

www.universitaria.cl

Dieudo LECLERCQ



Álvaro CABRERA MARAY



UNIVERSIDAD
DE CHILE



Directores de la publicación:

Dieudonné Leclercq
Universidad de Liège (ULg)

Álvaro Cabrera Maray
Universidad de Chile (UCH)

IDEAS e INNOVACIONES
Innovaciones en Dispositivos de Evaluación
de los Aprendizajes en la enseñanza Superior
2014

Se pueden bajar gratuitamente
desde <http://orbi.uliege.be>, después Leclercq D., o
desde www.evaluaraprendizajes.cl

- Los **resúmenes** de los 23 capítulos
del libro IDEAS <http://hdl.handle.net/2268/173543>
- El **índice** de este libro para buscar entre
entradas de 1500 conceptos y
400 de autores <http://hdl.handle.net/2268/180060>

Dieudonné Leclercq

Dr. en Educación (1975) en « La Metacognición vía la autoevaluación con grados de certeza » y con postdoctorales en las universidades de Pittsburgh y UCLA. Fue profesor en las Universidades de Namur (1975-1980) y de Liège (1980-2010). Es emérito desde 2010. Enseña como invitado en las Ues. de Liège y Paris 13. Recibió el título de *Honorary Member of the World Cultural Council* (México). Ha colaborado, en Chile, con la U de Chile (UCH -Santiago), la UMCE, la UCT (Temuco), la UC del Maule, la UNAB y la UCSC (Concepción). En Perú con la PUCP y el SINEACE (Lima), la UNSAAC (Cusco) y la UNTRM (Chachapoyas). En México con la U A Chapingo. En España con la U de Sevilla y la U de Deusto (Bilbao). d.leclercq@uliege.be

Álvaro Cabrera Maray

Licenciado en Artes mención Teoría de la Música, y Master en Pedagogía en Educación Superior de la U. de Liège (Bélgica). Ha sido profesor en la Facultad de Artes y en Cursos de formación General, trabajando en el Depto. Estudios de Pregrado de la U. de Chile a cargo del Área de Formación. Integró la Red nacional de Centros de Enseñanza-Aprendizaje y la de expertos SCT-Chile sobre sistema de créditos transferibles. Trabajaba en el Ministerio de Educación de Chile, coordinando los programas de la reforma educacional en Educación Superior. alvarocabreramaray@gmail.com

Contenidos del libro IDEAS:

ES: Calificación ; Evaluación ; Productos ; Meta-cognición ; Resolución de problemas ; Proyectos ; Trabajo de grupo ; Portafolio ; Vigilancia cognitiva ; Pruebas de Progreso ; Taxonomía de Bloom ; Auto-evaluación ; Grados de certeza ; Test de Concordancia de Script ; Retroinformación ; calidades ; validez

EN : Assessment ; Evaluation ; Outcomes ; OSCE ; MCQ ; PARMs ; Metacognition ; Problem solving ; Projects ; Group produced work ; Portfolio ; Cognitive vigilance ; Progress Tests ; Bloom's Taxonomy ; Self-assessment ; Confidence Degrees ; Concordance Script Test ; Feedbacks ; Edometrics ; Metacognitive Spectral Test ; ETIC PRAD ; quality ; validity

FR : Notation ; Evaluation ; Résultats ; ECOS ; QCM ; PARMs ; Métacognition ; Résolution de problèmes ; Projets ; Travail de groupe ; Portfolio ; Vigilance cognitive ; Tests de progression ; Taxonomie de Bloom ; Auto-évaluation ; Degrés de certitude ; Test de Concordance de Script ; Rétro-information ; Edumétrie ; Test Spectral Métacognitif ; qualités d'une évaluation ; validité d'une mesure

IDEAS = Innovaciones en Dispositivos de Evaluación de los Aprendizajes en la educación Superior

La lista de los capítulos y el resumen de cada uno

aparece a continuación después de este capítulo.

CAPÍTULO XV

Evaluar procesos cognitivos según la Taxonomía de Bloom

DIEUDONNÉ LECLERCO

A. La taxonomía de los procesos cognitivos según Bloom¹¹² *et al.* (1956)

A.1. Un apilamiento

Aunque se llama “Taxonomía de los *objetivos* (ámbito cognitivo)”, pienso que esta obra trata de los *procesos mentales cognitivos*, lo que admiten Anderson y Krathwohl (2001, p. 5). La clasificación o taxonomía de Bloom *et al.* (1956) se presenta como un apilamiento de seis niveles de procesos mentales. En 1971 Bloom *et al.* ilustraron aplicaciones de estos procesos mentales en varios ámbitos de la educación, mediante variados ejemplos.

La taxonomía se basa en el principio (la hipótesis) de que el nivel de abajo es necesario para alcanzar el nivel superior. Esta relación (dependencia) entre procesos mentales puede verificarse de varias formas. Por ejemplo, cuando un extranjero lee un diario de un país, no puede *comprender* (nivel 2) lo que significa la frase: “El señor X ha rechazado los argumentos de la señora Y en el asunto W, recordando que su partido prometió votar contra la ley 22”. El extranjero no puede comprender porque no *sabe* (nivel 1) quiénes son X e Y ni cuáles son sus partidos (que parecen antagónicos). Tampoco sabe qué son el asunto W y la ley 22.

Del mismo modo, una persona no puede *aplicar* (nivel 3) la fórmula de cálculo del área de un triángulo si no *comprende* (nivel 2) lo que son un triángulo, la altura y la base.

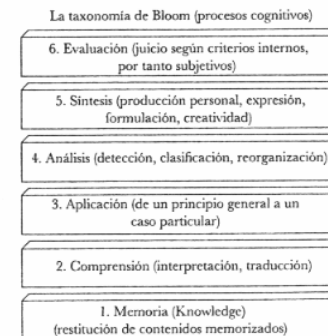


Figura 1: Apilamiento de los procesos mentales (del ámbito cognitivo) según Bloom *et al.* (1956)

¹¹² Benjamin Bloom (1913-1999), de la Universidad de Chicago, recibió en 1994 el Doctorado *Honoris Causa* (presentado por D. Leclercq) de la Universidad de Liège, con la cual trabajó durante años, especialmente desarrollando el IEA (International Association for the Evaluation of Educational Achievement), prefigurando PISA.

A.2. Un ensamblaje.

El mismo principio puede ser representado, ya no como un apilamiento, sino que en la forma de un ensamblaje.

La Figura 2 ilustra esta idea. Los procesos mentales aparecen en la mitad izquierda y constituyen capacidades, es decir, aplicaciones de procesos mentales (acciones) a contenidos (saberes) específicos. Las capacidades pueden ser entendidas como dominio de recursos o como resultados de aprendizajes. Lo primero ocurre cuando involucran procesos cognitivos simples (niveles 1, 2 y 3) en situaciones poco auténticas (alejadas de la realidad profesional o disciplinar; muy “escolares”). Lo segundo ocurre cuando demandan procesos cognitivos superiores (niveles 4, 5 y 6), exigen integración de saberes, y se desarrollan en situaciones experimentales o complejas organizadas con la intención de acercarse a la realidad profesional o disciplinar.

Las capacidades son a su vez componentes de las competencias. Para efectos de la evaluación, lo anterior implica que en general es la medición del logro de los resultados de aprendizaje lo que nos permite inferir sobre el desarrollo de las competencias. La evaluación de las competencias propiamente tales es un desafío mayor, debido al carácter complejo de estas y a su condición de orientador del programa global de la carrera, más que de cada una de las asignaturas. Aun así, las competencias pueden ser evaluadas con instrumentos como los mencionados en la sección derecha de la Figura 2, y que se abordan en otros capítulos de este libro.

El presente capítulo se dedica únicamente a la evaluación de los procesos mentales en performances aislables, es decir, no incluidas en situaciones complejas (como ocurre en la evaluación de los resultados de aprendizaje y las competencias).

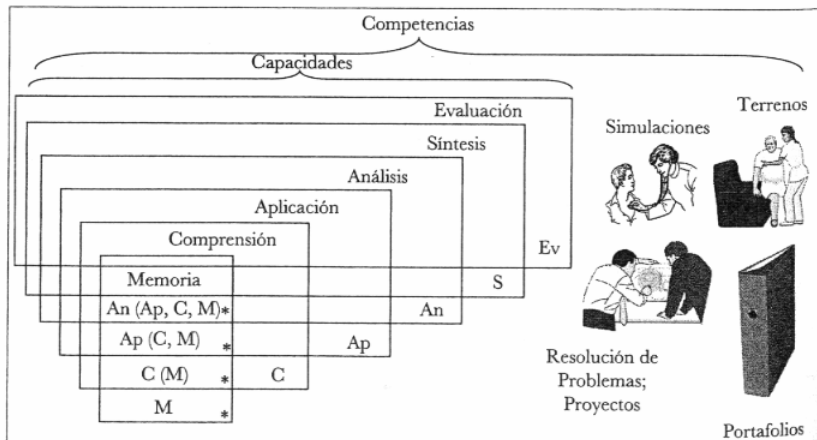


Figura 2: Representación en ensamblaje de los niveles de procesos mentales de la taxonomía de Bloom

La Figura 2 ilustra que:

- (1) Las competencias incluyen capacidades, y es un reto evaluar ambas en interacción (con frecuencia son evaluadas separadamente).
- (2) Aunque es fácil evaluar solo la memoria (celda M), es más difícil evaluar solo la comprensión (celda C), sin la memoria. A menudo se evalúa el conjunto de las dos (celda C (M)). Exámenes con libro abierto contribuyen a evaluar la comprensión aislada (celda C). Lo mismo con la aplicación: un fracaso en una situación “de aplicación” puede ser atribuido a una incapacidad de aplicar solo si se puede excluir que sea resultado de una falta de memoria o de comprensión.

La autoevaluación (por ejemplo con grados de certeza) ha sido representada (*) en todas las celdas porque se puede (¿se debe?) añadir en la evaluación de cada proceso (ver Capítulo 16).

B. Evaluar la memorización

En primer lugar, justificaremos la distinción entre el reconocimiento (la re-cognición) y la evocación, como procesos distintos en la categoría Memorización, tal como hace la taxonomía de procesos mentales de Bloom *et al.* (1956).

B.1. Reconocer (re-conocer)

A) ¿CUÁNDO SE NECESITA RECONOCER?

En muchas situaciones es suficiente reconocer una cosa, sin ser capaz de usarla o de producir una descripción detallada de ella. Los siguientes tres ejemplos ilustran este tipo de situación:

- (1) En geografía, muchas personas son capaces de reconocer el contorno del *mapa de África*, pero no sabrían dibujarlo (evocarlo) sin verlo, o cometerían grandes errores de distancia relativa. Esto no es importante pues en la vida real, para la enorme mayoría de las personas, es más importante viajar con los ojos, reconociendo formas en un mapa ya existente, que dibujar uno de memoria.
- (2) El vocabulario pasivo es aquel que una persona reconoce (entiende) pero no usa espontáneamente. Por ejemplo, la palabra “*antónimo*”; es importante que conozcamos su significado porque los diccionarios la usan, aunque en general no la produzcamos espontáneamente, prefiriendo decir “lo contrario”.
- (3) En la cultura general existen muchos hechos que las personas no son capaces de producir o evocar pero que sí reconocen, de modo que es suficiente para entender lo expresado en un diario o en la tv. Así, muchos europeos no sabrían decir cómo se denomina a la casa del Gobierno de Chile, pero cuando oyen (o ven) *La Moneda*, lo recuerdan.

El último ejemplo define para qué sirve el conocimiento pasivo o la memoria “de reconocimiento”: es un conocimiento *que permite entender* mensajes, no producirlos.

B) ¿CÓMO EVALUAR EL RECONOCIMIENTO (LA RE-COGNICIÓN)?¹¹³

En asuntos criminales, un problema que se enfrenta con frecuencia es que el testigo es incapaz de describir (producir, evocar) las características del victimario. En consecuencia, los métodos más conocidos consisten en *presentarle soluciones*, en este caso fotografías de varias personas. Por supuesto, existe un peligro inmenso: que el testigo (o el estudiante) conteste aunque no esté seguro de lo que dice (al azar). Es muy importante para este capítulo hacer notar que hay tres formas de tomar precauciones frente a esta amenaza:

- 1) *multiplicar el número (k) de soluciones ofrecidas*, de modo que la probabilidad de elegir correctamente al azar sea baja (mientras más grande es *k*, más pequeño es $1/k$)
- 2) usar el principio de la Solución General Implícita (SGI) *ninguna*, como otra forma de eliminar la eventualidad de que haya una respuesta al azar
- 3) pedir al estudiante que indique su nivel de *certeza* (ver Capítulo 16)

B.2. Evocar

A) ¿CUÁNDO SE NECESITA EVOCAR?

Existen muchas situaciones que exigen a la persona evocar de memoria, sin acceder a alternativas entre las cuales elegir. A continuación se presentan tres ejemplos.

- (1) *Un intérprete* de conferencias en lengua extranjera debe ser capaz de producir la traducción (las palabras adecuadas que corresponden a las del expositor). Y debe ser de memoria: no se puede interrumpir la conferencia a cada minuto para que el intérprete verifique o aprenda la palabra adecuada.
- (2) *Un cirujano* que va a penetrar un paquete básculo-nervioso (como el del cuello) debe saber con anterioridad la sucesión de órganos que encontrará (evocarlos sin verlos), como la vena yugular, la arteria carótida y después el nervio vago, para así anticipar los peligros de cortar elementos vitales. No es suficiente reconocerlos una vez que los encuentra... ya intervenidos.
- 3) *Un piloto de avión* está en una situación comparable a la del cirujano: debe saber de antemano que existe una velocidad límite bajo la cual el avión cae. No basta con “reconocer” el hecho una vez que se ha producido.

¹¹³ Ver también la “corrección por adivinación” en el Capítulo 13, sección C.5.

Cada educador debe decidir, en su ámbito, cuáles son los hechos y los elementos de conocimiento que los estudiantes deben ser capaces de evocar, y cuáles aquellos en que es suficiente que sean capaces de reconocer.

B) EL ASUNTO DE LA REACCIÓN VS LA ESPONTANEIDAD

En el ejemplo (1), del intérprete, no se presenta el problema de la espontaneidad de la evocación, pues las frases del conferencista exigen al intérprete reaccionar; son como preguntas que debe contestar. Esta situación se llama *evocación estimulada*¹¹⁴.

En el ejemplo (2) el cirujano debe *anticipar*, pero estando en una *situación previsible*, porque está *delimitada* (una parte específica del cuerpo humano). De este modo su atención está (debe estar) enfocada en asuntos específicos; sabe que debe plantearse *espontáneamente* las preguntas apropiadas, y sabe cuándo plantearse. Un “protocolo” de intervención consiste en preguntas “de rutina” que transforman la performance de *evocación espontánea* en una performance de *evocación estimulada*.

El ejemplo (3) del piloto de avión es similar al (2). Las famosas “checklist” sirven para estimular la evocación. Los instrumentos de a bordo están programados para enviar señales y compensar una posible ausencia de espontaneidad en la detección de una anomalía.

Los ejemplos (2) del cirujano, y (3) del piloto de avión, plantean otro asunto además de la evocación: la espontaneidad de la evocación, lo que se aborda en el Capítulo 13, secciones A4 y A5.

C) UNA PRECAUCIÓN DEONTOLÓGICA (DE JUSTICIA) ANTES DE EVALUAR LA CAPACIDAD DE EVOCAR

Ya sea de reconocimiento o de evocación, *los estudiantes deben ser avisados acerca de los contenidos que tienen que memorizar*. A continuación se muestra un ejemplo de lista (Figura 3) y otro de mapa conceptual (Figura 4), que describen a los estudiantes de un curso de D. Leclercq en segundo año de la Universidad de Liège, los contenidos (nombres, conceptos, vínculos) que deben memorizar. La Figura 3 lista los contenidos siguiendo el orden de los capítulos y páginas del libro que sirve de base al curso.

¹¹⁴ *Stimulated recall.*

CURSO PSICOLOGÍA 058 CONTENIDOS A MEMORIZAR			
CAPÍTULO	SECCIÓN	PÁGINAS	AUTORES
INTRODUCCIÓN	B1. Discursos sistémicos (sobre los objetivos)	17-19	
	B2. Discursos sobre los dominios de intervención	19-22	
	B3. Discursos sobre los mecanismos técnicos	23	Bloom (p. 23)
CAPÍTULO 1	A. Un cúmulo de desequilibrios	29-35	Piaget (pp. 30-32) ELKIND (pp. 33-35)
	H. El desarrollo moral	51-52	Kohlberg (p. 51)
	K. La relación padres-adolescentes	54-57	Berne (p. 56)
CAPÍTULO 2	A. Los tipos de memoria y sus funcionamientos	62-67	Anderson (p. 64) TULVING (p. 65)
	C. Leyes del recuerdo de corto plazo	69-74	Los diferentes principios
	D. Capacidad de MCT y CHUNKING	75-77	Miller (p. 75)
	E. Índices de evocación / recuerdo	78-83	Lieury (p. 78) LOS DIFERENTES PRINCIPIOS
	H. Medias mnemotécnicas	88-91	Lieury (p. 89)
	I. Curvas de desarrollo	92-93	
	J. Métodos para medir la retención	94-95	
CAPÍTULO 3	A1. Pavlov: los disparadores aprendidos	102	Pavlov (p. 102)
	A2. Explicación de los reflejos pavlovianos	102	
	A3. Condiciones de aprendizaje de los reflejos	102	
	A4. ¿Dónde y cuándo se aprenden los reflejos condicionados?	103	
	A5. ¿Cómo se constituyen las cadenas de disparadores?	103	
	B9. Reducir el estrés mediante información procedimental y sensorial	108	
	B10. El modelamiento	109	Bandura (p. 109)
	D1. La pirámide de Maslow	116	Maslow (p. 116)
	D2. La clasificación de Mc Clelland	116	Mc Clelland (p. 116)
	D4. Necesidades y deseos	118	
	D5. ¿Está Maslow en lo cierto?	118	Albertini (p. 118)
F. El modelo de Viau	125-127	Viau (p. 126)	

Figura 3: Lista de contenidos a memorizar

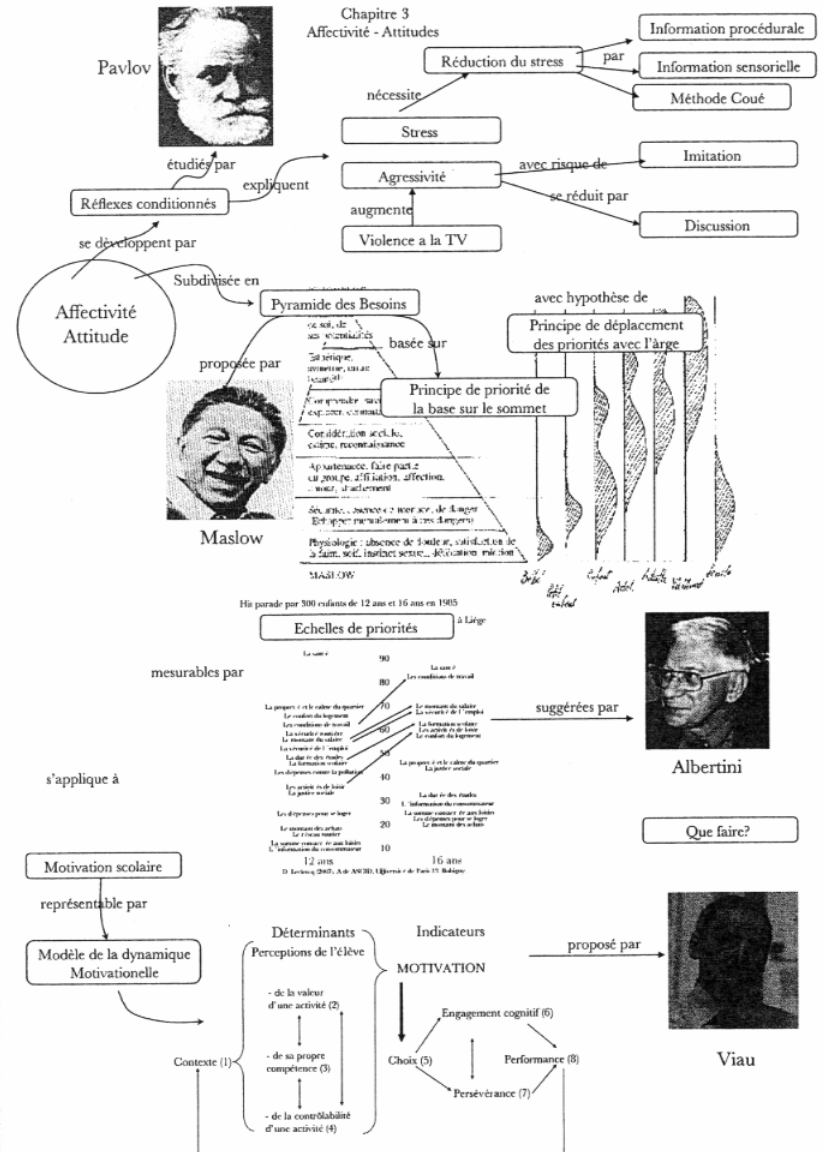


Figura 4: Mapa conceptual de los contenidos a memorizar (se entrega esta imagen para ilustrar la idea y no con la intención de explicar los contenidos involucrados; por esta razón se mantuvo el idioma original y se admitió imágenes de baja resolución).

D) TÉCNICAS DE EVALUACIÓN DE LA EVOCACIÓN

Tabla 1: Modos clásicos de evaluación de la evocación

Para verificar la capacidad de evocar...	...se usa la técnica de	Ejemplo(s)
...hechos aislados (y aislables)	Frase que completar o dibujo	Julio César fue asesinado en el año... Dibuje el sistema digestivo de un ser humano.
...principios	Frase que completar o dibujo	El peso de un objeto puede definirse como..... Dibuje el sistema digestivo de un ser humano, indicando con flechas las interacciones de órganos con el bolo alimenticio
...secuencias	Descripción en orden	Describa, en orden, las etapas de... numerándolas.

E) LAS PSN (PREGUNTAS CON SOLUCIONES NUMEROSAS): EVALUAR LA EVOCACIÓN CON AYUDA DE LA LECTURA ÓPTICA

El día del examen los estudiantes reciben una extensa lista de las posibles respuestas, con varios cientos de opciones en orden alfabético (Figura 5). De este modo, no es posible para el estudiante reconocer de antemano la respuesta correcta entre las opciones (son demasiadas), por lo que se espera que este evoque una respuesta y luego la busque en el listado de posibilidades.

21	Abductor de agua	75	Conciencia	129	Ejemplos
22	Abstracción	76	Consecuencia	130	Etnología
23	Absurdo	77	Consolidación	131	Eventos de Aprendizaje
24	Acción (guiones)	78	Convencional	132	Exégeta
25	Actitud	79	Correction for guessing clásica	133	Experto
26	Adquisición	80	Cortex	134	Fábula personal
27	Adulto	81	Coué	135	Factores
28	Afasia	82	Cuerpo caloso	136	Fatiga
29	Afectividad	83	Curva de Gauss	137	Feeling of knowing (FOK)
30	Afiliación	84	Curva en J	138	Fiabilidad
31	Agresividad	85	Curvas de desarrollo	139	Fluidez
32	Albertini	86	Curriculum	140	Formales
33	Álgebra (eB dux)	87	D'Hainaut	141	Freud
34	Análisis transaccional	88	Damasio	142	Frustración
35	Análisis	89	De Finetti	143	Gagné
36	Anderson	90	Decisión	144	Géneros superiores
37	Anorexia	91	Declarativo	145	Gestión de sí mismo
38	Aprendizaje verbal significativo	92	Deductivo	146	Grados de certeza
39	Aptitud	93	Definiciones	147	Hereditarias
40	Áreas del cerebro	94	Descartes	148	Iceberg
41	Arquitecto	95	Desequilibrio	149	Identidad
42	Arquitectura de competencias	96	Diagnóstico	150	Imagen de sí mismo

Figura 5: Extracto de una lista de soluciones posibles en una prueba constituida de PSN

Muchas universidades poseen un Lector Óptico de Marcas (LOM), que puede leer números constituidos por una serie de códigos marcados sobre un formulario apropiado, como el que se ve en la Figura 6 (*formulom*). A la pregunta 12 un estudiante contestó 273 (la solución n° 273 en la lista de soluciones posibles para la PSN), con un grado de certeza de 80%.

Figura 6 muestra un formulario de lectura óptica de marcas (formulom) para la pregunta 12. El formulario tiene una cuadrícula de marcadores con columnas numeradas del 0 al 9. Los marcadores 2, 7, 8 y 9 están sombreados, lo que indica que el estudiante seleccionó el número 273. Debajo de la cuadrícula hay una barra de porcentaje con marcadores de 0, 20, 40, 60, 80 y 100. El marcador 80 está sombreado, lo que indica que el estudiante seleccionó un grado de certeza de 80%.

Figura 6: Extracto de un formulario de lectura óptica de marcas (*formulom*)

Por ejemplo, la pregunta era: “El científico ruso que, en 1904, recibió el premio Nobel por sus obras relativas a la salivación en el perro se llama.....”. El estudiante tiene que (1) evocar “Pavlov”; luego (2) consultar la lista y descubrir que Pavlov tiene el número 273, y (3) codificar 273 sobre el formulom. A este tipo de sistema, que permite que las respuestas sean corregidas por computador, Leclercq les llama *Preguntas con Soluciones Numerosas (PSN)*, inspirándose en *The Long Menu Questions* de Schuwirth *et al.* (1996b) y Schuwirth (1998).

B.3. Resultados comparados

Es una observación común que los resultados del reconocimiento son mejores que los de la evocación.

Utilizando fotografías de grupos de estudiantes diplomados hace 50 años, Bahrck (1984) demostró que las personas identifican mejor (dan el nombre y apellido correctos) a sus antiguos compañeros cuando se les da la lista de sus nombres (aunque esté en un orden diferente al de la fotografía), que cuando deben evocar esos nombres y apellidos. Bahrck también testeó la memorización de contenidos de lenguaje español estudiados años antes, y de ambos modos: evocación, después reconocimiento.

Schuwirth *et al.* (1996a) midieron de las dos maneras conocimientos de varios grupos de estudiantes y profesionales en medicina. Se ve que, otra vez, el modo “reconocer” (PSM) obtiene mejores resultados (tasa de éxito) que las PRB o Preguntas de Respuesta Breve (Open Ended Questions), que exigen evocar.

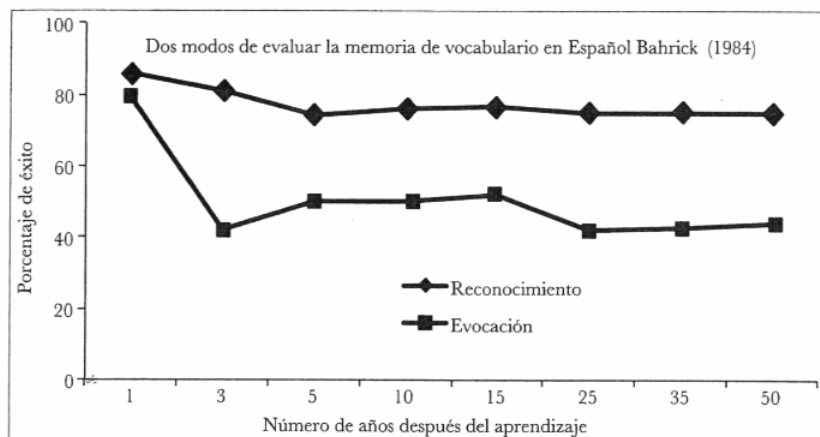


Figura 7: Comparación de resultados de las mismas personas con dos modos de evaluación

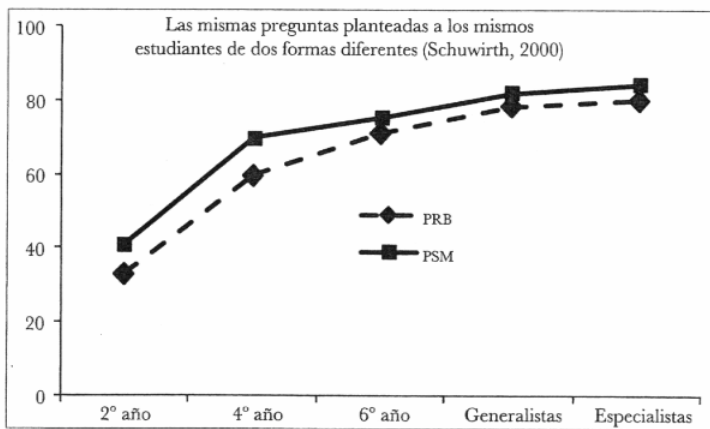


Figura 8: Comparación de resultados de las mismas personas con dos modos de evaluación

C. Evaluar la comprensión

C.1. Una definición de "Comprensión"

El lógico Jan Smedslund (1997) propuso una fecunda definición de lo que es "comprender":

Una persona E (como Estudiante) entiende un mensaje X producido por una persona D (como Docente) si E y D alcanzan acuerdo sobre

- lo que es equivalente a X
- lo que es lo contrario de X o que niega a X
- lo que implica X
- lo que no tiene ninguna relación con X

La falta de acuerdo en uno solo de estos criterios indica una mala comprensión por parte de E. Lo anterior postula que E es lógico. Como no se puede distinguir si un error de E proviene de una falta de lógica o de una falta de comprensión, Smedslund recomienda que nuestro postulado sea que los estudiantes son lógicos y que cuando se equivocan es por falta de comprensión.

La definición de Smedslund me parece fecunda porque:

1. Incluye el hecho de que no se comprende "un mensaje", sino "un mensaje emitido por un emisor en una circunstancia". En nuestro caso, cuando el docente D intenta verificar la comprensión, lo que verifica no es si el estudiante E ha comprendido "la ley del efecto descrita por Thorndike (1913)", sino que "la ley del efecto descrita por Thorndike, como la ha enseñado el docente D".
2. Por supuesto, una vez que el estudiante E ha al menos comprendido el mensaje X, puede añadir otras comprensiones del mismo concepto, siempre que el docente D se haya asegurado de que E comprende su presentación. En este punto, el profesor puede aceptar que son posibles otras comprensiones del mismo concepto. Este es el sentido de la expresión "alcanzan acuerdo": deben discutir, intercambiar.
3. Esta necesidad de intercambiar para entenderse ilustra que el debate no es solo necesario, sino que se espera que ambos actores, estudiante y profesor, puedan aprender de este debate, y, si es público, que cada estudiante pueda aprender del resto de sus compañeros (ver un ejemplo en el Capítulo 9, donde ocurre un debate en una gran aula con 400 estudiantes).

C.2. Un modelo: comprender es traducir

A) EL TRÉBOL DE LAS TRADUCCIONES

Una de las subcategorías de la comprensión consideradas por Bloom *et al.* (1956) es "traducir". Esta palabra se utiliza a menudo cuando se trata de traducir entre dos idiomas, pero se puede extender a la traducción entre lenguajes que no se expresan en palabras; que no son lenguajes verbales.

Por ejemplo, en matemáticas puede tratarse de traducir una tabla de números en un gráfico, un gráfico en una fórmula, y todo lo anterior (tablas, gráficos y fórmulas) en palabras (idioma español u otro). La Figura 9 muestra que hay 16 caminos posibles. Las direcciones son "reversibles" y también intra-modales (por ejemplo, simplificar fórmulas es quedarse en fórmulas; calcular valores nuevos a partir de una tabla de datos es quedarse en números). Un modelo como este permite concebir preguntas.

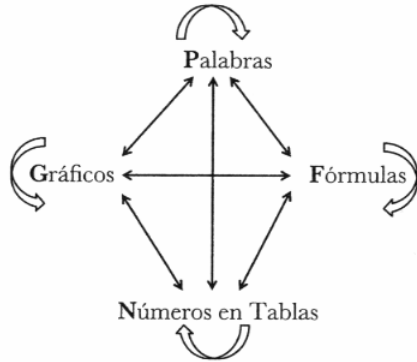


Figura 9: Trébol de las traducciones

B) EJEMPLO DE TRADUCCIÓN DE GRÁFICOS A PALABRAS (G→P)

Una respuesta; una solución correcta simple, o
6. Ninguna 7. Todas 8. Faltan datos 9. Absurdo

Los puntos rodeados representan niños con...

1. Una buena predicción y un fracaso
2. Una buena predicción y un éxito
3. Una mala predicción y un fracaso
4. Una mala predicción y un éxito

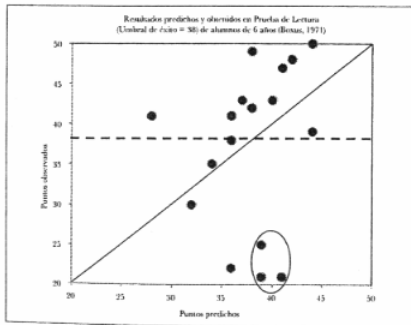
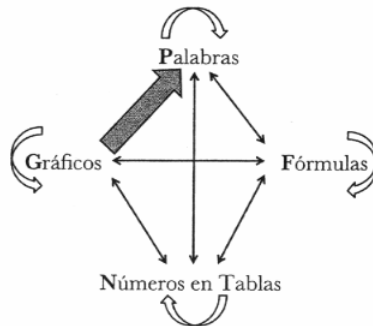


Figura 10: Ejemplo de Pregunta de traducción de Gráficos a Palabras (G→P)



C) EJEMPLO DE TRADUCCIÓN DE PALABRAS A PALABRAS (P→P)

Una respuesta; una solución correcta simple, o
6. Ninguna 7. Todas 8. Faltan datos 9. Absurdo

En sus conclusiones de la página 51, los autores sugieren que

1. Se debe introducir una cuota restrictiva
2. Se debe limitar los cambios de carrera
3. La incidencia de la repetencia es alta
4. Los esfuerzos en favor de los estudiantes que fracasaron es una inversión perdida

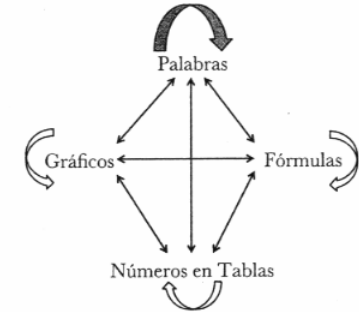


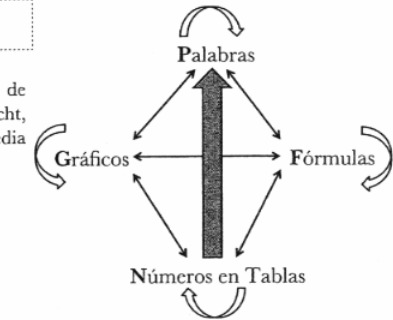
Figura 11: Ejemplo de pregunta de traducción de Palabras a Palabras (P→P)

D) EJEMPLO DE TRADUCCIÓN DE NÚMEROS A PALABRAS (N→P)

Una respuesta; una solución correcta simple, o
6. Ninguna 7. Todas 8. Faltan datos 9. Absurdo

En esta hoja de retroalimentación de una Prueba de Progreso en medicina a un estudiante de Maastricht, ¿en cuál ámbito es este estudiante superior a la media de su cohorte?

1. Anatomía
2. Bioquímica
3. Farmacología
4. Fisiología
5. Cirugía general



Descripción	NdP	Cor.	Inc.	?	C-I
1. Anatomía/embriología	13	54+	0-	46	54+
2. Bioquímica	13	77+	8-	15	69+
3. Farmacología	7	57+	0-	43	57+
4. Fisiología	11	73	0-	27+	73+
5. Genética - Biol. Cel.	8	50	25	25	25
6. Inmunología	4	75+	25+	0-	50
7. Microbiología médica	6	83	0-	17+	83
8. Patología	10	60	0-	40	60
Total ciencias de base	72	65+	6-	29	60+
9. Medicina del trabajo	1	0	0	100	0
10. Cirugía general	16	38-	6-	56	31+

Figura 12: Ejemplo de pregunta de traducción de Números a Palabras (N→P)

E) EJEMPLO DE TRADUCCIÓN DE FÓRMULAS A GRÁFICOS (F→G)

Dibuja la curva de esta ecuación
 $Y = (X:2) - 1$

Respuesta correcta:

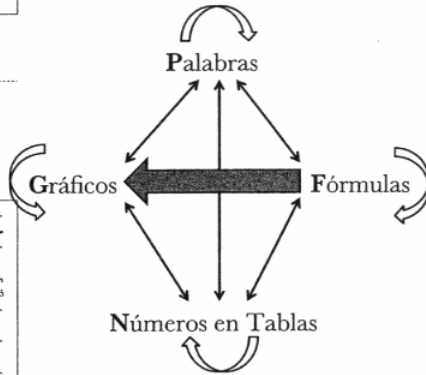
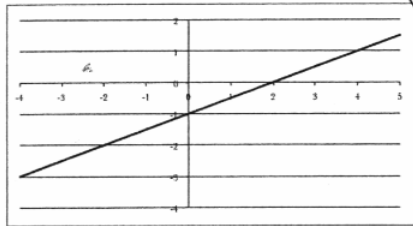


Figura 13: Ejemplo de pregunta de traducción de Fórmulas a Gráficos (F→G)

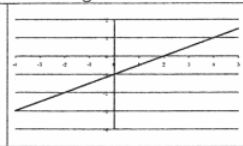
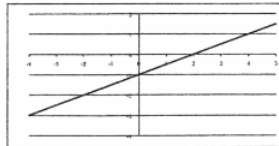
F) EJEMPLOS DE TRADUCCIÓN DE GRÁFICOS A FÓRMULAS (G→F)

Una respuesta; una solución correcta simple, u 8. Faltan datos 9. Absurdo

Una respuesta; una solución correcta simple, o 6. Ninguna 7. Todas 8. Faltan datos 9. Absurdo

Escribe la ecuación de la recta.

La misma pregunta:
Elige la ecuación



1. $Y = 2X - 1$
2. $Y = 2X - 2$
3. $Y = (X / 2) - 2$
4. $Y = X - 1$
5. $Y = (X / 2) - 1$

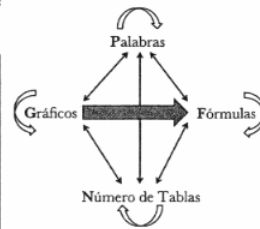


Figura 14: Ejemplo de preguntas de traducción de Gráficos a Fórmulas (G→F)

C.3. Un modelo que sirve para interpretar datos

En 1999, 4.000 estudiantes que se encontraban ingresando a las ocho universidades de la región francófona de Bélgica, rindieron una serie de 10 pruebas en varios ámbitos¹¹⁵. Uno de ellos era un test de matemáticas, con 22 preguntas concebidas de acuerdo al modelo del “trébol de las traducciones”.

Como se ve en la Tabla 2, cuando se ubica cada una de las 22 preguntas a la altura de su tasa (%) de éxito, calculada sobre los 2.539 estudiantes que rindieron esta prueba, las preguntas que demandan traducir desde otros lenguajes a *Palabras* tienen las tasas de éxito más bajas. Una de las explicaciones posibles es que en los cursos de matemáticas de la enseñanza secundaria los docentes no dedican suficiente tiempo al tipo de ejercicios que requieren traducir hasta el nivel de las palabras.

Las consecuencias en términos de regulación del currículo parecen evidentes: concebir e implementar un mayor número de situaciones de aprendizaje donde se guíe al estudiante en la traducción en palabras de gráficos, números y fórmulas matemáticas.

Tabla 2: Tasa de éxito de 22 preguntas de la Prueba de Matemáticas de MOHICAN, rendida por 2.539 estudiantes que se encontraban ingresando a las ocho universidades de la Bélgica francófona (Leclercq, 2003, p. 179)

Tasa de éxito	Tipo de traducción
89%	N → F
88%	N → N
87%	N → G
84%	F → G
79%	M → P
78%	G → F
75%	N → N
70%	M → N
70%	F → P
70%	M → P
64%	G → P
63%	G → N
63%	F → N
55%	M → P
55%	G → P
53%	F → N
51%	M → P
40%	F → P
27%	N → P
20%	G → N

¹¹⁵ En el Proyecto MOHICAN (Leclercq, 2003).

D. Evaluar la aplicación

D.1. Definición de "Aplicación"

(desde el punto de vista de la pregunta o problema)

- La pregunta es explícita (no hay trampas), el resultado a obtener está formulado en términos claros, habituales (a menudo técnicos); por ejemplo, "calcular el área"
- Todos los datos necesarios están disponibles
- El principio a aplicar (por ejemplo, utilizar la fórmula de cálculo de un cilindro) no se menciona
- La pregunta puede contener informaciones inútiles

D.2. Ejemplo de pregunta para evaluar la aplicación

Una respuesta; una solución correcta simple,
u 8. Faltan datos 9. Absurdo

Considera esta cabeza de martillo. ¿Cuál es su volumen?

Sabido que el martillo es de hierro, y tiene una densidad de 7.87; el agujero es circular y su diámetro es 2 cm; la altura es 3 cm, su ancho 4 cm y su largo es 5 cm

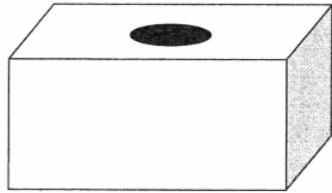


Figura 15: Ejemplo de pregunta para evaluar la aplicación

La pregunta puede ser formulada como Pregunta de Respuesta Breve (PRB) o como Pregunta de Selección Múltiple (PSM).

E. Evaluar el análisis

E.1. Definición de "Análisis"

(desde el punto de vista de la pregunta o problema)

- La pregunta no es explícita (o hay trampas)
- Puede ser que no estén disponibles todos los datos necesarios
- El principio a aplicar (por ejemplo, utilizar la fórmula de cálculo de un cilindro) no se menciona
- La pregunta puede contener informaciones inútiles

E.2. Ejemplo de pregunta para evaluar el análisis

Una respuesta; una solución correcta simple, o
6. Todas 7. Ninguna 8. Faltan datos 9. Absurdo

Considera esta cabeza de martillo. ¿Cuál es su volumen? Sabiendo que

- el martillo es de hierro, y tiene una densidad de 7.87, tiene un agujero circular cuyo diámetro es 2 cm; su altura es 3 cm, su ancho 4 cm y su largo 5 cm.

- 1) 45,38 cm
- ³
- 2) 47,44 cm
- ³
- 3) 50,58 cm
- ³
- 4) 56,64 cm
- ³
- 5) 62,38 cm
- ³

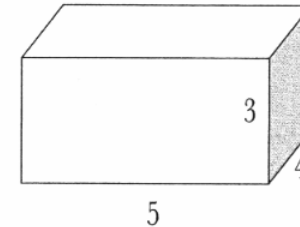


Figura 16: Ejemplo de pregunta para evaluar el análisis

Este ejemplo parece idéntico al de aplicación, pero no lo es. En este ejemplo no se dice cómo está posicionado el agujero cilíndrico en la cabeza del martillo, lo que afecta la respuesta. Si el agujero está posicionado como en el dibujo A de la Figura 17, la solución es 47,44 cm³. Si lo está como en B, es 50,58 cm³.

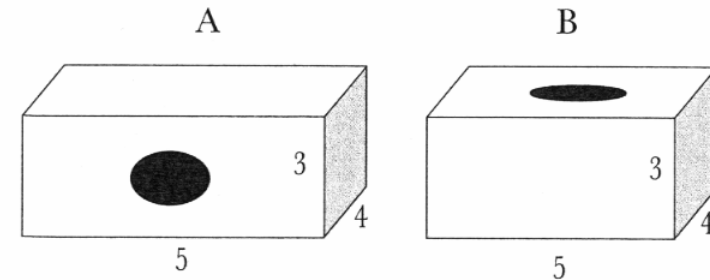


Figura 17: Solución Figura 16

Que el estudiante conteste con la alternativa 2 (47,44 cm³) o 3 (50,58 cm³) es una evidencia de que sabe aplicar. Pero si no contestó que las dos son correctas (o, como aquí, "8. Faltan datos"), es una evidencia de que no ha tenido suficiente vigilancia cognitiva en esta ocasión. Soluciones Generales Implícitas (las sGI) o explícitas (sGE) permiten evaluar y ejercitar la vigilancia cognitiva, especialmente si son implícitas, como se ilustra en el Capítulo 13, sección A7.

E.3. Preguntas PRIM-BIS para distinguir la comprensión del análisis (o vigilancia cognitiva)

El principio de este modo de evaluar (Leclercq, 2005) consiste en presentar dos veces la misma pregunta, cuando la respuesta correcta esperada es una *SGI*.

La primera vez (PRIM) la pregunta se presenta con toda su dificultad

Ejemplo:

PRIM Una respuesta; una solución correcta simple, o 6. Todas 7. Ninguna 8. Faltan datos 9. Absurdo

¿Cuál es el perímetro de un triángulo cuyos lados son respectivamente, 2 cm, 3 cm y 6 cm?
1) 10 cm 2) 11 cm 3) 12 cm 4) 36 cm

BIS Una respuesta; una solución correcta simple, o 6. Todas 7. Ninguna 8. Faltan datos 9. Absurdo

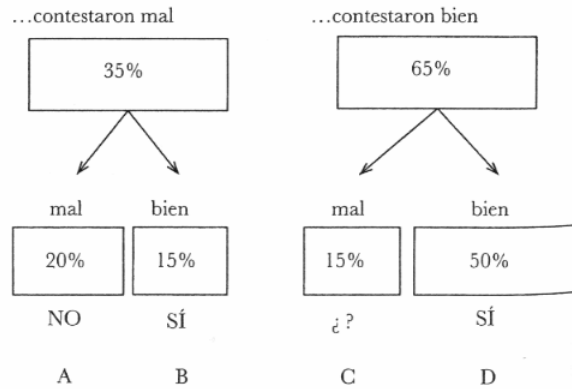
Había un Absurdo en la pregunta PRIM. ¿Cuál era ese absurdo?

Figura 18: Ejemplo de una pareja de preguntas PRIM - BIS

En este ejemplo, la segunda vez (BIS) se revela que hay un absurdo en la pregunta PRIM, pero no se dice cuál es (en este caso, que es imposible construir este triángulo, porque la suma de dos de sus lados (2 y 3) es más pequeña que el tamaño del tercer lado).

A continuación (Figura 19) se muestran los resultados (tasas de éxito) obtenidos por Leclercq (en 1998, no publicado) utilizando este modo de evaluar con 10 preguntas y 300 estudiantes de master que se preparaban para ser profesores.

A la pregunta 1 (PRIM)...



¿Comprensión?

Grupo:

NO

SÍ

¿ ?

SÍ

A

B

C

D

Figura 19: Tasas de respuestas en la combinación de preguntas PRIM y BIS en 1998

Aunque hay diagnósticos muy probables para tres grupos (A, B, D), queda pendiente profundizar en los casos C (entrevistándoles por ejemplo) para explicar este tipo de situaciones.

Tabla 3: Cuatro casos (A, B, C, D) en las respuestas a preguntas PRIM - BIS

	Respuesta correcta en la pregunta PRIM (ha descubierto que hay un absurdo)	Respuesta incorrecta en la pregunta PRIM (no ha descubierto que hay un absurdo)
	Retroalimentación: Hay un absurdo en la pregunta PRIM	
Respuesta correcta en la BIS	A	B
Respuesta incorrecta en la BIS	C	D

La situación A es la de los mejores estudiantes. Como ya han descubierto el absurdo en la pregunta PRIM (han estado vigilantes y han detectado un problema que entienden), la retroalimentación les conforta en su opinión y, consecuentemente, contestan a la pregunta BIS correctamente, rápidamente y con un alto grado de certeza.

La situación B es frecuente: esos estudiantes no han detectado el absurdo en la pregunta PRIM (no han estado vigilantes), pero cuando reciben la retroalimentación, leen otra vez la pregunta y descubren el absurdo (porque entienden el asunto).

La situación D significa que esos estudiantes no entienden el asunto porque, aunque se les ha dicho que hay un absurdo, no lo detectan en la BIS y tampoco pudieron detectarla en la PRIM.

La situación C es la más ambigua: un estudiante contesta correctamente la primera vez pero erróneamente la segunda. Una explicación posible es que en la pregunta PRIM habían elegido la solución correcta al azar y, cuando tienen que justificar esta respuesta, no pueden hacerlo.

En caso de una respuesta incorrecta para la Pregunta PRIM si la respuesta a la Pregunta BIS es correcta significa una *ausencia de vigilancia*, pero si la respuesta a la pregunta BIS también es incorrecta, significa una *ausencia de comprensión*.

F. Evaluar la Síntesis (Producción, Formulación, Creación)

F.1. Una definición de "Síntesis"

Evaluar la *síntesis*, según Bloom *et al.* (1956), es estimar la capacidad de un/a estudiante de *formular (producir, crear)* algo, ya sea en (i) un idioma natural (por ejemplo, español como lengua materna o inglés como segunda lengua), o (ii) un lenguaje artificial (de programación de computadores, por ejemplo), o (iii) en dibujos, o (iv) en gestos (de bailarina, por ejemplo), o (v) en expresiones faciales (un mimo, por ejemplo).

Con frecuencia lo anterior resulta en un desempeño complejo que puede ser evaluado de acuerdo con varias dimensiones (o criterios), para los que muchas veces se definen indicadores, pautas y rúbricas (ver Capítulo 6).

Pensando en el proceso metacognitivo de juicio, en este tipo de pregunta o problema no se debe pedir al estudiante que añada un grado de certeza a su respuesta o realización. Sin embargo, sí puede añadir varios grados de certeza: uno para cada dimensión (o criterio), que son los mismos en los que el profesor se basará para poder decidir si el/la estudiante obtiene un éxito o fracaso.

Para una pregunta de desarrollo (PDD) los *criterios* pueden ser (por ejemplo):

Tabla 4: Ejemplo de autoevaluación de un informe de algunas páginas por el estudiante y (columna derecha) los puntajes del docente

Criterios	Grado de certeza por el estudiante	Puntaje por el docente
C1. Pertinencia	80	1
C2. Estructuración	60	1
C3. Originalidad	100	0
C4. Nivel de detalle	60	0
C5. Uso de términos técnicos	40	1
C6. Etc.		

En este ejemplo se ve que el estudiante se ha sobrestimado en los criterios C3 y C4, y se ha subestimado en el criterio 5.

F.2. Una definición de creatividad

Paul Torrance (1966) propuso 4 criterios para evaluar la CREATIVIDAD (de respuestas a preguntas “abiertas”):

La *fluidez* (o abundancia): el número de respuestas.

La *flexibilidad* (o variedad): el número de respuestas que pertenecen a categorías diferentes.

La *originalidad* (o rareza) es un concepto estadístico: depende de las respuestas de un grupo de referencia.

Mientras menos aparezca entre las respuestas del grupo, más original.

La *elaboración* (o grado de detalle, o sofisticación).

Las palabras entre paréntesis son del autor

F.3. Un ejemplo de aplicación de tres de los criterios de Torrance

Durante 30 segundos escriba sustantivos comunes que empiecen con la letra “s”:

Consideramos las respuestas de tres estudiantes (A, B y C):

	A	B	C
	Sonrisa Señor Señorita Señora Sueño Secador Superficie	Silencio Señal Soltería Soliloquio Sobrino	Sicofante Servidumbre Servidor Servicio Servilismo Sorpresa
Fluidez :	7	5	6
Flexibilidad	4	5	3
Originalidad			
(1) En la clase	(1) 0	(1) 1=soliloquio	(1) 1 = sicofante
(2) En el grado	(2) 0	(2) 0	(2) 1 = sicofante

(1) Número de palabras que ninguna de las 30 personas de la *clase* usó en su respuesta.

(2) Número de palabras que ninguna de las 150 personas del *grado* (nivel) usó en su respuesta.

Figura 20: Ejemplo de pregunta y de respuestas de 3 estudiantes evaluadas según los criterios de Torrance

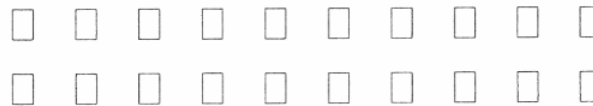
Este tipo de prueba no permite evaluar la “elaboración”.

Se ve que el estudiante que obtiene el puntaje más alto en Fluidez (A) no es el que obtiene el puntaje más alto en Flexibilidad (B), o el puntaje más alto en Originalidad (C).

También se ve que el puntaje en originalidad depende del grupo de referencia. La performance de B tiene un puntaje de 1 en originalidad cuando se refiere al grupo de 30 personas (ninguno propuso “soliloquio”), pero 0 en el grupo de 150 personas (al menos otra persona contestó “soliloquio”). La palabra “sicofante” (delator) es original en los dos grupos (de 30 y de 150 personas). Para entregar referencias a los psicólogos que quieren aplicar su test, Torrance estableció diccionarios de palabras con las estadísticas de los números de veces que han sido contestadas por miles de personas.

F.4. Un ejemplo de aplicación del criterio “elaboración” en una producción icónica

Utilice cada uno de los rectángulos para significar una cosa que se pueda reconocer:



Consideramos las cuatro primeras producciones de dos personas: A y B. No se distinguen en términos de Fluidéz, tampoco de Flexibilidad u Originalidad, pero B ha detallado mucho más sus respuestas, de modo que su puntaje en la categoría “elaboración” será más alta.

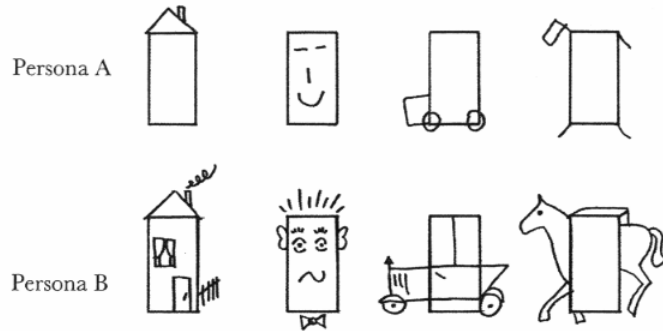


Figura 21: Ejemplo de pregunta y de respuestas de dos estudiantes evaluadas según los criterios de Torrance

F.5. Un ejemplo de aplicación del criterio “elaboración” en una producción verbal

Produzca nombres de marcas comerciales para una nueva pasta dentífrica

Si aplicamos los criterios de Torrance a las posibles respuestas a esta pregunta, soluciones como “blanca diente”, “clara diente” y “pura diente” podrían ser consideradas como una sola categoría (Fluidéz = 3; Flexibilidad = 1), y “Miramisdientes” de otra categoría (Fluidéz = 2). En términos de elaboración (detalle) se puede tener en cuenta, por ejemplo, que el creador haya considerado las connotaciones que una palabra o expresión puede tener en varios idiomas. Por ejemplo, que haya considerado que “Cheers” tiene, en inglés, la connotación de “saludo” o “brindis”, pero que en francés tiene la connotación de “costoso”.

G. Evaluar la evaluación o el juicio

G.1. La capacidad de aproximar¹¹⁶

Es la capacidad de producir, sin instrumento, estimaciones próximas a los valores reales medidos con instrumento(s). Estas producciones pueden estar expresadas en:

- valores numéricos absolutos en varias unidades
 - de tiempo (segundos, minutos, horas, días, semanas, meses, siglos, etc.)
 - de longitud (mm, cm, m, km, etc.)
 - de peso (mg, cg, dg, g, kg, etc.)
 - de área (cm², dm², m², km², etc.)
 - etc.
- proporciones de comparación (0,5 veces más largo, 2 veces, 10 veces, etc.)
- orden (el más largo es, el menos largo es..., etc.).
- especímenes próximos (mismo peso que..., misma duración que..., etc.)

G.2. Estimular la capacidad de aproximar: el sistema EU (Elicitation of Probabilities)

En 1993 Jelle Van Lenthe propuso un sistema que permite a los estudiantes aproximar un porcentaje como respuesta a una pregunta que tradicionalmente exigiría una cifra exacta como respuesta (y que muy frecuentemente sería presentada a los estudiantes como una de las alternativas en una PSM). Por ejemplo, frente a la pregunta ¿Cuál fue la tasa de niños nacidos fuera del matrimonio en el país, entre los años... y ...? (respuesta correcta = 38%). En lugar de pedir al estudiante que elija una de las cifras exactas que constituyen las alternativas, se le pide contestar entregando una distribución de probabilidades, a través del programa ELI (*Elicitation of Probabilities*).

ELI muestra sobre la pantalla del computador un eje horizontal que va de 0% hasta 100%, y donde el estudiante indicará su respuesta a la pregunta. El eje vertical va de -40 hasta +40, y se usará para calcular el puntaje otorgado a cada respuesta.

Además, se presenta una curva en forma de campana de Gauss, centrada (con su ápex) en 50% horizontalmente. Usando las 4 flechas del teclado, el estudiante puede:

- desplazar la curva hacia la izquierda (con la tecla ←) y la derecha (con →)

En los ejemplos de las figuras 22 y 23 el estudiante desplazó la curva hasta fijar la Media en 41% (en vez del 50% que propone ELI por *default*)

- disminuir la pendiente (o la concentración) de la curva (aplastarla) con la tecla ↓. Es lo que hizo el estudiante A en la Figura 22, de modo que la curva corta el eje horizontal en los valores 18% y 67% (rango amplio de distribución), estando la Media en 41%.
- aumentar la pendiente (disminuir la amplitud de la distribución o aumentar su concentración), usando la tecla ↑. Es lo que hizo el estudiante B en la Figura 23: la media es la misma (41%) que para el estudiante A, pero la Desviación Estándar es mucho más pequeña.

¹¹⁶ Este tema es tratado también en el Capítulo 7, sección D.4.

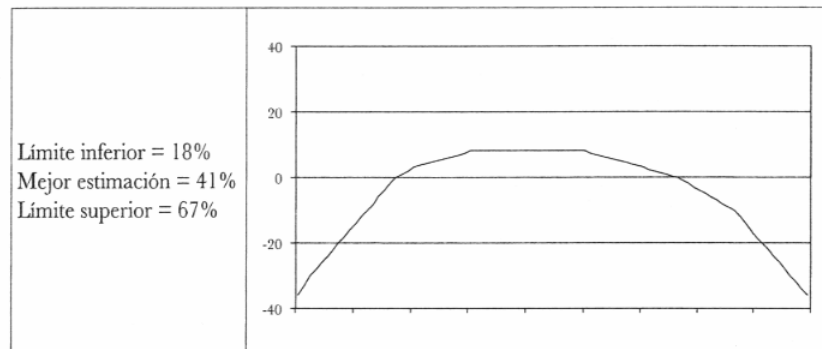


Figura 22: Respuesta de un estudiante A a la pregunta sobre el porcentaje de...

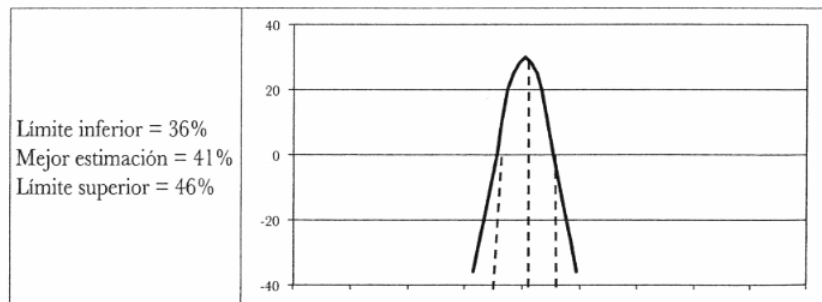


Figura 23: Respuesta de un estudiante B a la pregunta sobre el porcentaje de...

En un cuadro a la izquierda aparecen los valores llamados “*Bounds*” (límites) de la curva. Si la respuesta correcta se ubica dentro de esos límites, el puntaje de la respuesta será positivo. Si está fuera, el puntaje será negativo. El puntaje otorgado a la respuesta corresponde al valor del lugar del plano donde intersectan la línea horizontal de la respuesta correcta y la curva, según ha sido posicionada por el estudiante.

Por ejemplo, en la curva de la Figura 22 los puntajes serían negativos si la respuesta correcta estuviera bajo 18% o sobre 67%. Si el porcentaje real (respuesta correcta) fuera 5%, el puntaje para esta respuesta sería -20. En cambio, si el porcentaje real fuera 41% (que es justo la Media que el estudiante eligió), el puntaje sería +10. En este ejemplo el estudiante no obtiene un puntaje más alto (hasta +40) pues la amplitud de distribución de su respuesta es muy alta; el rango de posibles respuestas correctas es extenso, lo que implica una respuesta poco arriesgada.

En el ejemplo de la Figura 23 el estudiante B tomó más riesgo, pues bastaría que el porcentaje real (respuesta correcta) estuviera bajo el 35,9%, o sobre el 46,2%, para que su puntaje se volviera negativo. Sin embargo, si ha estimado perfectamente (es decir, que el porcentaje real sea 41%) recibirá un puntaje de +30.

G.3. Los fractiles¹¹⁷

Un modo de aproximación de un valor numérico (por ejemplo una fecha, una longitud, un peso) consiste en pedir al estudiante que conteste con un intervalo, es decir, con dos valores (un límite inferior y un límite superior) entre los cuales la respuesta correcta debe estar incluida con una probabilidad predefinida en la pregunta (por ejemplo, con 90% de probabilidad de inclusión). Eso se llama “*tertiles*” porque el *continuum* está cortado en tres categorías. Con este método, se supone que el valor que tiene la mayor probabilidad de ser el correcto es la Media del intervalo. La escala de cotejo puede inspirarse en la de Van Lenthe (ver sección G2).

También es posible pedir al estudiante contestar entregando tres valores: el más probable, el límite inferior y el superior, para que la distancia entre el valor más probable y el límite superior pueda ser diferente de la distancia entre el valor más probable y el límite inferior. Eso se llama “*quartiles*”.

Pitz (1974) utilizó varias consignas con fractiles (tertiles y quartiles) en sus evaluaciones, y observó una gran tasa de “sorpresa”, es decir, situaciones en las cuales el valor real está ubicado fuera de la zona entre los límites. Con la consigna “Define la zona de modo que haya solo un 10% de sorpresas” (es decir, que exista un 90% de probabilidad de que la respuesta correcta esté dentro de los límites definidos por el estudiante), observó más de 50% de ellas. De decepción en decepción, Pitz terminó utilizando la consigna “Conteste con dos límites de modo que en 1/3 de los casos el valor real esté bajo este intervalo, en 1/3 esté dentro y en 1/3 esté sobre”. Soll y Klayman (2004) también observaron este “exceso de confianza (*over-confidence*) en la estimación de intervalos”.

Se pueden imaginar consignas más sofisticadas, con 4 límites (y 5 zonas) por ejemplo.

Se debe recordar que siempre hay que precisar en la consigna cuál debe ser la probabilidad de éxito (inclusión del valor real dentro del intervalo). Esto funciona al revés de los grados de certeza, donde es el estudiante (y no la consigna) quien determina la probabilidad de exactitud de su respuesta.

G.4. Autoevaluaciones

Este tipo de juicio es frecuente y de importancia capital en la autorregulación de los aprendizajes. Por esta razón se puede considerar como un “nivel 7” de la taxonomía de procesos mentales de Bloom. Esta es nuestra opción, y en consecuencia dedicaremos los dos siguientes capítulos (16 y 17) a este tema.

¹¹⁷ Esta sección es complementaria con la sección D.4 del Capítulo 7.

H. Resumen

La Figura 24 intenta enumerar los criterios necesarios y suficientes para clasificar a los diferentes tipos de preguntas de evaluación, según los procesos mentales que evalúan, ordenados de acuerdo al proceso mental más alto¹¹⁸ evaluado. Para entenderla cabalmente precisamos algunos de los términos.

Vínculos entregados: Cuando el vínculo (ej: *la base*) entre dos cosas (ej: *un círculo y un cilindro*) es presentado (solo o entre otros: *superficie, perímetro, volumen*), el estudiante tiene que reconocer cuál vínculo es el correcto.

Proceso descrito: Cuando se dice cuál proceso mental aplicar (ej: *Evoca el nombre de...*, o *Calcula el perímetro de...*).

Términos definidos: Cuando los elementos esenciales son nominados por sus nombres familiares o técnicos habituales (ej: *en este triángulo la base es 5 m y la altura es 2 m*).

Pregunta clara: sin trampa, bien planteada.

Formulación entregada: Cuando el estudiante no tiene que formular (redactar) su respuesta.

Toma de riesgo: Cuando el estudiante debe expresar la seguridad en su estimación o la intensidad de su opinión (tomar partido), con el riesgo de que sea incorrecta.

Toma de riesgo	No	No	No	No	No	No	Sí
Formulación entregada	Sí	Sí	Sí	Sí	Sí	No	-
Pregunta clara	Sí	Sí	Sí	Sí	No	No	-
Términos definidos	Sí	Sí	Sí	No	No	No	-
Proceso descrito	Sí	Sí	No	No	No	No	-
Vínculos entregados	Sí	No	No	No	No	No	-

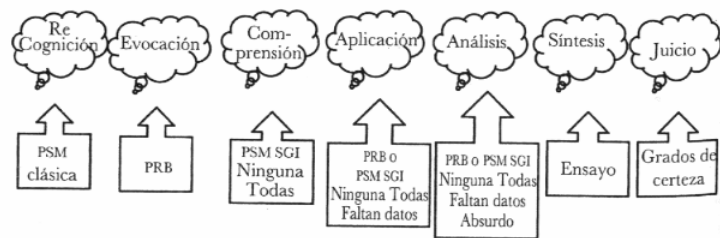


Figura 24: Ábaco de Criterios para vincular preguntas y procesos mentales evaluados

Por supuesto, un ábaco de este tipo necesita más matices, y su utilización demostrará sus fortalezas y debilidades. Así, habrá que mejorarlo... pero intentando mantener su concentración visual y eficiencia en el uso de palabras.

Referencias

- ANDERSON L. y KRATHWOHL, D. (2001). A taxonomy for learning, teaching and assessing. A revision of Bloom's taxonomy of Educational Objectives. New York: Longman.
- BAHRICK, H.P. (1984). Semantic memory content in permastore: 50 years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 120, 1-29.
- BLOOM, B.S. ENGELHART, M.D., FORST, E.J., HILL, W.H. y KRATHWOHL, D.R., (1956). Taxonomy of educational objectives. Handbook 1: Cognitive domain, New-York, McKay.
- BLOOM, B.S., HASTINGS, J.T. y MADAUS, G.F. (1971). Handbook on formative and summative evaluation of Student Learning, New York: Mc Graw Hill.
- LECLERCQ, D. (2005). Edumétrie et docimologie. Chap 4: TOISE. Editions de l'université de Liège.
- LECLERCQ, D. (Ed.) (2003). Diagnostic cognitif et métacognitif au seuil de l'université. Le projet MOHICAN mené par les 9 universités de la Communauté française Wallonie Bruxelles. Liège: Editions de l'Université de Liège.
- PITZ, G. (1974). Subjective probability distribution for imperfectly known quantities, in L. W. Gregg (Ed.), Knowledge and Cognition, New York: Wiley, 29-41.
- SCHUWIRTH, L.W.T. (1998). Computerized Case-based Testing: an approach to the assessment of medical problem solving. Ph.D. in Education, Maastricht: University of Maastricht.
- SCHUWIRTH, L.W.T., VAN DER VLEUTEN, G.P.M., y DONKERS, H.H.L.M. (1996a). A closer look at cueing effects in multiple-choice questions. *Medical Education*, 30, 44-99.
- SCHUWIRTH, L., VAN DER VLEUTEN, C., STOFFES, H. y PEPEKAMP, A. (1996b). Computerized long-menu questions as an alternative to open-ended questions in computerized assessment, *Medical Education*, 30, pp. 50-55.
- SMEDSLUND, J. (1997). The forgotten variable of understanding. *Cahiers de Psychologie Cognitive - Current Psychology of Cognition*, 16 (1-2), 217-221.
- SOLL, J. y KLAYMAN, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 30 (2), p. 299-314.
- THORNDIKE, E. (1913). *Educational Psychology: The Psychology of Learning*. New York: Teachers College Press.
- TORRANCE, E. P. (1966). *The Torrance Tests of Creative Thinking: Norms-technical manual*. Lexington, MA: Personal Press.
- VAN LENTHE, J. (1993). The Development and Evaluation of ELI, an Interactive Elicitation Technique for Subjective Probability Distributions, in D. Leclercq y J. Bruno (Eds), *Item Banking: Interactive Testing and Self-Assessment*, NATO ASI Series, Berlin: Springer Verlag, pp. 132-145.

¹¹⁸ Según la taxonomía de Bloom, un proceso mental necesita el dominio de los que están situados bajo él.

IDEAS E INNOVACIONES Dispositivos de Evaluación de los Aprendizajes en la educación

Dieudonné LECLERCQ y Álvaro CABRERA MARAY 2014

Resumen de cada capítulo

Los editores y autores principales del libro

p. 11-13

Prologo

Álvaro Cabrera &
Dieudonné
Leclercq

Parte 1: Conceptos clave en educación

p. 15-20

1	ATOME (Alineamiento en un Tablero de Objetivos, Métodos y Evaluaciones. Da una visión panorámica de los tres pilares de un programa de formación: los objetivos (y sus 4 niveles de alcance), los Métodos (y sus 8 Eventos de Enseñanza-Aprendizaje), las evaluaciones (y sus 4 niveles de profundidad), insistiendo sobre la Triple Concordancia (u alineamiento) O-M-E y dando ejemplos de inconsistencia.	D.Leclercq & Álvaro Cabrera p. 23-34
2	Los componentes de un dispositivo de evaluación de los aprendizajes (DEA) Da una visión de los vínculos entre las finalidades (formativas o sancionantes) de la evaluación, las competencias que desarrollar y los recursos que dominar, las condiciones de un dispositivo, las herramientas y los criterios de calidad de cada componente de un DEA.	D. Leclercq p. 35-50
3	El prisma de las características de un Dispositivo de Evaluación de los Aprendizajes (DEA) Presenta las características y las condiciones de un DEA como las facetas de un prisma: Quien (los agentes) evalúa, cuando (de manera definitiva o mejorable), quienes (individuo o grupo), para quienes (pública o confidencial), como (objetivamente o subjetivamente; estandarizada o adaptativa), que modifican la medición o su interpretación.	D. Leclercq p. 51-82
4	ETIC PRAD: Ocho criterios de validez de un Dispositivo de Evaluación de los Aprendizajes (DEA) Presenta 8 tipos de validez de un componente de un DEA: Ecológica (cerca de la situación real), Teórica (razonamiento o teoría que lo funda), Informativa (o diagnóstica), Consecuencial (lo que resulta del componente), Predictiva (correlada con otras mediciones), Replicabilidad (o fiabilidad), Aceptabilidad (para los profesores, los estudiantes, el público), Deontológica (equitativo).	D. Leclercq p. 83-92
5	Autodescribir y evaluar el Dispositivo de Evaluación de los Aprendizajes (DEA) de un curso Propone una secuencia que puede seguir un profesor para definir un DEA para su curso, es decir sus objetivos, sus métodos y sus evaluaciones, presentándoles en una tabla de modo que aparecen los vínculos y las ausencias de vínculos.	D. Leclercq & Álvaro Cabrera p. 93-102

6	<p>La calificación subjetiva de los desempeños complejos: Criterios y rubricas Presenta la docimología y sus evidencias de los efectos de notación o de calificación subjetiva (ley de Posthumus, ausencia de concordancia intra y inter-jueces, efectos de halo, de secuencia, de estereotipo, de confirmación (o de inercia). Además de esta docimología “negativa”, presenta principios de una docimología positiva y varios tipos de escalas (ej: la de Mercali) y rubricas.</p>	<p>D. Leclercq & Álvaro Cabrera p. 103-128</p>
7	<p>Evaluar la capacidad de resolver problemas Explica la diferencia entre una pregunta y un problema, el cono de la experiencia (Dale), y las heurísticas de Polya para resolver problemas. Da varios ejemplos de evaluaciones apropiadas a medir la capacidad y detectar los procesos utilizados en la resolución de problemas: las cascadas convergentes y divergentes, las análisis fraccionadas de casos (AFC), la facilitación progresiva, la medición de la búsqueda de información (Shannon, Rimoldi). Da ejemplos de medición de la creatividad, de la capacidad de aproximación y una teoría de la auto-fijación de la dificultad, como de la perseverancia.</p>	<p>D. Leclercq, S. Delcomminette (HERS) & A. Cabrera p. 129-152</p>
8	<p>ECOE: Exámenes Clínicos Objetivos y Estructurados Esta técnica consiste en una sucesión de estaciones en cada de cuales se juegan roles (simulaciones) donde el profesor juega el paciente (el estudiante jugando el del medico o de la enfermera) u el cliente (el estudiante jugando el del farmacéutico), o... para medir competencias, es decir capacidad de actuar en situación compleja. El sistema de notación incluye las actitudes, las destrezas, y la cognición. Las reacciones de los participantes como la predictividad de estas mediciones son presentadas.</p>	<p>G. Philippe (ULg), D. Leclercq & J-P. Bourguignon (ULg) p. 153-170</p>
9	<p>Meta cognición y Tests Espectrales Metacognitivos (TEMs) Para los docentes que quieren desarrollar y medir capacidades como la vigilancia cognitiva, el espíritu crítico, la auto-evaluación (y la meta cognición) y el desarrollo epistemológico es presentada el método “Test Espectrales Meta cognitivos” que combina PSM con SGI (cap. 13, 14 y 15), grados de certeza (cap. 15 y 16), debate y reflexión meta cognitiva. Presenta los aspectos técnicos como los resultados obtenidos en varios ámbitos (cognitivo, epistemológico, meta cognitivo).</p>	<p>D. Leclercq & Álvaro Cabrera p. 171-196</p>
10	<p>Evaluar los Aprendizajes en la Pedagogía Por Proyectos (PPP) La PPP permite de desarrollar y medir competencias complejas (incluido trabajar en equipo), con un enfoque sobre rubricas, tan como sus componentes (recursos) en términos de cognición, actitudes, destrezas. Se puede aplicar los principios de evaluación a 360° (por los pares, por su mismo, por los docentes, por el público). El capítulo plantea (y ilustra sobre un caso) el problema de la convergencia (o ausencia de congruencia) entre estas varias fuentes de evaluación, y el problema de la ponderación de los criterios.</p>	<p>Álvaro Cabrera p. 197-220</p>
11	<p>Evaluar la contribución de cada participante a un trabajo grupal Distingue colaboración y cooperación, presenta los elementos que deben ser parte de un contrato al inicio, y después presenta 6 métodos para evaluar el valor añadido de cada participante al trabajo de grupo. Ilustra el método 4 (declaraciones de participación) con un ejemplo, el de PARMs (Proyectos de Animación Reciproca Multimedia) y sus criterios DECLAR, el método 5 (observación continua con la simulación de actividad parlamentaria y el método 6 (observar la colaboración) con la pauta de Bales. .</p>	<p>D. Leclercq, P. Gillet (ULg), M. Erpicum (ULg) & A. Cabrera p. 221-242</p>
12	<p>Los Portfolios: Hacia una evaluación más integrada y coherente con el concepto de desempeño complejo Este principio (y método) de evaluación sirve no solo a evaluar desempeños complejos como estancias en terreno, sino de constituir una integración de varias evaluaciones. Es ilustrado en dos carreras de la universidad de Liège: Formasup o Master en Pedagogía Universitaria (con sus instrucciones o consignas de redacción del portfolio) y el Master en Logopedia (que permite de discutir de 4 niveles de calidad de evidencias).</p>	<p>M. Poumay (ULg) & Chr. Maillard (ULg) p. 243-260</p>

13	<p>Las Preguntas de Selección Múltiples (PSM): del currículo escondido a la vigilancia cognitiva Presenta los retos del currículo oculto y de la espontaneidad vs la limitación a respuestas sobre sollicitación. Explica como la vigilancia cognitiva se puede entrenar y medir con una consigna valida por las PRB (Preguntas a respuesta Breve) y las PSM (Preguntas a Selección Múltiple): las Soluciones Generales Implícitas (SGI) como “Ninguna, Todas, falta datos, Absurdo”. Da una definición muy precisa de PSM, sus formas de presentación, sus ventajas y desventajas y presenta los modelos mentales que cada de 8 consignas (instrucciones) favorece. Presenta la fórmula que vincula la fiabilidad de la nota final en la prueba, el número de PSM y el número de soluciones en ella.</p>	<p>D. Leclercq & Álvaro Cabrera p. 261-286</p>
14	<p>Reglas de redacción de las Preguntas de Selección Múltiples y la habilidad para responder pruebas Presenta 24 reglas (repartidas en 5 categorías) y los dispositivos experimentales (preguntas sobre contenidos ficticios) que permiten verificarlas, tan como los resultados de estas verificaciones en caso de transgresión de las reglas.</p>	<p>D. Leclercq p. 287-300</p>
15	<p>Evaluar procesos cognitivos según la Taxonomía de Bloom Presenta modalidades de evaluación apropiadas a cada de los 6 niveles de los procesos mentales descritos en la taxonomía de Bloom: la memoria (de re-cognición y de evocación), la comprensión (con la definición de Smedslund), la aplicación, el análisis (y las Preguntas PRIM-BIS para diferenciar entre análisis y comprensión, la síntesis y la creación (y los criterios de Torrance), el juicio(incluido la capacidad de aproximar).</p>	<p>D. Leclercq p. 301-328</p>
16	<p>Auto-evaluación con grados de certeza: un microscopio para la evaluación de los aprendizajes Presenta los retos del uso de grados de certeza: epistemológico (de definición de “dominio”), de medición en investigación (la necesidad de un microscopio del pensamiento), de caracterización practica (utilizable – inutilizable) de niveles de conocimiento) y de fijación de umbrales de éxito os resultados y de excelencia. Presenta las condiciones metodológicas de uso (3 principios), las distribuciones espectrales de calidad de les respuestas, las nociones de meta memoria y de meta comprensión (el JOC o juicio de comprensión).</p>	<p>D. Leclercq p. 329-356</p>
17	<p>Grados de certeza y docimología: como calificar Denuncia varios sistemas de cotejo inapropiados y la importancia (impredecible) de tener en cuanta el realismo de las respuestas acertadas por un estudiante en una prueba. Explica como verificar (con la ley binomial) la presunción de realismo, cálculo de un índice de calibración. Trata de la sobrestimación y de resolución (Discriminación y lucidez), tan como de una pauta innovadora de cotejo basada en ;los grados de certeza.</p>	<p>D. Leclercq p. 357-386</p>
18	<p>PdP: Pruebas de Progreso Presenta una modalidad de evaluación en cual la universidad de Maastricht se ha ilustrada como pionera: la Pruebas de Progreso que consisten en presentar el mismo día a todos los estudiantes de una carrera (que sean de primer o de ultimo año) una prueba sobre todos los contenidos de la carrera (centenas de preguntas), cuatro veces por año (con pruebas “paralelas”). Las ventajas y desventajas son revisitadas, como el modo de comunicar los resultados, original también. Estos principios son ilustrados por su aplicación en Maastricht desde cuarenta años.</p>	<p>D. Leclercq, A. Cabrera & C. Van der Vleuten (U. Maastricht) p. 387-408</p>
19	<p>TCS : El Test de concordancia de Script Esta técnica ha sido concebida para medir la capacidad clínica de tratar la información. Ha sido utilizada principalmente en medicina (revisión de opinión desde una información adicional). Es ilustrada con un ejemplo y resultados de su aplicación en la univ. de Liège.</p>	<p>V. Massart (ULg), A. Collard (ULg) D. Giet (ULg) p. 409-418</p>

20	<p>Concebir Dispositivos de Evaluación de los Aprendizajes (DEA) al nivel de un programa Presenta tres experiencias de desarrollo de un DEA al nivel de una facultad: la de Farmacia en Liège y las de medicina en Liège y en Maastricht.</p>	<p>D. Leclercq, C. Van der Vleuten & A. Cabrera p. 419-430</p>
21	<p>Retroinformaciones (Feedbacks) Empieza con el problema de la profundidad de penetración de una retroinformación, desde sobre los detalles de ejecución de la tarea hasta el <i>Self</i> (es porque son presentadas las teorías de William James sobre la auto-estima y la <i>FIT</i> o <i>Feedback Intervention Theory</i>). Un modelo integrador (llamado CAIRO) es presentado. Varios modos de presentación de las retroinformaciones después de una prueba son presentados. Una modalidad, utilizada en la UCH (Universidad de Chile) que se focaliza al esencial, es presentada con un ejemplo.</p>	<p>D. Leclercq, M. de la Fuente (UCH) & A. Cabrera p. 431-454</p>
22	<p>Los roles de un SMART: Servicio Metodológico de Apoyo a la Realización de Tests Un (SMART) ayuda docentes en la concepción y la realización de pruebas estandarizadas y en el procedimiento de las respuestas de los estudiantes (calcula de varios índices relativos a cada pregunta y cada solución de las PSM), como en las retroinformaciones automatizadas a los estudiantes. Un enfoque especial es dedicado al uso de cajas de voto a distancia (<i>clickers</i>).</p>	<p>D. Leclercq & P. Detroz (ULg) p. 455-476</p>
23	<p>Índices cuantitativos en Docimología Consiste en un catálogo de conceptos útiles para tratar cuantitativamente los datos resultando de evaluaciones estandarizadas como</p> <ul style="list-style-type: none"> -los tipos de categorías (nominales, ordinales, métricas). -los índices relativos a una distribución : índices de centración (Modo, Mediana, Media), de dispersión (rango, cuartiles, desviación estándar), de posiciones relativas o normativas (la nota z, los percentiles) de la forma de la distribución (asimetría o <i>skewness</i>). -las presentaciones gráficas de distribuciones. -índices de comparación o de progreso: la amplitud del efecto (AE), la ganancia relativa (GR). -la fiabilidad de la nota (<i>reliability</i>) al total de la prueba y el alfa de Cronbach. -el umbral de éxito, fijado a priori o a posteriori. -el índice de discriminación (correlación punto <i>biserial</i> o <i>rpbis</i>) de un modo de respuesta aplicado a cada de las soluciones de cada PSM -el análisis automática de una prueba -el valor heurístico de los nubes de puntos. 	<p>D. Leclercq, R. Roco (Chile) & A. Cabrera p. 477-543</p>
24	<p>Index de los autores 426 autores citados.</p>	<p>D. Leclercq & A. Cabrera p. 545-549</p>
25	<p>Index de los conceptos Se puede bajar gratuitamente via http://hdl.handle.net/2268/180060</p>	<p>D. Leclercq & A. Cabrera</p>