

Facultad de Psicología

Maquetación y Diseño:

Carlos Quintana Frías
Técnico Especialista de la Facultad (CSI)

Para citar:

Rodríguez de Miñón, P., Barbero, M. I., Navas, M. J., Suárez, J. C., Holgado, F. P., Villarino, A. y Recio, P. (2003). Recomendaciones para la elaboración de pruebas objetivas de evaluación y la interpretación de sus puntuaciones. Madrid: Facultad de Psicología de la UNED



Autores:

Pedro Rodríguez-Miñón Cifuentes
Isabel Barbero García
M^a José Navas Ara
Juan Carlos Suárez Falcón
Francisco Pablo Holgado
Ángel Villarino Vivas
Patricia Recio Saboya

RECOMENDACIONES PARA LA ELABORACIÓN DE PRUEBAS OBJETIVAS DE EVALUACIÓN Y LA INTERPRETACIÓN DE SUS PUNTUACIONES

Facultad de Psicología
Vicedecanato de Metodología.



1	<i>Introducción</i>	7
2	<i>Fases en la elaboración de un examen</i>	9
	2.1 <i>Objetivo de la prueba</i>	9
	2.2 <i>Diseño de la prueba</i>	9
	2.3 <i>Redacción de las preguntas</i>	10
	2.4 <i>Confección de la prueba</i>	10
3.	<i>Diseño del examen</i>	13
4.	<i>Formatos de preguntas del examen</i>	17
	4.1 <i>Respuesta abierta</i>	19
	4.2 <i>Respuesta cerrada</i>	20
5	<i>La redacción de las preguntas del examen</i>	23
	5.1 <i>Principios generales aplicables a cualquier tipo de formato</i>	23
	5.2 <i>La elaboración del enunciado: recomendaciones y sugerencias</i>	25
	5.3 <i>Preguntas de elección múltiple</i>	27
	5.3.1 <i>Redacción del enunciado</i>	27
	5.3.2 <i>Redacción de las opciones</i>	29
	5.3.3 <i>Redacción de los distractores</i>	33
	5.3.4 <i>Número óptimo de alternativas de respuesta</i>	34
	5.4 <i>Preguntas de elección binaria</i>	35
	5.5 <i>Preguntas de respuesta corta o completar</i>	36
	5.6 <i>Preguntas de respuesta extensa o ensayo</i>	37
6	<i>Asignación de puntuaciones</i>	39
7	<i>Interpretación de los resultados</i>	43
	7.1 <i>Interpretación por normas</i>	43
	7.2 <i>Interpretación por criterios</i>	44
	7.3 <i>Evaluación por normas o por criterios</i>	45
	7.4 <i>Dominio de evaluación</i>	46
8	<i>Bibliografía</i>	49



1.- INTRODUCCIÓN.

En la UNED, la calificación de los alumnos se apoya principalmente, y en muchas ocasiones exclusivamente, en el rendimiento obtenido en las Pruebas Presenciales o Personales realizadas de modo simultáneo por los alumnos en todos los centros asociados. Por ese motivo, la evaluación de la calidad métrica de esas pruebas y la posterior mejora de las mismas es uno de los objetivos prioritarios en los próximos años de la UNED. De hecho, el Plan de Mejora de la Calidad de la UNED (1997) señala la necesidad de lograr una evaluación del rendimiento de los alumnos rigurosa, acomodada a los objetivos de aprendizaje y técnicamente bien concebida y la necesidad de establecer controles de calidad internos y externos del servicio educativo y el diseño y aplicación de un sistema de control de calidad. Para ello, es fundamental desarrollar instrumentos que sean aplicables dentro de nuestro sistema educativo y que permitan evaluar diversos aspectos del mismo. Probablemente, como sugieren Jornet y col. (1995), el sector en el que es más patente la necesidad de desarrollar este tipo de instrumentos es en el de la evaluación del rendimiento educativo, tanto cuando el objetivo es la evaluación individual de los alumnos, como cuando se quiere llevar a cabo la evaluación del propio sistema (el rendimiento como indicador de calidad).

Esta necesidad también se puede apreciar en la Facultad de Psicología. Así, en el Informe Final de Evaluación de la Enseñanza en la Titulación de Psicología (2002), en el apartado de puntos fuertes y débiles y propuestas de mejora de la evaluación de aprendizajes se señala que "sería conveniente la creación por parte de la Facultad de una comisión de supervisión de exámenes que tuviera en cuenta el grado de dificultad, la adecuación a los contenidos de la asignatura, la relevancia de las preguntas, etc.", (Pág. 22).

En sintonía con los objetivos generales de actuación de la UNED y con el fin de cubrir la necesidad puesta de manifiesto en el informe de evaluación de la Facultad citado en el párrafo anterior, el vicedecanato de Metodología de la Facultad de Psicología, en colaboración con profesores del Departamento de Metodología de las Ciencias del Comportamiento propuso la creación de una *Unidad Técnica para la Elaboración y*

Evaluación de Pruebas Objetivas de Rendimiento. Los objetivos, funciones y características de esta unidad técnica se presentaron en una mesa redonda titulada *“Elaboración de pruebas objetivas: ventajas e inconvenientes de su utilización”* durante la IV Semana de Investigación de la Facultad de Psicología de la UNED, celebrada en noviembre de 2002. Las funciones principales de esta unidad técnica serían:

- a. Revisar las técnicas de construcción y escritura de pruebas objetivas y divulgación de las mismas para uso del profesorado.
- b. Diseñar y desarrollar un programa informático que facilite la elaboración y evaluación de los exámenes así como la gestión de las calificaciones.
- c. Impartir cursos de formación dirigidos al profesorado de la UNED en los que aprenderán el manejo de este programa y la interpretación correcta del informe de calidad del examen que ofrece el programa.
- d. Ofrecer un servicio permanente de orientación, asesoramiento y colaboración con el profesor de la UNED ante las dificultades técnicas que puedan surgir con el uso del programa y en la interpretación del informe de calidad cuando lo apliquen a los exámenes de su asignatura.

En este documento vamos a abordar la primera de las funciones de esta unidad técnica, ofreciendo una serie de recomendaciones para la elaboración de pruebas de evaluación que puedan proporcionarnos la información necesaria para que las decisiones que adoptemos en relación con el rendimiento de los alumnos tengan las suficientes garantías de objetividad y rigor científico.



2.- FASES EN LA ELABORACIÓN DE UN EXAMEN

Son cuatro las fases que hay que cubrir para elaborar una prueba de evaluación: (1) la definición del objetivo de la prueba, (2) el diseño de la misma, (3) la redacción de las preguntas que la van a conformar y (4) la confección de la prueba propiamente dicha. En este apartado se expondrá con cierto detalle las fases 1 y 4, apuntando solo el contenido de las fases 2 y 3, que serán desarrolladas en apartados posteriores del documento.

2.1.- Objetivo de la prueba: La primera fase supone definir de forma clara e inequívoca el objetivo que se persigue con esa nueva prueba o instrumento y, para ello, habrá que dar respuesta a las tres preguntas siguientes:

¿Qué se va a medir? Nos lleva a la variable de interés, a lo que queremos medir con esa nueva prueba, en nuestro caso, el rendimiento de nuestros alumnos en la asignatura que impartimos. Para ello, hay que tener en cuenta el programa y los objetivos de la asignatura en cuestión y acotar el conocimiento que podemos exigir al alumno.

¿A quién se va a medir? Nos lleva a la población en la que estamos interesados en conocer el nivel en el que los sujetos presentan o manifiestan la variable de interés. En nuestro caso, la población de los alumnos matriculados en nuestra asignatura.

¿Para qué se va a medir esa variable en esa población? Nos lleva a la utilización prevista de la prueba y, en este caso, se centra en la evaluación del rendimiento académico de los alumnos con el fin de realizar la correspondiente acreditación académica.

2.2.- Diseño de la prueba: El objetivo de esta fase sería definir con claridad la estructura del examen, atendiendo para ello a cuatro aspectos básicos: contenido y longitud del examen, formato y características métricas de las preguntas del examen.

2.3.- Redacción de las preguntas: Recomendaciones o normas a seguir para escribir buenas preguntas en los exámenes y que desarrollaremos más ampliamente en apartados siguientes.

2.4.- Confección de la prueba: Se debe construir la prueba de forma tal que facilitemos unas instrucciones sencillas de seguir, la lectura de sus preguntas y el registro de sus respuestas. En esta fase se toman decisiones sobre tres aspectos importantes: las instrucciones de la prueba, el ensamblaje de las preguntas en la prueba y el formato de registro de las respuestas de los sujetos.

- a. Con unas *instrucciones* adecuadas se trata de asegurar que todos los alumnos comprendan perfectamente la mecánica del examen para que su desconocimiento no afecte negativamente a la hora de responder a las preguntas. De esta manera podemos contribuir de forma positiva a su finalidad última de evaluación. Por ello, conviene indicar, con un lenguaje claro y sencillo, cómo se debe responder a las preguntas (rodeando con un círculo la alternativa que considere correcta o la opción que considere más acertada, o respondiendo exclusivamente en el espacio proporcionado al efecto, etc); dónde debe responder (en la misma hoja de examen o en hoja de respuesta aparte); cómo se van a puntuar las respuestas y si se penalizan los errores, por lo que no sería conveniente responder al azar cuando se tienen serias dudas sobre la respuesta correcta; el tiempo disponible para realizar la prueba y alguna sugerencia de cómo distribuirlo cuando se disponen de distintos tipos de preguntas (por ejemplo: teóricas o conceptuales y prácticas o de cálculo).
- b. El *ensamblaje* de las preguntas es importante para evitar cualquier tipo de confusión o desconcierto. Conviene agrupar las preguntas relacionadas con un mismo tema; las que comparten el mismo enunciado o material de referencia y aquellas con las mismas instrucciones o con el mismo formato. En situaciones como ésta, lo que debe hacerse es separar físicamente los grupos de preguntas y redactar un enunciado introductorio para cada bloque, del tipo: “*A continuación hay una serie de preguntas acerca de.....*”. Con esto



se trata de conseguir una prueba fácil de seguir, atractiva y con sentido.

- c. El *formato de registro* de las respuestas responde a dos posibilidades: directamente en la hoja del examen o en hoja aparte que, como se ha indicado, debe estar claramente especificado en las instrucciones.



3.- DISEÑO DEL EXAMEN

En la segunda fase del proceso de construcción de un examen vamos a definir o diseñar por completo la estructura de la prueba. Para ello, teniendo en cuenta el *qué*, el *a quién* y el *para qué* de la fase anterior, tendremos que tomar decisiones respecto al contenido y longitud del examen, así como acerca del formato y características métricas de las preguntas.

Vamos a ilustrar la forma de proceder en esta fase con un ejemplo, que servirá asimismo para evidenciar la estrecha dependencia que tienen las decisiones que tomemos sobre estos cuatro aspectos del diseño del examen. Dada la importancia que tiene el formato utilizado para las preguntas del examen, esta cuestión será tratada de forma independiente en el siguiente apartado mientras que en éste nos limitaremos al contenido y longitud de la prueba y a las características métricas de las preguntas, asumiendo en nuestro ejemplo que el examen va a estar formado por preguntas "tipo test".

Para definir el *contenido* de un examen se pueden utilizar muchas estrategias. Una de las más frecuentes en el ámbito cognitivo en el que nos movemos los docentes, es la de la *tabla de especificación* del examen.

		NIVEL DE DIFICULTAD			
		BAJA	MEDIA	ALTA	
CONTENIDO	BLOQUE 1 (2 temas)				17%
	BLOQUE 2 (6 temas)				50%
	BLOQUE 3 (4 temas)				33%
		20%	60%	20%	100%

Para definir el contenido de la prueba construimos una tabla de doble entrada y ponemos en las filas, por ejemplo, los distintos bloques temáticos, temas o unidades didácticas de nuestra asignatura. En función del peso que tenga cada uno de esos bloques en el programa de la asignatura (considerando bien su longitud o su importancia relativa), podemos decidir qué porcentaje de la prueba tiene que ver con cada bloque. En este caso, atendiendo al número de temas por bloque, vemos que el 17% contendrá material relativo al bloque 1, la mitad de la prueba tendrá que ver con el bloque 2 y el 33% restante del examen abordará contenidos relativos al bloque 3. De este modo, estamos poniendo los medios para tratar de conseguir que la prueba constituya una muestra representativa de los contenidos del programa de la asignatura.

Con esto tendríamos definida una dimensión de la tabla de especificación de la prueba. La segunda dimensión podría ser otro de los cuatro aspectos considerados en esta fase de la construcción del examen: *las características métricas* de las preguntas, en particular, su dificultad "a priori" que por experiencia profesional nos atrevemos a aventurar. (El análisis estadístico final nos dará o quitará la razón acerca de lo acertado de esta estimación).

Dado que el objetivo o utilización prevista de la prueba es la acreditación académica, conviene que tengamos un buen número de preguntas de dificultad media (entre un 50 y un 65% del total de la prueba), ya que son las preguntas que más discriminan, pero también necesitamos disponer de un cierto número de preguntas algo más fáciles y algo más difíciles para poder diferenciar entre los alumnos con un gran dominio de la asignatura y entre los que apenas han estudiado. En este ejemplo, la propuesta concreta es 20%-60%-20% de preguntas de dificultad baja, media y alta, respectivamente.

Con esta información, tendríamos estructurado el examen. Lo único que nos faltaría sería decidir qué longitud tiene que tener la prueba para traducir estos porcentajes que tenemos en los marginales de las filas y las columnas de la tabla de especificación en número de preguntas.



Para decidir acerca de la *longitud* de la prueba tenemos que tener en cuenta muchos factores, entre ellos, el formato de las preguntas (no se tarda lo mismo en resolver preguntas de elección múltiple que una pregunta tipo ensayo o una de relacionar), el objetivo o utilización prevista de la prueba (cuando ésta se va a utilizar para tomar decisiones importantes sobre los sujetos, es conveniente que la prueba tenga cuantas más preguntas mejor, ya que en ese caso el error de medida es menor) y el tiempo máximo de aplicación disponible. En el ejemplo que nos ocupa, trabajamos con preguntas de elección múltiple de tres opciones de respuesta para una prueba cuyo objetivo es la acreditación académica y disponemos como mucho de dos horas de duración para el examen. Por consiguiente, podemos poner perfectamente 50 preguntas en nuestro examen, aunque depende también del tipo de tareas que requieran esas preguntas de elección múltiple (no se tarda lo mismo en responder a una pregunta que supone resolver un problema en cuatro pasos que uno que requiere un único paso para su resolución).

Con una simple regla de tres calculamos cuántas preguntas de cada bloque tenemos que construir y cuántas de dificultad baja, media y alta tenemos que redactar. De este modo, tendremos $(17 \times 50 / 100) \approx 9$ preguntas del bloque 1, 25 del bloque 2 y 16 del bloque 3; asimismo, habrá $(20 \times 50 / 100) = 10$ preguntas de dificultad baja, 30 de dificultad media y los 10 restantes de dificultad alta.

Podemos quedarnos aquí o dar un paso más y trabajar con arreglo no ya a los marginales de la tabla sino a las celdillas de la misma. Esto es, podemos construir 9 preguntas del primer bloque del programa, pero 2 de dificultad baja, 5 de dificultad media y 2 difíciles. Y lo mismo para los bloques 2 y 3 del programa. La forma de proceder para determinar los valores concretos de las celdillas es idéntica a la seguida para calcular el total de preguntas en cada categoría de las dos dimensiones consideradas en la tabla de especificación. Para cada celdilla se indica el porcentaje de preguntas que debe contener el examen con esas características (por ejemplo, el 3'4% de las preguntas tienen que referirse al bloque 1 del programa y tener una dificultad baja) y el número de preguntas que representa en el total de la prueba $((3'4 \times 50) / 100 \approx 2$ preguntas).

Ésta es solo una de las muchas formas posibles que puede adoptar una tabla de especificación. Por ejemplo, podemos sustituir los bloques temáticos directamente por los temas del programa o por los objetivos definidos para la asignatura. La dificultad de las preguntas se puede sustituir por el proceso cognitivo despertado por la pregunta en su resolución (conocimiento, comprensión, aplicación, análisis, síntesis, evaluación, etc.).

		NIVEL DE DIFICULTAD			
		BAJA	MEDIA	ALTA	
CON TE NI DO	BLOQUE 1 (2 temas)	34 (2)	10'2 (5)	3'4 (2)	17 (9)
	BLOQUE 2 (6 temas)	10 (5)	30 (15)	10 (5)	50 (25)
	BLOQUE 3 (4 temas)	6'6 (3)	19'8 (10)	6'6 (3)	33 (16)
		20 (10)	60 (30)	20 (10)	100 (50)

De este modo, hemos definido la estructura del examen, decidiendo cuál va a ser el formato de las preguntas, su longitud, su contenido y características métricas, teniendo presente lo que queremos medir, a quién queremos medir y para qué queremos medir.

4.- FORMATOS DE PREGUNTAS DEL EXAMEN

Desde tiempo atrás viene la polémica entre los partidarios de las “pruebas tipo test” y los defensores de las “pruebas tradicionales de preguntas abiertas” consistentes estas últimas en el desarrollo de un tema general que permite evaluar aspectos como la expresividad oral o escrita, la creatividad, la riqueza de vocabulario, capacidad de síntesis, de solución de problemas, etc. Sus defensores critican que con las “pruebas tipo test” no se pueden conocer estos aspectos y que, además, se refuerza un tipo de aprendizaje más centrado en lo concreto que en las ideas generales y que los alumnos, en muchas situaciones, aprenden las respuestas sin entender las preguntas.



Por el contrario, los defensores de las “pruebas tipo test” se apoyan en las ventajas de su corrección “objetiva”, en la medida en que con preguntas abiertas de desarrollo extenso es más difícil precisar los criterios de corrección entre distintos jueces e incluso para el mismo juez con diferentes pruebas, por lo que a medida que aumenta la extensión de

la respuesta del alumno disminuye la fiabilidad de la prueba, entendiendo por fiabilidad la constancia de la medida obtenida. Una prueba de evaluación es fiable si, suponiendo que lo aprendido se mantiene constante, los mismos alumnos obtienen las mismas estimaciones del nivel de conocimientos al contestar a la misma prueba en condiciones similares. Una de las posibles causas de esta pérdida de fiabilidad de la "prueba tradicional" se debe, no a la mala calidad de la prueba, sino a la deficiente precisión en la corrección y al pobre acuerdo entre distintos evaluadores. El alumno puede enmascarar su carencia de conocimientos con una redacción y exposición que produzca equívocos sobre el objetivo final. Esto dificulta la labor del profesor para atenerse estrictamente a los criterios de corrección y le lleva a aplicar criterios propios que considere relevantes para cada caso.

Por tanto, la fiabilidad de la prueba debe ser independiente de quién sea el corrector. Pues bien, en la medida en que se cierre o se acorte la extensión de la respuesta del alumno es más fácil fijar unos criterios de corrección claros y precisos que mejoren sustancialmente el acuerdo entre los distintos componentes del equipo docente y, en consecuencia, es más difícil que la puntuación asignada por el profesor-corrector esté contaminada por factores subjetivos.

Por otro lado, una prueba de evaluación, en general, se dice que es válida cuando mide realmente la variable objeto de evaluación y no otra. Lógicamente las "pruebas tipo test", en la medida en que incluyen más preguntas, ofrecen más flexibilidad para evaluar una gran variedad de contenidos del programa docente y objetivos de aprendizaje, lo que redundará en mejorar su "validez de contenido". Con preguntas de respuesta extensa sólo se podrá preguntar sobre pocas cosas. Si, por otra parte, las "preguntas tipo test" están bien construidas, se pueden evaluar habilidades mentales más complejas y sofisticadas que la simple memorización, y se puede dirigir un estudio inteligente aunque el alumno conozca de antemano el tipo de prueba que debe realizar.

Entre sus aspectos negativos se apunta la posibilidad de acertar "al azar" y que el alumno sólo tiene que seleccionar la respuesta correcta



entre varias alternativas por lo que, como se ha comentado, inhibe la expresión de la creatividad, la exhibición de la originalidad y el pensamiento creativo y, según los más críticos, reduce todos los conocimientos importantes a hechos superficiales, vulgarizando la adquisición de conocimientos y de ideas más globales. Los críticos de las preguntas “tipo test” se apoyan en esta vulgarización del conocimiento en el sentido de que este tipo de preguntas, y concretamente las de elección múltiple, sugieren que existe una, y solo una, respuesta correcta para cada problema o situación obviando las posibles matizaciones.

Por tanto, en cuanto al formato de las preguntas, debemos decidir si vamos a diseñar preguntas con formato de respuesta cerrada en las que el sujeto selecciona la respuesta de entre una serie de posibles alternativas o formato de respuesta abierta en la que el sujeto produce la respuesta.

4.1.- Respuesta abierta

El formato de respuesta abierta –denominado también de producción de respuesta- se adapta mejor para la medición de procesos cognitivos complejos, y entre ellos podemos citar el formato de respuesta corta y el ensayo o formato de respuesta extensa.

Formato de respuesta corta:

Escriba en los espacios en blanco los estímulos correspondientes

En los ensayos excitatorios del procedimiento experimental se presenta el _____ seguido de _____, y en los ensayos inhibitorios se presenta el _____seguido de_____.

En los ensayos excitatorios del procedimiento diferencial se presenta el _____seguido de _____, y en los ensayos inhibitorios se presenta el _____seguido de _____.

Formato de respuesta extensa:

Exponga la evolución histórica del concepto de validez

4.2.- Respuesta cerrada

El formato de respuesta cerrada es el más utilizado, principalmente en evaluaciones masivas de sujetos, y es el que proporciona una medición más objetiva y fiable, además de una corrección eficiente. A los exámenes con este tipo de formato se les conoce también como pruebas objetivas y los tipos de preguntas utilizados con más frecuencia en los exámenes o pruebas presenciales de la UNED son los de elección binaria (verdadero-falso) y de elección múltiple, siendo también posibles las preguntas de relación, que responden a un formato como el que se muestra en el siguiente apartado.

Formato de relación

A este formato se le conoce también con el nombre de formato de emparejamiento y la tarea del sujeto consiste en relacionar lo que aparece en la columna de premisas con lo que aparece en la columna de respuesta de acuerdo a las instrucciones que se proporcionen.

Relacione, uniendo mediante flechas, los conceptos enumerados en la columna derecha con los que correspondan en la columna izquierda para aplicar el coeficiente de correlación más adecuado en función del tipo de variables que estemos analizando en cada caso:

- | | |
|---|----------------------|
| 1.- Dos variables cuantitativas | a.- Spearman |
| 2.- Dos variables cualitativas | b.- Pearson |
| 3.- Dos variable ordinales | c.- Biserial-puntual |
| 4.- Una variable dicotómica y otra cuantitativa | d.- Phi |



Formato de elección:

En las preguntas de elección se presenta al alumno un enunciado para que responda eligiendo la respuesta más adecuada. La respuesta puede ser binaria o bien de varias alternativas de respuesta o de elección múltiple. Entre las preguntas de elección binaria, la opción de verdadero-falso (V-F) es probablemente una de las más utilizadas por su aparente simplicidad, tanto desde un punto de vista conceptual como por su facilidad de construcción.

Formato de elección múltiple:

¿Qué técnica de Análisis de datos utiliza el diagrama de tallo y hojas como representación gráfica?

- a. El análisis exploratorio de datos
- b. El análisis bivariado
- c. Las técnicas inferenciales

Formato de elección binaria

El diagrama de tallo y hojas es un tipo de representación gráfica utilizado en el análisis exploratorio de datos	V	F
El diagrama de tallo y hojas es un tipo de representación gráfica utilizado en el análisis bivariado	V	F
El diagrama de tallo y hojas es un tipo de representación gráfica utilizado en las técnicas inferenciales	V	F

Una variedad de este tipo de preguntas se presenta cuando el alumno debe responder primero a alguna de las alternativas V-F y si opta por la opción de Falso, deberá proporcionar la respuesta correcta. De esta forma la probabilidad de acertar la respuesta correcta se reduce del 50% inicial a un valor menor.



5.- LA REDACCIÓN DE LAS PREGUNTAS DEL EXAMEN

A pesar de las dificultades que supone definir unos criterios absolutos para establecer lo que representa una buena pregunta o enunciado, se recomiendan ciertas normas que deben seguirse durante la construcción de la prueba. Estas recomendaciones van dirigidas a conseguir preguntas “representativas”, donde la respuesta del alumno deba atribuirse a su nivel de conocimientos de la materia objeto de evaluación y no a otras variables. En la práctica puede haber múltiples factores desconocidos e incontrolables que afectan a la respuesta, como la motivación del alumno, su práctica en responder a preguntas tipo test, la ansiedad, la adivinación, etc; pero otros factores relacionados con la propia construcción, por lo que a su redacción se refiere, sí pueden controlarse y mejorarse para procurar que esta condición se cumpla. En cualquier caso, la importancia de esta condición descansa en que, de no cumplirse, no disponemos de un procedimiento adecuado para inferir que la respuesta correcta de los sujetos se debe a su nivel de competencia o conocimientos.

Casi todo lo relacionado con los criterios de construcción de preguntas puede resumirse con la palabra *claridad*, en el sentido de cómo: a) está redactado, b) se relaciona con el dominio objeto de evaluación, c) “dirige” al estudiante competente hacia lo que se le pide.

5.1.- Principios generales aplicables a cualquier tipo de formato

- Adaptar la complejidad de las preguntas al nivel de los estudiantes. (El nivel de complejidad se refiere tanto a su redacción como a su contenido o incluso a los números que se utilicen en las preguntas de tipo matemático. Por ejemplo, salvo que se esté interesado en evaluar la habilidad para manejar números grandes, los números pequeños son igualmente apropiados para los problemas numéricos).
- Preguntar en concordancia con los procesos mentales que se desea que use el estudiante. Probar el conocimiento de los hechos, la comprensión, el razonamiento a partir de unos hechos, etc.

- Utilizar preguntas de diferente complejidad para discriminar entre todos los niveles de competencia del alumnado.
- Evitar las preguntas triviales. Es fácil preguntar por fechas, nombres, fórmulas y hechos simples que, sin ser nada ambiguas, no suelen ser importantes en detrimento de aquellas otras que sí lo son porque tratan de determinar cómo razonan los estudiantes respecto a la materia evaluada.
- Hacer las preguntas independientes unas de otras de modo que no existan preguntas que hagan referencia al problema planteado en otra o requieran conocer su respuesta correcta. Así, aunque dos preguntas estén relacionadas en cuanto a su contenido, la probabilidad de responder correctamente a una de ellas no debe condicionar la respuesta de la otra. Ejemplo de preguntas “dependientes”:

<p>¿Cuál de los siguientes insectos es un himenóptero?</p> <ul style="list-style-type: none">a. El mosquitob. La abejac. El escarabajo	<p>¿Por qué la abeja es un himenóptero?</p> <ul style="list-style-type: none">a. Recoge el polen de las floresb. Tiene dos pares de alas membranosasc. No tiene patas en el abdomen
--	---

- Evitar la doble negación. Ejemplos:

<p>Dos conjuntos de puntuaciones que NO tienen la misma puntuación verdadera y la misma varianza error, NO pueden ser medidas paralelas.....V F</p>	<p>Si una distribución de frecuencias NO es simétrica, la media:</p> <ul style="list-style-type: none">a. NO es igual a la modab. Es menor que la modac. Es mayor que la mediana
---	--



5.2.- La elaboración del enunciado: recomendaciones y sugerencias

Recomendaciones:

- Expresar con claridad, precisión y con un vocabulario adecuado la idea central objeto de evaluación. Si la pregunta se redacta de forma vaga, ambigua, de forma que origine distintas interpretaciones pierde validez para medir lo que pretende medir y se verá afectado por otras variables, como la riqueza de vocabulario del sujeto, la comprensión lectora, etc.
- Utilizar expresiones gramaticales sencillas y correctas.
- Utilizar preferentemente una frase o párrafo corto que presente al alumno una situación o problema específico que requiera una solución particular. Sólo en el caso de que la pregunta sea más compleja y en aras de la claridad, la extensión tendría prioridad sobre la concisión.
- Evitar en la redacción palabras superfluas o irrelevantes, tautologías, frases repetidas y palabras con doble significado
- No utilizar el enunciado para enseñar. La finalidad del enunciado nunca debe ser instructiva. En el siguiente ejemplo la frase señalada en cursiva es claramente instructiva e irrelevante para la finalidad evaluativa de la pregunta.

El coeficiente de fiabilidad de un test, *que hace referencia a la precisión y consistencia de la medida, a su capacidad para proporcionar la misma puntuación cuando se aplica varias veces a una misma muestra de sujetos*, recibe el nombre de coeficiente de equivalencia cuando se calcula por el procedimiento de:

- a. División en mitades
- b. Test-retest
- c. Formas paralelas

Sugerencias:

En la siguiente tabla se presentan distintos tipos de frases con palabra introductorias para estimular la imaginación cuando iniciamos el proceso de redacción de las preguntas.

Frase	Ejemplos de palabras introductorias	Ejemplo de pregunta
Comparativas	La diferencia entre----y ---- es.... El----y----necesitan.....	La diferencia entre las puntuaciones directas correspondiente a los percentiles 45 y 55 es probablemente _____ que la diferencia entre las puntuaciones directas correspondientes a los percentiles 5 y 15.
Condicionales	Si----entonces..... Cuando.....	Cuando trabajamos con una distribución normal la desviación típica y el error típico son equivalentes.....V F
Relacionales	El mayor.... El más alto.... El más bajo.... Al construir----es probable que....	Al construir un test más fiable es probable que sea: a) menos válido; b) más válido; c) más corto.
Explicativas o aclaratorias	La principal razón para.... El fundamento de.... Uno de los factores que.... Desde.... Después de... Si bien....	Uno de los factores que perjudica la fiabilidad de una prueba objetiva es la cantidad de aciertos al azar que se cometen en las respuestas.....V F
Ejemplos	Un ejemplo de.... Una situación de...	Un ejemplo de variable medida con escala nominal es: a) el peso; b) la ciudad de origen; c) la inteligencia.
Evidencias	Estudios de....revelan que...	Los estudios sobre la percepción de la profundidad revelan que los bebés de seis meses evitan gatear sobre un abismo visual.....V F
Predictivas	Es de esperar que.... El aumento en....producirá...	Cabe esperar que doblando el número de ítems de un test su coeficiente de fiabilidad inicial pase de 0,30 a 0,60.....V F
Cálculo	Ítems que incluyen números y requieren operaciones matemáticas.	El rango de las puntuaciones 2, 3, 4 y 7 es _____.



Procedimiento	Para encontrar---se debe... Con el fin de---se debe... Un método de---es... Un paso esencial para---es... La aplicación de.... El primer paso hacia.....	Una forma de asegurar que las puntuaciones obtenidas en un examen de 100 preguntas sean equivalentes a las obtenidas con otro de 25 preguntas es multiplicar estas últimas por cuatro.V F
Evaluativos	Un buen procedimiento para... Es mejor---que... Lo mejor es... La forma más fácil de---es... Es fácil demostrar que.... Es necesario que--- para... La mayor desventaja de--- es...	Es difícil obtener puntuaciones fiables en grupos con mucha variabilidad.V F

5.3- Preguntas de elección múltiple

Este tipo de pregunta tiene un enunciado que presenta una situación o una pregunta sobre un concepto o hecho con varias alternativas de posibles respuestas para que el alumno elija una, y solo una. Tanto el enunciado como la respuesta correcta determinan la clasificación del contenido de la pregunta, ya que las alternativas de respuestas incorrectas (denominados distractores) pueden referirse a varios contenidos o áreas sin determinar el contenido de la pregunta.

A continuación se presentan reglas y recomendaciones que hacen referencia a la redacción de: a) el enunciado, b) la opción correcta y c) los distractores.

5.3.1. Redacción del enunciado:

- Evitar el uso de negaciones. Los enunciados es preferible redactarlos de forma afirmativa. La experiencia ha puesto de manifiesto que esta forma de redacción facilita la comprensión y proporciona alternativas de respuesta más realistas. En todo caso, si se opta por incluir

preguntas redactadas en forma negativa, la práctica recomendada es resaltar el elemento negativo de la redacción y, bajo ningún concepto, incluir alternativas negativas ya que implicaría la doble negación. En cualquier caso la prueba en su conjunto no debe incluir más de un cinco, o como mucho un diez por ciento de este tipo de enunciados.

¿Cuál de estos animales NO es un mamífero?

- a. La vaca
- b. El delfín
- c. La jirafa
- d. El avestruz

- Poner la idea principal en el enunciado y no en las opciones. Asimismo, no es pertinente repetir información en las alternativas de respuesta cuando ésta pueda situarse en el enunciado.

Ejemplo de forma incorrecta y correcta de redacción:

<p>El espacio latente:</p> <ul style="list-style-type: none">a. Se denomina completo cuando contiene todos los componentes necesarios para describir el rasgob. Se denomina completo cuando contiene todas las puntuaciones de los sujetosc. Se denomina completo cuando contiene todos los valores de la escala del rasgo	<p>El espacio latente se denomina completo cuando contiene:</p> <ul style="list-style-type: none">a. Todos los componentes necesarios para describir el rasgob. Todas las puntuaciones de los sujetosc. Todos los valores de la escala del rasgo
--	--

- Cuidar la consistencia gramatical entre el enunciado y todas las opciones de respuesta. Es frecuente utilizar preguntas donde el enunciado queda incompleto y se cierra con las alternativas de respuesta. Hay que utilizar este formato de respuesta con cautela, evitando que una o más de las alternativas de respuesta no cierren el



sentido gramatical del enunciado, como ocurre con la alternativa *d* del siguiente ejemplo. Una posible solución es reescribir la pregunta de forma interrogativa. Otra razón para evitar las frases incompletas en la redacción de los enunciados es que pueden reducir la pregunta a una cuestión de memoria. Ejemplo de forma incorrecta y correcta de redacción:

<p>El diagrama de tallo y hojas es un tipo de representación gráfica utilizado en:</p> <ol style="list-style-type: none"> El análisis exploratorio de datos El análisis bivariado Las técnicas inferenciales Ninguna técnica de análisis de datos 	<p>¿En qué técnica de Análisis de datos se utiliza el diagrama de tallo y hojas como representación gráfica?</p> <ol style="list-style-type: none"> En el análisis exploratorio de datos En el análisis bivariado En las técnicas inferenciales No se utiliza en ninguna técnica de análisis de datos
---	---

5.3.2. Redacción de las opciones:

- Proporcionar una única alternativa que sea correcta. Esto puede parecer obvio, pero en ocasiones puede ocurrir que todas las alternativas sean tan creíbles que dificulte el acuerdo entre los expertos a la hora de decidir cuál es la alternativa correcta.

<p>¿Qué razón es la que mejor describe que los Sindicatos defiendan una huelga general?</p> <ol style="list-style-type: none"> Fortalecer la alternativa de la izquierda Parar la nueva normativa legal Desgastar la acción del gobierno No colaborar en la elaboración de un nuevo marco legal

En el ejemplo, la construcción de la pregunta es pobre porque incluso los expertos en el tema discreparían sobre la razón más importante de la huelga. La redacción debe hacerse con cuidado evitando

ambigüedades de este tipo. Por otra parte, la presencia de distractores creíbles perjudica a los sujetos más capaces en la medida en que disponen de conocimientos adicionales que trascienden el contenido de la pregunta e interfieren con la respuesta, que está basada exclusivamente en la información presentada en el enunciado de la pregunta. Esta razón justifica que se realice un análisis estadístico posterior que permita discernir estas situaciones.

- Mantener todas las alternativas, aproximadamente, con la misma longitud, estructura sintáctica y nivel de complejidad con el fin de evitar la introducción de variables extrañas que proporcionen pistas para descartar o seleccionar una de las opciones.

La impronta o troquelado:

- a. Es una conducta innata, no se aprende
- b. Se produce sin la existencia de ningún tipo de estimulación
- c. Puede ser de dos tipos: territorial y comunicativa
- d. Es el proceso por el cual un miembro de una especie dada adquiere preferencia por interactuar social y sexualmente con un determinado tipo de individuos como consecuencia de la experiencia

- Evitar el “efecto de resonancia”. Se produce cuando las palabras utilizadas en la redacción de la pregunta se asocian con una de las alternativas de respuesta facilitando la adivinación.

¿Cuál es el área cerebral cuya lesión afecta principalmente a la memoria?

- a. Hipocampo
- b. Exceso de dopamina
- c. Desarreglo hormonal

El alumno selecciona la alternativa de respuesta, no por sus conocimientos o habilidades, sino porque encuentra una asociación



entre las palabras utilizadas en la redacción de la pregunta y una de las alternativas de respuesta. Algunas veces es tan obvia, que solamente una de las alternativas de respuesta cumple la condición descrita en la redacción de la pregunta. Otras veces puede ser más sutil pero igualmente fácil de identificar.

- Ser cuidadoso al utilizar el recurso a *“Todas las anteriores”* o *“Ninguna de la anteriores”* como alternativa de respuesta. La utilización de este tipo de opción como alternativa de respuesta está muy extendida porque se ajustan fácilmente a casi todas las preguntas y facilita enormemente la tarea del que las construye. Sin embargo, como norma general se debe ser cuidadoso a la hora de incluirlas en nuestras preguntas y no abusar de ellas, utilizándolas sólo cuando está justificado por alguna circunstancia particular, ya que no se conoce ningún beneficio concreto distinto al objetivo final de discriminar entre los que conocen la respuesta correcta y los que no. Muy al contrario, este tipo de alternativas produce cierta contrariedad en el alumno que puede percibir las como un recurso fácil, por parte de un profesor desganado, para completar las alternativas de respuesta. Investigaciones recientes han encontrado que al presentar la alternativa *“Ninguna de las anteriores”* se incrementa el nivel de dificultad de la pregunta, disminuye su discriminación y también disminuye la fiabilidad de la prueba.

¿Qué exigencias debe cumplir una investigación para denominarla “experimental”?

- a. La manipulación de la variable independiente
- b. El control de las variables extrañas
- c. La asignación aleatoria de los sujetos a las condiciones experimentales
- d. Todas las anteriores

Cuando se redacta una pregunta con cuatro alternativas de respuesta incluyendo la opción *“Todas las anteriores”*, el alumno puede tener suficiente conocimiento de esa pregunta para saber que dos de las alternativas presentadas son correctas, sin saber absolutamente

nada sobre lo planteado en la tercera opción. En estas circunstancias, seleccionará la opción "*Todas las anteriores*" a partir del conocimiento parcial de las otras alternativas.

Basta con conocer dos de las tres primeras características, para que, sin tener absoluta seguridad respecto de la tercera, se seleccione la alternativa "*Todas las anteriores*" como correcta. Por esta razón, este tipo de alternativa de respuesta no debe seleccionarse como la respuesta correcta en una pregunta de elección múltiple bien construida.

Por otra parte si la respuesta considerada correcta es "*Todas las anteriores*" ello implica que cada una de las restantes alternativas es también correcta. Por ello, la ambigüedad y otros defectos en la construcción de la pregunta pueden proporcionar una puntuación que no refleje fielmente el nivel de competencia del alumno. Entonces, es importante resaltar en las *instrucciones* si se trata de seleccionar la respuesta correcta (una y solo una) de entre las distintas alternativas o la que respecto al conjunto sea la más verosímil.

Por el contrario, este formato de respuesta es especialmente útil en aquellos casos donde el sujeto puede resolverlo a partir del enunciado de la pregunta sin necesidad de buscar entre las alternativas de respuesta. Por ejemplo, cuando se requiere operaciones aritméticas, habilidades lógicas, pruebas de carácter memorístico como datos históricos, nombres o fechas:

¿Cuál es el resultado de la suma $54 + 14$?:

- a. 40
- b. 48
- c. 68
- d. ninguna de las anteriores



Otra de las razones para incluir este tipo de alternativa de respuesta se presenta cuando no se pueden presentar todas las alternativas de respuesta, como en el siguiente ejemplo:

En el Campeonato Mundial de Fútbol de Corea-Japón del año 2002 participaron, entre otras, las siguientes selecciones: Panamá, Alemania y Corea. ¿Cuál o cuales jugaron contra España?

- a. Panamá
- b. Alemania y Corea
- c. Panamá y Corea
- d. Ninguna de las anteriores es correcta

En situaciones como ésta, la lista de posibles opciones de respuestas es tan amplia que justifica recurrir a esta alternativa de respuesta.

- Las respuestas correctas deben colocarse al azar, distribuidas equitativamente entre las distintas posiciones.

5.3.3. Redacción de los distractores:

- Elegir distractores plausibles. Las opciones incorrectas deben ser igualmente atractivas y creíbles para el que no conoce la respuesta correcta. Para ello, es recomendable utilizar alternativas que son verdaderas para otras preguntas, o errores cometidos más frecuentemente por los alumnos y evitar alternativas “graciosas” que no serían elegidas por nadie. Ejemplo:

Cuál de las siguientes “Ilusiones Visuales” es la más conocida y estudiada?

- a. Muller-Lyer
- b. Claudia-Sheaffer
- c. Ponzo-Poggendorf

- Utilizar los distractores para modificar la dificultad. La proporción de aciertos varía en función de lo atractivo que sean los distractores, de modo que si se cambian éstos variará el índice de dificultad de la pregunta. De esta forma se puede adaptar el contenido de una misma pregunta para sujetos con distinto nivel de competencia. Véase en el ejemplo cómo cambia la dificultad de la pregunta modificando las opciones incorrectas.

<p>¿Cuál de las siguientes actrices ha sido la criada más famosa del cine español de los años 60?</p> <p>a. Maribel Verdú b. Gracita Morales c. Carmen Maura</p>	<p>¿Cuál de las siguientes actrices ha sido la criada más famosa del cine español de los años 60?</p> <p>a. Florinda Chico b. Gracita Morales c. Lina Morgan</p>
--	--

5.3.4. Número óptimo de alternativas de respuesta:

Las preguntas de elección múltiple puede contener tres, cuatro, cinco alternativas de respuesta o más y es importante fijar el número óptimo de alternativas ya que ello afecta a la fiabilidad del examen.

<p>Traducido a pesetas, ¿cómo es un euro en comparación con una libra esterlina?</p> <p>a. Mayor b. Menor c. Igual</p>	<p>¿Cuál es la solución de la ecuación $8 = 2X + 4$?</p> <p>a. 1 b. 2 c. 3 d. 4 e. 5 f. 6</p>
--	--



Aunque las investigaciones más recientes sugieren que lo mejor es utilizar tres alternativas de respuesta, también la estructura de la pregunta juega un papel importante a la hora de decidir el número óptimo de alternativas, ya que algunos tipos de preguntas dan pie a un número mayor de alternativas todas ellas igualmente creíbles.

5.4.- Preguntas de elección binaria

Son fáciles de construir en comparación con las de alternativas múltiples, ya que sólo requieren redactar una frase o enunciado y no hay que preocuparse por la complejidad que introduce la presentación de alternativas. Por el contrario, los críticos de este formato de preguntas afirman que tienden a favorecer una memorización rutinaria de hechos aislados, trivializando la importancia de comprender estos hechos y las relaciones entre ellos. Para que esto no suceda es necesario redactar buenas preguntas, que utilicen las palabras adecuadas y precisas y no trivialicen el conocimiento.

Hay que tener en cuenta además que dado que las preguntas de elección binaria son más cortas que las de elección múltiple, se requiere por lo general menos tiempo para responderlas. Por consiguiente, para un mismo tiempo de aplicación es necesario añadir al examen más preguntas de elección binaria o completar con preguntas de otro formato.

- Evitar expresiones como: todos, ninguno, siempre, nunca..., a no ser que sea clave para la pregunta en cuestión, ya que la introducción de estas expresiones conducen a preguntas que serán probablemente falsas. Igualmente, evitar palabras tales como: generalmente, debiera, algunos..., ya que se asocian con argumentos verdaderos.
- Evitar que la respuesta correcta siga una pauta determinada (V-V-F-V-V-F-...) y procurar que el número de respuestas verdaderas sea aproximadamente igual al número de respuestas falsas.
- Evitar que una pregunta pueda ser parcialmente verdadera o falsa dependiendo de algún factor no expresado en el enunciado. Las preguntas deben ser inequívocamente verdaderas o falsas.

Alfa se utiliza cuando los ítems del test no son dicotómicos y las ecuaciones de Kuder-Richardson cuando sí lo son.V F

En el ejemplo, es cierto que se utiliza KR20 cuando los ítems son dicotómicos pero simplemente porque su cálculo es más corto; perfectamente se podría utilizar Alfa y el resultado sería exactamente el mismo, ya que KR20 es un caso particular del coeficiente alfa. Aquí el alumno aventajado no sabrá si debe utilizar esa información para responder la pregunta.

- Formular cada pregunta de manera que contenga una única idea principal, para evitar que se afirme o se niegue en función de algún aspecto no fundamental.

El estudio de las características psicométricas de un test conlleva el análisis de su fiabilidad y validez, así como el alcance de sus predicciones.....V F

5.5.- Preguntas de respuesta corta o de completar

- Asegurarse de que la pregunta requiera una respuesta corta.
- Cerciorarse de que sólo una respuesta es aceptable o, en su defecto, elaborar un listado de las respuestas que se darán por válidas en su corrección (no es conveniente que haya demasiadas respuestas que puedan ser correctas).
- Asegurarse de que las palabras a completar en este tipo de preguntas sean conceptos claves.

El método general con el que la Psicología trató de abordar en _____ su objeto propio de estudio era el método de la ciencia natural.

En este ejemplo, habría que rellenar el hueco con "sus inicios", lo que en esta pregunta puede no ser tan clave como el preguntar al



alumno por el método utilizado. Para solucionarlo, se utilizaría la opción 1. Si por cualquier motivo no se desea variar el objeto de la pregunta, habría que transformar ésta a interrogativa, para que “el cuándo” sea lo principal de la pregunta y no haya lugar a equívocos. (Ver Opción 2).

Opción 1: El método general con el que la Psicología trató de abordar en sus inicios su objeto propio de estudio era el método de la ciencia_____.

Opción 2: ¿En qué momento de la historia de la Psicología se trató de abordar su objeto propio de estudio mediante el método natural? _____.

- Situar, preferentemente, los espacios a completar al final de la frase. Esto hará que el alumno no necesite terminar de leer el enunciado para entender de qué trata la pregunta. Véase en el ejemplo cómo se puede cambiar fácilmente el lugar inadecuado del espacio a completar de la formulación de la izquierda:

El _____ es la tendencia sistemática de un sujeto a responder a los ítems de un test de una determinada manera.

La tendencia sistemática de un sujeto a responder a los ítems de un test de una determinada manera se denomina _____.

5.6.- Preguntas de respuesta extensa o de ensayo

- Definir claramente la tarea que debe realizar el alumno evitando ambigüedades.
- Facilitar al alumno el tiempo suficiente para realizar el examen en función de la complejidad de la o las preguntas.
- Evitar las preguntas opcionales. Si se opta por incluirlas procurar que sean iguales en dificultad y extensión.
- Dejar claro a los alumnos los criterios de evaluación de acuerdo al contenido y a los niveles de complejidad, informándoles además si se considerará la ortografía, caligrafía y gramática en su puntuación.



6.- ASIGNACIÓN DE PUNTUACIONES.

Una vez diseñada la prueba y administrada a los alumnos existen distintos procedimientos para asignar una puntuación numérica a la ejecución del alumno en el examen. Centrándonos en las pruebas de elección múltiple, para obtener la puntuación directa, o nota, se ha tomado como criterio asignar un 1 a los aciertos y un 0 a los fallos y omisiones.

$$X = \sum_{i=1}^n w_i x_i$$

Donde:

X = puntuación directa, o nota en el examen.

w_i = peso asignado a la pregunta i dentro de una prueba constituida por n cuestiones.

x_i = puntuación de la respuesta del alumno a la pregunta, (1 ó 0) .

Habitualmente cada pregunta recibe la misma ponderación, es decir w igual a 1. Sin embargo, en función de la importancia de la pregunta se puede asignar un peso diferente a las preguntas del examen.

Aplicando esta expresión, si un sujeto ha contestado correctamente a 8 preguntas de las 10 que componen el examen su nota será de 8 puntos si todas tienen el mismo peso en la nota final, ya que no puntúan las preguntas en blanco o falladas.

El problema que plantea este tipo de asignación de puntuaciones es que no contempla la variabilidad que introduce las diferencias individuales ante preguntas cuya respuesta correcta se desconoce. En esta situación existen sujetos que optan por responder al azar y otros que prefieren no arriesgarse dejando la pregunta en blanco.

Para soslayar estos problemas se han propuesto distintas alternativas de las que presentaremos la más utilizada cuyo objetivo es eliminar de la nota del examen las preguntas que se hayan acertado por mero azar:

$$X_c = A - \frac{E}{(K - 1)}$$

Donde:

X_c = puntuación directa corregida, o nota en el examen tras contrarrestar el efecto del azar.

A = número de aciertos.

E = número de errores.

K = número de alternativas de los ítems.

Si una pregunta tiene 3 alternativas de respuesta, donde sólo una es correcta, la probabilidad de acertar es 1/3, mientras que la probabilidad de fallar es de 2/3. Se demuestra que los aciertos al azar se obtienen dividiendo el número de errores entre el número de alternativas menos uno.

Para un alumno concreto que haya acertado 6 ítems y fallado 3 en un examen de 10 preguntas con 3 alternativas de respuesta, su nota se calcularía -de acuerdo con la fórmula- de la siguiente manera:

$$X_c = 6 - \frac{3}{(3 - 1)} = 4.5$$

El alumno obtendría una puntuación de 4.5. Si el punto de corte entre el aprobado y el suspenso se coloca en 5 estaría aprobado si no hubiésemos contrarrestado el efecto del azar.

La aplicación de esta fórmula se basa en el cumplimiento de los supuestos teóricos que las avalan. Sin embargo hay ciertos aspectos como el conocimiento parcial de las preguntas que permite a los alumnos eliminar opciones como alternativas de respuesta, o la capacidad para asumir riesgos que suponen una crítica a las fórmulas de puntuaciones corregidas.

Una vez obtenida la puntuación directa, deberíamos transformarla en una escala de 0 a 10 con objeto de facilitar la comprensión de la nota



numérica obtenida por el alumno. Una manera fácil de hacerlo consiste en aplicar una sencilla regla de tres. Así por ejemplo, si la puntuación obtenida en el examen (ya sea corregida o directa) es de 15 sobre 20 tendríamos que:

$$\begin{array}{l} \text{Si} \quad 20 \text{ puntos equivalen a } 10 \\ \text{entonces} \quad 15 \text{ puntos equivalen a } X \end{array} \left. \vphantom{\begin{array}{l} \text{Si} \\ \text{entonces} \end{array}} \right\} X = \frac{15 \times 10}{20} = 7,5$$

Por tanto un alumno que haya obtenido 15 puntos sobre 20 obtendría un 7.5 en una escala de 0 a 10.



7.- INTERPRETACIÓN DE LOS RESULTADOS

Una vez que hemos corregido las pruebas de evaluación hemos de dar un significado a los resultados obtenidos por los alumnos. En la actualidad hay dos tendencias diferentes a la hora de interpretarlos: por normas y por criterios.

7.1.- Interpretación por normas

Muchos de vosotros recordareis aquellos sistemas que se utilizaban en los colegios, y que todavía se siguen utilizando, que ordenaban a los alumnos de manera que se sabía quien era el primero de la clase en matemáticas, el segundo en lengua, etc., y sólo los primeros estaban en el “cuadro de honor”; pues bien, el tipo de evaluación que se hacía en estos casos era una evaluación por normas.

¿Qué es lo que caracteriza a esta forma de evaluación e interpretación de las puntuaciones?. Pues que para poder llevarla a cabo es necesario conocer cuál ha sido el rendimiento de todos ellos, ya que la interpretación y la consiguiente calificación se hace comparando el rendimiento de cada uno de ellos con el de los demás.

Esta comparación puede dar lugar a varios tipos de interpretaciones; unas veces la interpretación de los resultados se hace en términos ordinales, tal y como hemos expuesto anteriormente, de manera que los alumnos queden ordenados desde el que ha tenido un mejor rendimiento al que lo ha tenido peor; otras veces es posible comparar el rendimiento o ejecución de un alumno en particular no con el rendimiento de cada uno de los miembros del grupo, sino con el rendimiento medio de los mismos.

De cualquier forma, cuando los resultados se interpretan por normas no se obtiene información de lo que el alumno puede o no puede hacer, sabe o no sabe, sino de lo que puede hacer o saber más o menos que los demás; es decir, en relación a su grupo. Todo esto implica que cuando lo que se está evaluando es, por ejemplo, el conocimiento de los

contenidos de una asignatura este tipo de interpretación no ofrece información acerca de si un alumno determinado conoce o no esos contenidos, sino si sabe más o menos que sus compañeros del grupo. ¿Qué problema tiene esta forma de interpretar los resultados?, que dado que los resultados se interpretan en relación a un grupo de referencia, si un alumno que ha obtenido la puntuación máxima en relación a un determinado grupo se cambia de grupo, no hay garantías de que siga obteniendo esa misma nota o puntuación; puede suceder, por ejemplo, que si el grupo es de muy bajo rendimiento, un alumno apruebe un examen que hubiera suspendido si el grupo hubiera tenido un rendimiento más alto.

7.2.- Interpretación por criterios

A diferencia de la interpretación anterior, en este caso, los resultados de cada uno de los alumnos se comparan con un criterio absoluto constituido por lo que, a juicio del profesor, cada alumno debe saber o conocer con independencia de lo que sepa el resto de sus compañeros.

A ese criterio absoluto se le denomina “dominio” y, a la hora de evaluar a los alumnos, los resultados que han obtenido se comparan con el dominio para constatar si su rendimiento ha sido superior o inferior al dominio con independencia de si es superior o inferior al rendimiento de los demás alumnos.

Una cuestión básica en la evaluación por criterios es que el profesor tiene que tener muy claros cuáles son los objetivos docentes y hacer una descripción precisa del dominio. Tiene que tener muy claros los conocimientos mínimos que va a exigir a sus alumnos para considerar que han superado ese dominio y, por lo tanto, van a obtener una evaluación positiva de su rendimiento.



7.3.- Evaluación por normas o por criterios

La elección de una forma u otra de interpretar los resultados obtenidos por los alumnos en las pruebas de evaluación dependerá de la información que se quiera obtener y de las decisiones que se vayan a tomar a partir de esa información. Teniendo en cuenta que estos aspectos hay que tenerlos muy claros desde la primera fase del proceso de elaboración de la prueba, es en esa fase cuando se debe de tomar una decisión acerca de la forma que se va a seguir a la hora de interpretar los resultados.

Cuando se quiere llevar a cabo una evaluación selectiva, por ejemplo en unas oposiciones, es conveniente hacer una interpretación normativa de los resultados ya que habrá que seleccionar a los mejores del grupo. Por el contrario, cuando se trata de evaluar los resultados que han obtenido los alumnos en la escuela o en la universidad, teniendo en cuenta que nuestra función es formativa no selectiva, es más conveniente hacer una interpretación criterial de los resultados, de esta manera podremos obtener información acerca de si nuestros alumnos han alcanzado los objetivos y dominios marcados.

Tanto si se opta por una interpretación por normas como si es por criterios, es necesario asegurar que el examen tiene una adecuada validez de contenido. Ahora bien, para conseguir esto hay que proceder de diferente forma. Si se realiza una interpretación por normas, entonces habrá que elaborar exámenes largos, dado que lo que hay que hacer es maximizar las diferencias entre *todos* los alumnos. Por el contrario, si se va a llevar a cabo una evaluación por criterios lo primero es definir claramente el dominio y, una vez definido, redactar preguntas que permitan discriminar al máximo en la *zona del dominio*, por lo que suelen resultar pruebas más cortas que en el caso anterior.

En este sentido, teniendo en cuenta que la evaluación criterial se lleva a cabo para conocer el nivel de rendimiento de los alumnos respecto a un dominio bien definido, las pruebas objetivas o de respuesta corta son más eficaces pues permiten cubrir un amplio espectro de contenidos. Es

difícil que una prueba de ensayo pueda tener validez de contenido ya que no se pueden poner muchas preguntas cuando hay tiempo limitado.

7.4.- Dominio de evaluación

En realidad el dominio no es más que un conjunto de criterios, aspectos o dimensiones que el profesor considera que son relevantes en el aprendizaje de sus alumnos; de manera que si éstos no alcanzan *el nivel mínimo de conocimientos* que implican esos criterios no podrán obtener una evaluación positiva. En definitiva, no llegarán al aprobado. Esos criterios fijados por el profesor deberán estar recogidos en la guía de la asignatura para conocimiento de todos los alumnos.

Para ser operativos, hay que marcar un punto de corte que diferencie entre los alumnos que han alcanzado el nivel mínimo de conocimientos y aprobarían el examen y los que no lo han alcanzado y, por lo tanto, suspenderían. Cada equipo docente deberá reflexionar acerca de donde situar ese punto de corte y la forma de evaluar cada uno de los criterios que forman el dominio.

Por otra parte, si se utilizan distintos procedimientos de evaluación (teoría y práctica), es necesario también determinar la forma en que se van a combinar las puntuaciones de los alumnos en ambas partes para proporcionar una puntuación única de cada uno de ellos.

Hay varios modelos que pueden ayudar, entre ellos merecen destacar:

- a. *Compensatorio*. Bajo este modelo la calificación final de los alumnos sería función de la combinación aditiva y ponderada de la puntuación obtenida en una serie de criterios. Por ejemplo, si un examen tiene una parte teórica y otra práctica, la puntuación final sería la suma de la puntuación obtenida en cada una de las partes. El problema que plantea este modelo es que si el punto de corte está en el 5, un sujeto puede tener un 5 en la parte teórica y un 0 en la práctica y aprobar. Si la ponderación consiste en hallar la



media entre las dos partes, un sujeto que obtuviera un 8 en la parte práctica y un 2 en la parte teórica también aprobaría.

- b. *Conjuntivo*. Este modelo implica que se ha de alcanzar el punto crítico exigido en cada uno de los criterios o indicadores. Siguiendo con el ejemplo del examen con una parte teórica y otra práctica, para poder aprobar el alumno deberá superar el 5 en ambas partes.
- c. *Conjuntivo – compensatorio*. Se aplicaría en primer lugar el modelo conjuntivo y a los sujetos que aprobaron el examen cuando se aplica ese modelo se les aplica a continuación el modelo compensatorio. Supongamos que un sujeto ha obtenido un 7 en la parte teórica del examen y un 5 en la parte práctica, al aplicar el modelo conjuntivo se le considera que está aprobado puesto que ha alcanzado en ambas partes el punto crítico establecido. Ahora, para obtener su puntuación final, se le aplicaría el modelo compensatorio hallando la media entre las dos partes del examen: $(7+5)/2 = 6$.
- d. *Disyuntivo*. Bajo este modelo se consideraría que un alumno estaría aprobado cuando hubiera superado el punto de corte en alguno de los criterios o indicadores.

Hemos puesto un ejemplo en el que se trataba de evaluar un examen formado por una parte teórica y otra práctica, y donde el punto de corte se había establecido en el 5. Esto es un mero ejemplo y no tiene porqué ser así. El punto de corte lo establece el profesor o el equipo docente en función de lo que ellos consideren que deben conocer sus alumnos para aprobar la asignatura.

Por otra parte, se puede considerar que los alumnos deben tener un conocimiento mínimo no de la asignatura en términos globales, sino de cada uno de los temas o bloques temáticos; en este caso se puede evaluar por separado cada uno de los bloques y aplicar a continuación el modelo seleccionado. Si un examen tiene 4 problemas, se pueden evaluar por separado cada uno de ellos y, a la hora de calificar al alumno, aplicar el modelo conjuntivo estableciendo unos mínimos para cada problema, de

manera que no pueda aprobar un alumno que haya dejado un problema en blanco.

La casuística es muy variada y la decisión de elegir un modelo u otro deberá ser tomada por el equipo docente, sin embargo, a simple vista, se comprende que alguno de ellos no es muy recomendable en nuestro contexto. Sea cual sea el modelo elegido, deberá quedar explicitado en la guía del curso para conocimiento de los alumnos.



8.- BIBLIOGRAFÍA:

Barbero, I. (2003). *Psicometría* (Capítulo 3). Madrid: UNED

Case, S.M. y Swanson, D.B. (2001). *Constructing Written Test Questions for the Basic and Clinical Science*. (3ª ed.). Philadelphia: National Board of Medical Examiners. www.nbme.org/PDF/2001iwg.pdf

Ebel, R.L. (1979) *Essential of Educational Measurement*. Englewood Cliffs, NJ: Prentice Hall.

Jorner, J.M.; González Such, J. y Pérez Carbonell, M.A. (1995). Líneas Emergentes de Investigación a partir de las Reformas Educativas en España en el Ámbito de la "Medición y Evaluación Educativas". 1ª Conferencia electrónica AIDIPE. <http://www.uv.es/aidipe/Ponenci2.html>

Haladyna, T.M. (1999). *Developing and validating multiple-choice test items* (2ª ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Haladyna, T.M.; Downing, S.M. y Rodríguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309-334.

Muñiz, J. (2001). *Teoría Clásica de los Tests*. Madrid: Pirámide.

Navas, M.J. (Ed). (2001). *Métodos, diseños y técnicas de investigación en Psicología*. (Capítulos 3, 4 y 8). Madrid: UNED.

Osterling, S.J. (1998). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats*. Massachusetts: Kluwer Academic Publishers.

Bibliografía

Suárez, J.C. (2001). Análisis de ítems. En Navas, M.J. (Ed). (2001). *Métodos, diseños y técnicas de investigación en Psicología*. Madrid: UNED.

Thorndike, R.M; Cunningham, G.K; Thorndike, R.L, y Hagan, E. (1991), *Measurement and evaluation in psycholgy and education*. (5ª ed.). Nueva York: Macmillan.