

LA EVALUACION DEL ALUMNO EN LA UNIVERSIDAD: EL PROYECTO CERT¹

Juan De PABLOS PONS²

Rafael GARCÍA PÉREZ

M^a Angeles REBOLLO CATALÁN

M. Ignacio ALGARÍN FIESTAS

Angeles MARTÍNEZ LÓPEZ

Laurentino SALVADOR BLANCO

M^a Paz PRENDES ESPINOSA

Este artículo presenta un instrumento para la evaluación del aprendizaje basada en pruebas objetivas de elección múltiple. Se contempla la utilización de algunos constructos como «las Soluciones Generales Implícitas» y los «Niveles de Confianza». Para aplicar este instrumento se ha diseñado un soporte informático que facilita el diseño y corrección de las pruebas. Este programa ha sido aplicado por un grupo de profesores de la Universidad de Sevilla y el resultado de esta experiencia se muestra también en el trabajo. Este proyecto ha sido financiado por los programas EUROTECNET y EUROFORM de la Comunidad Europea.

This paper presents an instrument for learning evaluation based on multiple-choice test. The use of some structures as «Implicit General Answers» and «Degrees of Confidence» characterizes this proposal. To apply this instrument a software has been designed; this computer component facilitates the design and correction of the tests. This material has been used by professors in the University of Seville and this work also presents some results of the experience. The programs EUROTECNET and EUROFORM from the European Community had contributed to the financing of this project.

LA ENSEÑANZA UNIVERSITARIA

La evaluación recibe actualmente la consideración de ser una parte fundamental en el proceso de enseñanza. Des-

de el punto de vista del desarrollo académico de los estudiantes resultan fundamentales los juicios y valoraciones que se hacen sobre sus actividades académicas, pero además otros aspectos relacio-

¹ Proyecto Cooperativo sobre Evaluación de Resultados de Aprendizaje.

² Los autores del artículo forman parte del Proyecto CERT, y pertenecen a las universidades de Sevilla, Barcelona, Cantabria y Murcia.

nados con la educación también se ven influidos por fenómenos evaluativos más o menos explicitados. La exigencia social sobre la necesidad de mejorar la calidad de la enseñanza ha ido ampliando los objetivos y ámbitos de la evaluación, de manera que además de los alumnos y profesores, sucesivamente también van siendo evaluados proyectos y programas, materiales de enseñanza, o instituciones educativas.

Sin embargo, las aproximaciones a la problemática universitaria desde perspectivas que pretendan una mejora de la calidad de la enseñanza, son relativamente recientes. Es sólo a mediados de los años setenta cuando se concretan iniciativas significativas, provenientes del ámbito anglosajón (Altbach, 1985). Este relativamente corto desarrollo de la investigación sobre pedagogía universitaria redundaba en una insuficiente fundamentación teórica que permita una praxis justificable a partir de un corpus científico (Cannon, 1983). La *enseñanza universitaria* como un campo de estudio específico de las Ciencias de la Educación, es actualmente una línea de trabajo en expansión.

La institución universitaria representa una realidad compleja en la que desde el punto de vista evaluativo nos enfrentamos a aspectos muy variados y que en muchos casos entrecruzan sus campos de influencia. El alumno universitario adquiere informaciones, conocimientos, destrezas, capacidades, todo lo cual le exige poner en juego procesos cognitivos de distinto orden. Pero también aprende hábitos, actitudes, formas de actuar tomando como referencia a sus profesores, y a sus compañeros, todo lo cual también supone una aportación cara a un apren-

dizaje, que en último término debe llegar a dar forma a una ética profesional. Facta ésta imprescindible si estamos pretendiendo la formación de profesionales. En este sentido, una primera cuestión de interés que puede formularse consiste en preguntarnos hasta dónde llega la responsabilidad evaluadora del docente universitario.

La enseñanza universitaria a semejanza de otros niveles del sistema educativo genera unas necesidades evaluativas de tipo académico, pero su responsabilidad institucional en la formación de profesionales de alto nivel, añade una característica específica a la trascendencia del hecho evaluativo. En este sentido, cabe afirmar que la experiencia proveniente de otros niveles educativos en relación a la evaluación, en muchos casos no son trasvasables a la enseñanza universitaria.

En cualquier caso, toda propuesta de trabajo formativo formulada por un profesor y dirigida a sus alumnos se apoyará implícita o explícitamente en un determinado modelo de enseñanza. Cualquier modelo siempre implica determinadas opciones metodológicas, fórmulas concretas en la secuenciación de los contenidos objeto de estudio, tipos de tareas, actividades, etc. Lógicamente, determinadas opciones de enseñanza suponen también opciones concretas desde el punto de vista evaluativo.

Una aproximación básica a este planteamiento nos lleva, si seguimos la propuesta de Santos Guerra (1991), a dos concepciones divergentes: a) un modelo técnico y b) un modelo práctico. Ambos reflejan respectivamente un *enfoque instrumental*, en el primer caso, preocupado de forma prioritaria por la calidad de los resultados, y un *enfoque ético*

(Pérez Gómez, 1991) centrado en la calidad de los procesos educativos.

Bien es cierto, que estas opciones pedagógicas deben aplicarse en un contexto -el universitario-, caracterizado por una serie de rasgos institucionales (socio-económicos, políticos, ideológicos, organizativos, etc.), que se concretan en realidades como: la masificación estudiantil en muchas de las carreras, gestión burocratizada de los recursos, renovación de la oferta formativa sobre la base de criterios tomados de la economía de mercado, desdoble del profesorado en tareas docentes e investigadoras, etc.

Por lo tanto, la combinación de concepciones pedagógicas con la necesaria búsqueda de soluciones prácticas para las realidades mencionadas, exige una elaboración de fórmulas donde el profesor universitario encuentre apoyos razonables.

LA EVALUACION DEL ALUMNO UNIVERSITARIO

En la universidad resulta habitual que el profesor tenga una gran autonomía a la hora de planificar su trabajo docente. En este sentido, cada vez parece más evidente que el docente universitario debe esforzarse en hacer llegar a sus alumnos de manera explícita su plan de enseñanza, lo que normalmente toma forma mediante el programa de la asignatura, en el que figuran componentes pedagógicos tales como objetivos educativos, metodologías, y opciones evaluativas, además de la propuesta de contenidos que formulada de manera organizada constituye el temario.

La necesaria diversidad de actividades a realizar por los alumnos en una

enseñanza compleja, como la universitaria, parece exigir en coherencia, una diversidad de aspectos objeto de evaluación.

Desde una perspectiva pedagógica la metodología de la evaluación ha experimentado un notable desarrollo en años recientes (De Miguel, 1992). La progresiva sustitución de una concepción evaluativa instrumental, por una visión de la evaluación concebida como una estrategia científica (investigación evaluativa) que permite al profesor obtener informaciones útiles para poder tomar decisiones docentes. Todo proceso evaluativo, independientemente de su vertiente metodológica, debe ser capaz de generar credibilidad, tanto para evaluados como para evaluadores. Y a su vez, esa credibilidad puede tener un soporte técnico y también ético.

DIMENSIONES A EVALUAR

Desde una concepción procesual de la enseñanza, resulta clave plantear qué tipo de información queremos obtener, puesto que en función de esa decisión habrá que decidir, en coherencia, el tipo de evaluación a aplicar.

Es evidente que, si bien, no existe todavía una amplia tradición de estudios sobre enseñanza universitaria, sí que goza de una larga trayectoria el concepto de evaluación, fundamentalmente en su acepción de «medición», en el campo científico en general. Y con antecedentes más específicos en la docimología o ciencia de los exámenes (dokimé, prueba y logos, ciencia), con referencias documentadas desde 1922. Desde entonces, la evolución del concepto ha sido constante. De hecho, hoy la evaluación ya no

cumple las mismas funciones que hace veinte años. Es una de las variables que determinan el estilo de enseñanza de un profesor como parte de su plan de trabajo docente.

Desde una perspectiva contemporánea, **evaluar** significa «examinar el grado de adecuación entre un conjunto de informaciones y un conjunto de criterios adecuados al objetivo fijado, con el fin de tomar una decisión» (De Ketele, 1984, 17).

Los procedimientos técnicos fiables que arbitremos para obtener el tipo de

información buscado constituirán la base de las estrategias de evaluación educativa a utilizar.

LAS ESTRATEGIAS DE LA EVALUACIÓN

Siguiendo a De Ketele (1984, 3) podemos establecer una referencia inicial diferenciando dos parámetros básicos para caracterizar las principales estrategias de evaluación. Por una parte, estrategias tipo *longitudinal* (se utilizan a lo largo del proceso de enseñanza), y por otra

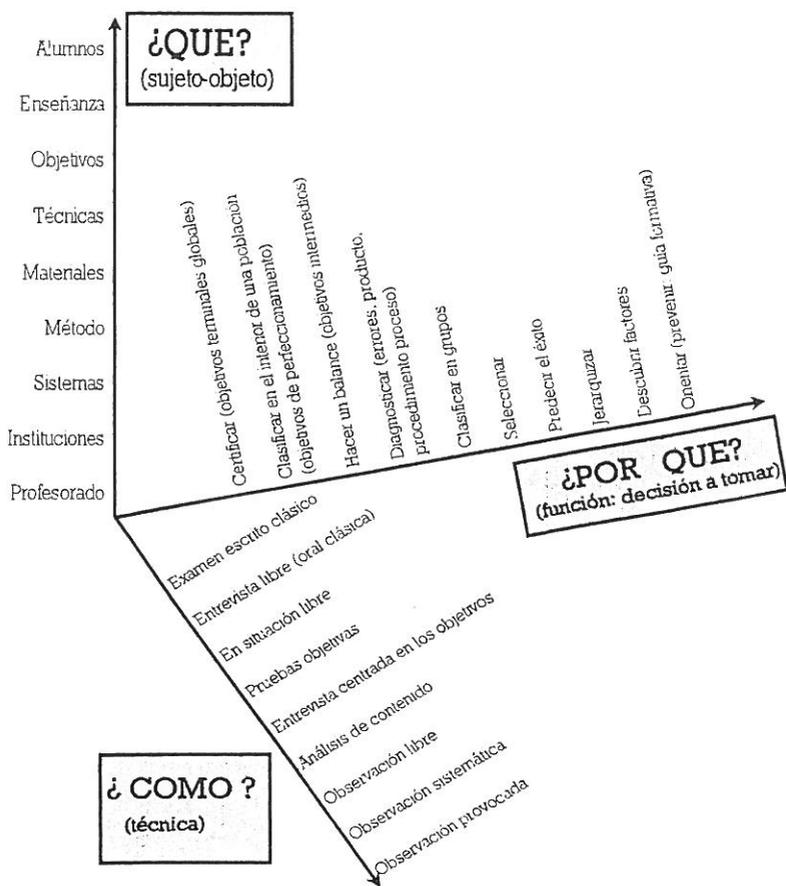


Figura 1

estrategias de carácter *puntual* (empleadas en momentos específicos de la enseñanza).

Pero a la hora de decidir una estrategia no basta con establecer el *cuando*, tendremos que tratar de despejar otros interrogantes como el *qué*, el *cómo* o el *por qué* evaluamos. Teniendo en cuenta además que los factores que están respaldando estas respuestas actúan de manera interdependiente. Una síntesis de todos estos aspectos queda plasmada en la figura siguiente elaborada a partir de una propuesta de De Ketele (1984, 39).

Uno de los posibles instrumentos a utilizar que están presentes en la propuesta anterior son las pruebas objetivas. Este trabajo se centra en dicha opción a partir del desarrollo y aplicación del Proyecto CERT.

PROYECTO COOPERATIVO SOBRE LA EVALUACIÓN DE LOS RESULTADOS DEL APRENDIZAJE (C.E.R.T.)

El origen de este proyecto hay que buscarlo en la Universidad belga de Lieja, (Leclercq, 1989), en la cual se diseñó y desarrolló un programa informático (CERT) que apoyara y facilitara el uso de pruebas objetivas en la evaluación educativa, sobre la base de una serie de concepciones que sustancialmente este artículo presenta. Tras una investigación inicial financiada por el programa europeo EUROTECNET, la Universidad de Barcelona se vinculó al proyecto y decidió iniciar un nuevo estudio con la cooperación de otras universidades del Estado español y así nació la presente investigación.

Esta segunda fase está siendo llevada a cabo con la colaboración de las univer-

sidades de Cantabria, Murcia, Oviedo y Sevilla, coordinadas por la universidad de Barcelona, bajo los auspicios y la financiación del programa EUROFORM de la Comunidad Europea, programa cuyo objetivo principal es promover la innovación en el campo de la formación profesional de base y continua con el fin de valorar los cambios tecnológicos actuales y futuros así como su impacto sobre el empleo y las competencias para el desempeño del mismo.

El programa CERT posibilita gestionar gran cantidad de información recogida a través de pruebas objetivas de opción múltiple y analizarla, agilizando el proceso de manipulación y cálculos manuales de los formadores para evaluar aprendizajes. Este software ha sido apoyado por una serie de materiales formativos. Un «Manual del software CERT», un documento sobre «Introducción a las pruebas objetivas», una «Guía Rápida» de consulta sobre conceptos y software, dos videos, así como materiales complementarios.

LAS PRUEBAS OBJETIVAS

El uso de pruebas objetivas para el desarrollo de determinadas funciones evaluativas es hoy día práctica habitual para un número creciente de profesores universitarios, fundamentalmente ante determinadas situaciones como la masificación del alumnado o la realización de exámenes sobre contenidos básicos. Como instrumento es evidente que se trata de una opción que tiene debidamente perfilado su papel evaluador. Son bastantes los autores con trabajos sobre el uso de pruebas objetivas con aplicaciones educativas (Rodríguez Diéguez, 1980; Sampascual, 1982; De Ketele, 1984; entre

otros). En general, puede decirse que están claramente desarrollados los siguientes aspectos: 1) Posibilidades evaluativas en relación a otros instrumentos; 2) Características generales de este tipo de prueba; 3) Ventajas y limitaciones de su uso. En este sentido, Cabrera y Espín (1986) realizan un exhaustivo análisis de las características, forma de construcción y modos de uso de las pruebas objetivas. También en esta línea, Colás y Rebollo (1993) abordan su utilidad y adecuación en diferentes momentos de la evaluación.

En este trabajo, trataremos de aportar una visión sintética de las características fundamentales de estos instrumentos, analizando su utilidad para la evaluación de los aprendizajes. Asimismo, examinaremos esta cuestión en torno a la posibilidad de proponer el manejo, por parte del profesor universitario, de un recurso didáctico más. Y de manera específica de aplicación al ámbito universitario.

La evaluación entendida como un proceso sistemático de búsqueda de información al servicio de una toma de decisiones requiere de un juicio basado en la información recogida. En este proceso de búsqueda de información son factibles el uso de diferentes técnicas. Cabrera y Espín (1986) las sintetizan en cuatro tipos fundamentales:

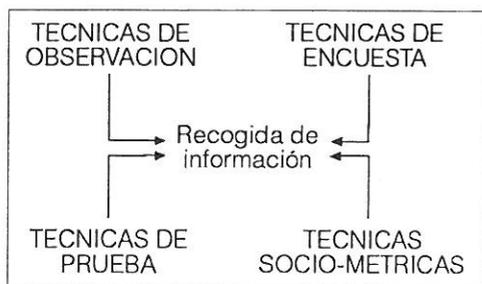


Figura 2

En este sentido, podemos situar las pruebas objetivas entre las «Técnicas de prueba». También se clasifican como técnicas de prueba: las pruebas escritas (respuesta breve, las pruebas objetivas sencillas (verdadero/falso), las pruebas orales (formales, informales, ensayos) los trabajos extensos, etc. (CERVA, EUROFORM: Introducción a las pruebas objetivas, 1993, p.19).

La característica más destacable de las pruebas objetivas es el afán por objetividad en la evaluación; «sobre todo en lo que respecta a la corrección y cómputo de resultados de las pruebas» (SAMPASCUAL, 1982, P.8).

Una prueba objetiva habitualmente está constituida por un amplio número de preguntas que requieren de una respuesta muy breve o incluso la selección entre un conjunto de respuestas preestablecidas. Si estas preguntas están bien formuladas las respuestas del alumno deben reflejar aquellos procesos mentales que esperamos que él haya puesto en juego abarcándose un dominio de conocimientos concreto que es el que se está evaluando.

Según el grado de estructuración que se llegue se pueden obtener diferentes tipos de pruebas que conforman un continuum de posibilidades en cuyos extremos hallamos las pruebas de respuesta libre o pregunta abierta, en uno, en el otro extremo las pruebas objetivas. Por lo tanto, éstas son instrumentos de evaluación muy estructurados, destinados a dimensionar el grado de conocimiento, comprensión, interpretación, etc. que tiene el alumno sobre unos contenidos concretos.

Dadas las características de este tipo de pruebas se suelen usar en la evaluación

diagnóstica y con referencia al producto (rendimiento o grado de consecución de objetivos). (Colás y Rebollo, 1993, 125).

En definitiva, para ayudar a comprender claramente el tipo de instrumento evaluativo que estamos describiendo puede decirse que:

...las pruebas objetivas son un conjunto vertebrado de preguntas o ítems breves, claros y precisos que demandan del alumno una respuesta también breve (puede ser una palabra, un número, marcar entre varias posibilidades de respuesta la correcta o la incorrecta, ejecutar una orden, etc).

(CERT-EUROFORM: Introducción ..., 1993, 22).

LAS PRUEBAS OBJETIVAS DE OPCIÓN MÚLTIPLE

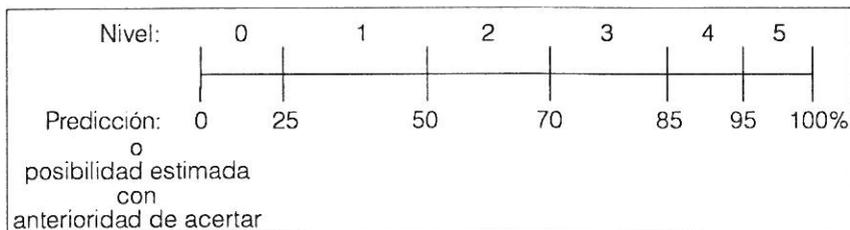
La característica fundamental de esta modalidad de prueba es que incorpora diversas posibilidades de respuesta entre las cuales el sujeto evaluado debe elegir la más correcta. Con el objeto de realizar una exposición más exhaustiva de este particular incluimos posteriormente (tras el apartado dedicado a la construcción de pruebas) un apartado específico dedicado a las Cuestiones de Opción Múltiple como unidades básicas de este tipo de pruebas.

En las pruebas objetivas de opción múltiple pueden incluirse dos aspectos que añaden posibilidades a este tipo de evaluación.

En primer lugar, la inclusión de respuestas con carácter implícito (no aparecen escritas en el cuestionario, aunque el alumno sabe que debe tenerlas en cuenta). Las cuatro modalidades más frecuentes de respuestas implícitas son las siguientes: NINGUNA de las soluciones propuestas es correcta, TODAS son correctas, FALTAN datos en el enunciado y, por último, el enunciado es INCOHERENTE). A estas opciones las llamamos Soluciones Generales Implícitas (S.G.I.). El aspecto más relevante es que amplían el abanico de respuestas que el alumno tiene que manejar.

En segundo lugar, también pueden usarse lo que denominamos «NIVELES DE CONFIANZA» que son una estimación que el alumno realiza (en base a una escala de 0 a 5) del grado de confianza o seguridad con que responde a cada una de las cuestiones del test. Para la presentación, de la escala de estimación de la confianza el proyecto CERT recomienda utilizar la repartición no simétrica de las zonas de probabilidad.

Esta repartición asimétrica permite a los alumnos presentar las respuestas más matizadas en el extremo superior de la escala, siendo ésta la zona de probabilidades más elevadas. (Alonso, 1992, 57).



LA CONSTRUCCION DE LAS PRUEBAS

Para diseñar correctamente una prueba objetiva deben tenerse en cuenta los siguientes aspectos básicos: 1) El contenido y la estructura de la prueba; 2) Los items que la constituyen y; 3) Presentación de la prueba.

Respecto al primer apartado, contenido y estructura de la prueba, es preciso delimitar claramente el área de conocimientos (materias) y el tipo de procesos cognitivos (memoria, comprensión, aplicación, análisis, ...) que se van a contemplar en la evaluación, formulándose objetivos que se especificarán y operativizarán. Este paso delimita el alcance de la evaluación, condicionando la estructura, los contenidos, la redacción de los items y el formato de presentación de la prueba.

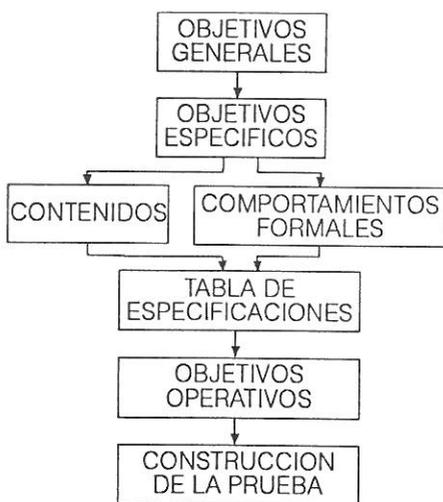


Figura 3: Diagrama sobre la formulación de objetivos que dan origen a la prueba, según Rodríguez Diéguez, (1980) y (BISQUERRA, 1985, 359).

Es necesario delimitar de forma correcta las materias y los procesos mentales que se van a evaluar dando, además, un peso específico a cada apartado en el conjunto de la prueba. En el marco del CERT hay diversas posibilidades para realizar esto. Por un lado, existe el recurso clásico de elaborar mayor número de items para los contenidos de mayor importancia o peso específico o, incluir mayor puntuación a los items de preguntas sobre dichas materias. Por otro lado, como aportación más novedosa, el CERT introduce la posibilidad de estimar en porcentajes el peso de cada una de las materias y de los procesos mentales en el conjunto de la prueba.

Este proceso de distribución de procesos por contenidos y habilidad cognitivas ha de estar justificado y argumentado en función de teorías de propia materia o disciplina.

Una tabla de especificación de objetivos por temas y procesos mentales y determinación del número de preguntas, como la que proponemos a continuación, puede ayudar para optimizar y controlar operativamente las tareas de delimitación de los contenidos y estructura de la prueba, y asignación del número de items a construir para cada apartado.

En relación al segundo apartado, las preguntas de la prueba objetiva con opción de respuesta múltiple están compuestas por un conjunto de unidades (items), que se vertebran estructuralmente en torno a los contenidos y procesos mentales que tratamos de evaluar.

Un item es una proposición o declaración que en forma de afirmación, negación, pregunta o instrucción acerca de u

| TEMA (1) | % POND (2) | DESCRIPCION OPERATIVA (3) | %POND (4) | Nº DE ITEMS (5) |
|--------------|------------|---------------------------|-----------|-----------------|
| Informática | 40 | Conocer... | 25 | 1 |
| | | Comprender... | 25 | 1 |
| | | Aplicar... | 25 | 1 |
| | | Analizar... | 25 | 1 |
| TOTAL TEMA 1 | | TOTAL TEMA 1 | | 4 |
| Carpintería | 10 | Conocer... | 100 | 1 |
| TOTAL TEMA 2 | | TOTAL TEMA 2 | | 1 |
| Automoción | 30 | Comprender... | 33 | 1 |
| | | Aplicar... | 33 | 1 |
| | | Sintetizar... | 33 | 1 |
| TOTAL TEMA 3 | | | 100 | 3 |
| Economía | 20 | Conocer... | 50 | 1 |
| | | Analizar... | 50 | 1 |
| TOTAL TEMA 4 | | | 100 | 2 |
| TOTAL PRUEBA | 100 | TOTAL PRUEBA | 100 | 10 |

Figura 4: Tabla de Determinación. (CERT-EUROFORM: Introducción..., P. 54)

contenido de la materia de evaluación, se dirige al estudiante debiendo éste ejercitar los procesos mentales oportunos para resolver correctamente la sollicitación que se le realiza.

Como concreción dentro del marco del CERT, estos items serán preguntas de elección múltiple, que incorporan múltiples opciones de respuestas que pueden ser explícitas (al menos tres y no más de

cinco) e implícitas (las cuatro Soluciones Generales Implícitas que hemos propuesto (rechazo, todas, carencia e incoherencia).

En relación al tercer aspecto, la presentación de la prueba, habitualmente se propone en formato de cuestionario con soporte de papel, aunque no es la única forma posible, ya que en el marco de CERT se han desarrollado propuestas:

para el uso interactivo con ordenador de pruebas de evaluación partiendo de estos mismos conceptos.

En general, es conveniente tener en cuenta los siguientes criterios para el formato final de la prueba (adaptado de CERT-EUROFORM, 1993, 43):

- a) La prueba sólo debe incluir los contenidos programados y que se hayan hecho explícitos a los alumnos.
- b) Los items deben abarcar todos los puntos del programa o, en caso de ser muy extenso, una muestra representativa del mismo (obtenida mediante técnicas de muestreo).
- c) Es aconsejable distribuir los items según su dificultad de resolución; así, pondremos inicialmente los más sencillos y a continuación los más complejos para finalizar con items de menor dificultad.
- d) La prueba se elaborará por un equipo de expertos (p.e.: profesores de las materias que se van a evaluar) o se expondrá al juicio de otros colegas especialistas en la materia.

Ya comentamos anteriormente que con el objeto de guiar la construcción de las pruebas objetivas de opción múltiple parece adecuado incluir algunas referencias acerca de la definición y formulación de este tipo de cuestiones. El apartado siguiente desarrolla esta propuesta.

LAS CUESTIONES DE OPCIÓN MÚLTIPLE COMO UNIDADES BÁSICAS DE LA PRUEBA.

1. DEFINIENDO LAS CUESTIONES DE OPCIÓN MÚLTIPLE.

Parece importante, una vez que se ha introducido el concepto de item, definir

claramente lo que son las Cuestiones de Opción Múltiple y sus diferencias con otros tipos de items.

Inicialmente diremos que son Preguntas de Elección Múltiple aquellas e las que, independientemente de la persona que contesta, se presentan varias soluciones entre las cuales ha de escogerse la correcta rechazando los distractores. Por lo tanto, no serán Cuestiones de Opción Múltiple las preguntas o items de opción múltiple en las que los sujetos pueden responder señalando su posición, identificándose, opinando o realizando estimaciones y valoraciones sobre el tema propuesto. (CERT-EUROFORM Guía Rápida, 1993, 20).

2. COMPONENTES DE UNA CUESTIÓN DE OPCIÓN MÚLTIPLE.

Este tipo de items están constituido por tres partes claramente diferenciable (CERT-EUROFORM: Guía Rápida, p.21)

- 1) La consigna.
 - 2) El cuerpo.
 - 3) Las opciones de respuesta.
- 1) La consigna indica al alumno la acción que debe realizar (si debe: elegir la respuesta correcta, relacionar términos, rechazar determinadas opciones etc.); el modo en que debe hacerlo (por ejemplo, indicar el número de la opción elegida en la hoja de respuestas) y las normas de valoración fundamentales (cada respuesta incorrecta restará un punto. Es mejor no responder que hacerlo por azar, etc). La consigna puede ser válida para un conjunto de cuestiones e incluso para toda la prueba. En este último caso sólo será necesario formular la consigna al comienzo del cuestionario.

En el marco del CERT es necesario presentar en la consigna, añadiéndose a la consigna clásica, una referencia hacia las Soluciones Generales Implícitas y los Niveles de Confianza.

- 2) El cuerpo es la parte donde se define el problema y se plantea, en definitiva, la cuestión.
- 3) Las opciones de respuesta constituyen la última parte que hemos distinguido en la Cuestiones de Opción Múltiple. Entre las opciones debe encontrarse la solución correcta, quedando el resto de soluciones propuestas como distractores que pareciéndose a la respuesta correcta obligan al alumno a discriminar. En el marco del CERT el número de opciones explícitas se ve incrementado por las cuatro opciones implícitas lo cual, además de determinar la vigilancia cognitiva del sujeto evaluado, permite obtener un mayor control sobre los efectos de la respuesta al azar.

3. TIPOS DE CUESTIONES DE OPCION MÚLTIPLE

En cuanto a los tipos de cuestiones que podemos encontrar o elaborar pueden ser equivalentes a cualquiera de los tipos de items que clásicamente encontramos en las pruebas de rendimiento, siempre que se ajusten a los preceptos anteriores. La ventaja que ofrece la inclusión de las Soluciones Generales Implícitas es que tal vez no se pierda validez a pesar de que sean items con muy pocas soluciones reales o explícitas.

Lo único que parece adecuado recomendar es el ajuste que debe existir entre la configuración de la cuestión y el dominio de materia y proceso mental que debe representar. En este sentido, se proponen

dos estrategias alternativas para la concepción de los items (ALONSO, 1992, 33) y para la posterior corrección de la prueba (DE KETELE, 1984, 81), estos son:

El método «Top Down», o nivel básico, en el que se determina previamente para cada objetivo lo que debe ser conocido, comprendido, aplicado, analizado, sintetizado..., de manera que posteriormente se puedan generar las cuestiones de opción múltiples ajustadas a cada uno de estos procesos mentales que podrán ser evaluados por separado.

El método «Bottom-up», o nivel de maestría, consistente en señalar las situaciones problemáticas que el estudiante debe ser capaz de resolver, creando a continuación cuestiones relevantes para la solución de dichos problemas que implicarán de forma simultánea a diferentes procesos mentales.

VALIDEZ Y FIABILIDAD DE LAS PRUEBAS OBJETIVAS DE OPCION MULTIPLE

Cuando se manejan pruebas objetivas para la evaluación del aprendizaje estamos en última instancia utilizando herramientas específicamente construídas para medir los resultados cognitivos del aprendizaje. Estas herramientas o instrumentos deben por tanto poseer una serie de características que nos aporten la confianza necesaria en los datos que recogemos cara a la toma de decisiones. Es decir, es muy importante, cuando se diseña o usa una prueba, conocer algunos indicadores de calidad de la misma para, de esta manera, estar seguros de la eficacia del trabajo que realizamos cuando evaluamos con ellas.

Los indicadores de validez y fiabilidad son características técnicas de los instru-

mentos de medida; también lo son por tanto en nuestro caso, para las pruebas objetivas de opción múltiples construidas en el marco del CERT.

«Una prueba es válida cuando mide lo que dice medir. Si se elabora una prueba para medir la adquisición de unos determinados objetivos, será válida si mide realmente esos objetivos y no otros.» (Sampascual, 1982, 45).

Existen diferentes indicadores de validez de una prueba, pero el indicador que nos parece más apropiado establecer para este tipo de pruebas es el que denominamos «validez de contenido». La prueba tiene validez de contenido en la medida en que los items que la componen son una buena representación del contenido que se va a evaluar. Como ya hemos expuesto anteriormente en el apartado de construcción de las pruebas, para conseguirse la validez de contenido de una prueba de rendimiento ha de planificarse cuidadosamente la cantidad y cualidad de los items que representan a cada bloque de materia y a cada proceso mental que se va a medir. Es por esto que proponíamos ayudarnos de una Tabla de Determinación como instrumento facilitador de la planificación del contenido de la prueba.

Sin embargo, no es éste el único indicador de validez que podemos aplicar para conocer la calidad de las pruebas objetivas. También proponemos el uso de indicadores de validez de criterio, concretamente, Bisquerra (1985, 361), con el objeto de validar una prueba objetiva de lengua catalana analiza la correlación existente entre las puntuaciones obtenidas por los sujetos en dicha prueba y las que ellos obtienen en pruebas alternativas de composición escrita y comprensión lectora.

Así mismo, otro indicador que es necesario conocer sobre la prueba objetiva es la fiabilidad de la misma. En este sentido: fiabilidad se refiere a la constancia de medidas. Una prueba es fiable si, al aplicada a alumnos de características no similares en circunstancias parecidas, produce resultados semejantes.» (Sampascual 1982, 45).

Evidentemente validez y fiabilidad son aspectos interrelacionados. Son factores que influyen en la fiabilidad de las pruebas objetivas, entre otros, los siguientes (CARRERA, y ESPIN, 1986): La longitud de prueba, heterogeneidad del grupo a que aplica, similitud del contenido de los items y dificultad intrínseca de la prueba. El método empleado para la obtención de este indicador de fiabilidad también es señalado como fuente de variación en el cálculo de este indicador. Todo ello refleja que es necesario realizar una apreciación completa de este aspecto basada en diversos indicadores, y además, apreciando cada uno de ellos en su justa medida.

Como estrategias para el cálculo de los indicadores de fiabilidad (que suelen ser coeficientes de correlación) pueden destacarse (Colás y Buendía, 240): 1) el método test/retest, 2) la construcción de formas equivalentes o paralelas de una misma prueba, 3) la división de las pruebas en dos mitades (división aleatoria o de algún modo equivalente) y, 4) El análisis de la consistencia interna de una prueba. Las tres primeras estrategias implican la aplicación de las mismas pruebas (la misma en dos ocasiones, las formas paralelas o las dos mitades) a una misma población y el cálculo de la correlación entre ellas será nuestro indicador de fiabilidad. La cuarta estrategia implica el desarrollo de indicadores a partir de coeficientes de correlación hallados entre pun-

tuciones de un mismo sujeto. Para ello puede procederse según diversas fórmulas clásicas (Coeficiente Alpha de Cronbach, Kuder-Richardson).

Las pruebas desarrolladas bajo el soporte informático que nos ofrece el CERT disfrutará del cálculo automatizado de un conjunto de estadísticos. Entre estos destacamos: porcentajes de acierto y error (aplicables al cálculo del índice de dificultad de los ítems), correlación biserial entre las opciones y la puntuación total obtenida en la prueba, aplicables para el cálculo de grado de discriminación y, medias y porcentajes de aciertos según Niveles de Confianza, para la estimación de la coherencia y el realismo de la autoevaluación del alumno.

Todos estos indicadores nos permiten elaborar conclusiones sobre la calidad de cada ítem (pistas sobre lo que sí/no funciona bien) y sobre la prueba en su conjunto. En función de esta información podremos tomar decisiones acerca de las pruebas objetivas que usamos. Así, éstas podrán ser mantenidas en su actual configuración, mejoradas o incluso abandonarse su uso para construir otras más apropiadas.

APLICACION DE LAS PRUEBAS OBJETIVAS DE OPCION MULTIPLE CONSTRUIDAS PARA EL PROGRAMA CERT

Respecto a la realización de este tipo de pruebas, la introducción de las Soluciones Generales Implícitas y los Niveles de Con-

fianza suelen suponer una novedad para el evaluado. Por ello, es conveniente explicar a los sujetos de la evaluación, los objetivos formativos que se persiguen con la introducción de dichos aspectos, procurándose su introducción gradual, especialmente en aquellos casos en que no exista costumbre o tradición en el uso de pruebas objetivas.

El CERT ofrece diversas posibilidades de uso tales como la aplicación de pruebas colectivas a grandes cantidades de alumnos o la interacción personal de estos con el programa informático CERT para la autoevaluación. En todos los casos se requiere un análisis previo de las características intrínsecas de cada situación educativa y de las medidas que requiere cada una de ellas.

Con respecto a la corrección de las pruebas, y en coherencia con el enfoque evaluativo planteado, puede utilizarse un procedimiento que ofrece el CERT consistente en la combinación de dos medidas: el número de respuestas correctas y los Niveles de Confianza seleccionados.

El programa ofrece inicialmente un baremo para la corrección, - que «se basa en un modelo teórico evolucionado (CHOPPIN, 1970; LECLERCQ, 1986), conforme a la teoría de decisiones y a las probabilidades subjetivas» (ALONSO, 1992, 55)-, que puede ser cambiado según el criterio del usuario o ser usado con diferentes niveles de rigurosidad (de 16 a 20). Este baremo está representado en el cuadro siguiente:

| | NIVELES DE CONFIANZA | | | | | |
|--------------------------------|----------------------|----|----|----|----|-----|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| Para las respuestas CORRECTAS: | 13 | 16 | 17 | 18 | 19 | 20 |
| Para las respuestas ERRONEAS: | 4 | 3 | 2 | 0 | -6 | -20 |

- Línea 1: En esta se señala el porcentaje de alumnos que han seleccionado ese modo de responder (o tasa de elección de cada solución propuesta) lo que permite identificar el grado de atractivo de cada una de las soluciones planteadas.
- Línea 2: R. BIS. En esta segunda Línea se presenta un coeficiente de discriminación en relación a cada una de las soluciones propuestas. Este coeficiente se basa en el cálculo de una correlación biserial (r.bis) para cada uno de los modos en que se puede responder un determinado ítem. Su valor esta comprendido entre -1 y +1. Si es positivo, significa que este modo de responder ha sido adoptado por alumnos cuya puntuación total en el test resulta más elevada que la puntuación en el total del test realizado por el resto de los alumnos (aquellos que han seleccionado otra respuesta). Este coeficiente debe ser positivo en relación a la respuesta correcta y nega-

tivo para las respuestas incorrectas y entre los dos en referencia a la omisión. Siempre que uno de los modos de responder no sea adoptado por nadie (o por todos), automáticamente deberá ser considerado nulo.

- Línea 3: C. moy. En esta se presenta el nivel de confianza medio de los alumnos que han seleccionado cada solución (o tasa media de confianza atribuida a cada solución). Es un índice que varía entre 12,5 y 97,5% (valores centrales de los niveles de confianza 0 y 5).

Este debe ser superior para los alumnos que eligen la respuesta correcta que para los que se han equivocado. De no ser así resultaría conveniente cuestionarse que ha conducido a los alumnos a ser confiados en relación a una respuesta incorrecta o poco confiados en base a una correcta.

Para obtener este índice en base a una solución dada, se calcula la media

| Procesos: No selección | | Materias: No selección | | | | | | | | | |
|------------------------|------|------------------------|-------|------|------|------|---|------|------|------|-------|
| | IT : | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| IT 1 | | 0.00 | 50.00 | 0.00 | 0.00 | | | 0.00 | 0.00 | 0.00 | 50.00 |
| Rbis | | 0.00 | -1.00 | 0.00 | 0.00 | | | 0.00 | 0.00 | 0.00 | 1.00 |
| Caed | | 0.00 | 80.00 | 0.00 | 0.00 | | | 0.00 | 0.00 | 0.00 | 80.00 |
| IT 2 | | 0.00 | 50.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 50.00 |
| Rbis | | 0.00 | -1.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 1.00 |
| Caed | | 0.00 | 80.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 80.00 |
| IT 3 | | 0.00 | 50.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 50.00 |
| Rbis | | 0.00 | -1.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 1.00 |
| Caed | | 0.00 | 80.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 80.00 |
| IT 4 | | 0.00 | 50.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 50.00 |
| Rbis | | 0.00 | -1.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 1.00 |
| Caed | | 0.00 | 80.00 | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | 80.00 |

Pulse cualquier tecla para continuar

Figura 6

de probabilidades (posibilidades de acertar) predecidas por los alumnos que han seleccionado esta solución. Se trata pues de la media de los valores centrales en los intervalos de confianza.

En relación a los datos y puntuaciones que se ofrecen sobre los alumnos, éstos se calculan en base al baremo, anteriormente reseñado, y el grado de rigurosidad que se aplique. Ambos son modificables, aunque se recomienda el uso de este baremo (Alonso, 1992, 59). En cuanto al nivel de rigurosidad se entiende que cuando se utilizan los grados de confianza la puntuación máxima corresponde a un nivel de exigencia muy elevado («injusto»), ya que es necesario responder correctamente todo y tener el máximo nivel de confianza también en todo (Alonso, 1992, 81). Por lo tanto, se recomienda no utilizarlo para definir el nivel óptimo ya que se trata de una exigencia muy elevada, cuando normalmente basta con contestar bien a todo.

INFORMACION SOBRE LOS ALUMNOS

El programa ofrece una serie de posibilidades con respecto a los datos puntuaciones de los alumnos, ofreciendo resultados de los alumnos ya sea en forma individual, por secciones o en forma colectiva.

1.- INFORMACIÓN INDIVIDUAL

En cuanto a la información individual que el programa proporciona por cada alumno tenemos los siguientes tipos de datos que se presentan a través de cuatro pantallas:

- A) Datos de los alumnos: Se presentan los datos personales del alumno, sus respuestas, información sobre el test, etc.
- B) Valores numéricos: (Figura 7)
 - a) Número y porcentaje de respuestas correctas: El programa presenta para cada nivel de confianza (columnas 0 a 5) y porcentaje de aciertos anteriormente en relación al conjunto de la prueba (columna TODO).

| Grado de confianza | 0 | 1 | 2 | 3 | 4 | 5 | TOT |
|--------------------|-------------------|--------------|--------------|----------------|--------------|-------|-----------|
| Nº de resp. | 0 | 4 | 4 | 4 | 4 | 4 | 20 |
| Nº correct. | 0 | 4 | 4 | 4 | 4 | 4 | 20 |
| Nº incorrect. | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| % resp. correct. | 0.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Val. centr. | 12.5 | 37.5 | 60.0 | 77.5 | 90.0 | 97.5 | 72.5 |
| PUNTAJES | | | | | | | |
| por resp. corr. | 0 | 64 | 68 | 72 | 76 | 80 | 360 |
| por resp. incorr. | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | | | | | | | 360 / 400 |
| Riguros. | 16 | PUNT. | 23.5/20 | SC. EV. | 20.0/20 | | |
| C | Coherencia | 0.000 | Insuficiente | TEST | ACCESO | | |
| | Realismo | 0.388 | Insuficiente | ALVINO | EXPOSITO J. | | |
| | | | | PM | No selección | | |
| | | | | HAT. | No selección | | |

Figura 7

* Nº respuestas: número de respuestas proporcionadas por el alumno y en base a ellas:

a.1 Nº correctas: número de respuestas correctas.

a.2 Nº incorrectas: número de respuestas incorrectas.

a.3 Porcentaje de respuestas correctas o tasa de exactitud observada.

a.4 Val. central: tasa de aciertos esperada que supone de hecho el valor central del intervalo de probabilidad correspondiente al nivel de confianza. Si compara con la anterior permite detectar dónde el alumno se sobrevalora o se infravalora. En caso de que el de respuestas correctas sea superior al valor central correspondiente, el alumno se ha infravalorado, si resulta inferior se ha sobrevalorado.

b) Puntuaciones brutas: Se presentan también una serie de puntuaciones brutas calculadas en función del baremo, atribuyendo a las respuestas correctas e incorrectas una puntuación que depende del nivel de confianza proporcionado.

En relación a cada uno de los niveles de confianza y para el conjunto de la prueba.

- Por respuestas correctas: puntuación bruta obtenida en relación a las respuestas correctas.
- Por respuestas incorrectas: Puntuación obtenida en base a las respuestas incorrectas.
- Total bruto: Puntuación total bruta obtenida por el alumno, calculada en base a sus respuestas y en función a los niveles de confianza proporcionados. Esta se relaciona con la puntuación máxima que se calcula en base a : Nº de ítems X 20.

c) Puntuación total adaptada: Una línea se reserva a esta puntuación, que se trata de la puntuación total bruta adaptada en relación a una puntuación de referencia, más justa, más habitual. Esta puntuación se calcula en función de la fórmula:

$$\boxed{\text{PUNTUACION CERT}} = \frac{\text{Puntuación Bruta} \times 20}{\text{Nº de Ítems} \times \text{Rigurosidad}}$$

d) Puntuación SC.GU: En este caso se ofrece una puntuación de corte más clásico en la que se elimina el azar.

e) Índice de auto-evaluación. Se proporcionan dos índices, el índice de realismo y el de coherencia (ver punto siguiente: Gráfica de realismo)

C) Gráfica de realismo. Se construye en base a los dos ejes siguientes:

- Sobre el eje horizontal (abcisas) se representan las zonas de probabilidades (probabilidad de acertar predicha por el alumno) que corresponde a cada una a un nivel de confianza.
- En el vertical (ordenadas) se sitúa la tasa de exactitud (de 0% a 100%) de las diversas respuestas proporcionadas por el alumno que han acompañado a cada grado de confianza (0,1,2...).

Por cada uno de los niveles de confianza que haya sido utilizado por el alumno, podremos encontrar en este gráfico un punto, cuya abcisa corresponde al valor central de la zona de probabilidad correspondiente a este nivel y cuya ordenada supone la tasa de exactitud observada en relación a este nivel (porcentaje de respuestas correctas).

Este punto es representado por un cuadrado para facilitar su visualización.

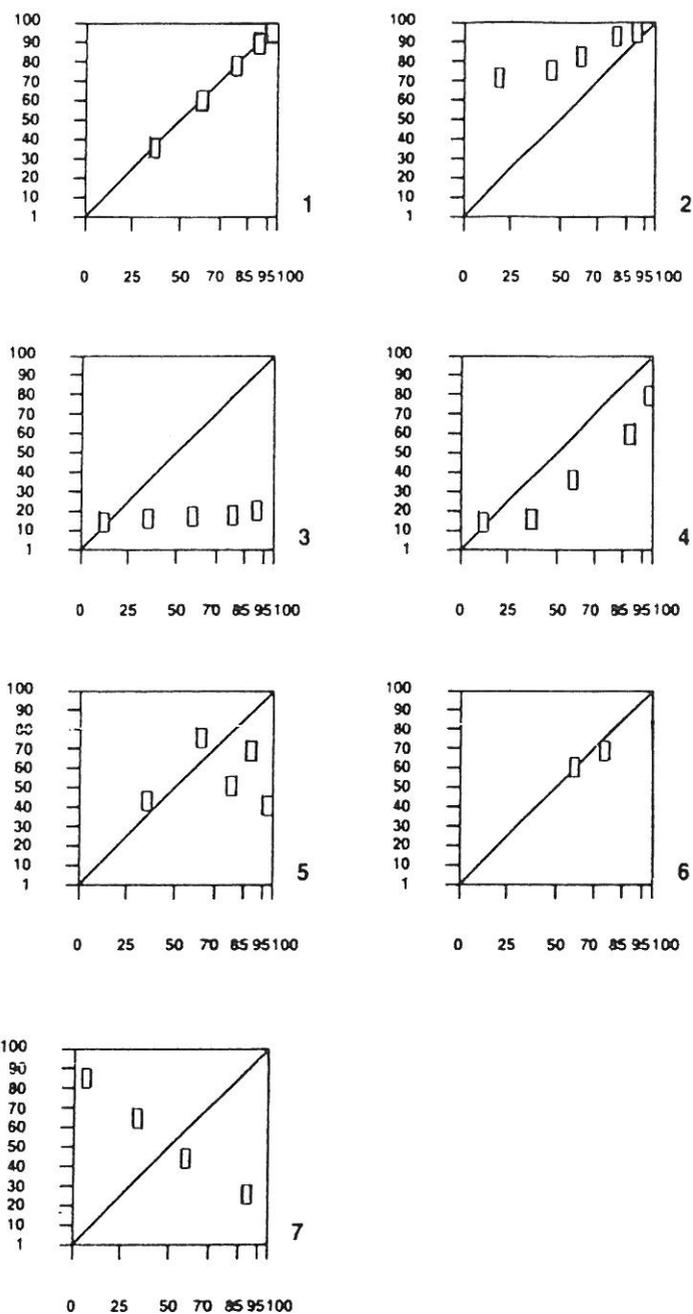


Figura 8

Estos se espera que estén sobre la diagonal, es decir, se espera que cada valor central de la zona de probabilidad (probabilidad de acertar predicha por el alumno) corresponda con las tasas de exactitud obtenidas. Pero en la realidad estos se pueden presentar de dos formas diferentes:

- Por debajo de la diagonal, significara que los éxitos son inferiores a las predicciones, es decir, que el alumno se sobreestima.
- Por encima de la diagonal, significara que este se subestima.

* Índice de realismo: Indica en qué medida la confianza que el alumno tiene en sí mismo se corresponde o no a la realidad. Es decir, si la probabilidad de acierto predicha por el mismo se acerca o no a los aciertos obtenidos en la realidad (tasa de exactitud).

Se calcula en base a diferencias entre la tasa de exactitud observada para cada nivel de confianza y el valor central de intervalo de probabilidad correspondiente a dicho nivel.

En la figura 8 se ofrecen una serie de ejemplos, en relación a los cuales podemos observar:

- Si las tasas de exactitud obtenidas se encuentran conformes a las esperadas por el alumno (valor central del intervalo de probabilidad), los rectángulos se sitúan exactamente sobre la diagonal (ver alumnos 1 y 6). El índice de realismo toma el máximo valor: 1.
- Si las tasas de exactitud obtenidas se encuentran alejadas en relación a los valores esperados, los rectángulos se alejan de la diagonal (ver alumnos 2,3 y 5). El índice de realismo puede incluso ser negativo y descender hasta -3.75 (ver alumno 7).

El índice de realismo no precisa en qué sentido se produce el alejamiento siendo necesario, por ello, observar en el gráfico el emplazamiento de cada rectángulo. Este índice es el que mejor revela la calidad de la autoevaluación, aunque resulta insuficiente en relación a ciertos aspectos sin embargo importantes: ¿cuál es la tendencia mayoritaria del error: infravaloración o sobrevaloración? ¿el individuo discrimina bien entre sus diversos niveles de confianza?.

Por otro lado, se calcula y añade a la tabla el error standar de medición de cada tasa de exactitud. Este error tiene mayor importancia si la tasa de exactitud ha sido calculada en base a pocas respuestas, debiéndose evitar entonces el conceder demasiada importancia a las observaciones calculadas para niveles poco usados o generalizar interpretaciones. como vemos en el siguiente ejemplo (figura 9), tomado de Alonso (1992, 66):

Si un alumno ha utilizado el nivel de confianza 3, dos veces de modo correcto, su tasa de exactitud vale el 50%. Si un alumno B ha empleado 20 veces la confianza 3, diez veces de forma correcta, su tasa de exactitud vale igualmente el 50%.

Es evidente que el valor de la tasa de exactitud resulta más fiable en el caso B que en el A, particularmente porque el error de medición es más fiable en el caso del alumno b.

- Índice de coherencia (personal): Indica si el alumno utiliza niveles de confianza más elevados cuando tiene mayor éxito o menos elevados cuando responde peor. Es decir, muestra si el alumno tiene un comportamiento coherente en relación a sí mismo o no.

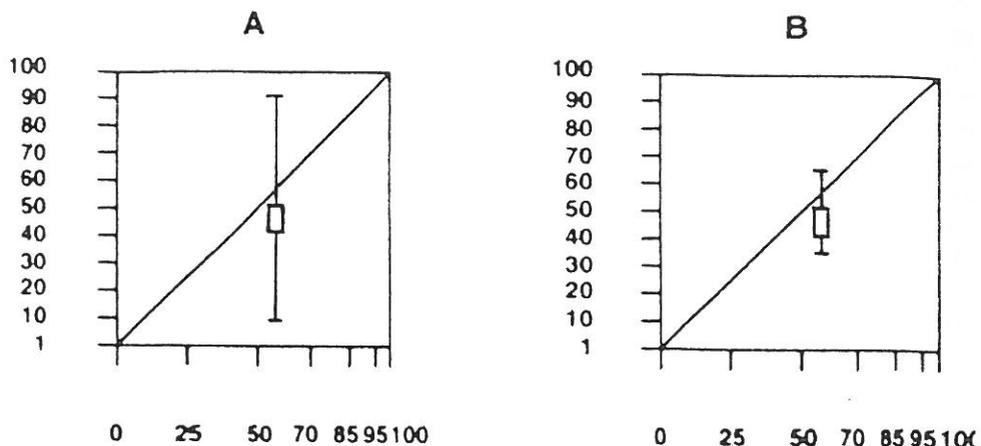


Figura 9

Debe de responder mejor cuando estima tener más posibilidades de contestar correctamente y peor cuando estima tener menos posibilidades de acertar.

El índice se calcula en base a la correlación entre las tasas de exactitud (TE) y los valores centrales de los X niveles de confianza (en este caso seis niveles) (Alonso, C. pag.124; 1992) . Los valores que puede tomar este índice oscilan desde:

- a) +1 Si la coherencia es perfecta, alineamiento perfecto y creciente (alumno 1)
- b) -1 Si la coherencia es completamente inversa, alineamiento perfecto y decreciente. Cuanto más piensa el alumno que tiene posibilidades de acertar menos acierta (alumno 7).
- c) 0 Si no hay coherencia, alineamiento horizontal o nube de puntos en la que no es posible deducir una progresión en las autoevaluaciones del alumno.

La fórmula de base para el cálculo del coeficiente de correlación de Bravais-

Pearson (también llamada r) puede localizarse en cualquier manual de estadística.

Este índice resulta mucho más fiable que el anterior, aparece positivo y elevado cuando el alumno se comporta de modo sistemático, coherente. Desgraciadamente, esta coherencia puede ser que comporta la infravaloración o sobrevaloración.

Como ejemplo de estos índices podemos tomar los siete casos de la figura 8 que:

- El alumno 1 resulta ser muy realista y coherente consigo mismo.
- El alumno 2 es muy coherente en relación a su infravaloración.
- El alumno 3 no resulta coherente y sobrevalora.
- El alumno 4 resulta coherente en sobrevaloración.
- El alumno 5 es realista pero no resulta coherente.
- El alumno 6 es realista y coherente.
- El alumno 7 resulta irrealista e incoherente.

2.- INFORMACIÓN COLECTIVA

También el programa posibilita la presentación de la información de forma colectiva. Existiendo diversas posibilidades:

- a) Información de los alumnos en relación a las diferentes secciones definidas: Sobre este apartado el programa ofrece dos posibilidades, una lista de los alumnos con las puntuaciones obtenidas en las diferentes secciones sin tener en cuenta el grado de confianza (sin CERT) y otra teniéndolo en cuenta.
- b) También se puede obtener información sobre las puntuaciones totales, ofreciendo la posibilidad de visualizar o imprimir un listado con las puntuaciones totales adaptadas, de las puntuaciones SC.GU. o ambas.

Por último el programa ofrece la posibilidad de almacenar los datos para su exportación a otros programas de análisis del tipo SPSS, SPAD-N, SYSTAT, BMDP, SAS, etc.

APLICACIÓN DEL PROGRAMA CERT EN LA UNIVERSIDAD DE SEVILLA

Esta herramienta evaluativa ha sido utilizada con carácter experimental por un grupo de profesores de la Universidad de Sevilla, con una alta motivación por la temática de las pruebas objetivas y su posible utilización con un soporte informático.

Inicialmente se impartió un curso para dar a conocer este instrumento y su aplicabilidad. Los profesores implicados en esta experiencia han sido diez. Respecto a estos docentes que han utilizado el programa CERT, cabe resaltar algunos

aspectos tales como la diversidad en cuanto a los perfiles profesionales y académicos, y las materias que imparten. Estos responden en algunos casos al perfil de «usuarios habituales» de las pruebas objetivas y en otros al de profesores iniciados en esta opción evaluativa a partir de su contacto con el Proyecto CERT. En cuanto a su origen institucional, estos profesores provienen tanto de Escuelas Universitarias como de Facultades. En este sentido, han participado profesores de las Facultades y Escuelas Universitarias de Psicología, Ingeniería Técnica, Educación, y Ciencias Económicas y Empresariales. Las materias a cargo de estos profesores quedan recogidas en el siguiente abanico de asignaturas: Psicología Experimental, Matemáticas Aplicadas, Didáctica de las Matemáticas, Física General, Sociología de la Empresa, Planificación Estratégica de Recursos Humanos, Teorías del Aprendizaje, Educación Física y del Deporte. La incidencia en la población estudiantil de la Universidad de Sevilla ha supuesto evaluar con este instrumento a un total de 2.000 alumnos.

En el caso de profesores que habían usado con anterioridad las pruebas objetivas de elección múltiple lo habían hecho por tres razones: el alto número de alumnos a su cargo, la objetividad en la evaluación y la rapidez y facilidad para la corrección. Estos aspectos se ven reforzados al incorporar el software CERT.

Respecto a la aplicación realizada con esta herramienta, hay que distinguir dos momentos: 1) diseño de la prueba para el programa CERT y 2) modo de uso dentro de la evaluación.

En el primer momento, se elaboraron distintas pruebas que contenían las soluciones generales implícitas y los niveles

de confianza. Con carácter general, los profesores han administrado dos pruebas con los elementos CERT. La inclusión de estos elementos dentro de las pruebas ha sido gradual. En la mayor parte de los casos, la primera prueba diseñada por los profesores sólo incorporaba dos de las cuatro respuestas implícitas posibles.

La incorporación del segundo elemento CERT, los niveles de confianza, en la evaluación de los estudiantes ha variado en función de la experiencia del profesor en el uso de pruebas objetivas y de la costumbre del alumnado en ser evaluado de esta forma.

Respecto a las modalidades en que se ha usado la herramienta informática, la mayoría de los profesores la han utilizado como un modo de extraer información objetiva y fiable del aprendizaje del alumno, con objeto de calificarlo más adecuadamente. En algún caso, la evaluación realizada con esta prueba ha tomado la forma de autoevaluación por parte del alumno, a través del ordenador de manera interactiva. Esta modalidad ofrece la posibilidad al estudiante de analizar sus errores y orientar su estudio.

El proyecto CERT ha utilizado para su aplicación la cobertura institucional del Instituto de Ciencias de la Educación de la Universidad de Sevilla.

PRIMERAS CONCLUSIONES DE LA APLICACIÓN

Las conclusiones iniciales que se han extraído de esta experiencia giran en torno al propio programa informático y sus aportaciones para la evaluación de los estudiantes universitarios.

En general, esta herramienta se ha mostrado eficaz para la evaluación rápida

y fiable de los conocimientos de los alumnos sobre la asignatura, en función de los resultados contrastados por los profesores. El programa ha proporcionado información sobre la validez y adecuación del enunciado de los ítems, lo que ofrece orientaciones para la construcción de la prueba. Aunque, sin duda, la principal virtud es que el programa y los elementos que incorpora permite juzgar valorar el trabajo de los alumnos con criterios predeterminados (objetivos).

Respecto a la introducción de las Soluciones Generales Implícitas y los Niveles de Confianza en la evaluación han resultado de utilidad a los profesores para valorar el aprendizaje de los estudiantes. El uso de estos elementos pueden contribuir a una mejor evaluación de los procesos formativos, tanto el referente al conocimiento como a la aplicación de determinados procesos cognitivos. La utilización del programa puede contribuir a orientar, por un lado al profesorado, para evaluar más adecuadamente, no limitándose sólo a valorar al alumno a partir de la recuperación o identificación de la información, sino, pudiendo orientarlo, en relación al uso de técnicas de estudio más adecuadas para su aprendizaje.

En cuanto a la manejabilidad del software Usuario, es necesario destacar que debido a las características del entorno informático dominante en el que se ha aplicado, se ha experimentado la versión para PC del CERT, lo que ha determinado en gran medida el contacto con esta herramienta. Esta versión es lo menos flexible de entre el conjunto de opciones de uso de este programa (Windows, Macintosh, Multimedia, etc). En este sentido, debemos señalar lo

existencia de versiones posteriores que mejoran las posibilidades de manejo del programa y sus prestaciones para la evaluación.

En general, los resultados obtenidos con esta aplicación inicial en un contexto universitario resultan positivos, (en el marco del Proyecto CERT se han realizado aplicaciones en otros niveles educativos). Parece oportuno continuar experimentando el programa, para explorar sus posibilidades evaluativas y formativas.

Queremos dejar constancia de nuestra gratitud a los profesores de la Universidad de Sevilla que han estado vinculados a este proyecto por la ilusión e interés mostrados en todo momento.

REFERENCIAS BIBLIOGRÁFICAS

- ALONSO, C. y Col. (1992): *Principios comunes para la evaluación de los resultados cognitivos de la formación*. Comisión de las Comunidades Europeas. Programa EUROTECNET.
- ALTBACH, Ph. (1985): Perspectives on comparative higher education: a survey of research and literature. En Ph. ALTBACH y D. KELLY (Eds.): *Higher education in international perspective. A survey and bibliography*. Mansell Publishing, London.
- BISQUERRA, R. (1985): Elaboración de la POLC (Prueba Objetiva de Lengua Catalana). *Revista de Investigación Educativa*. Vol. 3, Nº. 6., pp. 358-363.
- CABRERA, F. Y ESPIN, V. (1986): *Medición y evaluación educativa*. PPU. Barcelona.
- CANNON, R. (1983): The Professional Development of Australian University Teachers: an act of faith?. *Higher Education*, nº. 12, pp. 19 a 33.
- CERT- EUROFORM (1993): *Guía Rápida*. Centro de Video interactivo. Universidad de Barcelona.
- CERT- EUROFORM (1993): *Introducción a las pruebas objetivas*. Centro de Video interactivo. Universidad de Barcelona.
- COLAS, P. y BUENDIA, E. (1992): *Investigación Educativa*. Alfar. Sevilla.
- COLAS, P. y REBOLLO, M.A. (1993): *Evaluación de programas. Una guía práctica*. Kronos. Sevilla.
- DE KETELE, J.M. (1984): *Observar para educar. Observación y Evaluación en la práctica educativa*. Visor. Madrid.
- DE LANDSHEERE, G. (1978): *Los test de instrucción*. Oikos-Tau. Barcelona.
- DE MIGUEL, M. (1992): L'avaluació de l'ensenyament universitari des d'un enfocament institucional. *Temps d'Educació*, nº. 8, 2º semestre, pp.99 a 121.
- GARCIA BARBERO, M y Col.(1987): *Cómo confeccionar preguntas test y prac*. Alicante. ICE de la Universidad de Alicante.
- LECLERCQ, D. (1983): *Evaluation in education: An International Review Series*. Oxford, England. Pergamon Pres.
- PEREZ GOMEZ, A.(1991): Calidad de la enseñanza y desarrollo profesional del docente. En VV.AA.: *Sociedad, Cultura y Educación*. C.I.D.E./Universidad Complutense, Madrid.
- RODRIGUEZ DIEGUEZ, J.L. (1980): *Didáctica General I. Objetivos y Evaluación*. Cincel-Kapeluz, Madrid.
- SAMPASCUAL MAICAS, G. (1982): *Las pruebas objetivas. Un procedimiento para evaluar el rendimiento escolar*. Anaya/2. Madrid.
- SANTOS GUERRA, M.A. (1991): Criterios de referencia sobre calidad del proceso de enseñanza/aprendizaje en la Universidad. *Revista de Enseñanza Universitaria*, Vol.1, nº. 1., pp. 25 a 47.