

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 III

El Prisma de las características de un Dispositivo de Evaluación de los Aprendizajes (DEA).

La (compleja) rosa de los vientos de las características de un DEA (¿tan compleja como para perder el norte?)

DIEUDONNÉ LECLERCO

Preámbulo: la necesidad de precisiones conceptuales

Dimensiones

Bloom⁷ *et al.* (1971) popularizaron la oposición entre “evaluación sumativa” y “evaluación formativa”, los términos exactos del título de su *Handbook of formative and summative evaluation*. Ahora bien, durante los últimos 40 años la investigación y las prácticas en evaluación se han ampliado a la par con la necesidad de contar con precisiones conceptuales, resultando en que la oposición entre formativa y sumativa ya no parece ni adecuada ni suficiente. En primer lugar, pues se trata de dimensiones diferentes: lo formativo corresponde a la función de la evaluación; lo sumativo refiere a la precisión. Para concebir, diseñar e implementar Dispositivos de Evaluación de los Aprendizajes (DEAs) que den cuenta de la complejidad actual, debemos tener una visión global de todas las dimensiones y ser capaces de describirlas con palabras adecuadas.

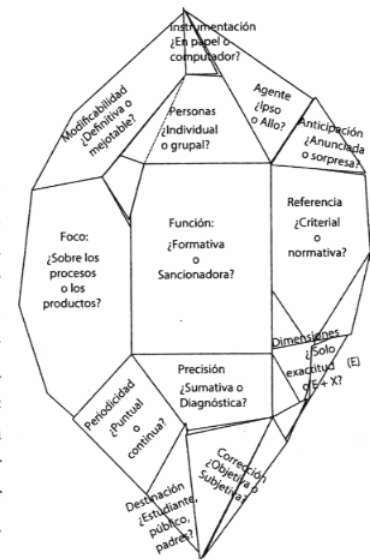


Figura 1: El Prisma de las Condiciones de un DEA

Prisma

Sin hablar todavía de los instrumentos (Segunda Parte de este libro), un Dispositivo de Evaluación de los Aprendizajes (DEA) presenta facetas múltiples (como un diamante o un prisma) a través de las cuales se miden las competencias y los desempeños de los estudiantes. Cada faceta es una dimensión, como se muestra en la Figura 1.

⁷ Ver detalles sobre Prof. Benjamin Bloom en Capítulo 15.

Rosa de los vientos

Hemos propuesto otra imagen, la de una "Rosa de los vientos" (Figura 2) con muchas agujas, cada una de ellas indicando algo diferente y constituyendo una dimensión particular que no debe ser confundida con las demás. Esta imagen ilustra también que estas dimensiones se pueden concebir como ejes bipolares. Puede que el lector llegue a crear nuevas dimensiones, pues el campo de la evaluación de aprendizajes se mantiene en constante evolución

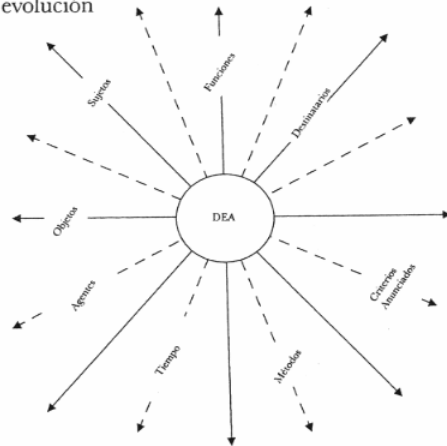


Figura 2: La Rosa de los Vientos de las Condiciones de un DEA

Profesionalismo en el uso de los términos

Se recomienda evitar la "mutilación" del lenguaje. Por ejemplo, no diremos "Evaluación formativa" sino "Evaluación con intención formativa". Tampoco diremos "Evaluación criterial" sino "Evaluación con referencia criterial", las palabras destacadas indicando la dimensión concernida. Dos dimensiones diferentes no se excluyen sobre el mismo eje, pues refieren a características diferentes que se pueden combinar; no están sobre la misma faceta del prisma. Además, la expresión "evaluación formativa" parece implicar que la evaluación ha sido efectivamente formativa, ha provocado cambios significativos, mientras que la expresión "evaluación con intención formativa" no presupone que sea eficaz.

Con frecuencia utilizaremos definiciones del diccionario (1979) de Gilbert de Landsheere (1921-2001), primer premio "Vasconcellos" del Congreso Mundial de Educación (México).

A menudo en un curso (y con mayor razón en un programa) un Dispositivo de Evaluación de los Aprendizajes (DEA) está constituido por varios componentes, a su vez compuestos por instrumentos de evaluación. Para cada uno de ellos el docente ha tomado —conscientemente o no— decisiones. Los conceptos analizados en este capítulo sirven para enfrentar dos desafíos:

- ayudar en la descripción de DEAs que ya existen y mejorarlos
- ayudar en la concepción de nuevos DEAs

Por esta razón comenzamos con una visión global; con una pauta que se puede completar para que cada decisión sea consciente y reflexionada. El Capítulo 5 da un ejemplo de uso de esta pauta.

Tabla 1: Plantilla para describir en forma sintética el prisma (condiciones) de un DEA y sus componentes

		COMPONENTES DEL DEA		
		COMP. 1	COMP. 2	COMP. 3
A. Las 3 facetas prioritarias				
A1. Función (PARA QUÉ)	Sancionadora			
	Formativa			
A2. Referencias	Criteriales (absolutas)			
	Normativas (relativas)			
A3. Precisión	Sumativa			
	Diagnóstica			
B. Objeto (QUÉ) B1. Foco				
	Productos			
	Procesos			
B2. Dimensiones	Exactitud + Concisión			
	Exactitud + Certeza			
	Exactitud + Rapidez			
	Exactitud +			
C. Sujetos (QUIÉN)				
	Individuos			
	Grupos / equipos			
D. Destinatarios (PARA QUIÉN)				
	Confidencial			
	Intermedia			
	Pública			
E. Agentes (POR QUIÉN) y en cuál ámbito (360°)				
	E1. Ámbito Escolar			
	Docente/experto/ipso/pares			
	E2. Ámbito profesional			
	E3. Ámbito Intermedio			
E4. En una relación	Arbitraria			
	Transaccional (confrontada)			
F. Tiempo (CUÁNDO) Periodicidad				
	F1. Continua			
	F2. Puntual			
	F3. Única			
	F4. Repetida			
	F5. Mejorable			
G. Estandarización (CÓMO)				
	G1. de la corrección subjetiva / objetiva			
	G2. del contexto: natural vs estandarizado			
	G3. adaptativo según contenido, exigencias, desempeños anteriores, lugar, momento.			
H. Previsibilidad de los criterios				
	H1. Anunciados			
	H2. Escondidos			
	H3. Intermedios			

A. Las FUNCIONES de un DEA: ¿POR QUÉ? ¿PARA QUÉ?

Tanto esta sección (A1) como la siguiente (A2) han sido tratadas en el Capítulo 2. Las evocamos aquí para resaltar su importancia como dos ejes capitales de las condiciones de la evaluación.

A.1. Evaluación con FUNCIÓN (o intención) FORMATIVA VS SANCIONADORA

A) LA EVALUACIÓN CON FUNCIÓN FORMATIVA

Inspirándonos en De Landsheere (1979, p. 113), definimos una EVALUACIÓN CON FUNCIÓN FORMATIVA como una “Evaluación que interviene, en principio, al fin de cada tarea de aprendizaje y que se dedica a informar al aprendiz y al docente del *nivel de dominio* [de capacidades y/o del *nivel de desarrollo* de las competencias] logrado y, eventualmente, permite descubrir dónde y en qué un aprendiz tiene dificultades de aprendizaje, para que descubra estrategias que le permiten progresar. ...La expresión ‘evaluación formativa’ propuesta por Cronbach y Scriven marca bien que la evaluación es ...parte integrante del proceso educacional normal, entendiendo los ‘errores’ como momentos del aprendizaje, y no como debilidades reprobables o manifestaciones patológicas. ...La evaluación con función formativa permite también determinar si un aprendiz domina los prerrequisitos”.

En la definición original, que fue escrita antes del movimiento de la formación orientada por competencias, hemos

- añadido “[de capacidades y/o del nivel de desarrollo de las competencias]”.
- subrayado los tipos de nivel para insistir en que una capacidad se puede dominar perfectamente, pero una competencia no tiene límite superior de desarrollo. Los estudiantes, una vez titulados, son competentes pero aún no experimentados, tampoco expertos.
- reemplazado “alumno” por “aprendiz”.

Esta expresión “Evaluación con función formativa” sirvió al comienzo para designar el proceso de mejoramiento de los programas (o del currículo). Fue Glaser (1963) quien la aplicó en un sentido orientado al aprendizaje.

La evaluación con función *formativa* puede ser autoadministrada (este es el reto de la metacognición y del aprendizaje autorregulado). A menudo tiene un carácter privado (no público), repetitivo (tantas veces como sea necesario) y efímero (el resultado cambia cada vez). Frecuentemente su referencia es criterial (no debe preocuparse del resultado de los otros aprendices, con excepciones como en el ámbito de los deportes de competición).

B) LA EVALUACIÓN CON FUNCIÓN SANCIONADORA

La evaluación con FUNCIÓN SANCIONADORA está al servicio de una regulación exterior al aprendizaje (funcionamiento del sistema escolar, contratación en una empresa, etc.). Esta evaluación puede ser sancionadora de dos maneras: la certificación (de cara al pasado) y la selección (de cara al futuro).

A.2. Evaluación con referencia NORMATIVA (o Relativa) VS CRITERIAL (o Absoluta)

A) LA EVALUACIÓN CON REFERENCIA CRITERIAL

La expresión *Criterion Referenced Tests* fue propuesta por Robert Glaser⁸ en 1963.

Esta evaluación tiene en cuenta *criterios de referencia* absolutos, fijos, que en principio son los objetivos a lograr. Así, una pregunta justifica su presencia en un test, si es que el objetivo correspondiente la necesita, aun cuando no discrimine (es decir, no sirva para medir las diferencias entre aprendices).

La nota mínima para aprobar es fijada sin importar el número (la tasa) de estudiantes que la lograrán. Por supuesto, como esta tasa no se conoce de antemano, el examinador se expone tanto a que todos la logren como a que todos fracasen.

Para evitar esta incertidumbre hay evaluadores que califican en un primer momento de forma criterial (y dan a cada aprendiz los puntos que merece) y, después, reinterpretan esas notas de manera normativa, y aprueban a los que se ubican en el porcentaje superior o a los *x* mejores, o a los que tienen una nota *Z* superior a -1 , etc. Esa práctica hace depender el éxito de una persona de la performance de otras, de modo que para un candidato es mejor presentarse un año en el cual los otros concurrentes sean débiles.

B) LA EVALUACIÓN CON REFERENCIA NORMATIVA

Esta evaluación expresa los resultados de una persona *X* en términos de su posición entre los resultados de un grupo que sirven de norma. Así, se dirá de *x* que su resultado se ubica en el cuarto lugar, o que es superior al promedio, o que es 2 desviaciones estándar más alto que el resultado promedio o que su nota *Z* vale $+2$. Este tipo de evaluación es típica de concursos. En esta perspectiva, se privilegia aquellas preguntas que discriminan mejor (no demasiado fáciles, tampoco demasiado difíciles), siendo el ideal 50% de éxito en la pregunta.

C) SITUACIONES MIXTAS

- (1) *Evaluaciones con referencias semicriterial y semialeatoria*: Un ejemplo consiste en escoger al azar entre todos los que han logrado una nota suficiente. Es lo que se practica al ingreso de la Facultad de Medicina de la Universidad de Maastricht para selec-

⁸ Robert Glaser, Director del Learning Research y Development Centre (LRDC) en Pittsburgh.

cionar los n entrantes (siendo n fijado cada año, con diferencias según necesidades económicas y políticas).

- (2) *Evaluación semicriterial y seminormativa*: Un ejemplo es un curso de una escuela secundaria que elige a sus mejores candidatos para la emisión de TV "Concurso inter-escuelas". Hay 9 candidatos que satisfacen los criterios, pero solo cuatro pueden estar presentes en el estudio de TV, por lo que se debe elegir a los mejores dentro de ese grupo.

A.3. Evaluación con PRECISIÓN SUMATIVA VS DIAGNÓSTICA

A) LA EVALUACIÓN CON PRECISIÓN SUMATIVA (O GLOBAL)

Esta evaluación resume en una *nota única* (por ejemplo 5,8 sobre 7,0 o 12 sobre 20 puntos) o en un nivel (por ejemplo "Con Distinción") un conjunto (o *total*) de sub-medidas. Es el balance sobre un conjunto de tareas que constituyen un todo, correspondiendo, por ejemplo, a una calificación que representa el conjunto de todas las asignaturas de un semestre (De Landsheer, 1979, p.115). Es decir, no se hace al final de cada tarea de aprendizaje. Este índice numérico no indica dónde el aprendiz se destaca, tampoco dónde están sus debilidades (y en consecuencia tampoco indica las causas de estas debilidades).

B) LA EVALUACIÓN CON PRECISIÓN DIAGNÓSTICA (O EXPLICATIVA)

Esta evaluación especifica los puntos de fortaleza y de debilidad, detalla los procesos erróneos e, idealmente, sus causas. A continuación siguen varios ejemplos de técnicas para aumentar el valor informativo de las evaluaciones.

Técnica 1: Las Preguntas de dos caras: PRIM-BIS

Este tipo de pregunta, presentada por computador, opera siguiendo tres etapas:

1. Una pregunta PRIM que funciona con las instrucciones "Puede ser que haya absurdo en la pregunta o que falten datos para que sea posible elegir UNA de las soluciones". El estudiante contesta a esta pregunta.
2. El estudiante es informado de una parte de la repuesta correcta. Por ejemplo: "Hay un absurdo en la pregunta", pero no se dice cuál es este absurdo.
3. La pregunta BIS continúa: "¿Cuál es este absurdo?" y propone varias soluciones (incluida "ninguna de la precedentes").

Esto permite establecer un diagnóstico diferencial respecto a las respuestas. Un ejemplo detallado se presenta en el Capítulo 15, sección E.3.

Técnica 2: La descripción del pensamiento en voz alta

Ejemplo: Causas de errores en aritmética. Anne Bonboir (1960) dirigió un proyecto interuniversitario que sometió a 3.000 alumnos belgas de sexto grado (12 años) a una prueba

en aritmética de 300 preguntas, cubriendo 13 áreas de matemáticas (fracciones, porcentajes, geometría, etc.). Los investigadores no solo contaron cada uno de los tipos de errores, sino que además *entrevistaron* (grabando los razonamientos en voz alta) a alumnos que se habían equivocado y a otros que habían dado la respuesta correcta. El resultado fue un libro con las listas de errores y los *razonamientos que las producen*.

Ejemplo:

Tabla 2: Respuestas con sus diagnósticos y tipos de error

PREGUNTA: ¿CUÁL ES EL 6% DE 200 EUROS?		
La respuesta siguiente	Indica:	Tipo de error
6 euros	Olvido de multiplicar por el número de centenas.	A (proceso parcial)
20 euros	Olvido de multiplicar por el número de %	A (proceso parcial)
1.200 euros	Olvido de dividir por 100	A (proceso parcial)
12	Olvido de la unidad	B (imperfección)
206 euros	Adición en vez de calcular el porcentaje	C (incomprensión)
16 o 18 euros	Error de tabla de multiplicación	D (error aleatorio)

Establecer los diagnósticos permite idear remediaciones pertinentes. Así, a partir de esta base fue fácil construir un PACELBRO (Leclercq, 1969), es decir, un Programa de Autocorrección con Encuadernación en forma de Libro Barajado con Respuestas abiertas. En un libro como este el estudiante, después de haber contestado, gira la hoja y compara su respuesta con una serie de respuestas posibles (cada una revela una causa diferente de error de razonamiento) que direccionan a páginas que tratan del error detectado. La primera cosa que hace esta página es mencionar que "hay un error", dejando así al alumno la posibilidad de descubrirlo y de autodiagnosticarse (identificar la causa de su error), antes de remediar. Ahora, la misma función es ejecutada por un programa informático, todavía dejando el diagnóstico al aprendiz. Las remediaciones con el PACELBRO permitieron un mejoramiento en los casos A y B, pero no en los casos C y D de la Tabla 2.

Leclercq realizó la siguiente situación experimental: 10 cursos de 6^{to} grado de educación primaria, de 10 docentes diferentes en diferentes escuelas, cada uno de aproximadamente 20 estudiantes, rindieron durante la mañana un examen de matemáticas. Estos fueron corregidos durante el almuerzo, y por la tarde cada curso se dividió en dos grupos de 10 estudiantes en salas distintas. Un grupo trabajó corrigiendo los errores detectados en la prueba con el docente, mientras el otro dedicaba el mismo tiempo a la autocorrección usando un PACELBRO. Una evaluación posterior indicó que este sistema logró equiparar un 92% de la eficacia promedio de los 10 docentes en la remediación de los errores.

Consideramos que dar a los estudiantes las soluciones correctas a las preguntas, después de un test, es un paso importante en la dirección de una evaluación auténticamente diagnóstica, pero es por sí solo insuficiente como para que se le pueda llamar así.

¿Y EN SU CURSO?

¿Piensa Ud. que en su curso hay contenidos en los cuales sería fecundo listar los errores más frecuentes y entrevistar a los estudiantes que se han equivocado, para ayudar a los futuros estudiantes a autodiagnosticar sus tipos de errores y, consecuentemente, autocorregirse?

Técnica 3: El mapa conceptual

LA TÉCNICA CLÁSICA

El principio clásico del mapa conceptual consiste en pedir a una persona escribir una palabra clave en el centro de una hoja de papel (por ejemplo, “azúcar”, en un mapa construido por un paciente diabético), y entonces conectar otros conceptos que, desde su punto de vista, estén relacionados con esta palabra. Esto se puede hacer de una manera “pre” o “post” un proceso de formación (por ejemplo, educación en salud para el autocuidado). Este método tiene varios inconvenientes. Primero, el paciente no escribe sus conceptos en el mismo lugar de la hoja en el pre y post-test, lo que dificulta la comparación. Segundo, el paciente no indica las significaciones de los vínculos entre los conceptos, lo que se transforma en la razón más importante de la dificultad para interpretar.

EL MAPA ACOMPAÑADO

Este método ha sido elaborado por C. Marchand (2000), J-Ph. Assal y J-F. d'Ivernois en la Universidad de París 13, con pacientes diabéticos. En este método el educador terapéutico del paciente (ETP) escribe, en un gran papel (A3), lo que le dice el paciente. Como en la entrevista clínica de Piaget (1936), o en el entretenimiento de explicitación de Vermersch (1994), la interrogadora improvisa las subpreguntas para que se complete el mapa. Por ejemplo, pregunta la significación de cada vínculo entre dos conceptos. Esos vínculos pueden ser: “...es un..., ...tiene como ejemplo..., no se aplica con ..., se hace solo en caso de ..., etc.”.

Durante el post-test la interrogadora escribe las modificaciones de modo que sean muy visibles. La Figura 3 presenta un extracto de un mapa completo realizado por Claire Marchand (2000, 113); los nuevos conceptos se representan en globos de línea punteada, y los desaparecidos (en comparación con el pre-test) son rayados con una cruz.

INTERÉS DE LA TÉCNICA

Esta técnica permite observar si el crecimiento del mapa conceptual después de una actividad de formación resulta en un aumento (1) del número de conceptos o (2) del número de vínculos o (3) de la precisión de los conceptos o (4) de la precisión de los vínculos.

Se ve que la elaboración de un mapa acompañado necesita mucho tiempo del formador con cada paciente (estudiante). Por supuesto, en este ejemplo, el tema (azúcar) es central en la enfermedad (diabetes) y este es un reto vital.

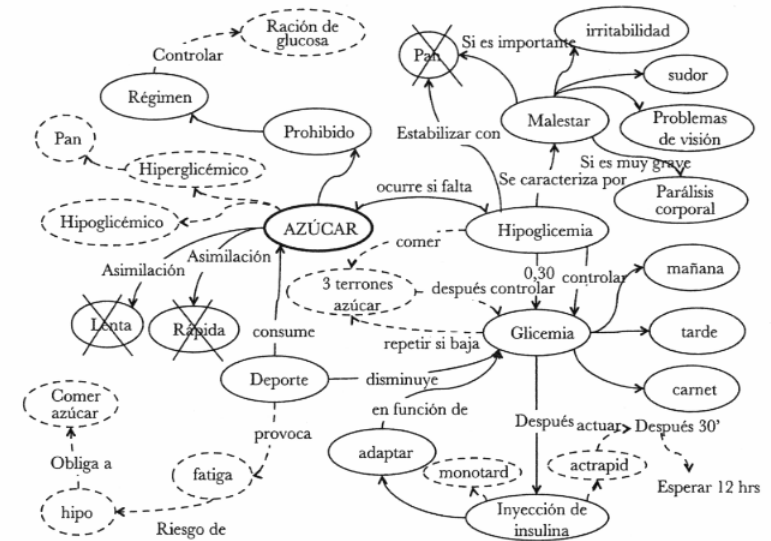


Figura 3: Extracto de mapa conceptual acompañado en formación de pacientes diabéticos por Claire Marchand (2000)

¿Y EN SU CURSO?

¿Piensa Ud. que, en su curso, hay conceptos que son mal entendidos por una gran parte de los estudiantes y, si es así, que el mapa conceptual podría ser fecundo y posible (en términos del tiempo necesario), individualmente o en grupo (la clase entera)?

Técnica 4: RESSAC o Resultados de Exámenes Estandarizados al Servicio del Aprendizaje en Candidaturas⁹

Durante el semestre N° 1 (de septiembre a enero) del primer año de la carrera de Psicología en la Universidad de Liège, 4 docentes responsables de 4 asignaturas diferentes decidieron calcular dos puntajes separados en sus exámenes de enero: un puntaje por Memorización y otro por Utilización-Comprensión, de modo que cada estudiante tenía 8 notas. Se calculó el promedio del grupo de más de 200 estudiantes¹⁰ (expresado como 0 (cero) en números de desviación estándar -ED), y cada puntaje fue transformado en una nota Z, que es la distancia al promedio. Por ejemplo, el estudiante G (Figura 5), en la asignatura A, en memoria obtuvo un puntaje que se puede expresar en Z = -0,5 (bajo, o peor, que el promedio en 0,5 ED).

⁹ Leclercq et al., 2003, pp. 155-170.

¹⁰ Los números varían de una asignatura a otra porque no era obligatorio presentarse a cada uno de los exámenes (se pueden pasar también en junio o en septiembre).

Aquí abajo se ven las “Radiografías Z” de tres estudiantes (D, G y H) que han fracasado en su examen de enero, sus respuestas al cuestionario de fin de año (septiembre), y el resultado en términos de éxito o fracaso.

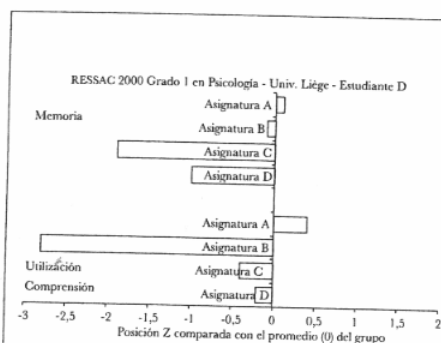


Figura 4: Radiografía Z del estudiante D

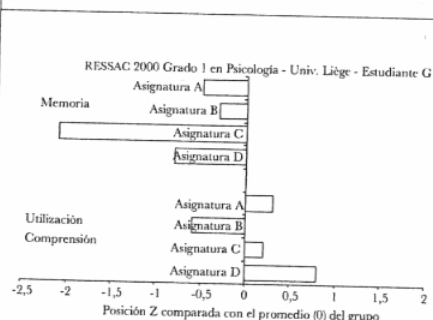


Figura 5: Radiografía Z del estudiante G

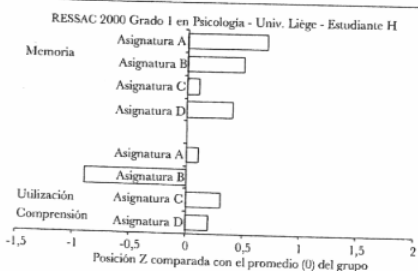


Figura 6: Radiografía Z del estudiante H

“¿Qué aprendiste de esta “radiografía Z”?”

D: “Tenía importantes lagunas en la comprensión de las preguntas”, lo que es una atribución causal interna (es mi culpa) y cambiante (estudiar con más profundidad).

“Esta observación ¿ha cambiado en algo tu modo de estudiar?”

D: “En vez de memorizar sin entender, he traducido el curso en mis términos, lo que resultó en mejor memorización también”.

A fin de año D tuvo un éxito.

“¿Qué aprendiste de esta “radiografía Z”?”

G: “Comprendo bien, pero no he memorizado bastante”, lo que es una atribución causal interna (es mi culpa) y cambiante (memorizar más).

“Esta observación ¿ha cambiado en algo tu modo de estudiar?”

G: “Sí. He redactado resúmenes insistiendo sobre los puntos más importantes”.

A fin de año G tuvo un fracaso.

“¿Qué aprendiste de esta “radiografía Z”?”

H: “Tengo peor comprensión que memoria”, lo que es una atribución causal interna, pero incambiable (tengo).

“Esta observación ¿ha cambiado en algo tu modo de estudiar?”

H: “No”.

A fin de año H tuvo un éxito.

La Figura 7 muestra las tasas de éxito en cada una de las cuatro asignaturas, y (a la izquierda) el promedio de todas. En paréntesis, los números de estudiantes concernidos. Por ejemplo, en la asignatura A fueron 28 estudiantes quienes declararon que *no* habían introducido cambios en su modo de estudiar (y su tasa de éxito fue 32%) y 34 estudiantes quienes dijeron que *sí* habían introducido cambios (y su tasa de éxito fue 56%).

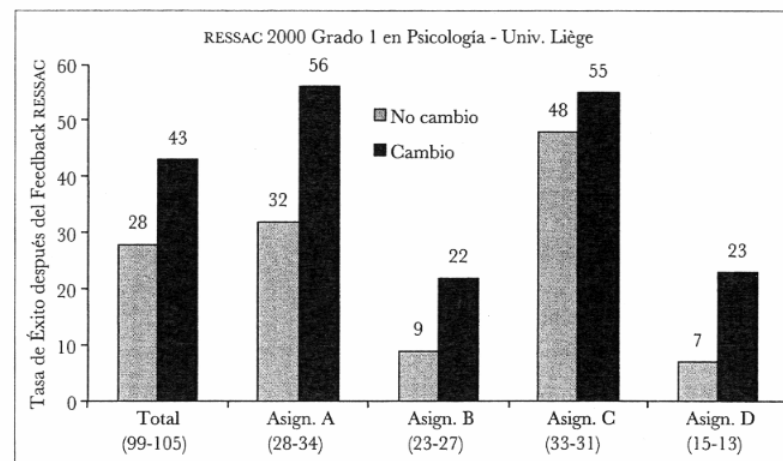


Figura 7: Tasa de éxito en cuatro cursos de estudiantes que declararon (sí o no) haber hecho cambios en sus modos de estudiar

Se puede ver que en cada una de las cuatro asignaturas la tasa de éxito de los que han declarado que habían introducido algún cambio es mucho más alta que la tasa de éxito de los otros (que declaran que no hubo cambio).

Este tipo de retroalimentación tiene tres importantes características: (1) la distinción es sencilla (solo dos categorías: memoria vs utilización-comprensión), (2) esa distinción está vinculada al modo de estudiar, y (3) la información es repetida (el estudiante ve que el fenómeno ocurre en más de una asignatura, de modo que no es ocasional sino sistemático).

¿Y EN SU PROGRAMA?

¿Hay un grupo de docentes que deberían y podrían concebir un RESSAC?

B. Los OBJETOS de la evaluación: ¿QUÉ observar o medir?

B.1. Los PRODUCTOS (o Resultados) vs los PROCESOS (o Trayectos)

A) LA OBSERVACIÓN DEL OBJETO: PRODUCTOS FINALES

La frase “Todo está bien si termina bien” considera únicamente el resultado final y no los procesos necesarios para lograrlo.

Ejemplo: Understanding Spoken English (USE)

Para practicar la comprensión del inglés oral el software USE presenta sobre la pantalla un texto donde cada letra ha sido reemplazada por un punto, como se ve en el ejemplo.

.....
.....

Después, el aprendiz escucha el texto

The magical number seven.
All my life I have been persecuted by an integer.

y es invitado a introducir, vía el teclado, las palabras que ha comprendido escuchando. Por ejemplo:

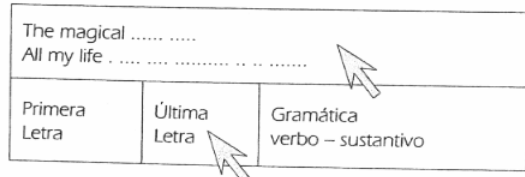
The magical
All my life

Evaluación del PRODUCTO (resultado): El corrector cuenta el número de palabras reemplazadas después de 5 minutos.

B) LA OBSERVACIÓN DE LOS PROCESOS, DE LOS RECORRIDOS MENTALES

Ejemplo 1: El software EPEL

Este software es parecido a USE, pero además permite al estudiante pedir ayuda sobre una palabra, indicándola con el mouse. Puede preguntar la primera o la última letra de la palabra, y su naturaleza gramatical (verbo, sustantivo).



Evaluación del PROCESO (trayecto)

Las trazas de las ayudas solicitadas por el estudiante permiten ver si son de tipo lógico (gramática) o de sonidos (letras), es decir, el PROCESO mental que ha guiado al estudiante.

Otros ejemplos:

- (2) *Modos de resolución de ecuaciones.* Para evaluar el nivel de conceptualización algebraico logrado por los estudiantes, Burton (1999) observa los *modos de resolución* (intuitivos o algorítmicos) practicados por estudiantes en la resolución de ecuaciones
- (3) En una prueba, con frecuencia ocurre que algunos estudiantes no alcanzan a llegar a las últimas preguntas: es un índice de que no han manejado bien el orden para contestarlas en el tiempo fijado.
- (4) Arreglar una tv: ¿Leyendo antes las instrucciones o intentando al azar?

¿Y EN SU CURSO?

¿Existen temas donde hay necesidad y posibilidades de observar el PROCESO y no solo el resultado? Si es así, ¿cuáles?

B.2. Evaluación con dimensión ÚNICA o con dimensiones MÚLTIPLES

A menudo los evaluadores se contentan con evaluar la exactitud de la respuesta, sin tener en cuenta la duración de la producción de cada una, los apoyos que fueron necesarios, el grado de certeza en la respuesta, entre otras posibles dimensiones.

B.2. A) EVALUACIONES BIDIMENSIONALES: EXACTITUD + CONCIENCIA.

En la resolución de ecuaciones, Burton (1999) tiene en cuenta la exactitud de la respuesta final y el número de *transformaciones* efectuadas para llegar a ella. Calcula una nota de *eficacia* sobre la base del número mínimo de etapas con las cuales un experto puede resolver el problema.

B.2. B) EVALUACIONES BIDIMENSIONALES: EXACTITUD + CERTEZA

Entre otros, De Finetti (1965), Shuford *et al.* (1966) y Leclercq (1982, 1993, 2003), tienen en cuenta el grado de certeza en cada respuesta. Eso les permite calcular varios índices de *realismo* en la autoestimación (¿Subestimación? ¿Sobreestimación?), establecer el *espectro de calidad de los desempeños (de las respuestas)*, y calcular una nota apropiada (en inglés *proper scoring rule*).

B.2. C) EVALUACIONES BIDIMENSIONALES: EXACTITUD + RAPIDEZ

Hay una relación entre la exactitud y la rapidez. En matemáticas, Burton (1999) ha observado que las respuestas correctas son emitidas más rápidamente que las inco-

rrectas. Fischer (1996) también ha observado que la detección de expresiones correctas (por ejemplo: $7 \times 6 = 42$) es más rápida que la de incorrectas (por ejemplo: $8 \times 6 = 54$). En algunos ámbitos el tiempo de reacción es parte de la calidad de una respuesta. Por supuesto, un test en tiempo limitado tiene la rapidez en cuenta, pero aquí nos interesamos en la medición de la rapidez en cada respuesta.

Ejemplo 1: Merciadri (1982) ha testeado a jóvenes parturientas (que recientemente han dado a luz) y ha medido el tiempo que necesitan para encontrar el número de teléfono del centro anti-veneno en la guía telefónica.

Ejemplo 2: Leblanc (1990) ha medido el tiempo entre la visión de una tira humorística sin palabras y el "Ahá" indicando la comprensión. Eso sirve para evaluar tanto la sutilidad de la persona como la eficacia del dibujo.

Ejemplo 3: La Reanimación Cardíaco-Pulmonaria (de vez en cuando llamada "boca a boca"), para ser un éxito debe ser realizada en menos de 5 minutos.

Ejemplo 4: Dehousse (1981) y Mathues (1981) midieron el tiempo necesitado por una enfermera para reparar la avería de un pulmón artificial (aparato de respiración artificial).

Ejemplo 5: Hay muchas razones médicas para reducir la duración de la amputación de un miembro; reducir la pérdida de sangre y la duración (en términos de años) de los dolores posoperatorios (que se llaman dolores fantasmas¹¹). Larrey, el quirúrgico de Napoleón, amputaba una pierna en 10 minutos.

Ejemplo 6: Foucambert¹² (1980) ha inspirado conceptos y softwares (por ejemplo ELMO) de entrenamiento en la lectura rápida. Ha detectado que, en francés,

- 1- Bajo 550 signos/minuto el lector está en la fase de exploración silábica y tiene muchas dificultades para comprender lo que lee.
- 2- Entre 550 y 820 signos/minutos el lector logra, aun con muchas dificultades, construir el sentido.
- 3- Entre 820 y 1.300 signos/minuto se puede verdaderamente hablar de lector. La comprensión se hace sin problema.
- 4- Sobre 1.300 signos/minuto, la lectura está bien instalada.

Ejemplo 7: Ackermans, Ernould y Leclercq (1985) han considerado la rapidez de la respuesta al trabajar con un software (MASTER DIAB) para entrenar niños y adolescentes diabéticos a reaccionar en diferentes situaciones (inyecciones, actividades físicas, análisis de sangre).

Kagan (1965) ha estudiado lo que ha llamado el "tiempo cognitivo". En sus términos, en un test, los estudiantes *reflexivos* cometen menos errores que el promedio de los estudiantes, pero en un tiempo superior comparado a este promedio. Los *impulsivos* co-

¹¹ J-Ph. Assal, Director del Centro de referencia de la OMS sobre Educación del Paciente Diabético, en Ginebra, nos ha enseñado que mientras más largo ha sido el tiempo de la amputación, más largo es el tiempo que perduran aquellos dolores.

¹² En <http://bigonnet.club.fr/Foucambert.html>. Consultado en agosto de 2005.

meten más errores en menos tiempo que el promedio. En los otros dos grupos, sin nombre, hay uno en que los estudiantes obtienen los peores resultados (más errores en más tiempo) y en el otro obtienen los mejores (menos errores en menos tiempo):

Tabla 3: Comparación exactitud/rapidez de las respuestas de cuatro tipos de estudiantes

TIEMPO DE RESPUESTA	-----▶	< MEDIA	> MEDIA
% respuestas	< Media	Impulsivos	Peores
correctas	> Media	Mejores	Reflexivos

Estas categorías tienen referencia normativa (la duración no está indicada en términos de minutos, sino en relación al tiempo promedio del grupo, tiempo que no está definido antes del test).

B.2. D) LOS EXÁMENES EN TIEMPO LIMITADO

a) *El tiempo global es limitado*

Muchas pruebas escritas se hacen en tiempo limitado. Los estudiantes hábiles en pasar tests (en inglés "test wiseness") analizan rápidamente cada pregunta y contestan en primer lugar las que conocen bien, dejando las otras para más tarde, en lugar de bloquearse sobre una sola pregunta.

b) *El tiempo es limitado en cada parte*

En los EOCÉ, también llamados OSCE (ver Capítulo 8) cada estación tiene una duración limitada (7 minutos, por ejemplo).

¿Y EN SU CURSO?

¿Hay pertinencia y posibilidad de que se evalúen otras dimensiones además de la exactitud (como la concisión, certeza, rapidez u otras)?

C. Los SUJETOS de la evaluación: ¿QUIÉN?

C.1. La evaluación de personas individuales (solos o en grupo) a través de la producción INDIVIDUAL

Es la más común de las evaluaciones en la Educación Superior. A través de tests y exámenes que se resuelven de forma individual, se infiere el desarrollo de competencias de los individuos.

C.2. La evaluación de la producción GRUPAL

Los criterios para evaluar un trabajo realizado por un grupo son los mismos que para un trabajo individual, pero con una calificación común a todos los estudiantes de un equipo (ver un ejemplo en el Capítulo 10). El docente también puede desear añadir nuevos criterios, como sobre

- El proceso. Por ejemplo, si existe un cronograma de trabajo, o el tipo de recursos humanos que han sido utilizados, o la lógica de armonización con que se han combinado las responsabilidades o tareas de cada miembro.
- El producto. Por ejemplo, si existen partes diferenciables que puedan ser atribuidas a individuos, o con qué tipo de equilibrio o pertinencia han sido distribuidas (ver un ejemplo en Capítulo 11: PARMs).

C.3. La evaluación de la contribución de cada persona en la producción GRUPAL

Evaluar lo que cada individuo sabe, moviliza y aporta en una producción grupal es un enorme desafío. El Capítulo 11 presenta varios métodos para enfrentarlo.

D. LOS DESTINATARIOS de la evaluación: ¿PARA QUIÉN?

Evaluación con destinación CONFIDENCIAL-PRIVADA VS PÚBLICA

D.1. Resultados con destinación privada-confidencial

La evaluación con función formativa es a menudo privada: “*es una forma de diálogo particular entre el educador y su alumno*” (De Landsheere, 1979, p. 115).

D.2. Resultados con destinación pública

La comunicación de las notas es pública cuando se proclaman los resultados durante una ceremonia o cuando se publican en un periódico. Los resultados de las evaluaciones con función certificativa tienen generalmente una destinación pública.

D.3. Situaciones intermedias (confidencialidad parcial)

El boletín escolar tradicional de la enseñanza primaria o secundaria es semipúblico porque es conocido al menos por los padres del alumno (es hecho para eso). En la universidad, como los estudiantes son mayores de edad (tienen más de 18 años), en muchos países sus padres no pueden acceder a los resultados sin el permiso del estudiante. Hay algunos resultados que son entregados al estudiante pero que las institu-

ciones pueden solicitar, entre otros casos, al ingreso en una universidad. Por ejemplo, los resultados en el TOEFL (Test Of English as a Foreign Language), o para los médicos los Medical Boards Examinations (en EE.UU.).

D.4. Adaptación de los mensajes según el destinatario

Aquí sigue un ejemplo de destinatarios múltiples. En la Universidad de Liège los cursos son evaluados por los estudiantes con cuestionarios, en un proceso conocido como EEE (Evaluación Estudiantil de la Enseñanza). Variados *feedbacks* tienen diversos grados de confidencialidad.

Todo el mundo (incluyendo los estudiantes) puede ver, para cada pregunta (por ejemplo “Las presentaciones orales están bien estructuradas”) la distribución de las notas promedio de las 8 asignaturas del semestre (las notas variando de “-3 Nunca” hasta “+3 Siempre”). En este ejemplo se puede ver que el promedio de las notas de dos asignaturas es máximo, y que no hay asignaturas evaluadas con el mínimo

Tabla 4: Feedback público para una pregunta de la EEE

Las presentaciones orales están bien estructuradas	-3	-2	-1	+1	+2	+3
		2	1	1	2	2

Cada docente recibe la misma distribución con la posición de su propia asignatura coloreada.

Tabla 5: Feedback para el docente de una pregunta de EEE

Las presentaciones orales están bien estructuradas	-3	-2	-1	+1	+2	+3
		2	1	1	2	2

En este ejemplo el profesor puede ver que otra asignatura ha recibido la misma nota promedio que la suya, pero no sabe cuál. El Decano de la Facultad puede ver la posición de la nota promedio de cada asignatura, con el apellido del docente responsable. El Decano felicitará a B y E, y pedirá un esfuerzo extra a C y H.

Tabla 6: Síntesis que recibe el Decano de los resultados para una pregunta de la EEE en una carrera

Las presentaciones orales están bien estructuradas	-3	-2	-1	+1	+2	+3
Asignaturas		C H	G	D	A F	B E

¿Y EN SU PROGRAMA?

¿Hay razones y posibilidades de utilizar, con las calificaciones de los estudiantes, esta forma de comunicar resultados?

E. LOS AGENTES de la evaluación: ¿POR QUIÉN? ¿En 360°?

La *Evaluación en 360°*¹³ tiene su origen en los departamentos de Recursos Humanos de las empresas, e implica que una persona es evaluada desde todos los posibles actores que se involucran con él/ella en una determinada función, acción o servicio. La evaluación se produce desde los superiores, los pares, los subalternos, los usuarios y el propio trabajador.

Desde este origen en la empresa, la Evaluación en 360° ha sido adaptada al ámbito escolar en general, y la educación superior en específico.

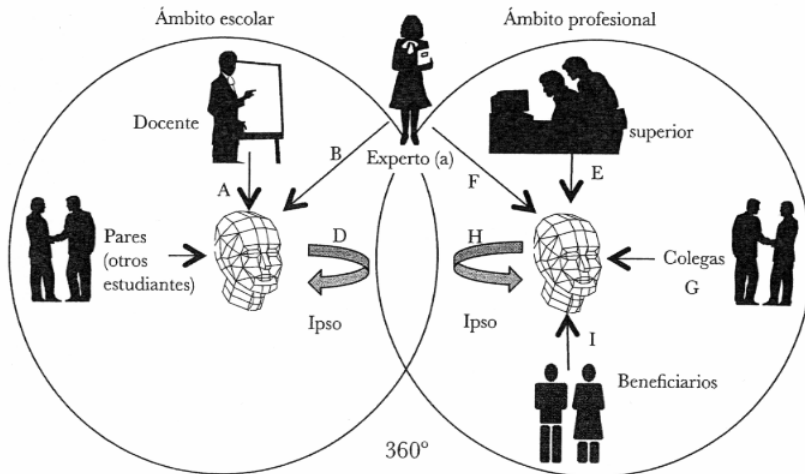


Figura 8: Evaluaciones 360° en dos ámbitos: escolar y profesional

E.1. En el ámbito escolar

La *situación A: Evaluación por el docente* que ha dado el curso. Es la más clásica, por la persona mejor informada de lo que pasó (el contenido transmitido, los ejemplos, los métodos, las circunstancias, etc.).

La *situación B: Evaluación por un(a) experto(a) externo(a)* que no ha dado los cursos y es invitado(a) a ser miembro *adicional* del tribunal. Cuando son *únicamente* personas

¹³ Ver también, en el Capítulo 10, sección B, el *Esquema Multidimensional de Evaluación*.

externas al programa de formación los que conciben y corrigen las copias, se habla de exámenes externos. Ejemplos son la PSU (prueba de selección universitaria) en Chile, el *Bachillerato* en Francia, los concursos de entrada en las Grandes Escuelas en Francia (Polytec, ENA, ENS, EMP, EMN¹⁴), y el SAT (*Scholastic Aptitude Tests*) en EE.UU.

La *situación C: Evaluación por los pares*. Este proceso puede ser eficaz siempre que los criterios de evaluación sean bien explicitados (Falchikov, 1995, p.175). Los contextos de "supervisión" (en una estancia en terreno) o de *mentoring* (por un estudiante más avanzado) utilizan este tipo de evaluación.

A menudo, lo que los pares evalúan son las performances y las actitudes, más que los conocimientos. Por ejemplo, cuatro grandes criterios para la evaluación por pares de una presentación oral pueden ser: (1) estructura y coherencia, (2) dominio del contenido, (3) cantidad de información transmitida, (4) claridad y expresividad en la presentación.

Korman y Stubblefield (1971) observaron que la evaluación (*rating*) por los pares constituye el mejor predictor de la performance médica de internos en un hospital, probablemente porque los estudiantes se conocen desde un ángulo al que el docente difícilmente puede acceder (Linn *et al.*, 1976).

D'Augelli (1973) constata que la evaluación por los pares está bien correlacionada con las de los expertos en lo que concierne a los "comportamientos interpersonales". Boud y Tyree (1979) observaron correlaciones pares / expertos altas (0,75 y 0,83), pero correlaciones ipso / expertos más bajas.

Los beneficios experimentados, declarados por los estudiantes que han sido evaluados por sus pares, se expresan en comentarios del tipo "un mejoramiento de mi performance, entendiendo mejor lo que los evaluadores consideran en las respuestas" (Margin y Helmore, 2001), y "esto me fuerza a estructurar mis respuestas y a estudiar más" (Falchikov, 1995).

Hay restricciones deontológicas y legales en la utilización de puntajes viniendo de pares *en una función certificativa*: no es responsabilidad de los estudiantes juzgar a los otros y entregar o denegar un diploma.

La *situación D: la ipso-evaluación*. Hemos preferido no hablar de *autoevaluación* (ambos términos significan "evaluarse a UNO MISMO") porque autoevaluación puede *también* significar "de manera autónoma". La *ipso-evaluación* puede ser *AYUDADA* (o apoyada, y, en este sentido, resulta menos "autónoma") en varios modos, dependiendo de quién define: (1) los objetivos a evaluar, (2) las preguntas que hacer, (3) el momento para evaluar, y durante cuánto tiempo, (4) los criterios de calidad, (5) de quién corrige y diagnostica, y (6) de quién decide las consecuencias (regulación). Mientras más definiciones son hechas por el estudiante mismo, más autónoma es una evaluación (auto-servucción¹⁵).

¹⁴ Ecole Polytechnique (de l'armée), Ecole Normale d'Administration, Ecole Normale Supérieure, Ecole des Mines de Paris, Ecole des Mines de Nancy.

¹⁵ Esta nueva palabra ha sido propuesta por economistas franceses para designar "la producción de servicios" y pone el foco sobre el hecho de que cada vez más el beneficiario de un servicio es invitado a ejecutar una parte de este (*self banking, self banking, self scanning, etc.*)

Poner a disposición de los estudiantes *bancos de preguntas* (con acceso a las respuestas correctas) favorece la autoevaluación autónoma permitiendo que los aspectos (1), (2), (3) y (6) sean decididos por el aprendiz y que el acto (5) sea ejecutado por el aprendiz mismo.

La evaluación es totalmente AUTÓNOMA cuando el estudiante decide todo y hace todo, aun en ausencia de bancos de preguntas, de criterios de exactitud, etc.

E.2. En el ámbito profesional

La *situación E* es donde el superior (el N+1) es el observador y el juez.

La *situación F* ocurre cuando un miembro externo a la empresa evalúa.

La *situación G* es cuando colegas evalúan.

La *situación H* es paralela a la D (ipso), pero evaluándose como un profesional, no como estudiante.

La *situación I* es particular al ejercicio profesional. Los beneficiarios (que dan su opinión) pueden ser clientes, pacientes, estudiantes, según la profesión.

E.3. Intersecciones entre los dos ámbitos

Cuando un estudiante pasa meses en terreno (lo que ocurre en casi todas las carreras), ¿estamos en el ámbito escolar o en el profesional? Si la evaluación se hace por colegas profesionales sobre las competencias profesionales del estudiante, estamos en el ámbito profesional. Pero si matizan, diciendo "hemos tenido en cuenta que aún es un estudiante", ¿dónde estamos? En una situación intermedia.

E.4. Evaluación en una relación ARBITRARIA VS TRANSACCIONAL

a) La evaluación ARBITRARIA

Esta evaluación "viene de arriba" y deja poco lugar para una discusión entre el evaluador y el evaluado.

b) La evaluación TRANSACCIONAL

Esta evaluación se produce cada vez que el evaluado puede expresar su punto de vista, que él/ella es escuchado/a, y tiene oportunidad de que su opinión sea aceptada, como se ilustra con el debate en el Capítulo 9. En acciones - investigaciones, un profesor¹⁶ puede pedir a sus estudiantes autoevaluar su propio trabajo (un informe escrito sobre una experiencia) en varios aspectos. Paralelamente, el docente evaluará los mismos aspectos de los mismos informes sin conocer las autoevaluaciones. Un diálogo transaccional se concentrará sobre las notas divergentes entre las ipso y las allo-evaluaciones. Revisiones de la notas se pueden producir en ambas direcciones.

¹⁶ Leclercq lo hizo con ayuda de asistentes (Jans, Serrurier, Delcomminette) en varias ocasiones.

Por ejemplo, en una experiencia de confrontación sistemática (Jans, 2000), cada estudiante tuvo que poner sus notas de autoevaluación en un sobre. Los profesores hicieron lo mismo con las suyas. Los dos sobres se intercambiaron a ojos de un testigo. El docente (D. Leclercq) había anunciado que el encuentro se desarrollaría como sigue: *Los 17 aspectos serán considerados, pero discutiremos más sobre aquellos en que no tenemos acuerdo.*

La Figura 9 es el gráfico que combina, para los 17 aspectos, las autoevaluaciones de un estudiante (Estudiante 07), en el eje horizontal y la allo-evaluación (por el docente) en el eje vertical. Este gráfico es la base de la confrontación docente / estudiante, que se realiza en presencia de una tercera persona. Este gráfico individual ha sido entregado a cada estudiante antes del encuentro, para que lo prepare. Durante este encuentro lo que importa son los argumentos, y las modificaciones de calificaciones que se produzcan deben ir con justificaciones escritas adjuntas. Además, las conversaciones son grabadas.

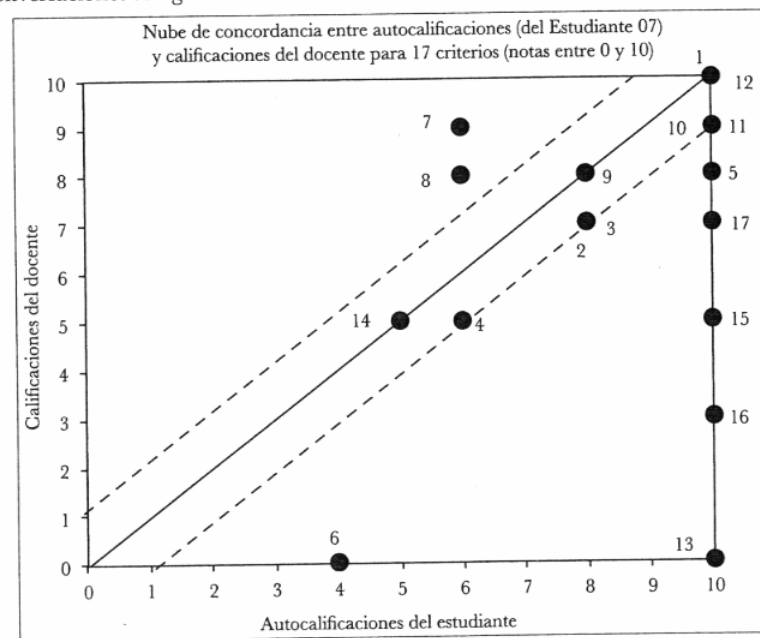


Figura 9: Correlación entre evaluación del estudiante (ipso) y del docente (allo) para 17 criterios

En el gráfico se ve que este estudiante es más optimista que el docente, pues en varios criterios se evalúa con mejor puntaje que el docente.

- un solo criterio (el 6) fue calificado con menos que 5 puntos por el estudiante; el docente lo calificó con un 0.

- algunos criterios (1, 2, 3, 4, 9, 10, 11, 12 y 14) se ubican en la zona de concordancia, es decir, con una diferencia máxima de un punto.
- solo dos criterios (7 y 8) son mejor calificados por el docente que por el estudiante.
- varios criterios (5, 13, 15, 16 y 17) fueron calificados como perfectos (nota 10) por el estudiante, y el docente les otorgó menos puntaje.

El encuentro entre estudiante y profesor se enfocará en los criterios de estos últimos dos grupos.

Jans (2000, 336) cita la frase de Boud (1995, 204-205) sobre un “feedback eficiente”:

Si Usted quiere dar un feedback eficaz, tendrá que: ser realista, preciso, sensible a los objetivos del estudiante, contestar en tiempo apropiado, ser descriptivo, no hacer juicios, no comparar, ser directo, positivo, consciente de su estado mental. Jans continúa: Y al receptor del feedback, Boud le aconseja ser explícito¹⁷, atento, consciente de su estado mental¹⁸, silencioso.

c) Situaciones intermedias

Hay legislaciones que estipulan que el estudiante tiene el derecho de ver su hoja (pero no la de otros estudiantes) corregida por el docente, y de exigir una re-corrección en caso de existir un error de juicio.

¿Y EN SU CURSO?

¿Hay interés y posibilidad de organizar este tipo de confrontación?

F. Los tiempos de la evaluación: ¿CUÁNDO?

Un docente (o, por ejemplo, un entrenador deportivo como un seleccionador nacional) pueden estimar la calidad de una performance (actuación, desempeño), ya sea:

- en *predicción*: antes de efectuar la acción (¿Debe X ser seleccionado?), y si la respuesta es NO, la calidad de esta predicción no se podrá verificar.
- en *per-dicción*: durante la performance, pero antes de que sea medida por un instrumento preciso (por ejemplo durante el match).
- en *post-dicción*: después de que la acción ha sido efectuada. Ej.: Gracias al video se puede mirar muchas veces la performance de un jugador de fútbol y contar su cantidad de “pelotas perdidas”, de los duelos en que salió vencedor o perdedor, de sus tiros (al arco). Hoy existen software de análisis detallado de estos y otros comportamientos.

¹⁷ Dejar en claro el tipo de feedback que está buscando.

¹⁸ Tomar conciencia de sus propias reacciones, tanto intelectuales como emocionales.

F.1. La evaluación con una periodicidad continua

Colección sistemática de scores o de apreciaciones... llegando a una nota final. Pueden intervenir en la elaboración de esta nota final: las tareas ejecutadas en clase o en casa, los resultados de pruebas escritas, de tests, así como de los juicios subjetivos del enseñante. La evaluación continua es... un proceso acumulativo que sigue el razonamiento del estudiante y refleja los cambios que ocurren en sus reacciones en el curso (De Landsheere, 1979).

Al contrario de la evaluación repetida (ver F4), *no son las mismas pruebas* las que se utilizan en la evaluación continua (o diluida o fraccionada o repartida en el tiempo). Hay docentes que organizan esas pruebas sin precisar el momento, para que los estudiantes se preparen permanentemente.

F.2. La evaluación con una periodicidad puntual o concentrada

Evaluación efectuada en un momento dado para contestar a una cuestión relativa a una persona o un grupo o un programa (promoción a una clase superior, atribución de una beca, etc.). Este tipo de evaluación no se justifica enteramente más que cuando una evaluación continua es imposible (De Landsheere, 1979, p.114).

Es el caso de la tesis de fin de programa de formación, porque los estudiantes nunca antes han tenido la oportunidad de llevar a cabo un trabajo de tal tamaño y con tales exigencias. Sin embargo, nos esforzamos en prepararlos con tareas similares pero “en versión reducida”.

F.3. La evaluación con una periodicidad única

Ocurre que una performance puede ser evaluada una sola vez en la vida de una persona y en un momento único (no hay posibilidad de un segundo intento). Es el caso de la defensa de un doctorado en algunos países (Bélgica, por ejemplo), aunque en otros (como EE.UU.) es un proceso más continuo.

En educación los ‘retests’ son difíciles porque el simple hecho de pasar el test produce un aprendizaje”. Con “retest”, De Landsheere (1979, p.238) entiende “el test repetido de forma idéntica: las mismas condiciones y las mismas preguntas.

F.4. Evaluaciones REPETIDAS

Se puede volver a dar el mismo test

- cuando se estima que la demora (han pasado muchos meses) hace olvidar las respuestas correctas. En el Capítulo 6, sección A.2., se encuentra un ejemplo de este tipo (un test de lectura en primer año de primaria).
- cuando la prueba contiene la población total de las preguntas posibles (por ejemplo, recitar los tiempos pasados de los verbos en inglés, las tablas de multiplicación).

ción, los conocimientos sobre la diabetes -para pacientes), y que se dispone del tiempo necesario para hacer todas las preguntas.

Es el caso de la batería de tests de lectura de Inizan (1963) compuesta de cuatro tests: en su investigación PREDIC, Boxus (1971, 48) presentó los cuatro tests a los mismos alumnos de primer año de escuela después de 3, 6 y 9 meses de iniciado el aprendizaje de la lectura. Los resultados se presentan en distribuciones en *i*, campana (Gauss) y en forma de J, como muestra la Figura 10.

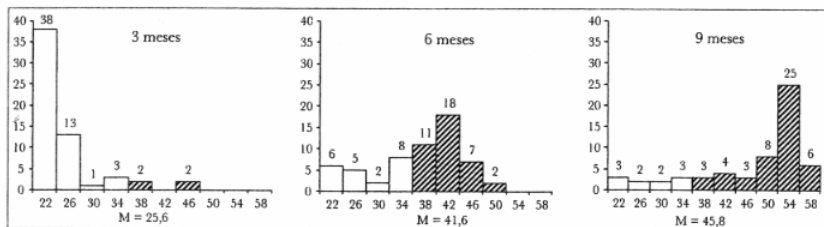


Figura 10: Distribuciones en *i*, Gauss y J. Mediciones 3, 6 y 9 meses después de iniciado el aprendizaje

F.5. Evaluación con oportunidad de MEJORAR el puntaje

a) Performances mejorables antes de la calificación

A menudo la calificación se hace sobre una performance que el estudiante no podrá modificar. Este carácter definitivo tiene su punto más alto de expresión en las consignas que no permiten las tachaduras o que, en las PSM, no permiten modificar la respuesta.

Al contrario, el SMART (Sistema Metodológico de Apoyo a la Realización de Tests) de la Universidad de Liège permite modificar fácilmente la respuesta y el grado de certeza, sin borrar y durante el test, mediante una línea de “arrepentimiento” (re) para cada uno. La Figura 11 ejemplifica un extracto de FORMULOM (Formulario para Lectura Óptica de Marcas) utilizado en el proyecto MOHICAN. En la primera y tercera filas el estudiante marca su respuesta y grado de seguridad. En caso de querer modificar su respuesta durante la prueba, en la segunda y cuarta filas debe marcar la nueva respuesta y grado de seguridad. El programa informático hace que el lector óptico lea las filas de arriba, en caso de que sean las únicas marcadas, o solo las fijas de abajo en el caso que se hayan utilizado.

R	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
R 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cert 1		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cert 2		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figura 11: Extracto de FORMULOM

Wood (1977, p. 234) ha elaborado sobre la modificabilidad de la evaluación, basándose en los estudios de Pippert (1966), Foote y Bellinks (1972), Reiling y Taylor (1972), Jacobs (1971), y Lynch y Smith (1975). Wood concluye que es mejor permitir a los estudiantes cambiar sus respuestas durante la prueba. Para profundizar esta conclusión sería interesante calcular las diferencias entre las respuestas originales y las modificadas, lo que hoy es posible mediante CAT (Computer Assisted Testing).

b) Performances y puntajes mejorables después de la calificación (doble oportunidad)

Esta modalidad se presenta en varias formas:

1°. Versión “revisada” (o bis) del trabajo

Un profesor puede permitir al estudiante modificar un informe, por ejemplo, después de que ha recibido críticas (escritas) relativas a la versión original. Es lo que han hecho Leclercq y Jans en 1999, y Delcomminette y Leclercq en 2012, otorgando como nota final el promedio de las dos versiones. Este principio evita que el estudiante presente una versión original chapucera, porque pesará la mitad de los puntos. Además, se debe tener en cuenta la capacidad de un estudiante de sacar beneficio de los consejos y críticas que ha recibido. La única situación dramática es aquella en que la primera versión es tan mala que la primera nota (débil) es “irrecuperable”. En la acción-experiencia de Delcomminette *et al.* (2012), los promedios de los puntajes de las versiones 1 han pasado de 11,3 (sobre 20) a 15,3, es decir, una Ganancia Relativa de 45%. Esos docentes-investigadores han estimado también el costo de la operación para dar referencias en términos de eficiencia (eficacia / costos).

2°. Defensa oral para mejorar un documento escrito

El docente puede ofrecer esta oportunidad solo si el documento alcanza un mínimo fijado antes. Es el principio de la defensa oral de las tesis.

3°. Feedback mínimo

El docente no puede limitarse a señalar una nota (sin precisión alguna) o el número de errores (sin indicar cuáles son) antes de permitir una nueva versión.

G. El grado de estandarización

G.1. Estandarización de la corrección

a) Corrección SUBJETIVA

No existiendo la objetividad, lo que podemos hacer es intentar acercarnos lo más posible a ella. La objetividad es con frecuencia definida como “el consenso entre expertos”. Piéron (1963), que inventó la palabra “docimología” (ciencia de los exámenes), estudió los efectos de corrección (ver Capítulo 6) y demostró que hay inconsistencias:

- de un mismo evaluador corrigiendo una misma hoja unos días después.
- entre varios correctores corrigiendo la misma copia.

En evaluación, para disminuir este tipo de efectos, se hacen esfuerzos para que las notas otorgadas a la misma performance no difieran de un juez (experto) a otro, y para que esos jueces se pongan de acuerdo sobre los instrumentos y criterios de evaluación.

Utilizar *escalas descriptivas de evaluación* (donde cada uno de los escalones está descrito) es mucho más objetivo puesto que la descripción de cada escalón tiene menos ambigüedad (ver Capítulo 6).

b) Corrección OBJETIVA

En EE.UU. se habla de "tests objetivos", especialmente cuando la corrección es automática. Es típico de las Preguntas Verdadero - Falso (PVF) o las Preguntas de Selección Múltiple (PSM). Eso no significa que esos procesos garanticen la objetividad y la equidad absoluta, como lo enseña el principio de los debates instaurado en los Tests Espectrales Metacognitivos (Capítulo 9): hay "errores" que pueden justificarse... si el estudiante tiene la oportunidad de hacerlo.

G.2. Estandarización del contexto

a) Evaluación en contexto NATURAL

Esta evaluación ocurre sobre casos y con las presiones (y azares) de la vida real, y en consecuencia es difícil asegurar estandarización. Por ejemplo, cuando los estudiantes se van a terreno durante meses, los lugares, las circunstancias y las personas que les acogen no son idénticos. Otra vez, escalas descriptivas y evidencias (como en un portafolio) son soluciones a este problema.

b) Evaluación en contexto ESTANDARIZADO

Se toman precauciones metodológicas para asegurar la comparabilidad de los resultados entre estudiantes o entre grupos. Todos los estudiantes reciben el mismo día y a la misma hora las mismas preguntas, contestan en las mismas condiciones (lo que es garantizado por vigilantes), y son evaluados con los mismos criterios y los mismos procesos.

G.3. Evaluaciones ADAPTATIVAS

a) Evaluaciones adaptativas *por el CONTENIDO*

Ocurre cuando, por ejemplo, se permite a cada estudiante elegir su pregunta o el tema de su ensayo. El *examen oral* es otro ejemplo que evoluciona en el tiempo, porque el interrogador modifica sus preguntas según las respuestas precedentes (principio de Wald, 1943). Un tercer ejemplo lo constituye el *testeo automatizado por computador*. Wald (1943)

propuso una fórmula que permite, a partir de una serie de preguntas (con algunas respuestas correctas y otras incorrectas), calcular la probabilidad de que el estudiante pueda ser declarado "competente" (y también la que tiene de ser declarado incompetente). Cuando un umbral fijado arbitrariamente (por ejemplo, 95% de certeza o 0,05 de probabilidad) no se logra (es decir, cuando se mantiene la incertidumbre), se continúa preguntando al estudiante.

b) Evaluaciones adaptativas *por el LUGAR*

En la formación a distancia es posible concebir exámenes que permiten al estudiante realizarlos desde su lugar, vía internet y/o vía videoconferencia. Esto plantea el problema de la estandarización de la vigilancia. La solución, que en inglés se llama Trusted Third Parties, o asociados fiables, consiste en que en el lugar distante de los examinadores existe un proceso y un responsable (de la universidad extranjera o de la embajada) que asegura esta vigilancia. Por ejemplo, cuando se hace una defensa de tesis a distancia por videoconferencia, un responsable académico de la otra universidad debe estar presente y visible en la pantalla.

c) Evaluaciones adaptativas *por el MOMENTO*

A menudo el estudiante no tiene la posibilidad de elegir el momento del test. El principio inverso operaba en el sistema IPI (Individually Prescribed Instruction) desarrollado por el LRDC (Learning Research y Development Center) de la Universidad de Pittsburgh, dirigido por Robert Glaser (1977). En este sistema cada alumno (de la enseñanza primaria) progresa a su propia velocidad con la ayuda de una combinación de libros y cintas de sonido, y se presenta a un test por computador cuando se siente preparado. Según su resultado, el programa le *prescribe* pasar a la unidad siguiente (si ha logrado el 85% o más de éxito) o profundizar en el tema, quizás con un método diferente. En este último caso pierde el privilegio de ser un *self reliant student*, un título que se puede ganar y perder.

En el sistema IPI se observan grandes diferencias de rapidez de *progresión* entre los alumnos de una misma edad. Los gráficos de las figuras 12 y 13 muestran la progresión de dos alumnos (Ann y Johnny) durante tres años de aprendizaje en la asignatura de matemáticas.

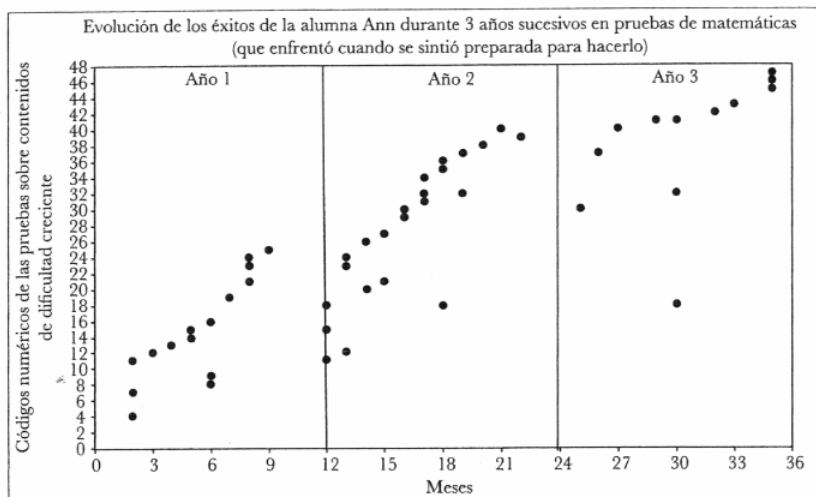


Figura 12: Evolución de estudiante Ann con el sistema IPI (Individually Prescribed Instruction)

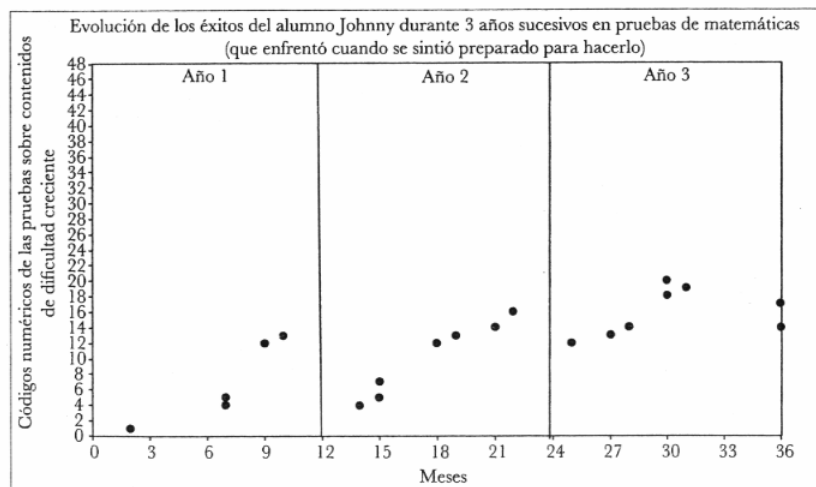


Figura 13: Evolución del estudiante Johnny con el sistema IPI (Individually Prescribed Instruction)

En el eje vertical se ubican los distintos tests de dificultad creciente. Por ejemplo: del test 1 al 3 tratan sobre operaciones aritméticas con números hasta 10. Del 12 a 16 se tratan de contar hasta 200 y las operaciones aritméticas hasta el 20. Del test 37 al 42 tratan sobre algoritmos de multiplicación y división y fracciones equivalentes.

Se observa que después de los periodos de vacaciones de verano baja el nivel de las pruebas logradas con éxito, logrando fuertes alzas en los meses siguientes. También se puede ver una gran diferencia entre las pendientes (el ritmo) de adquisición de aprendizajes de Ann y Johnny durante estos tres años.

Estimando que un sistema educativo democrático no puede aumentar tanto las diferencias inter-estudiantes en la escuela primaria, este sistema fue interrumpido. En la formación de posgrado hay sistemas de módulos que permiten a los estudiantes presentar el examen cuando quieren.

¿Y EN SU CURSO?

¿Es pertinente y posible que la evaluación se haga cuando el aprendiz lo desee?

H. La previsibilidad de los criterios de evaluación ¿CON CUÁL PREVISIBILIDAD?

H.1. Evaluación con Criterios ANUNCIADOS en la consigna

En esta evaluación todos los criterios son anunciados al comienzo.

Por ejemplo, en las PSM, la consigna precisa:

- la naturaleza de la pregunta (por ejemplo: solo una solución es correcta);
- cómo contestar (por ejemplo, acompañando su respuesta de un grado de certeza);
- las *consecuencias* en caso de Respuesta Correcta (RC) o Incorrecta (RI), aplicando una *escala de tarifas* anunciada (τ_C = Tarifa en caso de respuesta correcta; τ_O = en caso de omisión; τ_I = en caso de Respuesta Incorrecta).

Con las performances complejas, se anuncian los criterios que serán aplicados.

H.2. Evaluación con Criterios ESCONDIDOS

La consigna es "abierto" de modo que el estudiante tiene una gran libertad de expresión: "haga lo mejor que pueda". Es el caso del test "del hombre", en el cual se pide a niños de varias edades dibujar un hombre "lo más completo posible". Se aplica después una pauta (de Goodenough) que permite situar el nivel de desarrollo de la representación mental del esquema corporal (presencia del tronco, del cuello, de la palma de la mano, del talón, etc.).

H.3. Situaciones intermedias

El ensayo es una situación intermedia, porque se anuncian algunos criterios, pero de una manera imprecisa, como "Originalidad de las ideas" o "Grados de detalle", o "Coherencia", etc.

¿Y EN SU CURSO?

¿Es pertinente y posible anunciar los criterios o no anunciarlos?

I. Conclusiones

Existen muchas ventajas al organizar Dispositivos de Evaluación de los Aprendizajes (DEAs) con componentes múltiples, combinando instrumentos variados, en comparación con DEAs compuestos por un solo tipo de instrumento de evaluación. El conjunto de los capítulos intenta ilustrar estas ventajas y dar soporte a esta afirmación.

El presente capítulo intentó entregar una vista amplia de los aspectos en que los variados componentes de un DEA pueden diferenciarse. Se ilustró lo numerosas que son las opciones y la necesidad de tomar muchas decisiones, adecuadas a la complejidad de la situación educativa de cada docente. Todas estas decisiones deben ser tenidas en cuenta al momento de definir el programa.

Variaciones en el "prisma" (las condiciones) pueden influenciar el "reflejo" de las competencias o capacidades. Por esta razón, para interpretar un resultado (una observación, una performance) es importante conocer en cuáles condiciones (con cuál prisma) fue obtenido.

Referencias

- ACKERMANS, A., LECLERCQ, D. y ERNOULD, Ch. (1985). Un didacticiel d'évaluation des connaissances sur le diabète chez l'enfant et l'adolescent: Master-Diab. In D. Leclercq, M-P. Deghaye y M. Marlair: Handicaps et Technologies. Innovations et prototypes au service des personnes handicapées. Collection Santé-Education du CERES-Université de Liège, pp. 165-178.
- BLOOM, B.S., HASTINGS, J.T. y MADDAUS, G.F. (1971). Handbook on formative and summative evaluation of Student Learning, New York: Mc Graw Hill.
- BONBOIR, A. (1960). L'arithmétique au niveau de la 6^e primaire. Bruxelles: C.C.U.P.
- BOUD, D. y TYREE, A. (1979). Self and Peer Assessment in Professional Education: A preliminary Study in Law, Journal of the Society of Public Teachers of Law, 15, 1, 65-74.
- BOUD, D. (1995). Enhancing Learning through Self Assessment. London: Kogan Page.
- BOXUS, E. (1971). Une méthode de prédiction du rendement en lecture en première année. Education -Tribune Libre. 127, 63-78.
- BURTON, R. (1999). Approche psychologique de la résolution des équations du premier degré à une inconnue et de leur rôle dans la résolution de problèmes en physique, mémoire de licence en Sciences de l'Education, Université de Liège, 1999.

- D'AUGELLI, A.R. (1973). The Assessment of Interpersonal Skills: a comparison of Observer, Peer and Self Ratings, Journal of Community Psychology, 1, 177-9.
- DE FINETTI, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item, British Journal of Mathematical and Statistical Psychology, 18, 87-123.
- DE LANDSHEERE, G. (1979). Dictionnaire de l'évaluation et de la recherche en éducation. Paris: PUF.
- DEHOUSSE, M. (1981). Elaboration d'un cours multimédia sur l'Engström ER 312, Mémoire de Licence en Sciences Sanitaires, Université de Liège.
- DELCOMMINETTE, S., CHAPPELLIER, B. y LECLERCQ, D. (2012). L'évaluation longitudinale par étapes de la compétence (ELEC) à mener une recherche et à en produire el rapport. Actes du 24^e Colloque ADMEE-Europe. Luxembourg
- FALCHIKOV, N. (1995). Peer Feedback Marking: Developing Peer Assessment. Innovations in Education and Training International (IETI), 32, 2, 175-187.
- FISCHER, J-P. (1996). La mesure du temps de réponse aux calculs numériques élémentaires en situation scolaire. In J. Grégoire (Ed.). Evaluer les apprentissages. Le apports de la psychologie cognitive. Bruxelles: De Boeck, p. 81-97.
- FOOTE, R. y BELLINKS, C. (1972). It pays to switch? Consequences of changing answers on multiple-choice examinations. Psychol. Reps, 31, p. 667-673.
- FOUCAMBERT, J. (1980). La manière d'être lecteur. Paris: OCLD: Sermap.
- GLASER, R. (1963). Instructional technology and the measurement of learning outcomes. Some questions, American Psychologist, 18, pp. 519-521.
- GLASER, R. (1977). Adaptive Education: Individual Diversity and Learning, New York: Holt, Rinehart and Winston.
- INZAN, A. (1963). Le temps d'apprendre à lire. Paris: Bourrelier.
- JACOBS, S. (1971). Correlates of unwarranted confidence in responses to objective test items. Journal of Educational Measurement, 8, 1, pp. 15-20.
- JANS, V. (2000). Confrontations instrumentées et dialectiques des jugements auto/alloévaluatifs. Ph. D. Dissertation in Education, Liège University.
- KAGAN, J. (1965b). Reflexion-impulsivity and reading ability in primary grade children. Child Development, 36, pp. 609-628.
- KORMAN, M. y STUBBLEFIELD, R. (1971). Medical school evaluation and internship performance. Academic Medicine. 46, 8.
- LEBLANC, M.C. (1990). A la recherche de la synecdoque optimale dans l'humour dessiné. Mémoire de licence en Communication. Université de Liège.
- LECLERCQ, D. (1969). PACELBRO, Une technique originale d'enseignement programmé, in Education-Tribune libre, Bruxelles, n°118.
- LECLERCQ, D. (1969). Application d'un programme PACELBRO, in Education, n° 120.
- LECLERCQ, D. (1982). Confidence marking, its use in testing. in Choppin y Postlethwaite (Eds). Evaluation in Education, vol. 6, 161-287, Oxford: Pergamon Press.
- LECLERCQ D. y BRUNO J. (1993). Item Banking: Interactive Testing and Self-Assessment, NATO ASI Series, F 112, Berlin: Springer Verlag.
- LECLERCQ, D., DETROZ, P., DUPONT, Ch. y GILLES, J.L. (2003). Changer de méthode d'étude suite aux feedbacks: l'opération RESSAC (Résultats d'Épreuves Standardisées au service de l'Apprentissage en Candidatures). Chap 8 de 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. pp. 155-172.
- LINN, B., AROSTEGUI, M. y ZEPPA, R. (1976). Peer and self assessment in undergraduate surgery. Journal of Surgical Research. 21, 6, 453-456.
- LYNCH, D. y SMITH, B. (1975). Item response changes: Effects on test scores. Meas. Eval. In Guidance, 7, p. 220-224.

- MARGIN, D. y HELMORE, P. (2001). Peer and Teacher Assessments of Oral Presentation Skills: How reliable are they? *Studies in Higher Education*, 26, 3, 287-298.
- MARCHAND, C. (2000). Les connaissances antérieures des patients sur leur maladie dans le contexte de leur éducation : analyse de leur rôle, de leur nature, de leur configuration cognitive et de leur évolution, par l'utilisation de cartes conceptuelles. Thèse de doctorat. Paris: Sorbonne, U.F.R. des Sciences de l'Éducation.
- MATHUES, A. (1981). Expérimentation et essais dévaluation d'un cours multimédia sur l'Engström ER 312, Mémoire de Licence en Sciences Sanitaires, Université de Liège.
- MERCIADRI, L. (1982). Etude exploratoire de la lisibilité des images illustratives de textes. Mémoire de licence en Sciences de l'Éducation. Université de Liège.
- PIÉRON, H. (1963). Examens et docimologie. Paris: Presses Universitaires de France.
- PIPERT, R. (1966). Final note on the changed answer myth. *Clearing House*, 38, p. 165-166.
- REILING, E. y TAYLOR, R. (1972). A new approach to the problem of changing initial responses to multiple-choice questions. *Journ. Educ. Measurment*, 9, p. 67-70.
- SHUFORD, E., ALBERT, A. y MASSENGIL, N.E. (1966). Admissible probability measurement procedures. *Psychometrika*, 31(2), 125-145.
- VERMERSCH, P. (1994). L'Entretien d'explicitation en formation continue et initiale. Paris, ESF.
- WALD, A. (1943). *Sequential Analysis of Statistical Data*. New York: Columbia University Press.
- WOOD, R. (1977). Multiple-choice: A state of the art report. In Choppin y Postlethwaite (Eds). *Evaluation in Education*. International Progress. Oxford: Pergamon.

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 solicitud. 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>