



EURTECNET

**PRINCIPIOS
COMUNES
PARA LA
VALUACION DE
LOS RESULTADOS
COGNITIVOS DE
LA FORMACION**

Citar:
Bartolomé, Antonio R. (1992). Principios comunes para la evaluación de los resultados cognitivos de la formación. (colab. con J. Horgan, D. Leclercq, E. Boxus, C. Alonso). Bruselas: Comisión de las Comunidades Europeas. 1992.
Descargable desde:
http://www.lmi.ub.edu/personal/bartolome/libros/1992_Bartolome_Alonso_CERT.pdf

PRINCIPIOS COMUNES PARA LA EVALUACION DE LOS RESULTADOS COGNITIVOS DE LA FORMACION

Responsable del proyecto:

Jack HORGAN

Coordinador general de la

Oficina de Asistencia Técnica

EUROTECNET

Con la coordinación:

A. BARTOLOME C. ALONSO

Departamento de Didáctica y Organización Escolar

Universidad de Barcelona

ESPAÑA

Esta obra (principios comunes) está destinada a toda persona que en el terreno educativo tiene la obligación de tomar ciertas decisiones. Presenta IDEAS esenciales en relación a la presente propuesta.

Otro documento (un instrumento común) proporciona a los formadores los detalles operacionales necesarios para el buen funcionamiento del software CERT condebido para facilitar su trabajo de corrección.

Este proyecto cooperativo ha sido puesto en marcha por el equipo Eurotecnnet bajo la responsabilidad de Jack Horgan. Reagrupa a cuatro equipos provenientes de Bélgica, España, Gran Bretaña e Italia.

Cada equipo nacional participante esta compuesto por uno o dos investigadores o responsables de formación que, en el marco del presente proyecto, trabajan en estrecha colaboración como mínimo con una empresa.

Coordinador de Proyecto: Jack Horgan

ESPAÑA

<i>Departamento de Didáctica y Organización escolar Universidad de Barcelona</i>	<i>Departamento de recursos humanos NISSAN, Motor Ibérica</i>
A. BARTOLOME PINA	D. RIVEROLA
C. ALONSO CANO	

BELGICA (Coordinación científica)

<i>Servicio de Tecnología de la educación Universidad de Lieja.</i>	<i>Formaciones STE</i>
E. BOXUS	H. Wuidar
D. LECLERCQ	
S. OSTERRIETH	
P. DE BROGNIEZ	

© **Comisión de las Comunidades Europeas 1992**

Este documento ha sido elaborado por la Comisión de las Comunidades Europeas en el marco del programa EUROTECNET. No representa necesariamente las posiciones oficiales de la Comisión Europea.

ISBN 90-74193-02-1

ITALIA

*Tecnología educativa
y formación a distancia.*

A. GAMMALDI

V.A. DELLE PIANE,
CN/ITE
M. MAZZOTTA, *consultor*
G.PINI, *consultor*

REINO UNIDO

*Astra Training services
Product development Branch
Section*

*Sills Training Agency
Sheffield Skillcentre
Carpentry/Joinery*

G. TAYLOR

B. TROUT

INDICE

PROLOGO

CAPITULO I

LA EVALUACION EN LA FORMACION

1. El papel de la evaluación en la formación
2. La evaluación formativa
3. Limitaciones del presente proyecto
4. Conceptos e instrumentos fundamentales
5. En resumen

11
15
16
18
22

CAPITULO II

UNA TAXONOMIA

1. Objetivos establecidos
2. Taxonomía adoptada
3. En resumen

23
23
34

CAPITULO III

EL CUESTIONARIO Y LAS QCM

1. La evaluación
2. Características de las QCM
3. Ventajas y desventajas de las QCM
3. Un ejemplo
4. En resumen

35
36
39
43
44

CAPITULO IV

SOLUCIONES GENERALES IMPLICITAS

1. Finalidad fundamental de las S.G.I.
2. Nivel de Solicitud implícita por las QCM
3. Consignas concierntes a las soluciones generales
4. Soluciones generales y procesos mentales
5. El entrenamiento en el cuestionario QCM en relación a las S.G.I
6. En resumen

45
46
48
49
50
52

CAPITULO V NIVELES DE CONFIANZA

1. La postura social
2. La medida del conocimiento
3. condiciones metodológicas de utilización
4. En resumen

53
54
55
69

CAPITULO VI

LA NECESIDAD DE UN SOFTWARE

1. Razon de ser de un Software
2. Principales etapas del trabajo
3. La introducción de información en el software
4. La información proporcionada en relación a la calidad de las preguntas.
5. La información proporcionada en relación a cada alumno
6. La información colectiva
7. En resumen

71
71
72
74
77
85
86

CAPITULO VII

PERSPECTIVAS

1. Una primera puesta en marcha
2. Futuros desarrollos

87
88

CONCLUSIONES GENERALES

91

GLOSARIO

93

REFERENCIAS BIBLIOGRAFICAS

101

ANEXOS

1. Discusiones en relación a la taxonomía de Bloom
2. La r.bis
3. El índice de realismo
4. El índice de coherencia

103
109
119
123

PROLOGO

EUROTECNET: UN PROYECTO TRANSNACIONAL

Los sistemas de evaluación de competencias y calificaciones requeridas por el mercado del trabajo evolucionan rápidamente, éstos, deben ser capaces de adaptarse a las nuevas exigencias.

Los métodos tradicionales utilizados para evaluar los resultados de la formación fueron establecidos en relación a un programa escolar y universitario relativamente fijo y estable.

El mercado de trabajo moderno revela nuevas y diferentes necesidades a medida que la tecnología progresa. Estos avances continúan precisando variaciones en el seno de la organización del trabajo.

Este doble fenómeno engendra una nueva serie de competencias y de calificaciones respecto a las cuales los trabajadores deben estar formados.

Por otra parte, tanto la nueva y rápida progresión de las competencias tecnológicas, como una formación sobre cualificaciones inter e intrapersonales son igualmente necesarias, como seguimiento de la evolución de las prácticas del trabajo; por ejemplo, resolución de problemas y toma de decisiones, planificación de tareas, capacidades de comunicación, responsabilidad personal en materia de control de calidad, etc.

Varios Estados miembros elaboran actualmente sistemas pedagógicos y de evaluación de cara a aportar respuestas eficaces a estas cuestiones.

El trabajo científico presentado en esta publicación EUROTECNET resume algunos de estos desarrollos y presenta el nuevo cuadro conceptual adoptado como modelo de trabajo. Este nuevo modelo Responde a un cierto número de intereses manifiestos: satisface las necesidades de los aprendices pero también las de las

empresas, instituciones o responsables de formación; puede responder rápidamente a las demandas del mercado de trabajo y se basa en la tecnología.

Resulta interesante señalar que la experimentación de base de dichos conceptos se realiza en el seno de las empresas. Así ha ocurrido con trabajadores de la industria del automóvil en Barcelona (España), trabajadores de Bélgica, que siguen una nueva formación tecnológica con finalidad de conversión, y finalmente, trabajadores de Sheffield (Royaume-Uni) (largo tiempo en situación de paro). Consecuentemente, la utilización de este cuadro conceptual sera también válida tanto a nivel de formación profesional inicial como a nivel de formación continua.

Desarrollado sobre la base de una colaboración transnacional entre cuatro Estados miembros (Bélgica, Italia, España y Reino Unido) bajo los auspicios del programa EUROTECNET, ha recibido el nombre de "Principios comunes para la evaluación de los resultados cognitivos de la formación".

El principal objetivo del programa EUROTECNET es promover la innovación en el campo de la formación profesional de base y continua, a fin de considerar los cambios tecnológicos actuales y futuros así como su impacto sobre el empleo, el trabajo y las cualificaciones y competencias necesarias.

El programa EUROTECNET se centra por este motivo en la difusión y la proliferación de nuevas aproximaciones permitiendo obtener un provecho máximo, en términos socioeconómicos, de las nuevas tecnologías.

Una de las medidas comunitarias del programa EUROTECNET es el desarrollo de proyectos transnacionales. Por su naturaleza, los proyectos transnacionales acrecientan en gran medida sus posibilidades de ser inmediatamente pertinentes así como de facilitar la transferencia de un Estado miembro a otro, dando por supuesto que las diversas hipótesis han sido examinadas y experimentadas por todos los colaboradores en el seno de dicho proyecto.

LA EVALUACION EN LA FORMACION

Para esta tarea particular, el programa EUROTECNET dispone de un grupo de colaboradores dinámico, profesional y entusiasta. Han conservado sus modelos profesionales respetando la exigencia que supone orientar los trabajos en base al aprendizaje individual.

Los colaboradores belgas, quienes han asumido la coordinación científica, conjuntamente con los investigadores provenientes de Italia, España y Reino Unido, han demostrado un grado de compromiso elevado en relación a este proyecto, demostrando de manera clara e inequívoca que los Estados miembros son completamente capaces de cooperar de manera productiva en un plan transnacional.

Esta publicación muestra igualmente que el efecto sinérgico de dicha colaboración entre Estados miembros enriquece considerablemente el resultado final más que reducirlo a denominadores comunes de base.

Se trata de una innovación en el dominio de las nuevas tecnologías resultante del progreso tecnológico y en este caso utilizándolas. Hasta aquí la primera fase Aún falta mucho por hacer, pero en el caso presente, podemos mirar hacia el futuro con confianza.

Agosto 1991

Jack Horgan
Coordinador General
Oficina de Asistencia Técnica
EUROTECNET

1. LOS ROLES DE LA EVALUACION EN LA FORMACION

- 1.1. Posibles intenciones de la evaluación**
La evaluación puede servir para,
1. demostrar la eficacia de la formación (PROVING).
 2. mejorar las actividades de formación (IMPROVING).
 3. mejorar los aprendizajes de los participantes (LEARNING).
 4. ayudar a los formadores a comprender mejor los problemas generales (KNOW HOW).

Las intenciones 2,3 y 4 son el centro del proyecto. Para comprender mejor los principios de la evaluación, es necesario descomponer las "etapas" de un proyecto, o más concretamente las operaciones sucesivas del desarrollo del mismo.

1.2. Las cinco etapas de la regulación

Para asegurar su futuro, una sociedad debe formar profesionales competentes.

En principio, una formación se desarrolla en base a cinco grandes etapas. Son aquellas del proceso de regulación, característica de todo fenómeno viviente:

A. En función de las **NECESIDADES** expresadas o sostenidas, se concibe el **PROYECTO**: los responsables fijan los objetivos *generales*.

Ejemplo:
– *formar especialistas en los dominios X, Y y Z.*

B. Se lleva a término la elaboración de Los **PLANES** (desde la planificación del proyecto hasta la identificación y la definición de las tareas): los especialistas (metodólogos, expertos en contenido, etc.) traducen estos objetivos generales en objetivos específicos u operacionales.

Ejemplos:

- *un especialista del nivel 2 será capaz de ...*
- *esta competencia será adquirida en x años.*
- *la responsabilidad de la formación tendrá la calificación ...*
- *las modalidades de acción serán ...*

C. Las **ACCIONES** son ejecutadas: los formadores y los alumnos intervienen.

Ejemplos:

- *los cursos son impartidos, las emisiones de radio difundidas, los libros enviados;*
- *los alumnos leen, utilizan videos, el software, realizan los trabajos.*

D. Se recogen las **OBSERVACIONES**: los evaluadores describen o miden las competencias, los procesos, las actitudes, ...

Ejemplos:

- *los correctores anotan las copias;*
- *los jueces atribuyen los puntos, corrigen los errores;*
- *los cuestionarios recogen las opiniones.*

E. Se realiza la toma de **DECISIONES**: los gestores sugieren o imponen las “retroalimentaciones” (r) o *feedbacks*

Ejemplos:

- *a nivel de las observaciones (r4) (volver a las observaciones o a los instrumentos empleados);*
- *al nivel de las acciones (r3) (continuar / detenerse / recomenzar / cambiar / etc.);*
- *a nivel de planes (r2) (actuación a revisar);*
- *o igualmente a nivel del objetivo (r1) (el proyecto es contestado).*

a. **PROYECTO** de respuesta a las **NECESIDADES**

b. **PLAN** de intervención y de evaluación

c. **ACCION** de ejecución del plan

d. **OBSERVACIONES** (medición)

e. **DECISIONES** de retroalimentación

Cada uno de los papeles correspondientes a cada etapa puede ser asumido por una (o varias) persona(s) según las modalidades variables.

La evaluación - las etapas d (**OBSERVACION**) y e (**DECISION**) - es la resultante de las operaciones a (**PROYECTO**) y b (**PLANES**)!

El presente proyecto desarrolla conceptos e instrumentos esencialmente válidos para la fase d (**OBSERVACION**) y supone un soporte para la fase e (**DECISION**). Siempre considerando que aquí sólo se aborda un aspecto del proceso de regulación, mecanismo claramente más amplio.

1.3. Los niveles adoptados

Las cinco etapas de la regulación se reencuentran a diversos niveles según el promotor del proyecto :

- a. El ALUMNO (quién solicita, por ejemplo una promoción social), se fija un plan (por ejemplo conseguir una formación en dos años), etc. Este proyecto puede ser descompuesto en materias, en patres de materias y sesiones de curso.
- b. El FORMADOR (quién por ejemplo quiere impulsar x alumnos en la competencia y en el tiempo z).
- c. La INSTITUCION DE FORMACION (quién por ejemplo quiere satisfacer las demandas de formación en los plazos y con los formadores de que dispone).
- d. El RESPONSABLE DE LA FORMACION (quién por ejemplo quiere asegurar nuevas calificaciones profesionales en x empleados).

Este proyecto Cooperativo se centra fundamentalmente en los niveles:

- del alumno
- del formador
- de la institución de formación.

y propone conceptos, métodos e instrumentos de evaluación fecundos, es decir

- que mejoran la calificación;
- que profesionalizan formadores y formados.

Estos conceptos podrán servir :

- la evaluación sumativa: ¿se han conseguido las capacidades *al final* del proceso?
- la evaluación formativa: en curso del aprendizaje, verificación del avance del proyecto, de la formación.

2. LA EVALUACION FORMATIVA

2.1. La centración sobre el alumno

Las informaciones deberían ser

- proporcionadas al mismo alumno (primer beneficiario);
- presentadas a tiempo para influir en los aprendizajes posteriores.
- suficientemente diagnosticas para modificar las estrategias de aprendizaje.

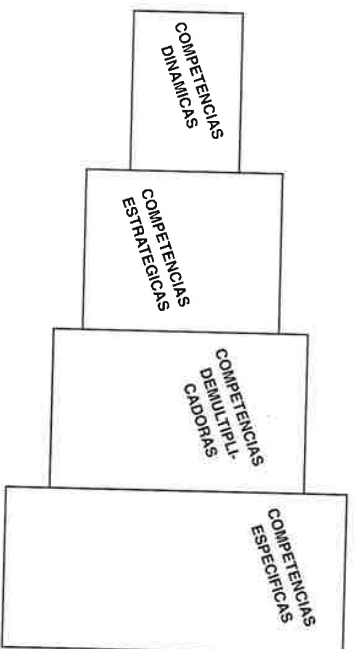
2.2. Los cuatro niveles del aprendizaje

Las informaciones se sustentan sobre cuatro aspectos del individuo:

- sus saberes o sus maneras de saber hacer en los campos específicos;
- sus capacidades demultiplicadoras (sus métodos y técnicas de trabajo y aprendizaje);
- sus recursos estratégicos (especialmente el conocimiento de sí mismo y su relación dinámica con el entorno);
- su motivación, sus actitudes, en particular, respecto al aprendizaje.(Competencias dinámicas)

¿Nuestros sistemas de formación aseguran un equilibrio suficiente entre estos 4 aspectos?. ¿No privilegian los conocimientos en detrimento de las capacidades de autoformación, lo que obliga a reiniciar la formación para los trabajadores cuyos conocimientos se han vuelto "obsoletos"?

Resulta importante proponer conceptos, instrumentos y técnicas, útiles en los campos *específicos* así como estimular las capacidades *demultiplicadoras* o *estratégicas*.

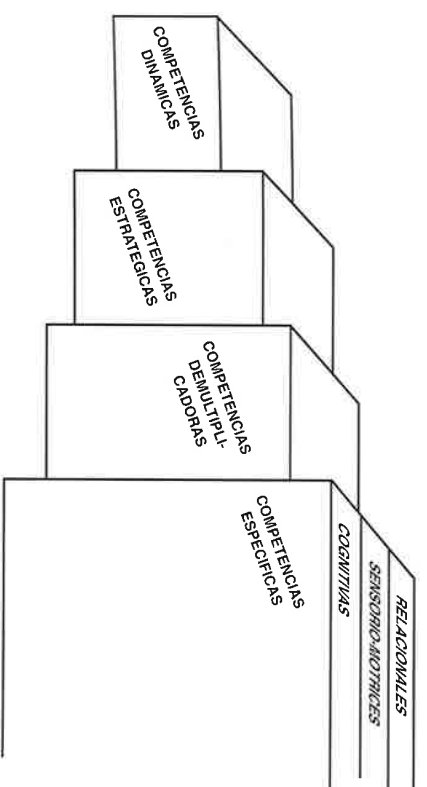


2.3. La optimización de las relaciones formador/alumno

La evaluación formativa no posee un carácter irreversible. El objetivo es mejorar el producto, mediante regulaciones sucesivas. El alumno por consiguiente no está preocupado porque sabe que tendrá la posibilidad de reintentar la prueba en varias ocasiones. Se siente ayudado más que juzgado.

3. LOS LÍMITES DEL PROYECTO

El proyecto CERT no puede cubrir todos los campos de la evaluación. Solamente pretende considerar una pequeña parte de esta última:



Este proyecto cooperativo está centrado en el dominio cognitivo de los alumnos. En el interior de este campo, el grupo ha decidido utilizar los niveles taxonómicos definidos por BLOOM y Otros (1956, 1969).

El objetivo general pretendido por los diferentes equipos es contribuir a aumentar la eficacia general del aprendizaje y de los sistemas de formación **ADOPTANDO modelos, métodos** e instrumentos **COMUNES** de medida de las competencias.

Esta cooperación debería desembocar en los conceptos así como en las **HERRAMIENTAS**

- *Compartidas* por los diferentes miembros de todos los equipos;
- *Flexibles* (es decir modificables en todo momento);
- *Adaptables* a cada cultura específica.

Las producciones actuales consisten en un software y diversos documentos escritos: el "Cuadro Conceptual" (el presente documento) y un manual de utilización (ilustrado por una guía que comprende un ejemplo).

4. CONCEPTOS E INSTRUMENTOS FUNDAMENTALES.

El proyecto desarrolla fundamentalmente cuatro conceptos:

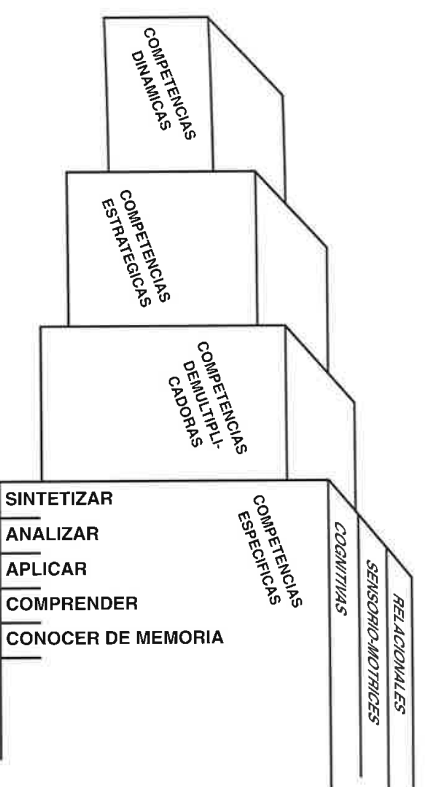
- una taxonomía,
- las Cuestiones de Elección Múltiple (QMC),
- las Soluciones Generales Implícitas (SGI),
- los niveles de confianza.

4.1 Necesidad de una taxonomía de niveles cognitivos.

Las exigencias profesionales se acrecientan. Los formadores (como los empleados) desean asegurarse de que las personas que ellos forman (o que ellos contratan) son capaces:

- de *ejecutar* las operaciones
- así como:
- de *comprender* los problemas, conceptos y principios en profundidad.
 - de *analizar* las situaciones originales;
 - de *producir* nuevas soluciones, dándoles una nueva forma; y aún:
 - de *asegurar las responsabilidades*; es decir, tomar decisiones en situaciones de incertidumbre.

Los formadores y los empleados deben por consiguiente ser capaces de, en un *dominio determinado*, evaluar estos grandes objetivos.



4.2. La utilización de las cuestiones de Elección Múltiple

La evaluación de las capacidades superiores (análisis, síntesis, creatividad...) parece en principio competencia de las cuestiones de respuesta abierta. Sin embargo, las cuestiones de elección múltiple (QCM) bien construidas permiten exploraciones mucho más sutiles de lo que nos parece.

Esta es la razón por la cual se recurre a la vez a cuestiones abiertas y a una nueva forma de QCM: las soluciones generales implícitas (SGI) han supuesto la solución adoptada.

Como las QCM, esta técnica de sollicitación tiene las ventajas:

- de la objetividad de la corrección;
- de la posibilidad de recoger numerosas respuestas en poco tiempo;

Además, permite desarrollar otro currículum, fuera del "currículum oculto" o implícito que condiciona las formaciones.

4.3. Las SGI ante el currículum implícito

Junto a los contenidos y los saberes que el alumno adquiere mediante el estudio, aprende mucho sobre él mismo, sobre el mundo, sobre el aprendizaje en general sin ser consciente de ello. Se habla de "currículum latente u oculto" (hidden curriculum) vehiculado por la escuela, es decir "que nadie enseña, pero que todo el mundo aprende".

¿Cuáles son estas cosas acerca del mundo, sobre él mismo y sobre el aprendizaje que la escuela "destila" y que un buen número de estudiantes aprenden sin saberlo?

Una de ellas es la de no responder, de no tomar iniciativas si no ha sido invitado a ello, si no se ha recibido la orden. En el mundo adulto, cuantas faltas por omisión son de este modo justificadas.

Un buen número de personas aprenden igualmente en la escuela que para toda pregunta existe una y sólo una respuesta. Si ésta no ha sido enseñada "dentro de la materia", se le dispensa de buscarla por sí mismo, tarea que, de todas formas, se es incapaz de realizar.

Una actitud tal es sumamente perjudicial pues resulta evidente que no se puede hoy en día enseñar todas las respuestas que serán necesarias y oportunas para el mañana, cuando el alumno deba "crear" sus propias soluciones a los problemas que no pueden ser más que vislumbrados hoy.

Las SGI apuntan a sistematizar en el alumno la capacidad de

- tratar las cuestiones
- sin respuesta correcta;
- donde todas las proposiciones son correctas;
- detectar:
- las incoherencias;
- las situaciones imposibles de resolver porque faltan datos.

4.4. El juicio y los niveles de confianza

Los programas de educación ponen cada vez más el acento sobre objetivos tales como "aprender a aprender", "comportarse de manera autónoma", "tomar iniciativas", etc. El análisis de estos objetivos generales muestra el juicio (algunos dirían el espíritu crítico) como un aspecto particularmente solicitado.

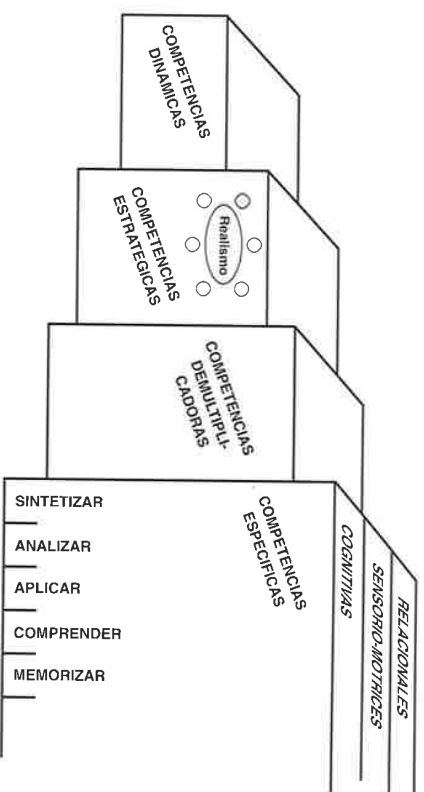
¿Aprender a aprender, no es, ante todo darse cuenta de todo lo que se sabe y de todo lo que se ignora, para después poner en marcha los medios que parecen más apropiados para adquirir el saber y al final hacerse una idea clara (la más objetiva posible) del resultado esperado por el conjunto de la gestión?

Tomar iniciativas comporta, entre otras cosas, una evaluación (forzosamente subjetiva) de las posibilidades de éxito y de fracaso de los proyectos examinados. Esto implica elecciones permanentes entre la reserva y la audacia, entre una actitud receptiva y una participación activa.

Comportarse de una manera autónoma supone que el individuo se remite a su propio juicio, que tiene confianza en sí mismo, se conoce lo suficiente para otorgar a los juicios que los demás hacen sobre él mismo la importancia que se merecen, no ser cerrado ni vulnerable.

Por estas razones, la autoestima es un objetivo importante. Un procedimiento general propio para la consecución de este objetivo en la formación consiste en utilizar los grados de confianza y los índices de realismo en la autoevaluación.

UNA TAXONOMIA



5. EN RESUMEN

Reunidos estos conceptos (taxonomía, QMC, SGI y grados de confianza) permiten suministrar, en una prueba, una visión *a la vez*:

- de los conocimientos y de las capacidades de aplicación;
- de la comprensión y de las capacidades de análisis;
- de las posibilidades de síntesis

de una persona (categorizando las diversas cuestiones de la prueba después de haber suministrado los niveles taxonómicos) así como

- de la capacidad de juicio (por la manera en que el alumno ha empleado los niveles de confianza respondiendo la prueba).

La combinación de estas cuatro opciones debería permitirnos obtener informaciones sobre los diversos niveles cognitivos del alumno. El capítulo siguiente intenta definir claramente los objetivos cognitivos que este proyecto transaccional pretende evaluar.

1. **OBJETIVOS QUE SE PRETENDE CONSEGUIR**
Antes de enfrentarse a cualquier acción, el formador debe interrogarse sobre los objetivos a conseguir, definir lo que va a medir así como los medios que empleará.

Una taxonomía sirve de referencia, de guía, al diseñador de la formación. La importancia atribuida a los diferentes aspectos de un problema permanece subjetiva, relacionada con aquella que el formador o el patrón le atribuye.

2. LA TAXONOMIA ADOPTADA

2.1 La jerarquización de objetivos

Se denomina taxonomía a una clasificación jerarquizada de objetivos. Las taxonomías han sido creadas para ayudar a los formadores a elaborar evaluaciones para cada nivel de aprendizaje y por efecto de retorno, favorecer el desarrollo de las enseñanzas y elevar cada vez más el nivel de la formación.

Globalmente, el éxito obtenido habiendo alcanzado un determinado objetivo situado en un nivel taxonómico dado, requiere haber superado aquellos pretendidos en niveles inferiores.

Por ejemplo, ser capaz de analizar y de diagnosticar una disfunción en un motor implica que se comprende el proceso y que se conocen los elementos que intervienen.

2.2 La taxonomía de Bloom

B.S. BLOOM y Ot. (1956, 1969) editaron una taxonomía de objetivos cognoscitivos. El proyecto CERT la ha seleccionado por su pertinencia y su gran difusión. Esta taxonomía ha sido utilizada por numerosos pedagogos. He aquí los diferentes niveles que comprende:

- Conocer de memoria
- Comprender
- Aplicar
- Analizar
- Sintetizar
- Evaluar

2.3 Debate a propósito de la taxonomía de BLOOM

a. Los niveles 1 al 4

Existe un gran acuerdo sobre estos niveles entre los evaluadores en el campo de la práctica.

b. El nivel 5

Existe alguna discusión sobre el nivel de Síntesis, en el que algunos desean incluir la resolución de problemas.

c. El nivel 6

En este nivel (evaluar o producir juicios cualitativos o cuantitativos) se está lejos de encontrar una acuerdo unánime (cfr. V. y G. de LANDSHEERE, 1976).

Lo mismo sucede en relación a la creatividad (producir algo utilizando o no elementos ya disponibles, creando un orden nuevo a partir del desorden). No existe consenso en base a esta definición este concepto, evidentemente una entre varias posibles. Uno de los trabajos posteriores de los equipos podría consistir en la puesta a punto de definiciones no ambiguas de los diferentes niveles de una taxonomía de los comportamientos cognitivos.

Finalmente, la toma de decisiones, la certeza, el conocimiento del propio conocimiento son elementos importantes en el dominio del aprendizaje y no están incluidos en los cinco niveles de esta taxonomía. CERT se limitará a un aspecto muy concreto: el realismo en la auto-apreciación de las propias competencias.

Debemos señalar que el anexo 1 presenta una discusión en relación a esta taxonomía.

2.4 Algunas definiciones y ejemplos

A continuación, se presenta para cada nivel una definición teórica, sus implicaciones prácticas y algunos ejemplos aportados por los diferentes equipos.

a. Conocer de memoria

Definición: El conocimiento supone la capacidad de reconocer y recurrir a la memoria.

- Datos específicos, términos, hechos tal que sucesos, fechas, personas, lugares;
- Modos de tratar los datos específicos, convenciones, (el lenguaje, usos...)
- tendencias y secuencias (evolución de los fenómenos), clasificaciones, criterios de evaluación de datos particulares, métodos;
- Representaciones abstractas como principios, leyes, teorías y estructuras.

En la práctica: en el texto o cuestionario se pretende medir la capacidad del estudiante de recordar o reconocer, informaciones orales o escritas, ya leídas o escuchadas en una situación idéntica o similar.

Ejemplos:

1. En Carpintería: (Conocimiento de métodos)

Se han colocado andamios en una obra en construcción. ¿Cuándo deben ser controlados a fin de asegurar una plataforma de seguridad?

1. Cuando termina la jornada diaria
2. Cada día, antes de comenzar el trabajo
3. Una vez al mes

Respuesta Correcta = 2

2. En Automoción: (conocimiento de datos)

El Freón es un:

1. Líquido fluido
2. Gas
3. Material combustible
4. Material aislante

Respuesta Correcta = 2

3. En Informática: (conocimiento de datos)

¿Mediante qué comando puede conocerse el tamaño de memoria central disponible?

1. VER
2. SYS
3. CHKDSK

Respuesta Correcta = 3

4. En Economía de base: (conocimiento de hechos)

Si dos sujetos pueden comunicar, si uno de ellos posee alguna cosa interesante para el otro, y ambos pueden aceptar o rechazar la oferta, se trata de

1. Un trueque
2. Un intercambio
3. Una compraventa de bienes
4. Una compraventa de servicios

Respuesta Correcta = 2

b. Comprender

Definición: Se trata de la comprensión al nivel más elemental. Se manifiesta por:

- **La transposición:** o habilidad para reproducir un conocimiento, cambiando su forma (por ejemplo elaborar una metáfora en lenguaje corriente) o transformar enunciados simbólicos bajo forma matemática o verbal.
- **Interpretación:** Es decir, la explicación o el resultado de cierto material.
- **Extrapolación:** Habilidad para explicitar tendencias, implicaciones, consecuencias del fenómeno descrito en el material presentado.

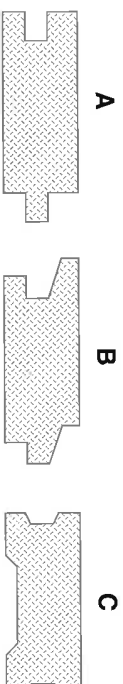
En la práctica: Todos los datos están disponibles. El estudiante debe trasladar, interpretar correctamente.

Ejemplos:

1. En Carpintería

La imagen muestra tres secciones de planchas empleadas para construir un parquet. ¿Cuál sería la más apropiada para un emplazamiento discreto, es decir, uno en que las juntas sean invisibles?

1. A
2. B
3. C



Respuesta Correcta = 2

2. En Automoción

Entre las bombas que se indican, ¿cuál se emplea para aumentar la presión?

1. Bomba de Aire
2. Bomba centrífuga
3. Bomba de pistones
4. Bomba de piñones

Respuesta Correcta = 4

3. En Informática

Un disquete se considera "de arranque" (no provoca mensaje de error salvo por caída de tensión si no se toca el teclado) si contiene al menos

1. El fichero AUTOEXEC.BAT
2. El controlador de teclado
3. El fichero CONFIG.SYS
4. Los ficheros del sistema

Respuesta Correcta = 4

4. En Economía de Base

Por la palabra servicio se entiende

1. un conjunto de bienes que pueden satisfacer necesidades precisas, pero que no están necesariamente a la venta
2. los bienes del sector terciario
3. un bien que no puede ser mostrado concretamente antes de ser alquilado

Respuesta Correcta = Ninguna

c. Aplicar

Definición: Aplicación supone la capacidad de utilizar en **situaciones particulares** y concretas las representaciones abstractas, es decir, las leyes y teorías adquiridas.

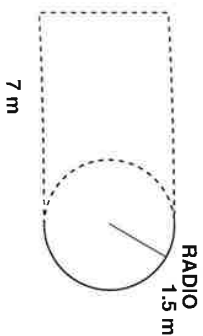
En la práctica: Se han mostrado Las reglas y principios aunque no todos los datos están necesariamente disponibles. Se mide la capacidad del alumno para resolver un

problema mediante la aplicación de principios (que debe haber memorizado previamente).

Ejemplos

1. En Carpintería

He aquí el esquema de un exhibidor circular. Este debe transformarse en una pieza rectangular terminada en una sección semicircular (fíjese en los trazos punteados para la transformación). ¿Cuántos metros cuadrados de plancha son necesarios?



1. 10'5
2. 16'5
3. 21'00

Respuesta Correcta = 2

2. En Automoción

Para perforar la madera deberé utilizar una broca de...

1. Wídia
2. Acero esmaltado
3. Acero a base de carbono
4. Acero inoxidable

Respuesta Correcta = 3

3. En Informática

Entre los identificadores de ficheros siguientes, ¿cuál no es válido?

1. alumno.DOC
2. A+B.CAL
4. TXT.DIR

Respuesta Correcta = 2

d. Analizar

Definición: El análisis comprende la capacidad de buscar elementos, relaciones entre elementos y los principios que los organizan en relación a un material o una situación presentada.

Los autores añaden a esta definición de BLOOM:

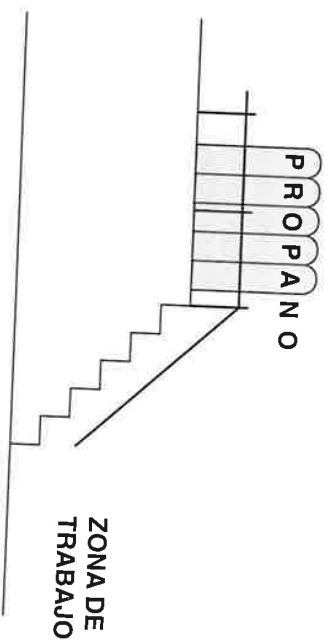
La capacidad de detectar incoherencias, absurdos, falta de datos. Especialmente en situaciones novedosas o complejas.

En la práctica: Se mide la capacidad del estudiante de identificar estructuras, examinar las relaciones que se establecen (pertinentes) entre los elementos que se problema planteado; verificar, establecer un diagnóstico; descubrir las incoherencias;... Es necesario aportar conocimientos externos.

Ejemplos:

1. En Carpintería

Varias bombonas de gas propano han sido aseguradas y sujetadas al muro adosado, sobre unas pequeñas escaleras. En la parte inferior de estas se encuentra una pequeña pieza en la que trabaja un soldador. Esta situación se ilustra en el esquema siguiente. ¿Por qué es peligrosa?



30

1. Las bombonas de gas pueden caer fácilmente por las escaleras
2. El gas propano es más ligero que el aire
3. Las bombonas de gas obstruyen el paso.

Respuesta Correcta = 2

2. En Informática

¿Qué comando no dará el resultado esperado sin ser por ello ilícito?

1. DIR?.COM
2. DIR/P/W
3. DIR *E.ESE
4. DIR DIR * *

Respuesta Correcta = 3

3. En Economía de Base

¿Qué se entiende por la palabra bienes?

1. Un producto que es el resultado de un proceso de transformación industrial, destinado al consumo
2. La materia prima destinada a ser trabajada
3. Un producto concreto, capaz de satisfacer exigencias específicas.

Respuesta Correcta = 3

e. Sintetizar

Definición: Supone la capacidad de organizar elementos (o partes de elementos) con la finalidad de componer un todo nuevo. Puede ser:

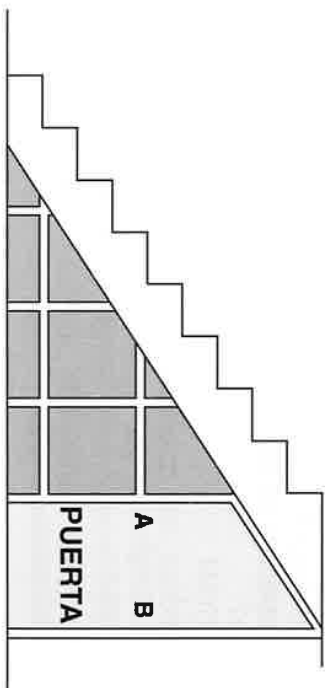
- Elaboración de un trabajo personal, (por ejemplo un poema, extraído de un fragmento de prosa presentado);
- Elaboración de un plan, de una proposición de operaciones organizadas (por ejemplo, proposición de métodos de verificación);
- Derivación de una serie de relaciones abstraídas (por ejemplo, la capacidad de establecer generalizaciones)

31

En la práctica: se mide la capacidad del estudiante de crear un orden a partir de un desorden. Organizar, formular, crear un plan, etc. estando disponibles todos los elementos.

Ejemplos:

1. En Carpintería



Debe construirse una puerta bajo la escalera del dibujo. Esta puerta debe abrir hacia el interior, bajo la escalera.

¿Cómo recomendarías fijarla?

Mediante una bisagra

1. espiral en A

2. normal en A

3. normal en B

Respuesta Correcta = 3

2. En Automoción

Una máquina utiliza un motor de 220 v. Sin embargo, la única fuente de energía eléctrica disponible es de 380 v.

¿Cómo poner en funcionamiento esta máquina?

1. No puede hacerse

2. Comprando una máquina de 380 v.

3. Conectarla mediante un transformador

4. Conectándola sobre dos fase

Respuesta Correcta = 3

2.5 ¿Cómo determinar el nivel taxonómico de una cuestión?

¿Cómo saber si una cuestión determinada mide la memoria o la analiza?

Todo depende de lo que se haya enseñado o de la experiencia anterior del sujeto. Si la solución a esta cuestión ya ha sido proporcionada, se está midiendo la memoria. Si se ha dado la respuesta a una pregunta similar, lo que se mide es sobre todo su capacidad de aplicar. Si ninguna respuesta ni problema de este tipo ha sido resuelto por o para el estudiante anteriormente, la tarea solicitada se sitúa al nivel de análisis o síntesis. En efecto, en este último caso deben concebirse soluciones personales.

El formador tan sólo puede formular hipótesis sobre los procesos mentales que intervienen y sobre las causas de los errores. La clasificación taxonómica deberá ser confirmada por diversos medios, por ejemplo, mediante una correlación entre las respuestas a diversas cuestiones y una entrevista.

2.6 Estrategias para la concepción de preguntas

El método *Top Down* consiste, para cada objetivo operacional, en determinar al principio el nivel o niveles taxonómicos exigidos, es decir, lo que debe ser conocido, comprendido, aplicado, analizado, sintetizado, ... Y posteriormente producir las cuestiones necesarias para cada uno de estos niveles de objetivos operacionales.

En esta perspectiva, se puede *fraccionar* la presentación de preguntas de modo que se aísle la medición de diferentes procesos mentales.

Así se presentará al principio todas las cuestiones de memoria, a continuación se señalarán las respuestas correctas **ANTES** de presentar las preguntas de comprensión, y así sucesivamente...

El método Bottom-Up consiste en señalar las situaciones-problemas que el estudiante debe ser capaz de resolver. Se crean cuestiones relevantes que, inevitablemente implican simultáneamente diferentes niveles taxonómicos.

3. EN RESUMEN

Se han conservado cinco niveles de la taxonomía de BLOOM. El conocimiento, la comprensión, la aplicación, el análisis y la síntesis.

En este proyecto CERT, sólo las aptitudes cognoscitivas han sido tomadas en cuenta; otras como el auto-aprendizaje y las aptitudes psicomotrices se retomarán posteriormente.

Los ejemplos de preguntas propuestas anteriormente cubren esos cinco niveles cognoscitivos.

EL CUESTIONARIO Y LAS QCM

1. EL CUESTIONARIO

En general, se puede solicitar la respuesta de un estudiante con ayuda de dos *tipos de cuestiones*:

- *abiertas*
En las que disfruta de una gran libertad de expresión, tanto en el contenido como en la forma de la respuesta;
- *cerradas*
En las que el número de posibles respuestas ofrecidas al estudiante es limitado.

Las respuestas a las preguntas cerradas están precodificadas: el estudiante selecciona números o letras correspondientes a las diferentes soluciones planteadas.

Las respuestas a las preguntas abiertas son postcodificadas.

Las QCM ("Preguntas de elección múltiple") parecen muy directas y poco matizadas. El recurso a este modo de cuestionario se explica fundamentalmente por la facilidad de corrección que implica, aunque ésta no es sino una ventaja entre otras. Las páginas que siguen tratan de demostrar que los cuestionarios cerrados pueden proporcionar una evaluación afinada de los conocimientos y de las estrategias del individuo.

El proyecto acepta los dos tipos de preguntas, abiertas y cerradas, y permite que sean utilizadas en una misma prueba. Es necesario recordar que las QCM no son sino una de las formas de evaluar, al lado de la observación directa, la entrevista, el análisis de producciones o de actividades complejas.

2. LAS CARACTERÍSTICAS DE LAS QCM

2.1 Definición

Llamamos "Pregunta de elección múltiple" (QCM, "Question à choix Multiples") a una cuestión en la que

- a. el estudiante debe escoger entre varias soluciones propuestas;
- b. las respuestas propuestas son correctas o incorrectas, independientemente de la persona que responde.

Así pues, *no son* QCM

- los cuestionarios de identificación

ejemplo:

- D está,*
- soltero
 - casado
 - viudo
 - divorciado

- los cuestionarios de opinión

ejemplo:

¿Está de acuerdo con las matanzas de focas?

Sí

No

- las escalas de evaluación

ejemplo

- Este trabajo es*
- excelente
 - bueno
 - regular
 - malo

2.2 Los componentes de una QCM

Los tres componentes de una QCM son:

- a. la consigna
- b. el cuerpo
- c. las soluciones propuestas

a. La consigna

La consigna se da al sujeto al comienzo de la prueba. Siempre deberá estar escrita. En general, es válida para todas las preguntas de un mismo cuestionario.

La consigna describe:

- 1. la cuestión

ejemplo:
Para las cuestiones que siguen, una única solución es correcta.

- 2. el modo de respuesta

ejemplo:
Escoger una única solución de entre las presentadas o dejar la respuesta en blanco.

- 3. las normas de notación

ejemplo:
Las soluciones correctas suman puntos, los fallos restan (Le interesa abstenerse de responder antes que hacerlo por azar).

4. el baremo de puntuaciones (También de-

nombrados notas, puntuaciones a ganar o a perder), debe precificar las notas obtenidas en cada caso de respuesta correcta, incorrecta o en caso de omisión.

b. El cuerpo

El cuerpo, también llamado tronco, base, artículo de base, premisa,

1. define el problema

ejemplo:
Se quiere diluir el ácido sulfúrico concentrado.

2. propone la pregunta

ejemplo:
¿Cuál de las siguientes acciones es más segura?

c. Las respuestas propuestas

Las soluciones propuestas, también llamadas opciones, sugerencias, etc., comprenden la solución correcta y las soluciones incorrectas, llamadas distractores.

Tradicionalmente se ha aconsejado durante años formular las QCM con un mínimo de 4 ó 5 opciones a fin de disminuir las oportunidades de emitir una respuesta al azar.

Puesto que este proyecto transnacional añade una componente nuevo a las QCM clásicas, las soluciones generales implícitas (ver capítulo IV), el número de soluciones propuestas puede ser cualquier (2, 3, 4, ó 5).

3. VENTAJAS Y DESVENTAJAS DE LAS QCM

3.1 Ventajas cuantitativas de las QCM

Las QCM permiten, en un tiempo breve obtener del estudiante un **gran número** de respuestas, y así

- verificar su competencia sobre numerosos contenidos.
- adquirir medias representativas válidas
- fijar las competencias haciendo intervenir diferentes procesos mentales.

Esto es consecuencia de las conocidas ventajas de las QCM:

- a. la **pregunta es clara** (lo que facilita la comprensión)
- b. la respuesta es **breve** (lo que evita al estudiante una formulación laboriosa)
- c. la corrección es **sencilla** (lo que evita al profesor tomas de decisión difíciles).

3.2 Las ventajas cualitativas de las QCM

Una QCM se caracteriza por:

a. la posibilidad de evaluar sistemáticamente niveles superiores de actividad mental.

La vigilancia cognitiva, la lógica, así como la comprensión profunda de principios y de conceptos pueden ser evaluadas *sistemáticamente* por las QCM.

b. la precisión de exigencias.

Como señalan Noizet y Caverni (1978), una QCM "no permite ser esquivada", a diferencia de las pruebas de producción (disertación, examen oral) que dejan siempre al candidato la posibilidad de disimular una ausencia de conocimiento. En general el candidato habla de lo que sabe y evita lo que conoce peor.

c. La objetividad de la corrección.

Una misma actuación del estudiante recibirá siempre la misma nota sea cual sea el corrector. Esta ventaja favorece la autorregulación. Los estudiantes tienen de esta manera un medio de evaluarse rigurosamente ellos mismos.

d. La concisión del código permite los análisis sinópticos y la autorregulación.

Las respuestas pueden ser tratadas mediante sistemas informáticos. Los formadores al igual que los estudiantes pueden recibir resultados muy detallados sobre diversos aspectos de la enseñanza que imparten (o de su aprendizaje).

3.3 El precio de las ventajas ofrecidas por las QCM

a. El trabajo de puesta punto de una QCM.

La construcción de una QCM requiere una real competencia sobre el fondo y una rica experiencia pedagógica. Esta construcción se debe hacer según una serie de reglas "de redacción" (LECLERQ, 1986).

Todo constructor de QCM puede verse reprochado por sus colegas por haber considerado como falsa una respuesta verdadera, o por haber construido una QCM en cuya respuesta correcta no haya unanimidad: a éstas se denominan preguntas indecisas.

Se comprende que uno de los problemas del constructor de QCM sea ponerse a cubierto de ítems indecisos. Esto tiene como consecuencia indirecta que los QCM hacen a menudo una llamada a la memoria mucho más que a la lógica. Sin embargo, la variedad de consignas posibles permite dejar atrás este problema. Por otra parte, el peligro del recurso a soluciones de facilidad disminuye a medida que los profesores enseñan las técnicas de sollicitación de la respuesta.

b. La imposibilidad de una presentación oral.

Una QCM comporta un cuerpo denso (cada palabra cuenta) y soluciones propuestas que difieren solo por uno o dos elementos. Reconsultar el todo al momento de responder es a menudo muy útil. El escrito lo permite, el oral no.

c. La tentación de realizar preguntas de detalle.

La respuesta correcta a una QCM debe conseguir la unanimidad de los expertos sobre el tema, lo cual es fácil cuando la QCM comporta detalles y entonces tiende a ponerse al abrigo de la crítica.

d. La facilidad de copiar las respuestas del vecino.

Se han practicado procedimientos de entredo del orden de las preguntas. Se puede solicitar al alumno que argumente su respuesta, que la justifique.

e. La posibilidad ofrecida a los estudiantes de presentar respuestas correctas respondiéndolo al azar.

Numerosos *procedimientos de corrección* (la famosa "corrección por adivinación" han sido utilizados. Existe una literatura abundante (sobre todo en inglés) en relación a esta cuestión. Este problema ha sido completamente superado por el recurso a los niveles de confianza (ver capítulo V).

f. Cuatro desventajas fundamentales de las QCM.

Las desventajas que acaban de ser expuestas son secundarias o evitables. No ocurre lo mismo con las siguientes:

1. *La impotencia de las QCM a la hora de medir cierto tipo de actuaciones.*

Es el caso, entre otros, de la expresión espontánea, de la aptitud de redactar y expresar su pensamiento, de la invención de soluciones nuevas e incluso la reproducción de la memoria sin soporte.

Las QCM son un instrumento entre otros: resulta importante recurrir al modo de evaluación más adecuado a cada situación. Las QCM resultan menos convenientes para las actuaciones complejas (respuestas largas donde la estructura y la expresión juegan un gran papel) que para las actuaciones aisladas.

2. *La presentación de soluciones erróneas a los estudiantes.*

SKINNER (1961) emitió la hipótesis de que un día el alumno se acordaría mejor de la respuesta incorrecta encontrada en un QCM que de la respuesta correcta.

Esta hipótesis ha sido verificada por PRESTON (1965), pero KARRAKER (1967) ha demostrado que este riesgo no existe si se revela a los alumnos las respuestas correctas a las preguntas. Se puede, sin embargo, temer que en ciertos campos, los errores se fijen en su memoria. Está, por ejemplo, contraindicado presentar las preguntas de ortografía de uso sobre la forma de las QCM. Pero otros terrenos no presentan la misma vulnerabilidad porque hacen más llamada al razonamiento (ejemplo: ortografía gramatical)

3. *La incapacidad de las QCM para entrenar en la formulación de respuestas.*

Es poco probable que la utilización de las QCM deteriore la capacidad del alumno para redactar su respuesta. Sin embargo, se puede objetar que toda pregunta debería ofrecer una ocasión para formular su pensamiento. Ahora bien, el recurso de las QCM no lo ofrece. En niveles cognitivos donde este objetivo es importante, se debe administrar al alumno un número suficiente de oportunidades para formular su pensamiento.

4. *La Contracción del campo cognitivo*

Se podría temer que las QCM contribuyan a simplificar la "visión del mundo" de los estudiantes, a reducirla a un esquema, y como consecuencia a uniformar a los individuos. Introduciendo las soluciones generales implícitas (ver capítulo IV), se hace de la QCM un instrumento de medida y de entrenamiento de la vigilancia cognitiva y de la relación entre las competencias y los riesgos sociales.

4. **UN EJEMPLO**

Se ha decidido ofrecer un ejemplo completo en relación a cada nuevo concepto presentado en las páginas siguientes. Para facilitar la lectura, éste, en cada ocasión estará cada vez rodeado de un cuadro como muestra este primer ejemplo.

LAS SOLUCIONES GENERALES IMPLICITAS

1. LA FINALIDAD ULTIMA DE LAS S.G.I.:

Más allá de las consideraciones en relación a la utilidad de las S.G.I. en la evaluación de los niveles elevados de procesos mentales y a fin de controlar una evaluación reducida a aquello que es explícitamente enseñado, analizamos aquí la aportación fundamental de tal sistema: entrenar y evaluar la vigilancia cognitiva.

Como ya hemos anunciado en el capítulo II, el alumno de hoy debe aprender a *producir* sus soluciones personales a los problemas que se le presentarán en su trabajo cotidiano y que nosotros podemos preveer actualmente.

La puesta en evidencia de las actitudes y del curriculum latente dependen en gran medida de las modalidades habituales de cuestionarse.

El comportamiento del alumno puede ser impulsado según diversos modos:

- El formador solicita *explícitamente* una respuesta.
- El formador solicita *implícitamente* una respuesta.
- El alumno emite *espontáneamente* un comportamiento.

ahora bien, el alumno tan sólo se comporta espontáneamente cuando se muestra vigilante, en otros términos si proporciona muestras de espíritu crítico, si se pregunta sobre aquello que oye o que ve, definimos esta competencia como "vigilancia cognitiva".

He aquí un formulario destinado a recibir las respuestas de un alumno a 20 preguntas de una prueba.

Preguntas 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Respuestas

5. EN RESUMEN

Hay un cierto número de ventajas al utilizar las QCM, particularmente la exactitud y la precisión. Incluso, las QCM pueden medir eficazmente los cinco niveles de la taxonomía y son cómodas a la hora de corregir. En el contexto de este proyecto cooperativo, el inconveniente más importante es el tiempo que se necesita para construir un banco de preguntas.

Si el objetivo último del educador es obtener un comportamiento espontáneo, las demandas tanto explícitas como implícitas son un medio para alcanzar este objetivo.

2. LA SOLICITACION IMPLICITA EN LAS QCM

La enseñanza tradicional procede por sollicitación explícita, mientras que la vida exigirá actores sociales de respuestas espontáneas.

Para paliar esta carencia, ha parecido útil desarrollar un sistema de **sollicitación implícita**, donde el formador presenta una "situación de interrogación" al alumno pero sin anunciársela como tal. Por ejemplo, el formador cometerá voluntariamente un error de razonamiento, esperando que los alumnos lo detecten. Se tiene noticia de las experiencias del Doctor FOX en los Estados Unidos, las cuales consisten en cometer el mayor número posible de estupideces en un discurso aparentemente serio antes de que un miembro asistente lo remarque.

¿Cómo sistematizar el principio de la sollicitación implícita? ¿Cómo situar por igual a todos los alumnos de un grupo ante tales sollicitaciones, de manera que se pueda evitar que uno de ellos, al descubrirla, impida a los otros ejercerla?

Es posible recurrir a las QCM, pero realizando algunas transformaciones, por ejemplo, introduciendo absurdos en el enunciado, no proporcionando suficiente información, proponiendo preguntas en las que todas las soluciones sean correctas o otras en las que ninguna lo sean. Claro está, se advertirá a los alumnos de estas eventualidades.

¿Es necesario imprimir las soluciones NINGUNA, TODAS, CARENCIA e INCOHERENCIA en la pregunta? no, porque hacerlas figurar entre las soluciones propuestas, es alertar al alumno. Para estas soluciones se quiere medir la *vigilancia cognitiva* del alumno.

Las soluciones NINGUNA (6) y TODAS (7), llevan la vigilancia cognitiva hacia las proposiciones de respuestas, mientras que las soluciones CARENCIA (8) e INCOHERENCIA (9) hacen llevar la vigilancia cognitiva sobre el enunciado (o introducción) de la pregunta. En estos casos (8 y 9), el alumno no debe leer incluso las soluciones propuestas.

En base a este principio se ha desarrollado la técnica titulada "las soluciones generales de preguntas de elección múltiple" (LECLERCQ, 1986). Se trata de comunicar al inicio de la prueba a todos los alumnos que ciertas soluciones existen, que son válidas para todas las QCM que serán presentadas, pero que estas soluciones no estarán impresas. Sobre esta base se pueden considerar otras soluciones generales.

3. LAS CONSIGNAS QUE CONCIERNEN A LAS SOLUCIONES GENERALES

He aquí un ejemplo de consigna presentando las soluciones generales:

Las preguntas que se presentan a continuación son de elección múltiple. Cada pregunta comporta una única solución correcta. Sin embargo ciertas soluciones (llamadas soluciones generales) requieren una cierta atención; estas preguntas se proponen bajo los códigos siguientes: 6-7-8 o 9 y no serán impresas en las QCM, aunque ellas pueden aplicarse a todas las cuestiones.

- 6 o RECHAZO = NINGUNA de las soluciones propuestas es correcta.
- 7 o TODAS = TODAS las soluciones propuestas son correctas.
- 8 o CARENCIA = Es imposible responder porque la información presentada en el enunciado de la pregunta no es completa (falta como mínimo un dato).
- 9 o INCOHERENCIA = Una incoherencia en el enunciado hace que la pregunta no tenga sentido (por ejemplo, CONTRADICCIONES en el enunciado).

¡Atención! con la finalidad de hacer que el alumno lea atentamente los enunciados, la respuesta 9 tiene prioridad sobre las soluciones 6, 7 y 8 y, evidentemente sobre las respuestas 1, 2 y 3, impresas, YA que si existe una incoherencia en el enunciado todas las soluciones propuestas son, evidentemente, no pertinentes.

Ejemplos:

Debe responderse

La capital de Francia es	1. Lille 2. Lyon 3. Paris	3
La capital de Italia es	1. Berlín 2. Praga 3. Tokio	6
Gran Bretaña comprende	1. Inglaterra 2. Escocia 3. País de Gales	7
¿Qué edad tenía Rimbaud?	1. 2 años 2. 10 años 3. 20 años	8
¿En qué año Julio Cesar encontró a Napoleón?	1. 1850 2. 1915 3. 1945	9

4. LAS SOLUCIONES GENERALES Y LOS PROBLEMAS MENTALES

Aunque aún no hayamos podido verificarlo experimentalmente, estamos convencidos de que esta técnica puede aclarar los problemas propios de las taxonomías.

El **RECHAZO** (6) obliga al estudiante a responder mentalmente a la QCM antes de examinar las soluciones propuestas, en segundo lugar a buscar su respuesta entre dichas soluciones y de esta forma evita parcialmente "el reconocimiento" de la respuesta esperada.

La solución **TODAS** (7) permite que el estudiante reflexione sobre la posibilidad de que hayan varias respuestas correctas para un mismo problema.

Partimos de la hipótesis de que la **CARENCIA** (8) de datos permite evaluarlo todo, particularmente los niveles taxonómicos de análisis y de aplicación (según BLOOM).

Ocurre lo mismo con la **INCOHERENCIA** (9). Creemos que es una buena manera de medir la comprensión (en el sentido que BLOOM da a este término en su taxonomía de objetivos cognitivos).

La posibilidad de ofrecer la respuesta correcta por una elección al azar es mínima. En efecto, si la QCM presenta K soluciones, existe automáticamente K+4 soluciones posibles.

Resumiendo, se evita el fraccionamiento de los conocimientos y se acentúa la medida de la atención cognitiva.

5. COMO ENTRENAR AL ALUMNO EN EL USO DE QCM Y SGI

5.1. La metodología

Antes de utilizar las preguntas de soluciones generales, es indispensable sensibilizar a los alumnos en la filosofía de este tipo de evaluación y en los objetivos que se pretenden con su utilización.

Los estudiantes pueden ceñirse a este género de consigna compleja, no se encuentran cómodos hasta haber realizado uno o dos ejercicios. Debe excluirse la presentación de todas las soluciones generales de una sola vez hasta que los estudiantes no estén suficientemente entrenados. Se recomienda proceder de forma progresiva (por ejemplo, introducir la solución general de RECHAZO en un primer test, añadir TODAS en un segundo, etc.).

Igualmente es necesario ejercitar a los estudiantes sobre los contenidos generales y sencillos.

5.2 Algunos ejemplos

Presentamos a continuación 5 ejemplos relativos al cálculo del área y del perímetro de los triángulos.

Geometría

Pregunta 1: Un triángulo isósceles tiene una base de 10 cm, ¿Cuál es su perímetro?

1. 28 cm 2. 30 cm 3. 32 cm.

Pregunta 2: Un triángulo equilátero tiene un lado de 10 cm, ¿Cuál es su perímetro?

1. 28 cm. 2. 30 cm. 3. 32 cm

Pregunta 3:

Un triángulo isósceles tiene dos lados de 5 un lado de 6 m. ¿Cuál es la superficie (en m y m. cuadrados)?

1. 8 m² 2. 10 m² 3. 30 m²

Pregunta 4:

En un triángulo equilátero, ¿qué línea recorta de cada vértice puede alcanzar el centro del lado opuesto?

1. la altura 2. la mediatriz 3. la bisectriz

Pregunta 5:

Los tres lados de un triángulo miden respectivamente 6 m, 2m y 3m. ¿Cuál es su perímetro?

1. 11 m. 2. 12 m. 3. 15 m. 4. 18 m.

Las respuestas correctas (RC) son las siguientes:

Pregunta 1:

RC = 8 (Carencia de datos: en efecto, no se conoce la longitud de los otros lados).

Pregunta 2:

RC = 2 (simplemente)

Pregunta 3:

RC = 6 (Ninguna es correcta, porque la superficie es 12 m² por la aplicación del teorema de Pitágoras la altura es 4 m.

Pregunta 4:

RC = 7 (Todas son correctas).

Pregunta 5:

RC = 9 (Incoherencia, ya que es imposible obtener un triángulo. Según una propiedad fundamental del triángulo la medida de un lado no puede ser superior a la suma de las medidas de los otros dos lados).

5.3. Un ejemplo de respuestas

Presentamos un ejemplo de respuestas proporcionadas por un estudiante a las 20 preguntas de un test:

Q. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

R. 5 3 2 1 1 3 4 7 6 1 9 8 8 9 1 2 5 3 6 7

Las respuestas correctas (esperadas por el formador) son las siguientes:

Q. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

R. 5 1 2 1 6 3 7 7 6 3 5 9 8 9 1 2 6 4 6 8

Se puede observar que el número de respuestas correctas es 11/20, siendo la tasa media de éxito de un 55 %.

6. EN RESUMEN

Las potencialidades de utilización de las QCM son reforzadas por la adición de posibles respuestas suplementarias como NINGUNA, TODAS, CARENCIA de información o INCOHERENCIA.

Como consecuencia de esta modificación, los niveles taxonómicos (desde el reconocimiento de memoria hasta la síntesis) pueden ser evaluados con una mayor eficacia.

LOS NIVELES DE CONFIANZA

1. LAS POSTURAS SOCIALES

1.1. Los ignorantes no son peligrosos

Cada uno de nosotros ignora múltiples hechos. La vida social está organizada según una "repartición" de saberes unidos a profesiones especializadas: los juristas, cirujanos, químicos, meteorólogos, etc. El sentido común nos lleva a no actuar en terrenos desconocidos.

Podemos llegar a ser peligrosos, para nosotros mismos y para nuestro entorno, cuando no confesamos nuestra incompetencia (en el pilotaje de un avión, en el conocimiento de una lengua, en el efecto de un medicamento, en la hora de salida de un tren...). Si alguien se dirige a nosotros para obtener una información es fácil y eficaz reconocer de inmediato nuestra ignorancia. Nuestro interlocutor se dirigirá hacia otra persona.

Saber que no se sabe, o hasta qué punto se sabe... es lo máspreciado.

Bertrand RUSSEL decía que "el problema, en este mundo, es que los idiotas están seguros de ellos mismos y la gente común llena de dudas".

1.2. El conocimiento de uno mismo

Para un alumno, una percepción (subjetiva) correcta de la calidad (objetiva) de sus respuestas participa en la

construcción de lo que G. DE LANDSHEERE llama "una imagen de sí mismo realista, completa y estable". La aceptación de sí mismo no tiene nada en común con el permanecer inactivo y carecer de motivación. Ser consciente de sus puntos débiles no implica la construcción de una imagen negativa de sí mismo; una característica evaluada negativamente puede ser aceptada en el cuadro de una personalidad, que en el conjunto llega a evaluarse de forma positiva.

Resumiendo, lo que tiene mayor importancia es saber aquello que no se sabe.

1.3 La apertura a la crítica

Aceptar la crítica sin perder confianza en uno mismo, sin sentirse afectada nuestra personalidad pero al mismo tiempo, sin mostrar indiferencia, es un objetivo fundamental para el futuro del alumno.

1.4. La Tolerancia al fracaso

Soportar el fracaso y la frustración sobreponiéndose a la dificultad es importante.

En esta perspectiva, las instituciones de formación deben impulsar las condiciones "objetivas" que permitan una evaluación de uno mismo, referencia más estable para el individuo que los juicios ajenos o que sus propios juicios.

2. La medida del conocimiento

El empleo de las QCM expone a la "adivinanza". El evaluador busca que la esperanza matemática de una respuesta al azar sea nula.

El procedimiento siguiente:

- El estudiante proporciona, además de su respuesta, un nivel de confianza (sobre una escala de probabilidades de exactitud).
- permite al evaluador anotar esta respuesta con la ayuda de un baremo de puntuaciones teniendo en cuenta, *al mismo tiempo*, la calidad (correcta/no correcta) de la respuesta y el nivel de confianza del evaluado.

Este procedimiento presenta las siguientes ventajas:

- 2.1. Se basa en un modelo teórico evolucionado (CHOPPIN, 1970, LECLERCQ, 1986), **conforme** a la teoría de decisiones y a las probabilidades subjetivas.
- 2.2. Es justo, ya que recompensa a cada uno en función del grado (sutíl) de su competencia y realismo.
- 2.3. Es **informativo**, ya que facilita al profesor (y al estudiante) un dato nuevo: su realismo (¿se sobrestima? ¿se subestima?).
- 2.4. Es **formativa**, ya que hace que el alumno viva la gestión de su duda, de su conocimiento particular, elementos que acompañan a todo proceso de aprendizaje.
- 2.5. Es **general**, porque se aplica tanto a las preguntas abiertas como a las QCM (por las cuales se resuelve el problema de la "adivinanza").

3. LAS CONDICIONES METODOLÓGICAS DE UTILIZACIÓN

La utilización de los grados de confianza presenta perspectivas de búsqueda y de desarrollo muy importantes

sobre la capacidad de los individuos de sacar partido de las informaciones, de utilizar técnicas... en resumen, de aprender.

Este proceso ha sido objeto de una serie de trabajos en los Estados Unidos (entre 1960 y 1975), los cuales globalmente poseen en común diversos fracasos, ya que los investigadores no se habían centrado suficientemente en las condiciones metodológicas de utilización.

Cinco de ellas, nos parecen importantes:

- una consigna presentando una escala métrica
- una baremación conforme a la teoría de decisiones
- una distinción entre medida y recompensa
- gráficos de realismo
- índices numéricos de comportamiento de autoevaluación.

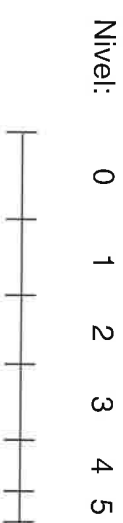
Estas condiciones serán presentadas en breve. Las informaciones técnicas profundizadas se encontrarán en LECLERCQ (1990).

3.1. Una consigna presentando una escala métrica

La mayor parte de las investigaciones sobre los niveles de confianza se han conformado mediante consignas confusas. Se solicita al estudiante que indique, además de su respuesta, si estaba "poco seguro, mínimamente seguro, o muy seguro". Estas expresiones vagas no autorizan sino los tratamientos ordinarios intra-sujetos, ya que nada garantiza que una misma persona guarde las mismas "referencias" (a la fuerza subjetivas y por otra parte desconocidas) a lo largo de un mismo test. Por lo que hace referencia a comparaciones inter-sujetos eran totalmente excluidas.

La solución, ya preconizada por SHUFFORD, ALBERT y MASSENGILL (1966) consiste en establecer, sobre el eje de probabilidades, zonas de confianza simétricas.

El proyecto CERT recomienda utilizar la repartición no simétrica de las zonas de probabilidad.



predicción: 0 25 50 70 85 95 100%
0
posibilidad estimada con anterioridad de acertar

Esta repartición asimétrica permite a los alumnos presentar respuestas más matizadas en la extremidad superior de la escala, siendo esta la zona de probabilidades elevadas. De esta forma, el estudiante es capaz de diferenciar entre 95% y 99% en relación al nivel de confianza. En valor absoluto, esta diferencia no es más que el 4%. Sin embargo, para el estudiante, dicha diferencia es identificable: en el primer caso (95%), tiene una oportunidad sobre 20 de equivocarse, es decir, cinco veces menos. La misma persona será incapaz de diferenciar entre dos niveles de confianza distantes un 4% en la mitad de la escala (ex: 43% y 47%).

Mostramos a continuación los niveles de confianza presentados por el estudiante además de sus respuestas:

Q. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

R. 5 3 2 1 1 3 4 7 6 1 9 8 8 9 1 2 5 3 6 7

C. 1 5 5 5 1 5 3 3 3 3 1 5 3 5 5 1 1 3 5

Se constata que los grados de confianza 0, 2 y 4 no han sido utilizados en ningún momento por el estudiante en estas pruebas.

El grado de confianza 1 ha sido utilizado 5 veces, de las cuales 1 ha sido correcta, siendo la tasa de exactitud de 20%.

El grado de confianza 3 ha sido utilizado 7 veces, de las cuales 3 han sido correctas, siendo la tasa de exactitud de un 42%.

El grado de confianza 5 ha sido utilizado 8 veces, de las cuales 6 han sido correctas, siendo la tasa de exactitud de un 75%.

Nota: Las respuestas correctas se proporcionan en el punto 5.3 del capítulo IV.

3.2. Baremos conformes a la teoría de las decisiones

La puntuación, donde la atribución de puntos debería hacerse según dos criterios: la exactitud (la respuesta es correcta o no) y el grado de confianza (débil o elevado).

Como en la realidad, el grado de confianza debe estar asociado a los riesgos que se acepten tomar y la persona debe estar expuesta a las consecuencias, buenas o malas, de sus actos.

Una respuesta correcta acompañada de un grado de confianza elevado debe ser gratificada con una puntuación superior a la misma respuesta con un grado de confianza menor. Por el contrario, en caso de respuesta incorrecta, la penalidad asignada a una confianza elevada debe ser superior a la asignada a una confianza menor.

Respetar este principio no es suficiente. Los puntos exactos deben ser calculados de forma que el alumno tenga interés en decir la verdad, es decir, expresar su íntima convicción (su confianza) sin deformarla.

El proyecto CERT propone el baremo siguiente:

	puntuación en caso de respuesta	
	Correcto	Incorrecto
entre 0 y 25%	13	4
entre 25 y 50%	16	3
entre 50 y 70%	17	2
entre 70 y 85%	18	0
entre 85 y 95%	19	-10
entre 95 y 100%	20	-20

3.3 Distinción entre medida y recompensa

La nueva puntuación no es una medida, sino la *combinación* de dos medidas:

- la medida de la competencia (el número de respuestas correctas)
- la medida de realismo (de la autoevaluación del alumno).

Se han elaborado diversos índices matemáticos además de una representación gráfica de este *realismo*.

En base al cálculo de todos los índices mostrados a continuación se encuentra la noción de **TASAS DE EXACTITUD**, es decir el tanto por ciento de respuestas correctas proporcionadas por un alumno en relación a un cierto nivel de confianza. Se trata, para un alumno dado, de calcular tantas tasas de exactitud, como niveles de confianza. Si un alumno no ha utilizado más que dos niveles de confianza, tan sólo se pueden calcular dos tasas de exactitud.

Se puede calcular, en relación a cada alumno, tantas tasas de exactitud (TE) como niveles de confianza haya utilizado. si un alumno ha empleado tan sólo dos niveles de confianza tan sólo se pueden calcular dos tasas de exactitud.

3.4. Las gráficas de realismo

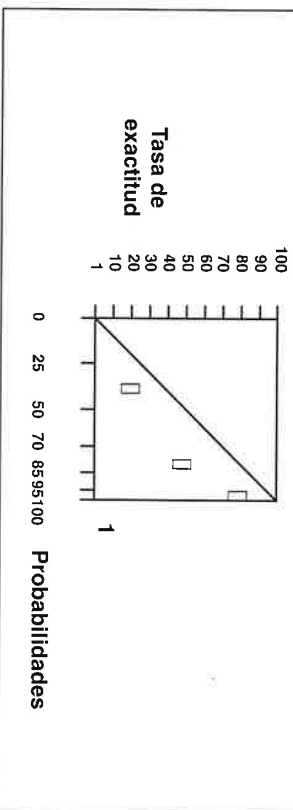
Las gráficas de realismo se construyen sobre los dos ejes siguientes:

- Sobre el eje horizontal (abscisas), representa las zonas de probabilidades (probabilidad de acertar predicha por el alumno) que corresponde cada una a un nivel de confianza.
 - Sobre el eje vertical (ordenadas), 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 nivel de confianza utilizado por el alumno es posible localizar en este gráfico un punto cuya,
- abscisa corresponde al valor central de la zona de probabilidad correspondiente a este nivel.

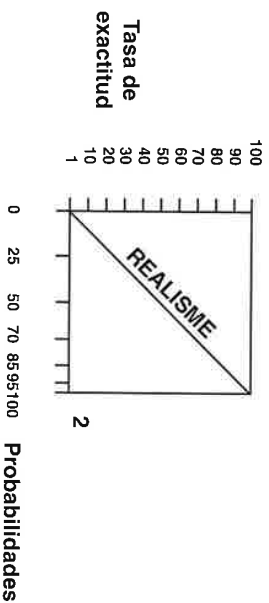
- ordenada supone la tasa de exactitud observada en relación a este nivel (porcentaje de respuestas correctas).

Este punto se encuentra representado por un rectángulo a fin de resultar más visible.

En nuestro ejemplo (gráfico 1), tres valores (20%, 42%, 75%) pueden ser representados en la vertical respectivamente los valores centrales de las zonas de confianza 1 (37,5%), 3 (77,5%) y 5 (97,5%).

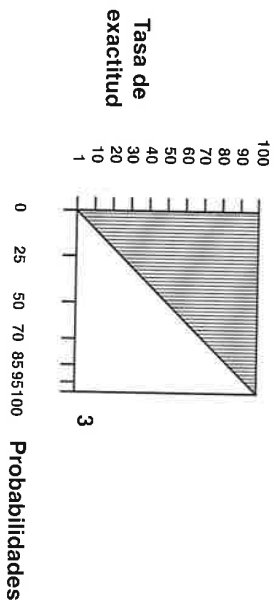


Se espera que los rectángulos estén sobre la diagonal (gráfico 2). En otros términos se espera que cada valor central de la zona de probabilidad (probabilidad de acertar predicha por el alumno) corresponda con las tasas de exactitud obtenidas.

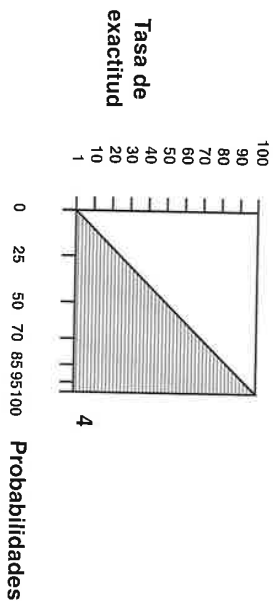


Pero la realidad se presenta en ocasiones de forma diferente. Ocurre que dos rectángulos estén:

- Sobre la diagonal. Esto significa que el alumno ha obtenido mejores resultados de los previstos por él mismo, es decir, se subestima (gráfico 3).



- Bajo la diagonal. Esto significa que el alumno ha obtenido peores resultados de los previstos, es decir, se sobrestima (gráfico 4).



En el ejemplo, se puede constatar que las tres tasas de exactitud se sitúan BAJO la diagonal, es decir, que los éxitos son continuamente inferiores a las predicciones del estudiante, que como consecuencia se sobrestima.

Resulta posible analizar de modo sistemático las características de la auto-evaluación del alumno, recur-

riendo a índices numéricos cuyo realismo y coherencia describimos a continuación.

3.5 Los índices numéricos del comportamiento de autoevaluación

Los índices numéricos de comportamiento de auto-evaluación permiten particularmente responder a dos preguntas:

- 1) las percepciones que el alumno tiene de sus posibilidades de éxito se encuentran lejos o cerca de sus éxitos reales (tasas de exactitud). En otros términos ¿Es realista?
- 2) ¿Existe una *progresión* en los aciertos del alumno cuando aumenta su confianza en sí mismo? En otros términos, ¿Es el alumno coherente consigo mismo en base a sus diversas auto-evaluaciones?

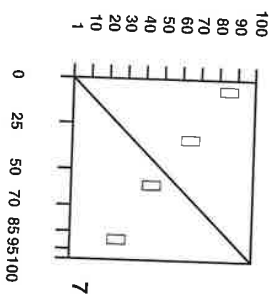
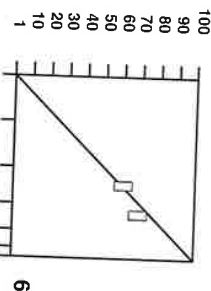
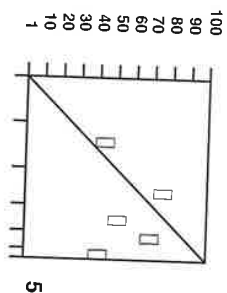
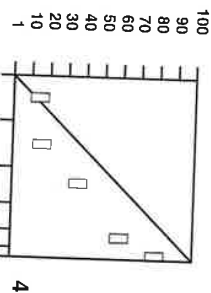
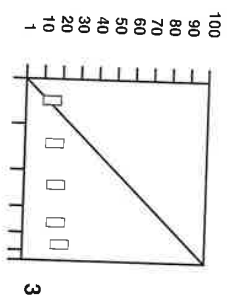
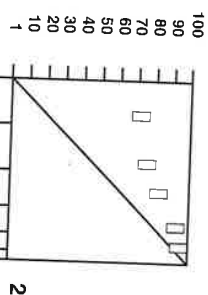
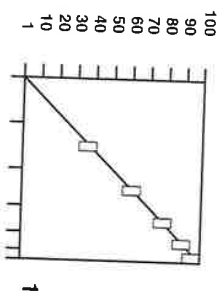
1) EL INDICE DE REALISMO

Este índice indica en qué medida la confianza que el alumno tiene en sí mismo corresponde o no a la realidad. La probabilidad de acierto predecida por él mismo ¿es cercana o lejana en relación a los aciertos obtenidos en la realidad (tasa de exactitud)?

Este índice ha sido calculado en base a diferencias entre

- La tasa de exactitud observada para cada nivel de confianza.
- El valor central del intervalo de probabilidad correspondiente a dicho nivel.

a continuación presentamos siete gráficos:



En relación a los mismos (alumnos del 1º al 7º), podemos observar los casos siguientes:

- 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 2,3,5 y cinco): el índice de realismo puede incluso ser negativo y descender hasta -3,75 (ver alumno 7).

El índice de realismo no precisa siempre en qué sentido se produce el alejamiento. Por ello, es necesario que observen el gráfico el emplazamiento de cada rectángulo:

- ¿Se encuentra situado bajo la diagonal? Si es así se trata de sobre-estimación. (Es el caso de los alumnos 3 y 4 sea cual sea el nivel empleado y por el alumno 5 cuando utiliza los niveles 3,4,5)

- ¿Se encuentra situado bajo la diagonal (intravaloración)? es el caso del alumno 2 y 5 cuando utilizan los niveles de confianza 1 y 2.

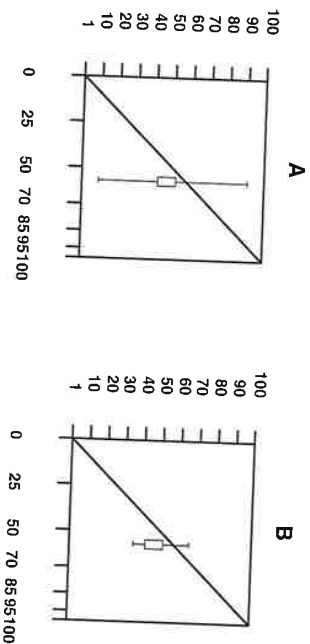
Por otra parte, es posible calcular (y añadir a esta gráfica bajo la forma de un trazo vertical) el error Standard de medición de cada tasa de exactitud. Este error resulta más importante si la tasa de exactitud ha sido calculada en base a pocas respuestas. se evitará desde ese momento dar una excesiva importancia a las observaciones calculadas para niveles de confianza escasamente empleados o de generalizar las interpretaciones.

ejemplo:

Un alumno A ha utilizado la confianza 3, dos veces de modo correcto. su tasa de exactitud vale el 50%

Un alumno B ha empleado 20 veces la confianza 3, 10 veces correctamente. 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 caso A, particularmente porque el error de medición es más fiable en el caso del alumno B.



Se debe notar que en el anexo se encuentran informaciones suplementarias que facilitan la interpretación de los valores de los índices de realismo.

2) El índice de COHERENCIA personal.

Este índice indica si el alumno utiliza niveles de confianza más elevados cuando tiene mayor éxito o menos elevados cuando responde peor.

En otros términos, muestra si el alumno presenta un comportamiento coherente en relación a sí mismo:

Efectivamente, responde mejor cuando estima tener más posibilidades de contestar correctamente y peor cuando estima tener menos posibilidades de acertar. (alumnos 1, 2 y 4)

Gráficamente, el índice precisa en qué medida las tasas de exactitud correspondientes a los niveles de confianza sucesivos se disponen en línea derecha sobre el gráfico.

Este índice se calcula de manera que vale,

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

¡Atención! Es posible ser coherente, aunque sobre o infravalorarse. Dicho de otro modo, se puede acertar más a menudo cuando se emplea un nivel de confianza elevado y menos cuando dicho nivel resulta menor, **pero no emplear los niveles adecuados.**

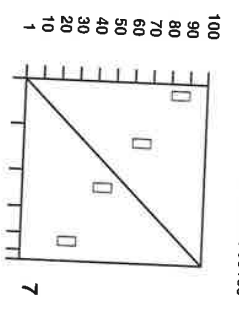
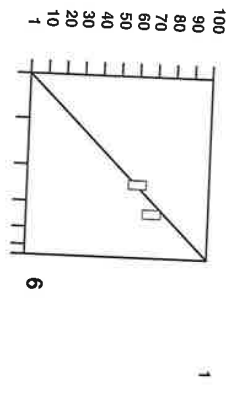
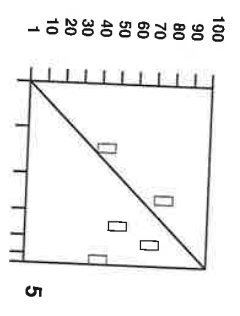
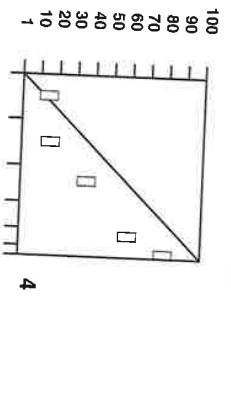
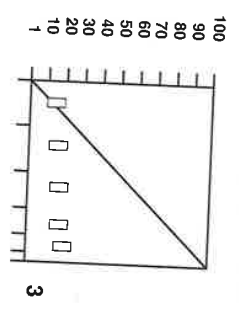
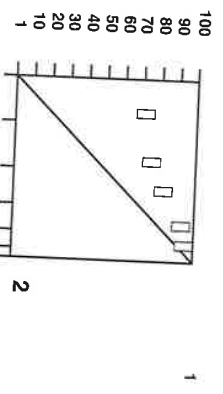
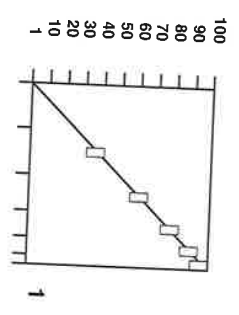
Ejemplo:

Tomemos el caso del alumno 4. en lugar de emplear los niveles de confianza 1, 2, 4 y 5, este alumno debería sin duda utilizar los grados 0, 1, 3 y tres. Resulta pues, coherente en relación a sí mismo aunque poco realista. Se sobreestima.

En el anexo se presentan indicaciones más precisas sobre la interpretación de los índices de realismo.

- Resumamos la interpretación de los gráficos de los alumnos 1 al 7.
- 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 se sobrevalora.
- El alumno 4 resulta coherente en la sobrevaloración.
- El alumno 5 es realista pero no resulta coherente.
- El alumno 6 es realista y coherente.
- El alumno 7 resulta irrealista e incoherente.



4. EN RESUMEN

Se ha integrado un nuevo elemento en la evaluación: el grado de confianza de un alumno en las respuestas que proporciona.

La puntuación total de la prueba es una combinación de niveles de confianza y de la exactitud de las respuestas dadas. Esta información permite medir la agudeza, la centralización, la coherencia y el realismo.

LA NECESIDAD DE UN SOFTWARE

1. LA RAZON DE SER DE UN SOFTWARE

El cálculo de la puntuación total, las tasas de exactitud y de los cuatro índices de auto-evaluación del mismo modo que la elaboración del gráfico requieren una media hora por alumno si se realiza manualmente.

El ordenador puede realizar estos trabajos en pocos segundos y proporcionar una versión en papel.

2. LAS INFORMACIONES PRESENTADAS POR EL SOFTWARE CERT

Las etapas fundamentales del trabajo relacionadas con el software CERT son las siguientes:

1) *concepción de la prueba*

El formador decide el contenido, el número de ítems, aquello que el test debe medir, las secciones de materias y las secciones de procesos mentales.

2) *Administración de la prueba a los alumnos*

Los alumnos responden y el formador recoge las respuestas.

3) *Introducción de información en el software CERT (codificación)*

- A) El formador diseña el test: número de ítems, respuestas correctas, secciones, consejos a los alumnos, peso de cada ítem...
- B) Recoge los datos de los alumnos (respuestas y nivel de confianza).
- C) Toma decisiones en relación al tratamiento (nivel óptimo, baremo de corrección)

4) *Análiza los resultados mediante CERT El software proporciona:*

- A) Análisis de ítems (para el formador)
- B) Análisis en relación a los resultados individuales obtenidos por los alumnos (Gráfica de realismo incluida).
- C) Listas de puntuaciones colectivas.

En este capítulo únicamente describimos los aspectos teóricos de las etapas (tercera y cuarta) realizadas mediante el ordenador. Los procedimientos prácticos del conjunto de trabajos son descritos en el manual de empleo.

3. INTRODUCCION DE INFORMACION EN EL SOFTWARE

3.1. **Recogida de datos de los alumnos**

El formador debe introducir dos códigos por ítem: el código de la respuesta proporcionada por el alumno y el nivel de confianza que le ha sido atribuido.

Recordemos que el Software CERT impone seis niveles de confianza (0 a 5) y deja libertad total en cuanto al tipo de ítems propuestos que pueden ser:

- ABIERTOS, pero la respuesta (literal, gráfica o numérica) debe ser codificada a posteriori por el profesor en:
 - 0 = omisión
 - 1 = respuesta correcta
 - 2 = respuesta incorrecta.

- VERDADERO-FALSO, acompañados de un código (con los códigos
 - 1 = verdadero 2 = falso)

- QCM CLASICOS, con un máximo de 9 soluciones posibles (el 0 indica omisión);

- QCM de soluciones generales implícitas.

Además del empleo de las QCM el formador puede solicitar al alumno la inclusión de una justificación en todas sus respuestas. Ello permite, entre otras cosas, solventar después de la prueba posibles errores de codificación cometidos por los alumnos.

ejemplo:

Si la respuesta correcta a un ítem es 3, un alumno elige la respuesta incorrecta 4, pero si la justificación muestra que ha "comprendido" el formador puede tener en cuenta dicha justificación modificando la respuesta seleccionada antes de introducir los datos en el ordenador: introduce el código de la respuesta esperada (3) y no el de la respuesta proporcionada (4).

3.2 **Decisiones en relación al tratamiento: cálculo de puntuaciones**

CERT impone un baremo de puntuaciones precisas, calculado según la teoría de las decisiones (ver capítulo V). Este puede ser modificado por el formador.

Si el alumno no acompaña su respuesta con un nivel de confianza, éste será considerado 0, o sea, el nivel de confianza memor. En efecto, el nivel de confianza 0 tiene como valor central 12,5% lo cual significa que el alumno tiene 12,5/100 o 1/8 posibilidades de acertar, es decir, aproximadamente 1 posibilidad entre el máximo de soluciones posibles (9)

Si el alumno omite alguna respuesta, la respuesta será considerada como incorrecta con un nivel de confianza 0 incluso en el caso de que haya proporcionado otro nivel.

Se ha de notar que en la versión actual del software las preguntas son, sea correctas o incorrectas, pero no parcialmente correctas (pues se obtendría una puntuación intermedia).

Finalmente, cuando se crea un test, el formador debe proporcionar un peso, una "importancia" en relación a cada ítem. Proporcionar el mismo peso a todos los ítems de un test, comporta proporcionarles la misma importancia en el cálculo de la puntuación total. Proporcionar un peso de 2 a un ítem del test supone doblar su peso en relación a los demás.

Supongamos un Ejemplo:

Si un test de nueve preguntas comprende ocho ítems de importancia 1 y uno de peso 2, la puntuación total será calculada como si el ítem de importancia 2 estuviera puesto dos veces, en otros términos, como si el test comprendiera 10 ítems de importancia 1.

4. INFORMACION PROPORCIONADA EN RELACION A LA CALIDAD DE LOS ITEMS.

El programa CERT posibilita evaluar la calidad de los ítems propuestos gracias a cierto número de indicaciones numéricas proporcionadas en cada versión del test.

SOL:	0	1	2	3	4	5	6	7	8	9
Q 1	6,25	18,75	*68,75	6,25			0,00	0,00	0,00	0,00
Rbis	-0,30	0,27	0,20	-0,52			0,00	0,00	0,00	0,00
Omoy	12,50	45,00	80,00	77,50			0,00	0,00	0,00	0,00

A continuación figura el primer ítem (Q1) de una prueba. Para cada posibilidad de respuesta (columnas 0 (u omisión) a 9) se indican tres informaciones.

— LINEA 1

El porcentaje de alumnos que han seleccionado este modo de responder (o tasa de elección de cada solución propuesta) permite identificar el atractivo de cada una de las soluciones planteadas.

En relación al ítem 1 hay que señalar, por ejemplo, que la solución (correcta) 2 ha atraído al 68,75% de alumnos, que las soluciones 1 y 3 han atraído respectivamente al 18,75 y 6,25% de alumnos y finalmente que un 6,25% de los alumnos no ha respondido. (código 0)

— LINEA 2: R.BIS

El índice calculado en este punto supone un coeficiente de discriminación en relación a cada una de las soluciones propuestas. Está comprendido entre -1 y +1.

Si este índice resulta positivo significa que este modo de responder ha sido adoptado por los alumnos "fuertes". Se entiende por alumnos "fuertes" aquellos cuya puntuación total en el test resulta más elevada que la puntuación en el total del test realizado por el resto de los alumnos (es decir aquellos que han seleccionado otra respuesta).

Dicho índice debería ser positivo en relación a la respuesta correcta (a grosso modo), negativo para las respuestas incorrectas y entre los dos en referencia a la omisión.

En el momento en que un modo de responder no sea adoptado por nadie (o por todos), automáticamente, el índice debe ser considerado nulo.

En el anexo 1 se encuentra representado de modo más detallado el modo de interpretar este coeficiente.

En nuestro ejemplo, la solución correcta 2 resulta elegida por alumnos "fuertes" en el conjunto del test (R. BIS= 0.20), aunque ocurre lo mismo con la solución 1 (R. BIS=0.27)

El formador debe preguntarse sobre la causa de esta anomalía, sobre el modo en que modificará la solución incorrecta 2 (puesto que ha sido escogida por los alumnos "fuertes")?

Las soluciones 0 y 3 comprenden un R. BIS negativo, lo cual significa que son más bien los alumnos "débiles" en el conjunto del test los que las seleccionan.

- LINEA 3: C.moy

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.

Normalmente debería esperarse que la confianza media obtenida por los alumnos que eligen la respuesta correcta sea superior a la de aquellos que se han equivocado. Cuando no se da este caso resulta conveniente cuestionarse a fin de saber qué ha conducido a los alumnos a ser confiados en relación a una respuesta incorrecta o poco confiados en base a una correcta.

A fin de obtener dicho índice en base a una solución dada, se calcula la media de probabilidades (posibilidades de acertar) prededidas por los alumnos que han seleccionado esta solución. Se trata pues de la media de los valores centrales en los intervalos de confianza.

Observemos la solución 3 del ejemplo. ¿Por qué los alumnos poco numerosos (6,25%) y "débiles" (-0.52) han elegido esta solución errónea con tanta confianza (77,8%)?

5. INFORMACION PROPORCIONADA EN RELACION A CADA ALUMNO

El software presenta una serie de información por medio de cuatro pantallas.

5.1. Datos

Test :	Archivo :
Alumnos :	Alumnos :
Nombre :	Apellidos :
Respuestas correctas :	
Peso de cada ítem :	
Respuestas :	
Nivel de confianza :	

5.2. Valores numéricos

Nivel de confianza	0	1	2	3	4	5	TOT
	0	25	50	70	85	95	100
No. respuestas	0	8	8	1	5	4	26
No. correct	0	3	4	1	3	3	14
No. incorrect	0	5	4	0	2	2	12
% resp. correct	0.0	37.5	50.0	100.0	60.0	75.0	59.8
Val. central	12.5	37.5	60.0	77.5	90.0	97.5	69.3
PUNTUACIONES BRUTAS							
por resp. correct	0	64	68	18	76	80	306
por resp. incorrect	0	15	10	0	-12	-40	-27
Niv. opt.	16	SC. AD.		11,3/20		Total	
Coherencia	0.838	Débil		Test : INFORMAT			
Realismo	0.991	Excelente		Alumno: DEGHAYE M.			
				PM: Pas de sélection.			
				Mat: Pas de sélection.			
				Total			
				279 / 620			

a) Número y porcentaje de respuestas

El programa presenta para cada nivel de confianza (columnas 0 a 5) y posteriormente en relación al conjunto de la prueba (columna TODO):

- (Nº resp.): el número de respuestas proporcionadas por el alumno y en base a las mismas,
 - Nº correct: número de respuestas correctas
 - Nº incorrect: Número de respuestas incorrectas;
- (% resp. correct.): Porcentaje de respuestas correctas proporcionadas por el alumno o tasa de exactitud observada.
- (Val. Central): tasa de aciertos esperada que supone de hecho el valor central del intervalo de probabilidad correspondiente al nivel de confianza.

Si se compara con la tasa precedente ésta permite detectar dónde el alumno se sobrevalora y dónde se infravalora. En el caso en que el porcentaje de respues-

tas correctas sea superior al valor central correspondiente, el alumno se ha infravalorado, si resulta inferior se sobrevalora.

b) Puntuaciones brutas

El programa presenta igualmente una serie de puntuaciones (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 se presentan las siguientes puntuaciones:

- (Por res. correct): Puntuación bruta obtenida en relación a las respuestas correctas.
- (Por resp. incorrect): 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 de los niveles de confianza proporcionados.

La puntuación total bruta se relaciona con una puntuación de referencia (denominador de la fracción) que supone la puntuación máxima matemáticamente posible. Esto último se calcula en base a la fórmula siguiente:

$$\text{PUNTUACION MAXIMA} = \text{NUMERO DE ITEMS X 20}$$

Es decir el número de ítems de la prueba comprende la máxima baremación posible.

- En una prueba de 20 items cuyo peso es 1 y al que se aplica un baremo en el que se obtienen 20 puntos en caso de respuesta correcta con máxima confianza. La puntuación máxima es $26X\ 20 = 520$
- En una prueba de 26 items de 20 puntos de los cuales 5 tienen una importancia de 2 y 21 un peso de 1. La puntuación máxima a alcanzar es $31X20 = 620$

c) Puntuación total adaptada.

Una línea de la tabla se reserva a la puntuación total adaptada: Se trata de la puntuación total bruta que se ha adoptado para situarla en relación a una puntuación de referencia: 1) Presentada de modo habitual 2) "Justa".

En efecto, una puntuación siempre debe ser expresada en relación a un baremo de referencia. A menudo la puntuación de referencia elegida se presenta bajo la forma de una referencia habitual (10, 20, 100...) y corresponde a la puntuación máxima matemáticamente posible (cf. puntuación máxima descrita a continuación).

C.1) Si la puntuación de referencia comprende un número "inhabitual", para transformarlo en referencia habitual (por ejemplo 100), resulta suficiente con adaptar la puntuación del alumno mediante una regla de tres.

$$\text{PUNTUACION TOTAL BRUTA X SEÑAL} \\ \text{PUNTUACION TOTAL ADAPTADA} = \frac{\text{PUNTUACION TOTAL BRUTA X SEÑAL}}{\text{PUNTUACION DE REFERENCIA}}$$

Ejemplo: Tomemos un alumno que obtiene 600 en una prueba de 60 items a 20 puntos cuyo máximo matemático resulta pues 1200:

80

$$\text{puntuación adaptada a la señal 100} = \frac{600 \times 100}{1200} = 50$$

$$\text{Puntuación total adaptada a la señal 20} = \frac{600 \times 20}{1200} = 10$$

C.2. Sin embargo cuando se emplean índices de confianza la puntuación máxima corresponde a un nivel de exigencia muy elevado ("Injusto").

La puntuación máxima (60 NQ X 20 = 1200) no puede ser utilizada como puntuación de referencia para definir el nivel óptimo. En efecto, se trata de una exigencia demasiado elevada, porque habitualmente resulta suficiente con proporcionar todas las respuestas correctas sin más.

Es pues necesario proporcionarse una puntuación de referencia más "razonable", más "justa", más "habitual". En lugar de seleccionar la puntuación máxima matemática, resulta suficiente con multiplicar el número de items por 20 (máximo obtenido en caso de confianza 5), si

- por 19 (máximo obtenido en caso de confianza 4) o
- 18 (máximo obtenido en caso de confianza 3) o
- 17 (máximo obtenido en caso de confianza 2) o
- 16 (máximo obtenido en caso de confianza 1)

Los valores seleccionados se denominan NIVELES OPTIMOS.

La puntuación máxima de referencia se calcula de la manera siguiente:

$$\text{PUNTUACION DE REFERENCIA} = \text{NUMERO DE ITEM X NIVEL OPTIMO}$$

81

EJEMPLOS:

1) En caso de un test de 60 ítems cuyo peso es 1.

Si el nivel óptimo elegido es 18, entonces la puntuación de referencia es

$$\text{PUNTUACION DE REFERENCIA} = 60 \times 18 = 1080$$

Si la referencia seleccionada es 100, si la puntuación total bruta del alumno es de 600, entonces la puntuación de referencia es:

$$\text{PUNTUACION TOTAL} = \frac{\text{PUNTUACION TOTAL BRUTA} \times \text{REFERENCIA ADAPTADA}}{\text{NUMERO DE ITEMS} \times \text{NIVEL OPTIMO}}$$

$$= \frac{600 \times 100}{60 \times 18} = 0.55 \times 100 = 55$$

Con un nivel óptimo 17, el mismo alumno habría obtenido con la misma señal (100):

$$= \frac{600 \times 20}{60 \times 17} = 0.62 \times 100 = 62$$

Con un nivel óptimo 17 el mismo alumno habría obtenido con referencia 20:

$$= \frac{600 \times 20}{31 \times 16} = 0.62 \times 20 = 12.4$$

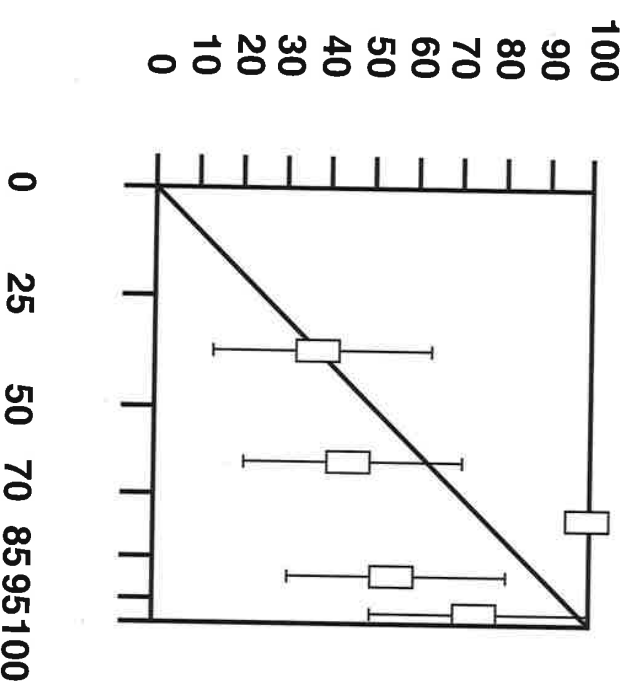
2) En el ejemplo presentado al comienzo del capítulo, "el número de ítems" es 31 (21 ítems importancia 1 y 5 de peso 2). La puntuación bruta (279) adaptada a la señal 20 y al nivel óptimo 16 se calcula del siguiente modo:

$$\frac{279 \times 20}{31 \times 16} = 11.25$$

d) Índices de auto-evaluación.

El programa presenta de igual modo dos índices de calidad del comportamiento de auto-evaluación descritos en el capítulo precedente.

5.3. Gráfica de realismo



5.4. Informaciones individuales relacionadas con las secciones

CERT presenta de igual modo informaciones relativas a secciones definidas por el profesor (quien debe indicar en qué sección se encuentra cada ítem). Los subconjuntos de ítems correspondientes a una preocupación precisa, por ejemplo, una sección materia o una sección de procesos.

El programa indica en relación a cada sección el número de respuestas correctas obtenidas en referencia al número de ítems propuestos.

Se presentan diversos consejos en función de un umbral previamente definido por el evaluador.

En el ejemplo siguiente el alumno ha obtenido 3/4 en aplicación. Alcanzado el umbral del 75% deseado por el evaluador el mensaje presentado es "satisfactorio". Sin embargo en la sección de materias, en relación al contenido del software, la puntuación del alumno es de 7/12. El mínimo del 75% requerido por el evaluador no ha sido obtenido. Recibe pues el consejo de repetir los ejercicios 12 a 15.

Procesos mentales	Nbr. corr.	Consejos
memoria	4 / 9	Procedimientos memotécnicos?
comprensión	4 / 8	venga a verme
aplicación	3 / 4	Satisfacción
síntesis	3 / 5	Venga a verme.
materias		
	Nbr. corr.	Consejos
material	3 / 5	Satisfacción
le logiciel	7 / 12	Haz los ejercicios 12 a 15
l'algorithme	4 / 9	Venga a verme.

6. LA INFORMACION COLECTIVA

El software proporciona igualmente información sobre

- Los aciertos de los alumnos en relación a las diferentes secciones definidas: cuadro retomando el número de ítems de cada sección acertada por cada alumno.
- Lista de puntuaciones adaptadas (SC. AD) en relación a la referencia 20 y en función de un nivel óptimo seleccionado por el formador.

Seccion de procesos :	1	2	3	4
Nombre	9	8	4	5
Apellido	4.5	4.8	2.8	3.5
Umbral	:	:	:	:
BIDDOCHON R.	4	3	0	2
BLOOM B.	9	7	4	5
DEGHAYE M.	4	4	3	3
FRENCH B.	1	7	4	5
JAURES J.	5	5	1	4
LABEVUE B.	0	2	0	2
LAGAFFE G.	8	6	4	4
MANDRIN A.	8	7	4	4
MARX K.	4	2	0	2
MASSON P.	4	5	1	3
O'CONNEL S.	5	5	4	5
ORBAN M.	4	5	1	0
SHAW B.	5	4	1	4
WASIEGE R.	7	5	4	5
WUIDAR H.	6	7	4	5
ZOLA E.	7	5	3	5

PERSPECTIVAS

1. UNA PRIMERA APROXIMACION

Una primera aproximación sistemática al empleo de CERT llevada a cabo por los cuatro miembros participantes en el presente proyecto es reveladora de la *variedad* de posibilidades que ofrece.

El equipo británico ha creado una prueba de carpintería. Se ha adoptado una aproximación "bottom up":

- los formadores se han planteado cuales son los problemas. En segundo lugar han elaborado la preguntas que revelan la capacidad (global) de responder;
- una vez elaboradas las preguntas, los autores han intentado clasificarlas según los niveles taxonómicos.

Ciertas preguntas podían ser contestadas satisfactoriamente siguiendo estrategias mentales muy diversas. Hay pues una dificultad de clasificación taxonómica a priori.

El equipo italiano ha elaborado una prueba sobre economía de base. Este equipo ha adoptado una aproximación "top-down":

- los formadores han decidido inicialmente sobre qué tema y sobre qué nivel (taxonómico) de proceso mental deseaban obtener una evaluación;
- a continuación, el equipo ha elaborado las preguntas para cada nivel elegido.

Nombre	Apellido	Grupo	Puntuación
BIDOCHON	ROBERT	CLASSE B	7.6 / 20
BLOOM	BENJAMIN	CLASSE B	22.8 / 20
DEGHAYE	MANON	CLASSE A	11.3 / 20
FRENCH	BOB	CLASSE A	14.1 / 20
JAURES	JEAN	CLASSE A	13.9 / 20
LABEVE	BERTRAND	CLASSE B	2.9 / 20
LAGAFFE	GASTON	CLASSE B	17.3 / 20
MANDRIN	ALBERT	CLASSE A	19 / 20
MARX	KARL	CLASSE B	2.7 / 20
MASSON	PIERRE	CLASSE A	6.6 / 20
O'CONNEL	STEVEN	CLASSE B	15.1 / 20
ORBAN	MICHEL	CLASSE A	10.7 / 20
SHAW	BERNARD	CLASSE B	13.5 / 20
WASIEGE	ROBERT	CLASSE A	19.2 / 20
WUIDAR	HENRI	CLASSE A	16.5 / 20
ZOLA	EMILE	CLASSE B	16.9 / 20

7. EN RESUMEN

El software CERT es un instrumento dirigido a facilitar información profunda en relación a:

- La calidad de los ítems de un test
- Las puntuaciones de los alumnos que han participado en el test y comprendido las informaciones relativas a su auto-evaluación.
- La repartición de los resultados de los alumnos según las secciones de materias y las secciones de procesos.

El tratamiento de datos sólo dura una fracción del tiempo que sería necesario si la acción fuese realizada manualmente.

El software CERT presenta un gran número de potencialidades interesantes y puede desarrollarse en diversas direcciones.

El equipo español ha elaborado un cuestionario sobre automoción, a partir de la transformación de una prueba existente. Ciertas preguntas han sido "diseñadas" de tal manera que la solución correcta se ha convertido en una solución general del tipo 6,7,8 y 9. Por otra parte también se han añadido nuevas cuestiones.

El equipo belga ha creado una prueba sobre **informática**. Se prevé que en un primer momento el alumno pueda manipular el ordenador para verificar sus hipótesis. A continuación y en el transcurso de la realización de la prueba, los alumnos disponen del manual (Manual de DOS). Se trata por lo tanto de una prueba a *libro abierto*.

2. DESARROLLOS FUTUROS

Los desarrollos ulteriores del proyecto pueden enmarcarse en una serie de directrices:

a) Validación de procedimientos

Será importante estudiar:

- las modalidades de utilización más apropiadas en contextos particulares;
- la aceptación del procedimiento por parte de los alumnos;
- la aceptación del procedimiento por parte de los formadores.

b) Validación de las mediciones obtenidas

- Correlación con los resultados profesionales sobre el terreno.
- Validez de constructo: poner a prueba la teoría subyacente y la coherencia interna de las pruebas.

- Profundización en ciertos aspectos, como por ejemplo la capacidad de decidir y el sentido de responsabilidad.
- Utilización del proyecto transnacional para medir los efectos de la formación.

c) Perfeccionamiento del proyecto

- Estudiar las necesidades de evaluación no cubiertas y las posibilidades de extensión de los materiales desarrolladas.
- Estudiar los problemas prácticos detectados al utilizar el programa CERT y proponer soluciones.

- Elaborar procedimientos experimentales comunes.

- Confrontación de los resultados entre los diversos países.

- Poner a punto una versión interactiva (sobre ordenador), del software CERT, añadiendo procedimientos de "testing" criterial secuencial (WALD) y/o de "testing" sobre medida (FASH por ejemplo).

- Producción de un CERT interactivo audiovisual (magnoscopio, videodisco,...) que permitan crear simulaciones próximas a la realidad.

- Construcción de bancos de items y de datos.

d) La difusión de CERT

- Seminarios de formación en el proyecto cooperativo.
- Presentación de los principios comunes en seminarios de formación sobre evaluación.

- Conexión del presente proyecto con otros existentes (CEE o no) en el campo del "testing" y del desarrollo del currículum, y especialmente, sobre la formación de formadores.
- La conexión del proyecto CERT con las andaduras de promoción de la innovación en la formación profesional dirigida a aumentar las competencias y niveles de cualificación.
- Participación en operaciones dirigidas a aumentar la motivación en la comprensión de los procesos de aprendizaje.

CONCLUSIONES GENERALES

Desarrollar las competencias y las cualificaciones de los trabajadores es un objetivo primordial de la política social de la Comunidad Europea.

Una evaluación más motivadora y más informativa para el alumno debería contribuir a la consecución de este objetivo.

Esta modesta investigación práctica de evaluación se centra sobre el alumno. En el presente responde a una serie de necesidades específicas, pero exige un desarrollo y una profundización posterior.

Se puede esperar, racionalmente, las consecuencias de la explotación directa sobre el mercado de trabajo europeo. Este proyecto necesita competencias suplementarias y apoyos comunitarios y nacionales.

GLOSARIO

Las definiciones que se presentan a continuación han sido aceptadas por los miembros de los diferentes equipos del proyecto CERT.

La mayoría de estas definiciones han sido tomadas del Dictionnaire de l'évaluation et de la recherche en éducation de G. DE LANDSHEERE (1979).

Los números entre paréntesis indican el número de la página donde se encuentran las definiciones.

El signo (#) indica que se han efectuado ligeras modificaciones a la definición.

El signo (*) indica que la definición recurre a otros términos definidos en el glosario.

ACTITUDES: *ATTITUDE*

Organización emocional, motivacional, perceptiva y cognitiva duradera de creencias relativas a un conjunto de referentes que predisponen a un individuo a reaccionar positivamente o negativamente ante los objetos o referentes (de la actitud). G. de Landsheere, p.22).

APRENDIZAJE: *LEARNING*

Adquisición o modificación de comportamientos o estructuras mentales o afectivas.

ACTITUD:

Organización emocional, motivacional, perceptiva y cognitiva durable de creencias relativas a un conjunto de referentes que predisponen a un individuo a actuar positiva o negativamente a los objetos o referentes (de la actitud)

AUTO-ESTIMACION: SELF-ASSESSMENT

Estimación de su propia competencia (que será verificada a través del "testing", por ejemplo) o de sus propias características. Los índices de realismo (*), de coherencia (*), de agudeza (*), de centralización (*) en la auto-estimación han sido desarrollados por D. LECLERCQ (1990).

BAREMO: SCALE OF MARKS

Modalidades de notación respecto a una prueba: lista de baremos aplicados en relación a las respuestas correctas e incorrectas. Diferencias cuando se utilizan niveles de confianza.

BOTTOM UP (aproximación... en la construcción de ítems):**BOTTOM UP** (... *approach in construction of questions*)

Método de elaboración de ítems que parte de los problemas o situaciones **CONCRETAS** a solucionar. A continuación la preocupación recae en el proceso mental que implica la respuesta.

CALIFICACION:

valor de un alumno referente a su formación, aptitudes profesionales y experiencia.

CAPACIDAD: ABILITY

Posibilidad de llevar a cabo un acto, de producir un comportamiento o un conjunto de comportamientos adecuados en una situación dada. (G. De Landsheere, p.31).

COGNITIVO (Dominio): COGNITIV

Dominio que se reduce al conocimiento o, más exactamente, a las capacidades y mecanismos de apren-

dizaje y de acceso al conocimiento de un saber (G. De Landsheere, p.48).

COHERENCIA (Índice de): **CONSISTENCY index**

Correlación entre las estimaciones de una persona por diversos grados de confianza y las tasas de exactitud medias de estos grados de confianza. Se puede ser coherente (con uno mismo) sin ser realista.

COHERENCIA INTERNA (Índice de):**INTERNAL CONSISTENCY Index**

Búsqueda de la medida en la cual todos los ítems de un test miden la misma variable (G. De Landsheere, p.147).

CONFIANZA (Grado de): **CONFIDENCE degree**

Probabilidad que un alumno otorga a su propia respuesta de ser correcta. Esta "confianza" debe ser expresada bajo la forma de una zona sobre el eje de probabilidades (ejemplo: zona 1 = de 0 a 25% de posibilidad de ser correcta) y no bajo forma de grados en una escala ordinal (muy seguro, medianamente seguro, poco seguro).

CONSECUENCIA: PAYOFF

Beneficio o penalidad que una acción (o una decisión) aporta a la persona referida.

CORRELACION: CORRELATION

Estimación numérica de la relación, de la unión lineal que existe entre dos o más variables (G. De Landsheere, p.59).

CURRICULUM: CURRICULUM

Conjunto de acciones planificadas para suscitar la instrucción. (G. De Landsheere, p.65).

CURRICULUM OCULTO O LATENTE:**HIDDEN CURRICULUM**

Lo que todo el mundo aprende y que nadie enseña. El currículum no declarado o no consciente.

DECISIONES (Teoría de las): *DECISION THEORY*

Teoría que permite decidir la acción o elección más favorable, teniendo en cuenta al mismo tiempo la utilidad (atractivo de las consecuencias (*)) y la probabilidad de las consecuencias.

DIAGNOSTICO (Test): *DIAGNOSTIC test*

Test que sirve para localizar las dificultades de aprendizaje en vías de solucionarlas. (G. De Landsheere, p.272).

DISCRIMINACION (Índice de discriminación de un ítem):

DISCRIMINATION index of an item

Correlación entre la respuesta a la pregunta y la puntuación total del test.

DISTRACTOR: *DISTRACTOR*

En una QMC, se llama distractores a las proposiciones falsas o menos apropiadas y más lejanas de la respuesta esperada.

ESPECIFICO (Objetivo): *SPECIFIC OBJECTIVE*

Combinación de una aptitud (o comportamiento) a desarrollar y un CONTENIDO. (G. De Landsheere, p.189, #).

EVALUACION: *ASSESSMENT*

Atribución de una categoría a un comportamiento o a un producto, esta categoría forma parte de una escala

- nominal (ejemplo: Clásico, Romántico,...)
- ordinal (ejemplo: Excelente, Muy bien, Bien, ...)
- métrica de intervalos (ejemplo: 10, 5, 0, -2)
- métrica de relación (ejemplo: 10, 5, 0)

EVALUACION FORMATIVA: *FORMATIVE EVALUATION*

Evaluación interviniente, en principio, al término de cada tarea de aprendizaje y teniendo por objeto informar al alumno y al profesor sobre el grado de aprendizaje logrado y, eventualmente, descubrir dónde y en qué un alumno muestra dificultades de aprendizaje, con intenciones de proponerle, o de hacer que descubra estrategias que le permitan progresar (G. De Landsheere, p.115, #).

EVALUACION SUMATIVA: *SUMMATIVE EVALUATION*

Evaluación que asume el carácter de balance e interviene tras de un conjunto de tareas de aprendizaje constituyendo un todo (G. De Landsheere, p.115, #)

FIDELIDAD: *RELIABILITY*

Las medidas repetidas de un objeto, realizadas en las mismas condiciones con un mismo instrumento, deben ser fieles, es decir, presentar el mismo resultado. (G. De Landsheere, p.125).

IMPORTANCIA:

peso dado a cada pregunta de un test a fin de acordar una mayor influencia de ciertas cuestiones respetando el mismo baremo (*) de corrección.

OBJETIVO: *OBJECTIVE*

Meta que se propone la acción.

OPERACIONAL (Objetivo): *OPERATIONAL*

Aquel que se define por comportamientos observables o medidas. (G. De Landsheere, p.193).

OPTIMO (nivel): *LEVEL OF EXCELLENCE*

Nivel al que se adapta una puntuación de referencia para convertirlo en equitativo.

ejemplo: proporcionar un nivel óptimo de 18 implica que la puntuación de referencia será el número de ítems X 18 (todas las respuestas correctas con nivel de confianza 4) en lugar de número de ítems X 20 (todas las respuestas correctas con nivel de confianza 5 = puntuación máxima matemáticamente posible).

PESO: *IMPORTANCE*

Importancia dada a cada pregunta del test con la finalidad de acordar una mayor influencia en relación a ciertos ítems respetando el mismo baremo (*) de corrección.

PREGUNTA ABIERTA: *OPEN-ENDED QUESTION*

Pregunta a la que el sujeto responde utilizando su propio vocabulario. (G. De Landsheere, p.223).

PREGUNTA CERRADA: CLOSE-ENDED QUESTION

El sujeto debe simplemente elegir una de las diferentes respuestas que le son propuestas (G. De Landsheere, p.222).

PROBABILIDAD: PROBABILITY

Medida de la tendencia de aparición de un acontecimiento. (G. De Landsheere, p.214).

PROCESO: PROCESS

Cambio no puntual en un organismo. (G. De Landsheere, p.215).

PUNTUACION ADAPTADA: ADAPTED SCORE

Puntuación bruta (*) del alumno que se adapta en parte con la finalidad de presentarla bajo una forma habitual (ver referencia (*)) y con la finalidad de convertirla en una puntuación más equitativa (ver nivel óptimo)(*).

PUNTUACION BRUTA: RAW SCORE

Puntuación de un alumno directamente calculada en base a un baremo de corrección (*) y sin adaptación.

PUNTUACION MAXIMA:

matemáticamente posible: Puntuación obtenida cuando todas las respuestas son a la vez correctas y proporcionadas con un nivel máximo de confianza (5).

PUNTUACION DE REFERENCIA:

Puntuación a la que se compara la puntuación del alumno. Puede tratarse de la puntuación máxima matemáticamente posible, pero igualmente de una puntuación calculada con otro nivel óptimo. Puntuación de referencia: número de ítems X nivel óptimo.

REALISMO: REALISM

Índice que muestra si la probabilidad de acertar que estima tener un alumno se encuentra cercana o lejana en relación a los aciertos obtenidos en la realidad.

REFERENCIA: REFERENCE

Valor arbitrario habitual al que se refiere la puntuación de referencia (*) en el caso en que ésta sea un número inhabitual. Se trata del denominador de la fracción que representa la puntuación del alumno (10,20, 100, etc).

REGULACION: READJUSTMENT LOOPS

Proceso por el cual un PROYECTO, dirigido a responder ciertas NECESIDADES, se traduce en un PLAN, después en ACCIONES, MEDIDAS, DECISIONES que consisten en bucles de RETROALIMENTACION*

RETROALIMENTACION: FEEDBACK

Información que "retorna" teniendo por objeto advertir al emisor de un comportamiento; de la desviación entre el efecto real y el efecto esperado, y de este modo permitirle adaptar mejor sus comportamientos, sus respuestas. (G. De Landsheere, p. 238).

"SAVOIR FAIRE": SKILL

Actuación general estable, centrada sobre la tarea. (G. De Landsheere, p.244).

SINOPTICO: SYNOPTIC

Que permite tener una visión de conjunto de un todo. (G. De Landsheere, p.253).

SUBJETIVO: SUBJECTIVE

Lo que está afectado por intuiciones íntimas, sin posibilidad de verificación directa por observación y medida. (SUTTER).

TASA DE EXACTITUD: RATE OF SUCCESS

Porcentaje de respuestas correctas de un alumno obtenido en base a las respuestas que ha acertado dado un determinado nivel de confianza. En relación a cada alumno se pueden calcular tantas tasas de exactitud como niveles de confianza haya utilizado.

TAXONOMIA: TAXONOMY

Clasificación jerárquica establecida según uno o más criterios explícitos. (G. De Landsheere, p.260).

TOP DOWN (Aproximación... en la construcción de ítems)
TOP DOWN (...*approach in construction of questions*)

Método de creación de ítems donde se parte de los procesos mentales implicados por los objetivos operacionales. Se crea la pregunta en función del nivel taxonómico que se quiere medir.

UTILIDAD: *UTILITY*

Número que mide el incentivo de una consecuencia sobre una persona dada, la medida se efectúa sobre una escala de probabilidad. (LINDLEY).

VALIDEZ: *VALIDITY*

Adecuación perfecta entre lo que se pretende medir y los ejercicios de medida.

VIGILANCIA COGNITIVA: *COGNITIVE AWARENESS*

Capacidad de plantearse espontáneamente preguntas, de detectar trampas en los razonamientos o en los datos.

REFERENCIAS BIBLIOGRAFICAS

Las tres obras siguientes deben ser siempre mencionadas ya que constituyen tres de los pilares de este cuadro conceptual.

BLOOM, B. and Coll., (1956), *Taxonomy of educational objectives. Handbook I: Cognitive Domain*. New York: McKay.

DE LANDSHEERE, G., (1979), *Dictionnaire de l'évaluation et de la recherche en éducation*, Paris: PUF.

LECLERCQ, D., (1983), *Confidence marking, its use in testing*, Oxford: Pergamon, 126 p.

ANEXOS

ANEXO 1: DISCUSION EN RELACION A LA TAXONOMIA DE BLOOM

a) *La taxonomía de partida*

La taxonomía de objetivos cognitivos de BLOOM y al. (1956) evoca la imagen de una escalera de dos tramos. Cada tramo posee tres peldaños. Primer tramo: *Conocimiento, comprensión, aplicación*. Mediante las dos primeras se conoce y comprende una realidad; mediante la tercera se resuelven problemas simples, clásicos, familiares; en base a la aplicación de leyes, reglas, principios conocidos y comprendidos.

En este lugar hay un descansillo: podríamos detenernos. Pero un aprendizaje encaminado al dominio de un contenido exige llegar más lejos, al segundo peldaño.

Mediante el análisis, se profundiza en la comprensión de la realidad. En base a la síntesis se produce de una manera creativa, se formula, se expresa una realidad análoga. Finalmente mediante la evaluación se emiten juicios en relación a la realidad según determinados criterios personales.

A continuación, presentamos el modo en que la taxonomía se propuso en un principio: estrictamente lineal, donde cada nivel encuentra su soporte en el precedente.

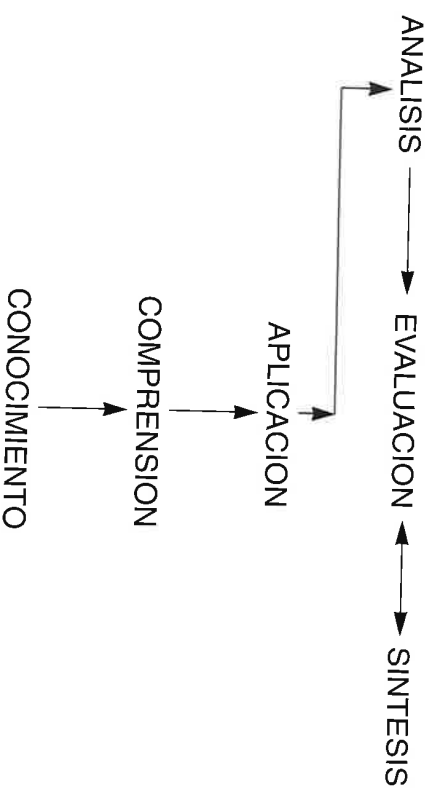
b) La taxonomía bifurcada

Más tarde en base a diferentes investigaciones, MADDAUS y al. (cfr. G. DE LANDSHEERE, 1977) se afirma que después de la *aplicación*, la taxonomía se difurca: por un lado hacia el *análisis*, por otro hacia la evaluación y la *síntesis*, siendo las tres capacidades de elevado nivel.

Recordando otros estudios, DE LANDSHEERE (op. cit) acepta inscribir sobre el mismo plano las tres capacidades, aunque las distingue del modo siguiente: de un lado sitúa el *análisis* y la *evaluación* como capacidades de tipo convergente que continúan de manera rígorosa la taxonomía. En el otro lado sitúa la *síntesis*, es decir la creatividad que se beneficia del pensamiento divergente y se sustenta en una capacidad convergente: la evaluación (o juicio personal). Esta interacción entre las capacidades de *síntesis* y *evaluación* (interacción que ha sido examinada por especialistas de la creatividad) nos convence mejor en relación al hecho de que al menos evaluación y *síntesis* se encuentran en el mismo plano (MAZZOTTA, 1990)

c) Nuestra visión actual

Considerando las encuestas experimentales de MADDAUS, las conclusiones de DE LANDSHEERE y algunos teóricos de la creatividad, hemos decidido adoptar el siguiente esquema en relación a la taxonomía.



El esquema muestra que:

- Las tres capacidades superiores se encuentran en el mismo plano.
- Existe como mínimo una progresión del análisis a la síntesis.
- La síntesis no se activa seguidamente a la aplicación (como hubiera demostrado MADDAUS, op cit).
- La síntesis requiere de un análisis previo.
- La evaluación y la síntesis interactúan (MAZZOTTA, op. cit)

d) Aportación de CERT en relación a la medición del análisis

El abordaje CERT contribuye a medir con mayor sistematicidad la capacidad de análisis, el cual ocupa un lugar "delicado" en la clasificación.

Análisis supone la capacidad de buscar elementos y sus relaciones no evidentes. Se trata del nivel más profundo de

comprensión. Es por ello que resulta en ocasiones difícil distinguir cuando se trata de un resultado de comprensión o de análisis.

La *comprensión* supone un "comprender" intuitivo pues todos los elementos se dan de modo evidente. Contrariamente el *análisis* supone un comprender consciente pues es necesario detectarlo, identificar los aspectos (no evidentes) del problema y las relaciones existentes entre ellos (MAZZOTA, op. cit)

Ejemplo 1: "Dejó caer la cabeza hacia atrás, posteriormente se mesó las sienes con las yemas de los dedos."

Del análisis de esta descripción se deduce que el participante se encuentra cansado (lo cual no se explicita).

Ejemplo 2: "Cierra la ventana, por favor".

El análisis de los aspectos implícitos en el texto nos hace suponer que: Existe una habitación que tiene una ventana, aunque es posible que haya más de una.

- Al menos dos personas se encuentran en la habitación aunque es posible que haya más.
- La persona que habla no puede o no desea cerrar la ventana por algún motivo.
- El que escucha puede, aparentemente, estar en condiciones de hacerlo, puede del mismo modo no estarlo o no desearlo y el que habla no lo sabe.

Sugiriendo el principio de las soluciones generales implícitas en las QCM el abordaje CERT proporciona un instrumento sistemático de medición de la capacidad de análisis.

e) El abordaje CERT en la medida de la evaluación

situando sistemáticamente al alumno en situación de juzgar la calidad de sus propios resultados, CERT proporciona a alumno y formador datos sobre la calidad de su juicio; en relación a un punto en particular este hecho es cierto: el conocimiento de su propia competencia.

ANEXO 2: EL R.BIS (1)

2.1. El r. bis "CLASICO"

Se calcula una correlación biserial ($r_{:bis}$) para cada uno de los modos en los que se puede responder a un determinado ítem: Omisión, respuesta correcta, error 1, error 2, error 3, etc.

Se trata de calcular en cada ocasión la concordancia entre el modo de responder a un ítem y las puntuaciones obtenidas en una prueba "de referencia".

Del mismo modo se trata de averiguar si la diferencia entre los que han respondido de un modo determinado (por ejemplo, seleccionando la respuesta correcta) y los que han respondido de otro modo a un determinado ítem concuerda con otras diferencias o clasificaciones reputadas más válidas que un único ítem.

a) El principio.

Imaginemos que un profesor de matemáticas diseña una prueba escolar de cinco ítems en relación a la materia del programa y la presenta a ocho alumnos a los que imparte las clases. Por otro lado, un mes antes, estos alumnos habían realizado un test standard de matemáticas de 100 preguntas. Dicho test standard se haya reputado como muy válido ya que cubre toda la porción de materia considerada, necesita cuatro horas para su aplicación, debe corregirse de manera objetiva y la clasificación resultante de alumnos puede considerarse como un criterio de referencia fiable. Cada uno de los ocho alumnos a obtenido en este test una puntuación que será utilizada como criterio exterior y retomada en una clasificación.

(1) Extracto adaptado de D. LECLERCQ. *Qualité des questions et signification des scores*, Bruxelles: Labor, 1987

Imaginemos que el profesor dispone en relación a cada alumno por una parte, de la puntuación total obtenida en el test standard (criterio externo) y por otra de su comportamiento en cada uno de los cinco ítems de su prueba escolar.

Puede verificar si la diferencia establecida por un ítem (repartición de los alumnos en dos grupos: por ejemplo los que han proporcionado otras respuestas) concuerda con la clasificación obtenida en el test standard.

De este modo, si uno de los cinco ítems es resuelto correctamente por un sólo alumno y si éste obtiene precisamente el mejor resultado en el test matemático (criterio externo) entonces la concordancia entre las dos puntuaciones es perfecta.

El r . bis puede tomar valores que oscilan entre -1 y $+1$.

Si los que han proporcionado la respuesta correcta a la pregunta son alumnos más competentes (en relación al criterio externo) que los que no aciertan, entonces el ítem concuerda con este test tomado como criterio (externo) de medición de la competencia, y el r . bis será positivo (superior a 0)

Si las diferencias en relación a la competencia en el criterio externo son elevadas, entonces el r . bis será tanto más elevado (tanto más cercano a 1).

Este razonamiento puede ser tomado en relación a cada una de las respuestas posibles (por ejemplo para cada una de las soluciones propuestas en una QCM). Para una respuesta incorrecta se espera un r . bis negativo más cercano a -1 cuanto más "débiles" (en relación al criterio externo) sean las respuestas de los alumnos que la han elegido.

b) La representación gráfica

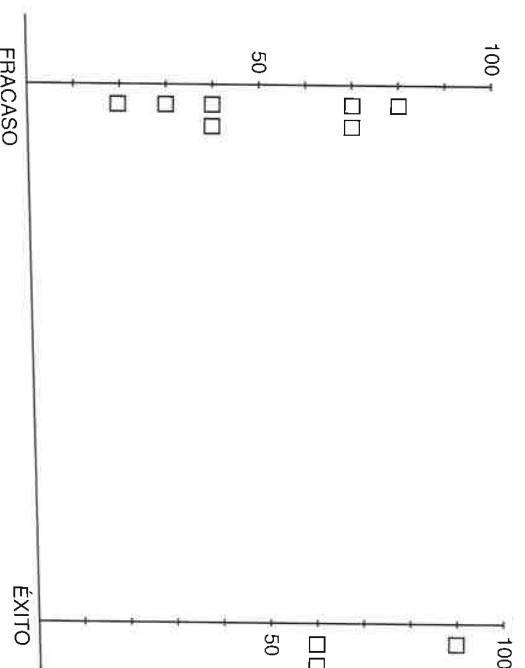
Esta relación de concordancia puede expresarse de modo gráfico aplicando el procedimiento siguiente en rela-

ción a cada ítem. Tomaremos como ejemplo el ítem X y la *respuesta correcta* en relación al mismo.

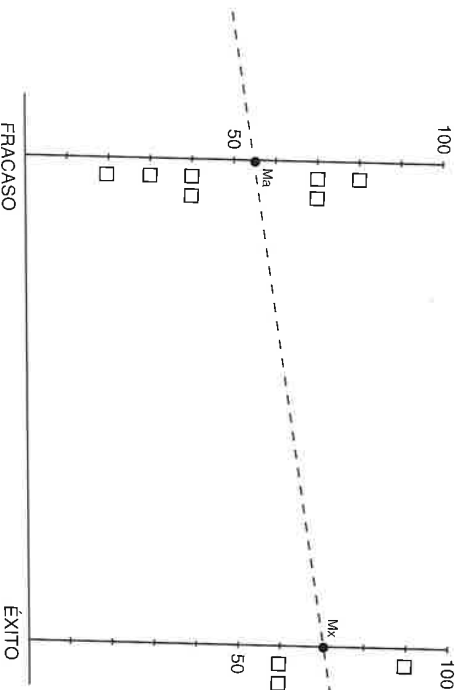
1. A partir de una **misma** línea de base horizontal, levantar dos escalas perpendiculares de **idéntica** altura. Esta altura corresponde a la puntuación **máxima** que resulta posible obtener en el **test criterio** (en este caso 100 puntos). Estas dos escalas son graduadas.

2. Considerar el ítem X puesto por el profesor. Sobre la escala perpendicular de la izquierda se colocan las puntuaciones obtenidas por los alumnos **que no han respondido correctamente al ítem x** en el test criterio y sobre la escala derecha las puntuaciones (en el mismo) obtenidas por los **que sí han contestado correctamente a dicha pregunta x**.

A título de ejemplo, supongamos un gráfico tal, establecido en base a un ítem acertado por tres alumnos y fallado por siete ($p = 0,3$ y $q = 0,7$). Para los tres primeros alumnos las puntuaciones en el test standard eran 60, 60 y 90. Los tres valores se colocan en la escala derecha. Las puntuaciones obtenidas por los alumnos que han contestado erróneamente al ítem 20, 30, 40, 40, 70, 70 y 80 se sitúan sobre la escala izquierda.



3. Se calcula entonces la media M_x de puntuaciones de la escala de la derecha y la media M_a de puntuaciones en la escala izquierda.
4. Se colocan ambas medias en el gráfico (M_x sobre la escala derecha y M_a sobre la de la izquierda).
5. Se unen mediante una recta los dos puntos (M_a y M_x). Si la recta obtenida asciende de izquierda a derecha hay concordancia entre el ítem y el criterio: se dice también que la relación resulta positiva pues el coeficiente de pendiente de la recta es positivo. Si la recta obtenida desciende de izquierda a derecha entonces hay discordancia entre el ítem y el criterio; la relación es negativa, y el coeficiente de pendiente de esta recta negativo. Si la recta es horizontal no hay concordancia, se dice también que la relación es nula (coeficiente de pendiente nulo).



En nuestro ejemplo, la media (M_x) de los tres alumnos que han acertado es 70, mientras que la de los alumnos restantes es de 50. Estos dos valores se representan mediante cruces sobre la gráfica. Si se une M_a y M_x se obtiene una recta **ascendente**, de pendiente positiva (dibujada mediante una línea discontinua). Existe pues concordancia entre las medidas obtenidas en el ítem y las obtenidas en base al test criterio.

El signo de la pendiente de la recta, o de la relación, viene dado por el signo de la expresión matemática $M_x - M_a$.

c) Interpretación de la representación gráfica

En el ejemplo precedente, la diferencia resulta positiva, lo cual significa que los que han acertado el ítem tienen una media (en el test criterio) más elevada que los demás. Este resultado resulta conforme con lo esperado, es lógico que la pregunta de la prueba escolar construida por el profesor proporcione resultados concordantes con los del test estándar.

En este ejemplo, las puntuaciones son muy dispersas: la desviación típica es elevada (D.T.). Cuando la desviación típica es elevada es posible observar entre M_x y M_a una diferencia más importante que si los resultados hubieran estado menos dispersos. Esto ocurre porque el valor numérico del índice de discriminación tiene en cuenta dicha desviación típica, expresando la diferencia $M_x - M_a$ en relación a la D.T. en la expresión:

$$\frac{M_x - M_a}{D.T.}$$

Por otra parte, la pendiente de la recta (en trazo discontinuo) no depende exclusivamente de la amplitud de la diferencia $M_x - M_a$, sino también de la distancia que separa las dos verticales. El índice de correlación biserial puntual toma también este aspecto en consideración. La fórmula del coeficiente de correlación punto biserial, o r_{pb} , donde aún r_{bis} es:

$$Mx - Ma \quad \sqrt{pq}$$

D.T.

dónde:

P vale la proporción de elección de la solución considerada (ej: 0.30)

Q vale la proporción complementaria de no elección (ej: 0.70)

2.2. EL TOTAL DEL TEST COMO CRITERIO

en la mayoría de los casos, el profesor no dispone de puntuaciones de alumnos en relación a un test-criterio válido. Se utiliza entonces la misma fórmula, pero se toma como *criterio* la puntuación *total* en el test en el que el ítem se encuentra comprendido. La concordancia entre la pregunta y el total de la prueba supone entonces un índice de coherencia interna entre el ítem y el test. La suma (poderada) de los r. bis puede entonces servir para calcular un coeficiente de consistencia interna o de *homogeneidad* del test en su conjunto.

a) La correlación automática

Cada ítem de un test contribuye a la puntuación total por una parte igual a 1/NQ cuando todos los NQ ítems resultan de igual facilidad. Se dice que se da "recubrimiento" o (inclusión) entre la puntuación del mismo y la puntuación total del test.

En el caso, puramente teórico en que los ítems de un test no tuvieran ninguna *correlación* entre ellos (ej: = 0), la correlación entre un ítem y el total del test vale automáticamente 1/.....NQ. De este modo para NQ=9, r. bis vale 1/3 o 0.33. Es lo que recibe el nombre de *Valor de referencia* del r. bis de la respuesta correcta a un ítem. (de hecho, se puede calcular tal valor referencia para cada solución de cada ítem)

En consecuencia, todo r. bis que se refiera a una respuesta correcta que no alcance al menos este valor-referencia no puede considerarse como índice de una buena correlación del ítem respecto a la prueba entera. Todas las correlaciones ítem-total son pues aumentadas en base a esta relación fundamental.

Partiendo de este mismo razonamiento HENRYSSON (1963) propone corregir (ver la parte inferior) la correlación "sobrestimada" mediante la fórmula siguiente:

$$C_{rpbis} = \left(\sqrt{\frac{n}{n-1}} \right) \left(\frac{r_{pbis}}{\sqrt{\sigma^2 - pq}} \right)$$

con crp.bis = r punto bis del ítem corregido por "recubrimiento" (Overlap)

σ^2 = Desviación típica de los resultados obtenidos en el total de la prueba.

b) Los r. bis de respuestas correctas y distractores

En una QCM de K soluciones, resulta posible calcular un índice de discriminación (r. bis) para cada una de los K+1 comportamientos posibles: (omisiones, elección de la solución 1,2...k).

El r. bis posibilita comparar alumnos que han adoptado uno de estos comportamientos (digamos x) con otros alumnos que adoptaron otros comportamientos (designados con la letra A).

En un test bien diseñado, el r. bis de la solución correcta resulta normalmente positivo y los r. bis de las soluciones incorrectas (distractores) son normalmente negativos.

Una QCM cuyo r . bis en relación a *la solución correcta* resulta *positivo* es un ítem que participa en la homogeneidad de la prueba pues mide lo mismo que las otras preguntas y armenuendo esta correlación (biserial puntual) es elevada.

Los distractores, presentan normalmente una correlación (r . bis) *negativa*.

La omisión presenta generalmente una correlación comprendida entre la presentada por los distractores y la de la respuesta correcta. Tanto sea positiva como negativa, la correlación (r . bis) en la omisión a menudo se aproxima a 0.

c) Explicación del valor r .bis

Un r . bis positivo en relación a una solución significa que son los mejores alumnos (en el total del test) han seleccionado esta solución. sería pues paradójal que el r . bis de un distractor resultara positivo. si se da el caso, resulta importante cuestionarse qué ha producido semejante resultado.

El resultado puede ser debido a:

- **Un defecto fundamental**
 - a) *Falta de validez*: El ítem no mide lo que se creía.
 - b) *Falta de homogeneidad* entre la prueba y el ítem, que mide otra cosa.

- **Un defecto formal**

Redacción: mala formulación, distractores no pertinentes, ambiguos, etc...

- **Un accidente**

Error de codificación: Se indicó como respuesta correcta una solución incorrecta.

Es el sentido *clínico* del profesor lo que le posibilitará realizar un diagnóstico y proceder a una adecuada remediación del ítem.

Un valor paradójal del índice de discriminación no és más que un *signo* de mal funcionamiento; no proporciona una explicación.

d) Ejemplos

Analicemos tres de los ítems de una prueba: (1)

Item 1:

SOL:	0	1	2	3	4	5	6	7	8	9
Q 1	0.0	0.0	15.8	0.0	0.0		*84.2	0.0	0.0	0.0
Rbis	0.00	0.00	-0.09	0.00	0.00		0.09	0.00	0.00	0.00
Cmoy	0.00	0.00	64.17	0.00	0.00		90.47	0.00	0.00	0.00

La respuesta correcta* (6 o NINGUNA) a atraido al 84,2 % de alumnos que han respondido con un nivel de confianza del 90%. Son alumnos ligeramente superiores a la media (0.09: resulta positivo, pero poco alejado de cero).

Tan sólo se ha observado otro modo de responder: la solución 2 (elegida por el 15,8% de alumnos, con un nivel medio de confianza más débil: 64%).

El r . bis es evidentemente - 0.09 pues los dos grupos de alumnos se comparan mediante la fórmula (Mx - Ma).

(1) Solución 0 = omisión
* Respuesta correcta

ítem 2:

SOL:	0	1	2	3	4	5	6	7	8	9
Q 1	0,0	0,0	0,0	10,5		0,0	10,5*	63,2	15,8	
Rbis	0,00	0,00	0,00	-0,23		0,00	0,26	0,23	-0,32	
Cmoy	0,00	0,00	0,00	87,50		0,00	87,50	82,92	88,33	

La respuesta correcta* (8) atraído al 63% de los alumnos con una confianza media del 82% y un r. bis positivo (como estaba previsto).

Respuestas (incorrectas):

- (3) Ha atraído al 10% de alumnos (cert = 87,5%) el r. bis resulta negativo (-0,23)
- (7) Ha atraído al 10,5% de alumnos (cert=87,5%) y su r. bis resulta positivo (0,26) lo cual resulta ANORMAL. Cuestionarse este hecho.
- (9) Atraído al 15,8% de alumnos (cert= 88,3%), su r. bis es negativo (- 0,32).

Esta pregunta tiene un (pequeño) defecto a nivel de la solución 7.

Item 3:

SOL:	0	1	2	3	4	5	6	7	8	9
Q 1	0,0	26,3	5,3	0,0		0,0	5,3*	63,2	0,0	
Rbis	0,00	0,50	-0,21	0,00		0,00	-0,02	-0,35	0,00	
Cmoy	0,00	93,00	77,50	0,00		0,00	97,50	92,08	0,00	

El r. bis de la solución correcta* (8) resulta negativo (0,35). Es ANORMAL y debería ELIMINAR el ítem de la prueba. Si se trata de mejorar dicho ítem a fin de realizar una mejor utilización posterior, es en base a la solución 2, que ha atraído a los mejores alumnos (r. bis = 0.50, así pues positivo) por el que resultaría conveniente inclinarse.

ANEXO 3: INDICE DE REALISMO (1)

a) Principio general

¿Cuál es la proximidad de CADA tasa de exactitud (TE) y del valor central (VC) en relación al intervalo de confianza correspondiente? La línea quebrada trazada entre los diferentes rectángulos del gráfico se encuentra cercana a la diagonal? Cada "desviación" (vertical) entre el rectángulo y la diagonal se considera como un error de estimación y es tan "grave" como importante en el sentido en que representa un gran número de respuestas (1).

b) Principio de la fórmula de cálculo

La fórmula de este índice está basada en los errores de medición (EM), es decir, para cada nivel de confianza, la diferencia (en valor absoluto) entre la tasa de exactitud (es decir la diagonal).

El hecho de que este error de medición se represente en valor absoluto convierte al índice de realismo sensible a todas las variaciones, sin posibilidades de compensación de un error por otro.

La media de estos errores de medición, MEM, representa el no-realismo.

$$MEM = \frac{\sum_i (TE_i - |VC|) \cdot NR_i}{N} \quad \text{O} \quad \frac{\sum_i |EM_i| \cdot NR_i}{N}$$

(1) Extracto adaptado de D. LECLERCO, auto-evaluación y conocimiento parcial, octubre 1990, Servicio de Tecnología de la Educación.

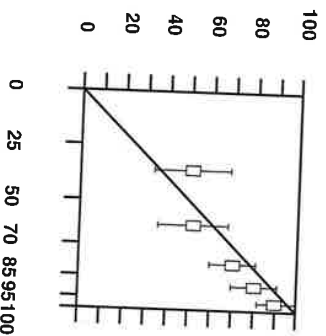
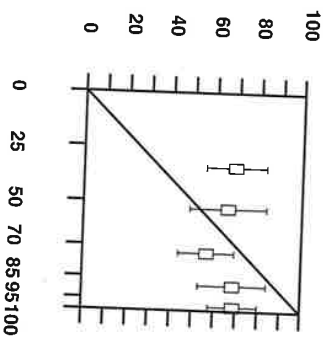
Este valor, multiplicado por 5 ha sido sustraído de 1 (la perfección). Así pues:

$$\text{REALISMO} = 1 - (5 \text{ MEM})$$

c) Dos ejemplos

Realismo: 0.759
 Coherencia: 0.748

Realismo: 0.961
 Coherencia: 0.941



d) Datos estadísticos (en base a 311 alumnos universitarios sujetos del estudio)

Máximo observado = 1
 Mínimo observado = -1,67
 Media = 0,78
 Desviación Típica = 0,26

e) Interpretación del índice:

Recomendamos interpretar del modo siguiente los valores numéricos de realismo:

REALISMO		% de 311 alumnos universitarios
IDEAL si	= 1	1%
EXCELENTE	> .95	15%
BUENO	entre .91 y .94	18%
DEBIL	entre .84 y .90	22%
MEDIO	entre .70 y .83	22%
INSUFICIENTE	si menos .70	22%
MINIMO si	-3.75	

El mínimo se alcanzaría cuando un alumno habiendo respondido todos los ítems de modo incorrecto con el nivel de confianza más elevado (97,5 % del valor central) cuyo error medio es de 0,975 y la MEM de 0.95.

f) Interés del índice

Este índice es el que mejor revela la calidad de la auto-evaluación, 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?

ANEXO 4: INDICE DE COHERENCIA (1)

a) Principio General

¿Las tasas de exactitud (TE) crecen regularmente?

¿Los rectángulos dibujados sobre el gráfico se alinean (aproximadamente) sobre una recta y en qué medida esta recta es paralela a la recta ideal? La TE de un nivel débil de confianza es inferior al TE de niveles de confianza más elevados?

Se hubiera podido denominar a dicho índice "nivel de ALINEAMIENTO" sobre la recta. Se trata en efecto de calcular la correlación entre los puntos, que puede visualizarse gráficamente. Se dibuja la recta (2) que pasa "lo más cerca posible del centro" de seis puntos por cada gráfica y se puede observar en qué medida estos seis puntos se acercan o se alejan de esta recta. ¡Atención! la interpretación visual puede conducir a error si se olvida que cada TE no tiene el mismo peso, éste último depende del número de observaciones sobre las que se calculó.

El índice podría resultar negativo si la pendiente de la recta fuera negativa. (si descende), lo que constituye una extrema falta de coherencia.

(1) *Estracto adaptado de D. LECLERCQ - Auto-evaluación y conocimiento parcial, octubre 1990, Servicio de Tecnología de la Educación*

(2) *Recordemos que cada recta puede expresarse mediante la fórmula $y = ax + b$ donde:*

x = posición sobre el eje horizontal (abscisa)

y = posición sobre el eje vertical (ordenadas)

a = coeficiente de pendiente de la recta. Si $a = 1$, la pendiente 45° , como la diagonal.

b = El lugar del eje vertical que la recta corta cuando $x = 0$

b) El principio de la fórmula de cálculo

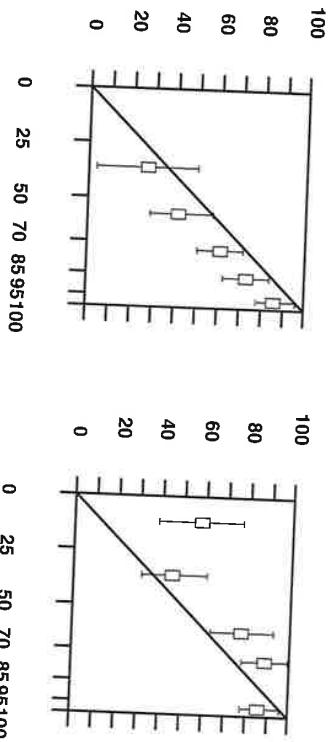
Se calcula la correlación entre las tasas de exactitud (TE) y los valores centrales de los X niveles de confianza (en este caso 6 niveles). El índice varía entre -1 (alineamiento perfecto, pero con pendiente hacia la derecha, lo que resulta inesperado) y +1 (alineamiento perfecto con pendiente positiva, es decir alineación "creciente").

La fórmula de base para el cálculo del coeficiente de correlación de BRAVAIS-PEARSON (también llamada r) puede localizarse en cualquier manual de estadística.

c) Dos ejemplos

Realismo = 0,941
Coherencia = 0,997

Coherencia = 0,893
Realismo = 0,861



d) Datos estadísticos (en base a 311 alumnos universitarios sujetos del estudio)

Máximo observado = 0,99
Mínimo observado = -0,92
Media = 0,78
Desviación Típica = 0,29

e) Interpretación del índice:

Recomendamos interpretar del modo siguiente los valores numéricos de COHERENCIA:

COHERENCIA	% de 311 alumnos universitarios
IDEAL si EXCELENTE	0%
BUENO	13%
DEBIL	18%
MEDIO	28%
INSUFICIENTE	18%
MINIMO si	23%
	0%

f) Interés del índice

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 la que comporta la infravaloración o la sobrevaloración.

INTERESES DEL PRESENTE PROYECTO COOPERATIVO PARA EL MUNDO DE LA FORMACION PROFESIONAL

Intereses del formador

Asegurar la formación de profesionales es una tarea compleja, quizá una de las más complejas que puedan existir, y que exige un alto nivel de profesionalismo.

Hay un réfran que dice: "Se reconoce al profesional por sus herramientas". En efecto, las herramientas reflejan la experiencia acumulada por los profesionales anteriores. Los formadores necesitan herramientas para CONCEBIR sus evaluaciones, CREAR SITUACIONES de evaluación, SOLICITAR respuestas, ANALIZARLAS e INTERPRETARLAS.

El presente proyecto procura facilitar herramientas (conceptuales, metodológicas, informáticas) a los formadores con vistas a ayudarles a llevar a cabo estas distintas operaciones.

Intereses del estudiante

Los procedimientos recomendados en este libro facilitan al estudiante datos valiosos sobre su aprendizaje. Esto es muy importante dado que necesita "situarse" para progresar mejor.

Al igual que el niño que aprende a andar recibe datos procedentes de los músculos, de los ojos, de la oreja interna (sentido del equilibrio), es esencial que el estudiante reciba mensajes que le permitan ajustar sus esfuerzos.

El presente proyecto puede ofrecer al estudiante datos para diferenciar entre su memorización y su comprensión de la asignatura, su capacidad para analizar un problema y aplicarle sus conocimientos, su capacidad para valorar sus aptitudes y arriesgarse.

* *

*

El formador y el estudiante que consideren que se debe mejorar tanto