuestas a los alumnos, considerando por un lado el número de items que contempla la prueba y por el otro el número de alternativas de cada uno de los items. Pudimos observar que, aunque podrían ser válidas para la evaluación, ningún examen contempló pruebas objetivas de asociación ni de ordenamiento. En la siguiente tabla se recoge este análisis.

TIPOS DE PRUEBAS OBJETIVAS (porcentajes)				
Nº ITEMS	NÚMERO ALTERNATIVAS			TOTAL
	2	3	4	
10 - 19	0	2,5	19,5	22
20 - 29	2,5	12	22	36,5
30 - 39	7	17	2,5	26,5
40 - 49	5	7,5	0	12,5
50	2,5	0	0	2,5
<b>TOTAL</b>	<b>17</b>	<b>39</b>	<b>44</b>	<b>100</b>

Se nos antoja excesivamente alto el porcentaje de pruebas que proponían sólo entre 10 y 19 items, un 22%. Ya se sabe que una prueba objetiva de escaso número de items es menos válida que otra que considera mayor número de ellos. Más de la tercera parte de los exámenes que proponían pruebas objetivas utilizaron entre 20 y 29 items para evaluar a los alumnos. Más del 40% de las pruebas objetivas propuestas se construyeron con más de 30 items. Por otra parte, el número de alternativas de cada item se nos presenta en orden creciente de las 2 alternativas (17%) a las cuatro alternativas (44%). Los dos bloques más representados en la tabla que ahora estudiamos son los de pruebas objetivas con un número de items entre 10 y 19 y 4 alternativas (19,5%) y pruebas entre 30 y 39 items y 3 alternativas (17%).

Ya hemos indicado que sobre estas pruebas no hicimos un estudio referido a qué capacidades trataban de medir los items de cada una de las pruebas por las razones que ya hemos tratado de explicar. En todo caso, para este análisis hubiese sido igualmente útil el cuadro de la *Taxonomía de Bloom* más arriba ofrecido. Conviene recordar que mediante unas pruebas objetivas bien elaboradas podemos medir, además de las capacidades referidas al primer tipo *conocimientos*, otras superiores como *comprensión*, *aplicación* y *análisis*. A los críticos de estas pruebas se les hace difícil aceptar estas posibilidades. No tenemos espacio para tratar de ejemplificar tipos de items de pruebas objetivas que pueden evaluar capacidades distintas a las de la mera memorización o repetición que, sin embargo son, como hemos visto, las más utilizadas en la evaluación de nuestros alumnos mediante las pruebas abiertas.



Aunque no llegamos a hacer el análisis referido, sin embargo sí nos detuvimos en cada uno de los exámenes que contemplaban estas pruebas y extrajimos los items de unos u otros exámenes que contemplaban algunos de estos defectos que los catalogaban como items deficientemente redactados.

- Abuso de la evaluación de contenidos meramente memorísticos.
- Ambigüedad en la formulación del item. Falta de claridad.
- Posibilidad de que exista más de una respuesta correcta entre las alternativas.
- Los distractores son poco plausibles por lo que son fácilmente eliminables.
- El pie de algunos items es excesivamente amplio con una información claramente irrelevante y no precisa.
- Unos items de la prueba dan la pista para resolver otros.
- Al redactar el pie se está dando la pista de la respuesta pedida.
- La falta de unidad gramatical con el pie delata a los distractores.
- Existe información reiterativa en todas las alternativas que deberían incluirse en el pie.
- Se abusa de las negaciones en el pie que desvirtúan el item.
- La respuesta correcta está redactada de forma distinta a los distractores.