



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.

P A R T E

5

Logística y
estadística en
evaluación

CAPÍTULO XXII

Los roles de un SMART: Servicio Metodológico de Apoyo a los docentes en la Realización de Tests

DIEUDONNÉ LECLERCQ Y PASCAL DETROZ

A. El SMART de la Universidad de Liège: motivaciones, funciones y estructura

A.1. Motivaciones y nacimiento

Todas las universidades tienen recursos limitados cuando se trata de destinar tiempo de académicos y académicas a la enseñanza. Los profesores de una universidad tienen varias misiones, y la función docente es solo una de ellas. La función de investigación les demandará dirigir equipos, promover y supervisar tesis, publicar artículos y libros, participar en congresos, y muchos profesores además participarán en acciones de extensión y vinculación con la comunidad, y/o gestión de la institución.

En consecuencia, es importante que la universidad se haga cargo de las tareas que puedan realizarse para preservar su tiempo. Si pensamos en el tiempo que es necesario dedicar exclusivamente a la evaluación de los aprendizajes de los estudiantes, existen tareas asociadas que en algunos casos superan las competencias habituales de los profesores (como concebir y realizar programas informáticos para la rendición de tests y/o el cálculo de puntajes e índices estadísticos de la prueba), o que no requieren de la atención directa del profesor pues podrían ser ejecutadas por otras categorías de personal (como imprimir hojas, manejar un lector óptico de marcas, vaciar datos).

Estas tareas pueden ser asumidas por una universidad a través de distintas fórmulas. Una de ellas es asignar a cada docente un asistente, pero es una fórmula costosa y que no necesariamente logra hacerse cargo de varias de las tareas asociadas a la evaluación, enumeradas más arriba, lo que mantiene la sobrexigencia de tiempo del docente.

Es por estas razones que muchas universidades se han dotado de lo que en la Universidad de Liège se denomina SMART: un Servicio Metodológico de Apoyo a la Realización de Tests (estandarizados). Es decir, un equipo con competencias variadas encargado de apoyar a los docentes en varias de las tareas asociadas a la evaluación de los aprendizajes de los estudiantes.

El ejemplo que utilizaremos en este capítulo es el SMART de la Universidad de Liège (ULg), institución integrada por unos 17.000 estudiantes en 10 Facultades, 350 profesores, 500 asistentes y 400 miembros del personal técnico y administrativo. El SMART-ULg nació en 1977, cuando Dieudonné Leclercq reunió a los docentes de varias

facultades que habían desarrollado sus propios programas informáticos para corregir las respuestas de los estudiantes a pruebas constituidas por PSM (Preguntas de Selección Múltiple). En aquel tiempo cada profesor (o su asistente o su secretaria) debía teclear cada respuesta para introducirla en el programa informático, lo que representaba una pesada carga cuando, por ejemplo, un profesor tenía más de 100 estudiantes (hay algunos con hasta 700 estudiantes). Si bien los programas desarrollados por cada profesor tenían componentes parecidos, solo uno de estos profesores había desarrollado análisis de ítems (incluyendo el cálculo de los índices de discriminación²⁰⁰ de cada solución de cada respuesta), una operación que debe realizarse antes de asignar calificación a los estudiantes y que exige un computador. Además, cada profesor tenía hábitos diferentes a los de sus colegas: otras formas de ponderar las preguntas y los errores; algunos restaban puntos en caso de error, otros no; algunos utilizaban grados de certeza, otros no; algunos utilizaban Preguntas Verdadero-Falso (PVF), otros no; algunos mezclaban preguntas de respuesta breve (PRB)²⁰¹ con las PSMs, otros no. Sin embargo, todos estimaron que existían muchas ventajas en utilizar un sistema común respetando los deseos de cada uno.

Durante un año Leclercq efectuó la lectura óptica de marcas llevando las pruebas de sus colegas a un lugar a 100 kilómetros de Liège. Luego redactó un proyecto de desarrollo de un servicio común y gratuito para los profesores (incluyendo un presupuesto pagado por la Universidad para el personal y los recursos técnicos permanentes y los insumos), que fue apoyado por las firmas de los profesores implicados. El rector de la época (A. Bodson) y el Consejo de Administración de la ULg lo aprobaron y otorgaron el presupuesto, con lo que nació el SMART-ULg. Durante años el SMART fue dirigido por J-L Gilles y actualmente lo dirige P. Detroz.

A.2. Las funciones del SMART-ULg

La evaluación de los estudiantes (EVAL ESTU) es la primera función del SMART-ULg. Consiste en apoyar a los docentes en la concepción y la realización de tests estandarizados, y será descrita en detalle en las secciones B y C. Sin embargo, rápidamente luego de su creación, se hizo evidente que semejantes recursos humanos y técnicos podrían ser utilizados para otras funciones. A continuación se presenta la lista actual de funciones del SMART, con las siglas con que estos servicios son conocidos en la ULg.

- (1) EVAL ESTU se presenta como un ciclo de calidad en 8 fases (ver sección B a continuación).

En 2011 apoyó cerca de 700 pruebas (comparadas con 350 en 2006), para 150 docentes diferentes, con “peaks” en enero, junio y septiembre, que son los periodos oficiales de exámenes en la ULg. En una encuesta contestada en forma

²⁰⁰ Las correlaciones punto biserial (rpbis). Ver Capítulo 23.

²⁰¹ De modo que eran ellos, los profesores, quienes debían codificar si la respuesta era correcta o no.

anónima por 76 docentes, los criterios acerca de la acogida, disponibilidad, eficacia y competencia del equipo fueron juzgados con valores entre 92,7 y 93,3 sobre una escala de 100 (donde 100 es igual a “perfecto”). Preguntas específicas de estas encuestas han provisto valiosas informaciones “formativas” utilizadas para mejorar los servicios del SMART.

- (2) EVAL ENS (Detroz y Mainferme, 2007) es el apoyo a la concepción y a la recolección de opiniones de los estudiantes sobre los cursos, y es la segunda función más recurrente e importante del SMART-ULg, en términos de volumen de hojas que imprimir y leer ópticamente. Se llama también EEE-Evaluación de la Enseñanza por los Estudiantes. Como no trata de la evaluación de los aprendizajes de los estudiantes no se abordará este tema en este capítulo. Alrededor de 1.500 cursos son evaluados cada año con el apoyo del SMART, con “peaks” en mayo para los cursos finalizados, y en octubre cuando los estudiantes son interrogados sobre los cursos del año anterior.
- (3) EVAL ENC (Detroz y Mainferme, 2007) trata de apoyo a encuestas administrativas o encuestas científicas, y no se verá en este capítulo. Dos ejemplos de EVAL ENC son la recolección (a menudo on line) de datos por parte de un departamento o una facultad
 - a) para preparar informes sobre calidad, periódicamente exigidos por organismos exteriores.
 - b) para preguntar a los estudiantes cuáles son sus elecciones cuando existen opciones en un currículum o cuando se trata de constituir grupos.
- (4) CVE (o utilización de Cajas de Voto Electrónicas o *Remote Control Devices* o “clickers”) es un servicio (Leclercq *et al.*, 1999) que ha sido ofrecido a partir de 1990 por el SMART-ULg a los profesores (ver sección D). El SMART asegura alrededor de 70 utilidades cada año, con grandes variaciones entre docentes (hay algunos que los utilizan semanalmente durante uno o dos meses).
- (5) TAO (Testeo Administrado por Ordenador) u *on line testing* o CAT (*Computer Assisted Testing*). Es ofrecido por el SMART a partir de 2007 (Gilles *et al.*, 2010) con un programa llamado EXAMS, que sirve para definir pruebas a ser rendidas por computador o en papel (formuloms), siguiendo los principios del ciclo de gestión de la calidad de una prueba estandarizada que se describe a continuación en sección B.
- (6) PCCD (Pantalla de Corrección Compartida a Distancia) es la más reciente oferta del SMART, desarrollada con el apoyo de la Facultad de Medicina de la ULg. Se trata de un sistema para facilitar la corrección de pruebas de desarrollo (de muchas hojas, con dibujos, etc.), de forma que varios jueces a distancia puedan corregir y tomar apuntes sobre la misma copia. Será descrito en la sección E a continuación.

La Tabla 1 resume los servicios actualmente ofrecidos gratuitamente por el SMART a los profesores de la ULg.

Tabla 1: Tipos de apoyo (a la docencia) del SMART ULg

Funciones	APOYO PEDAGÓGICO		APOYO TÉCNICO			
	Apoyo a... la concepción e interpretación de los resultados	LOM impresión y lectura óptica	Procesamiento de datos y comunicación	Web interactiva TAO y PCCD	Manejo de Clickers en el aula (CVE)	
(1) EVAL ESTU	X	X	X	X		
(2) EVAL ENS	X	X	X	X		
(3) EVAL ENCUESTA	X	X	X	x		
(4) CVE Cajas de Voto Electrónico	X		X		X	
(5) TAO : Tests Administrados por Ordenador	X			X		
(6) PCCD Pantalla de Corrección Compartida a Distancia [EXAM OROU]	X			X		

A.3. Los recursos humanos del SMART

- La Dirección.* El SMART-ULg es dirigido por una dupla: “Responsable Académico – Director Operacional”, con la supervisión de un consejo de pilotaje (del IFRES)²⁰². La dirección actual²⁰³ está compuesta por Ch. Monseur (RA - Profesor de Métodos cuantitativos en la Facultad de Ciencias de la Educación) y P. Detroz (DO - Doctor en Ciencias de la Educación y Profesor en el IFRES).
- Algunos investigadores* (actualmente 2) trabajan de forma permanente u ocasional en el SMART-ULg. Con frecuencia son Masters en Ciencias de la Educación o estudiantes cursando ese programa.
- Algunos informáticos* (actualmente 2) crean los programas, mantienen y mejoran las aplicaciones, y aseguran los procesos informáticos.
- Una secretaria* (a medio tiempo) mantiene los contactos (citas, peticiones de profesores, ciclo de calidad).
- Una operadora de correcciones* asegura los procesos informáticos de los datos.
- Un impresor* reproduce los *formuloms* y, como *operador*, maneja el lector óptico de marcas (LOM). *Formuloms* es una palabra inventada para referir a los “formularios destinados a la LOM (lectura óptica de marcas)”, que se describen a continuación en sección A4.

A.4. El lector óptico de marcas (LOM)

En una fotografía disponible en los complementos multimedia de este libro puede verse un lector óptico de marcas y su operador. Este coloca los *formuloms* en la zona de arriba y el lector electrónico los lee con una velocidad que alcanza los

²⁰² Instituto de Formación e investigación (Recherche) en la Enseñanza Superior de la ULg.

²⁰³ Después de las duplas D. Leclercq - J.L. Gilles, y J.L. Gilles - P. Detroz.

120 *formuloms* por minuto; o los rechaza si detecta errores que han sido pre-definidos con un programa especial. Por ejemplo, que existan varias marcas en una línea que ha sido definida como “uni-marcable”. Los *formuloms* pueden ser imprimidos (y leídos) usando ambos lados de la hoja (tiro / retiro). En ese caso los estudiantes pueden utilizar solo bolígrafos y lápices, pero no tintas que atraerían el papel.

A.5. La estructura de un formulom del SMART-ULg

La Figura 1 permite visualizar las distintas partes de un *formulom* clásico:

- La zona de más arriba presenta solo una línea con ocho celdas, que sirven para que el estudiante reproduzca ahí la marca que identifica la versión de la prueba que ha recibido. En efecto, para que los estudiantes no copien las respuestas de sus vecinos, el SMART produce versiones paralelas de la misma prueba, llamadas “forma A”, “forma B”, etc. Las preguntas son las mismas en todas las versiones, en el mismo orden, pero el número 1 es distinto en cada formato. La corrección de este desfase es realizada (automáticamente) por el SMART en el momento de la corrección.
- La zona de identificación del estudiante en escritura manuscrita.
- La zona de identificación del estudiante por el número de matrícula (códigos numéricos).
- Instrucciones sobre el modo de llenar las celdas para expresar sus respuestas.
- La zona de respuestas. En este *formulom*, en cada pregunta hay dos zonas: la primera está concebida para recibir la respuesta (R) a la pregunta, marcando una –y solo una– de las hasta nueve alternativas por línea. La zona de abajo está concebida para recibir el grado de certeza (C) de la respuesta, marcando una de seis opciones: 0, 20, 40, 60, 80 o 100% seguro de la respuesta entregada.
- Cada zona de una pregunta (para la respuesta y para el grado de certeza) tiene una segunda línea llamada “línea de arrepentimiento”. Si el estudiante quiere cambiar la respuesta que ha marcado en la línea 1 no tiene que borrarla o tacharla. Basta que marque la nueva respuesta en la segunda línea, y el programa informático sabe que, en ese caso, solo vale la respuesta de la segunda línea.

A.6. Algunos ejemplos de formuloms

Los *formuloms* que siguen (figuras 2, 3 y 4) ilustran su gran variedad de posibilidades. Esto deriva del hecho que el SMART puede concebir *formuloms* a petición de los docentes, y de la preocupación por concentrar muchas marcas en una sola hoja para economizar papel y tiempo de lectura óptica.

B. Una visión global del Ciclo de Gestión de la Calidad de una Prueba Estandarizada (CGCPE)

La Tabla 2 muestra una visión global del ciclo CGCPE: *Cycle de gestion de la qualité d'un Test Standardisé* (Gilles et al., 2005), cuyas etapas serán descritas en la siguiente sección. El SMART apoya en cada una de las ocho etapas a los docentes que lo solicitan.

Tabla 2: Implicación de cada grupo en las ocho etapas del CGCPE (Gilles et al., 2005). (a) = apoyo

	Docente	SMART	Estudiantes
B1. Análisis			
a) Puntos de Contenido a evaluar (PE)	X	(a)	
b) Categorías de Procesos mentales (CP)	X	(a)	
c) Objetivos	X	(a)	
B2. Diseño			
a) Elección de los tipos de preguntas (Items)	X	(a)	
b) Elección de la modalidad oral / escrita	X	(a)	
B3. Preguntas (Items)			
a) Construcción y criterios de corrección	X	(a)	
b) Instrucciones a los jueces	X	(a)	
c) Manejo de los bancos de preguntas	X	(a)	
B4. Entrenamiento de los estudiantes			
a) Versión papel	X	X	X
b) Versión informatizada	X	X	X
c) Clickers	X	X	X
B5. Rendición de la prueba			
a) Versión papel	X		X
b) Versión informatizada			X
c) Clickers		X	X
B6. Corrección			
a) Codificación manual	X		
b) Lectura óptica de marcas		X	
c) Versión 1 de los resultados		X	
d) Interacción Docente SMART	X	(a)	
e) Versión definitiva de los resultados		X	
B7. Retroalimentaciones			
a) Retroalimentaciones vinculadas a los PE y CP personalizados y confidenciales		X	
b) Índices metacognitivos		X	
c) Performance global		X	
B8. Regulación			
a) Recolección de las opiniones de los participantes		X	
b) Diagnósticos sobre el Proceso			
c) Discusión y mejora	X	X	

Las etapas 1, 2, 3 y 8 sirven a varios instrumentos de evaluación del docente, a partir de su banco de preguntas. Las etapas 4, 5, 6 y 7 están vinculadas a un momento de evaluación específico.

C. Las ocho etapas del ciclo CGCPE del SMART-ULG aplicado a EVALESTU

C.1. Análisis

- El docente decide los *Puntos de contenido a Evaluar* (PE), asegurando la validez de cobertura de su dispositivo de evaluación (ver Capítulo 4).
- El docente decide cuál(es) *Categorías de Procesos mentales* (CP), aplicadas a estos puntos de contenido deberán ser evaluadas por las preguntas. Se utilizan las categorías de la taxonomía de Bloom (ver Capítulo 15).

C.2. Diseño (con Tabla de Especificación - TdE)

- El docente elige la modalidad de las preguntas: oral (O) o escrita (E).
- El docente elige los tipos de preguntas (items) escritas: por ejemplo, PSM, PRB, PSM + SGI.

Estas dos etapas resultan en una "Tabla de especificaciones (TdE)", como la que ilustra la Tabla 3.

Tabla 3: "Tabla de especificación" (TdE) contenidos / procesos mentales. Las celdas indican las modalidades (oral/escrito) de las preguntas y cuántas de cada tipo debe contener la prueba

	Reconoci- miento	Evocación	Comprensión	Aplicación	Análisis	Juicio-Evaluación
Contenido 1		2E	2E		1E	
Contenido 2	1E	1E	2E + 1 Oral		1E	1 Oral
Contenido 3		1E		2E		
Contenido 4	1E			1E	1E	
Contenido 5		2E		1E + 1 Oral		1 Oral
Contenido 6			1E		1 Oral	1 Oral
Etc.						
Total = 20 + 3	2E	6E	5E+1 Oral	4E + 1 Oral	3E + 1 Oral	3 Orales
Tipo de pregun- tas	PSM	PRB	PSM + SGI	PSM + SGI	PSM + SGI	
¿Certeza?	sí	sí	sí	sí	sí	

C.3. Preguntas (items)

A) CONSTRUCCIÓN POR EL DOCENTE

- de las preguntas (Capítulos 13, 14, 15) y los criterios de corrección (Capítulo 17), según reglas de calidad de la redacción (Capítulo 14).
- de los comentarios que acompañan las respuestas incorrectas a modo de retroalimentación (introduciéndolas en el programa EXAMS cuando la prueba es aplicada en computador). Ver sección C.4 a continuación y Capítulo 21.
- de las instrucciones a los jueces, en caso de usar preguntas con respuesta breve (PRB).
- de las instrucciones a las personas que vigilarán el aula durante la rendición del test.

El docente maneja su banco de preguntas. Puede hacerlo con un programa especial del SMART llamado EXAMS, pero no es necesario tener un software especial para hacerlo. También puede tener el banco de preguntas de cada curso en un fichero Word, como el ejemplo a continuación.

B) GESTIÓN DE UN BANCO DE PREGUNTAS

La Tabla 4 a continuación presenta un extracto del banco de preguntas del curso ISE de Dieudonné Leclercq. En la segunda línea, en letras gruesas, está la *introducción* común a las cinco preguntas que siguen. Estas preguntas evalúan el mismo contenido, pero no el mismo proceso mental aplicado a ese contenido.

La última columna indica la fecha en que la pregunta fue utilizada en una prueba. Este banco de preguntas está escrito en Word, de modo que para introducir una de estas preguntas en la versión Word de una prueba basta que el docente la copie, incluyendo la columna central con la respuesta correcta. Eso constituirá su *prueba con las respuestas correctas*, las que desaparecerán en la versión (Word) de los estudiantes. Las dos últimas preguntas son nuevas (nunca han sido utilizadas).

C) PREGUNTAS ISÓTOPOS

La última pregunta muestra lo fácil que es cambiar una pequeña cosa (en este ejemplo, la alternativa n° 5 de una pregunta anterior), para transformar una pregunta con solución correcta "6. Ninguna" en una pregunta con solución correcta "5". Estas preguntas, muy similares pero con pequeñas variaciones, son llamadas *isótopos* por D. Leclercq.

Tabla 4: Extracto del Banco de preguntas del curso ISE de D. Leclercq: 5 formas de preguntar un contenido

Pregunta	Contenido: <i>Impacto del método ABP</i> Capítulo 8 Sección K3	RC	Fechas de utilización
Los metaanálisis enseñan que los médicos formados con el método ABP, comparados con los formados con métodos tradicionales, <i>en promedio</i> ,		RC	
...obtienen resultados: 1. Inferiores 2. Iguales 3. superiores		8 Faltan datos	12-01-1999 16-01-2002
...se caracterizan por performances inferiores en las pruebas estandarizadas performances superiores en los exámenes clínicos una duración inferior en los estudios de Medicina General una tasa inferior de deserción (renuncia) atender a menos pacientes por año		7 Todas	29-08-2000 17-01-2001 19-01-2006 18-06-2010
...se caracterizan por performances superiores en las pruebas estandarizadas performances inferiores en los exámenes clínicos una duración superior en los estudios de Medicina General una tasa superior de deserción (renuncia) atender a más pacientes por año		6 Ninguna	12-01-2000 27-08-2001 21-05-2002 08-06-2006 08-01-2010
...obtienen una tasa de mejoría más alta de los pacientes porque les dedican más tiempo les prescriben mejores curas la adherencia de los pacientes es más alta su comunicación con los pacientes es mejor		9 Absurdo	
...se caracterizan por performances superiores en las pruebas estandarizadas performances inferiores en los exámenes clínicos una duración superior en los estudios de Medicina General una tasa superior de deserción (renuncia) atender menos pacientes por año		5	

C.4. Entrenamiento de los estudiantes

Para minimizar el riesgo de que parte del fracaso de los estudiantes se deba a falta de familiaridad con el formato de la prueba y la forma de contestarla, existen al menos tres formas de permitir a los estudiantes ejercitar para estar preparados frente a las modalidades de un examen:

1. En clase, con el profesor y con *formuloms*
2. En clase, con cajas electrónicas de voto (ver sección D)
3. Vía internet utilizando el programa EXAMS. Esta ejercitación se puede realizar desde el domicilio o desde salas apropiadas como CAFEIM (Centro de Auto-Formación y Evaluación Informatizada Multimedia), de la Facultad de Psicología y Ciencias de la Educación de la Universidad de Liège.

El programa EXAMS del SMART ofrece a los estudiantes la posibilidad de ejercitar con pruebas anteriores y rendir un examen simulado desde su domicilio (internet) o desde un centro como el CAFEIM. Otros programas (como el del Campus Virtual de la ULg) también permiten ejercitar.

En los complementos multimedia de este libro se pueden ver fotografías del CAFEIM (Centro de Autoformación y Evaluación Informatizada Multimedia), de la Facultad de Psicología y Ciencias de la Educación de la Universidad de Liège. Se puede ver puestos individuales de trabajo, que en los momentos de examen se separan unos de otros mediante paredes móviles retráctiles.

C.5. Rendición de la prueba

- Versión papel:* El SMART envía por correo interno los *formuloms* solicitados por el docente, quien ha indicado en un formulario apropiado las respuestas correctas y otras especificaciones.
- Versión informatizada:* El programa EXAMS del SMART, y otros como el Campus Virtual de la ULg, permiten a los estudiantes rendir pruebas, generalmente desde centros como el CAFEIM vía la intranet de la Universidad.
- Con cajas de voto o clickers* (ver sección D a continuación).

C.6. Corrección

- Codificación manual:* Cuando se deben teclear respuestas a preguntas con respuestas breves (PRB) el SMART puede dar apoyo a través de una secretaria.
- Lectura Óptica de Marcas:* Para que este proceso pueda funcionar el docente debe previamente elegir y comunicar precisiones sobre *parámetros* presentados en listas:
 - A qué categorías de PE y CP (ver sección C1) pertenece cada pregunta
 - El peso (o importancia) de cada pregunta
 - El baremo de cotejo
 - La severidad (el umbral de éxito)
 - Los "bonus" a entregar por la calidad de la metacognición (ver Capítulo 17)

La *copia electrónica* de la lectura de las respuestas es enviada al docente. La Tabla 5 presenta un ejemplo de este tipo de resultado, relativo a la lectura de las respuestas de un grupo de estudiantes que han contestado a una prueba constituida de PSN (Preguntas con Soluciones Numerosas), cuyas respuestas pueden ir de 000 hasta 999.

Tabla 5: "Lectura electrónica" (copia) de las respuestas de estudiantes sobre un formulario para PSN

matricula	EXAMEN CON LIBRO CERRADO. 14 DE ENERO DE 2008: COPIA DE LAS RESPUESTAS																						
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23
999996	147	67	94	129	136	190	193	196	135	205	63	134	50	221	64	120	165	163	140	164	187	77	45
999997	147	67	94	129	136	190	193	196	135	205	63	134	50	221	64	120	165	163	140	164	187	77	45
74272	147	94	43	35	173	190	193	196	135	205	63	134	50	221	64	182	123	23	52	33	123	22	567
71266	147	67	94	126	136	190	193	196	135	205	63	134	50	221	64	54	165	163	140	164	112	77	45
73036	102	67	0	0	136	0	193	196	191	205	63	134	50	0	64	120	101	138	52	241	181	77	163
70160	27	120	67	133	131	190	193	196	115	232	63	0	50	218	73	116	88	211	44	0	33	77	225
72851	30	67	60	129	136	190	193	196	135	205	63	134	50	221	64	215	165	0	140	34	113	77	45
51008	147	116	60	129	136	190	193	196	135	205	63	134	50	221	64	251	165	163	140	164	187	77	45
62248	147	67	94	129	172	0	234	0	162	205	101	134	50	0	64	0	104	52	206	75	71	0	0
71958	147	60	94	129	0	190	193	196	135	195	63	134	50	221	64	120	101	163	140	241	0	77	45
74177	147	67	60	129	136	190	193	196	135	205	63	134	50	221	64	120	165	0	140	164	75	77	86
73999	147	67	94	129	172	190	193	196	39	205	63	134	50	221	64	120	174	163	140	164	113	77	86
72785	71	67	37	32	136	190	197	197	135	189	63	134	160	40	64	120	101	44	140	241	114	133	45
63720	147	67	94	129	136	190	193	196	135	205	63	134	50	221	64	182	165	163	140	241	114	77	86
70704	147	67	94	42	136	193	196	190	135	205	63	134	90	221	64	120	101	245	140	164	187	77	163
60231	147	67	45	129	136	190	193	196	135	205	63	134	50	221	64	60	165	152	140	164	114	77	145
72578	147	67	43	129	136	190	193	196	135	205	63	134	50	221	64	79	174	163	140	241	32	77	45
72572	147	67	60	110	136	190	193	196	233	250	63	134	50	220	64	182	174	152	140	164	114	33	45
62792	147	67	86	183	136	64	193	243	162	205	101	134	160	221	73	119	101	152	255	164	161	250	148
70566	147	67	94	129	136	190	193	196	0	205	63	134	50	221	64	120	174	163	140	164	187	77	45
71995	147	67	94	129	136	190	193	196	135	125	63	134	50	221	64	251	174	37	140	130	187	77	45
73698	233	67	94	129	131	190	193	196	39	205	63	134	50	38	60	60	165	163	140	160	113	77	45

El docente puede ver cuáles son las soluciones más populares en cada pregunta. Por ejemplo, en la Pregunta 7, con excepción de 3 respuestas, todos los estudiantes contestaron con la misma solución (193). En otros casos, como en las preguntas 18 y 23, hay una gran variedad de respuestas.

- c) *Versión 1 (provisoria) de los resultados* de cada uno de los ítems (cada pregunta): El SMART envía al docente los análisis de correlación *point biserial* –*rpbis*–, donde puede ser que aparezcan anomalías (*rpbis* positiva en lugar de negativa), como distractores que atraen a los mejores estudiantes, por ejemplo.
- d) *Interacción Docente-SMART*: en caso de anomalías, el docente puede pedir al SMART cambiar los parámetros de corrección, usando para ello otro formulario apropiado. Por ejemplo, el docente podría pedir:
- suprimir una (o más) pregunta(s)
 - cambiar la respuesta correcta
 - aceptar uno (o más) distractor(es) como respuesta(s) correcta(s) además de la oficial
 - cambiar el peso de una pregunta o de grupos de preguntas (a no ser que los pesos hayan sido comunicados a los estudiantes)
 - cambiar la severidad (umbral de éxito) o añadir puntos a todos los estudiantes (se prohíbe disminuir puntos)
- e) *La versión 2 (definitiva) de los resultados*: Normalmente esta versión presenta distribuciones de índices de discriminación de las soluciones conforme a las expectativas, de modo que estos datos y parámetros (modificados) pueden servir de base para calcular las calificaciones (puntajes) definitivas de cada estudiante, porque la fiabilidad de los resultados de la versión 2 (definitiva) es mejor que la de la versión 1 (provisoria). Crahay *et al.* (2006, p. 20) estudiaron 34 pruebas en las que se efectuaron modificaciones entre la versión 1 y la versión 2 de los resultados, totalizando 71 modificaciones (es decir, más o menos un promedio de 2 rectificaciones en cada una de estas 34 pruebas). La Tabla 6 indica la forma en que se distribuyeron estas 71 modificaciones:

Tabla 6: Distribución de los tipos de modificaciones (después de la versión 1 de los resultados) en 34 pruebas

	Número
– Valorización de un distractor	39
– Supresión de una pregunta	22
– Cambio de respuesta correcta	10
Total	71

C.7. Retroalimentaciones a los estudiantes por el SMART

Para tener una visión global de los tipos de retroalimentaciones posibles, invitamos al lector a ver el Capítulo 21 (Retroalimentaciones al estudiante), particularmente la Tabla 2 que constituye una taxonomía de las retroalimentaciones.

A) RETROALIMENTACIONES PERSONALIZADAS Y CONFIDENCIALES A LOS ESTUDIANTES

En primer lugar, cada estudiante recibe en su informe informatizado la copia de sus respuestas (ver Capítulo 21, Figura 5, tablas 7 y 8) para que pueda verificar que lo que ha sido leído por el LOM corresponde a lo que él/ella contestó. Si el profesor ha definido categorías PE (contenidos a evaluar) y CP (procesos mentales) y ha clasificado cada pregunta según ellas, las retroalimentaciones son entregadas organizadas por cada categoría. Los resultados se parecen a los que se pueden ver en las tablas 2 y 3 del Capítulo 18.

B) ÍNDICES METACOGNITIVOS

La retroalimentación se entrega en distribuciones espectrales que muestran a cada estudiante sus grados de certeza para las respuestas correctas e incorrectas, indicando su Confianza e Imprudencia. Ejemplos de estas distribuciones se pueden ver en el Capítulo 16, Tabla 5, figuras 4, 8 y 9, y Capítulo 9 (TEM), Tabla 5.

C) PERFORMANCE GLOBAL

Por supuesto, la retroalimentación al estudiante también incluye la entrega de índices sumativos (puntaje / calificación final).

C.8. Regulación

A) RECOLECCIÓN DE LAS OPINIONES DE LOS PARTICIPANTES

Automáticamente, junto con las retroalimentaciones, el SMART invita a los docentes y a los estudiantes a comentar la prueba, su contenido, sus condiciones de rendición y sus resultados. El SMART ha reunido las respuestas de estudiantes a varias pruebas de una facultad (la de Psicología y Ciencias de la Educación de la Universidad de Liège). Las opiniones promedio eran particularmente bajas en relación con cuatro preguntas:

- ¿Fue suficiente la preparación para afrontar el formato de la evaluación?
- ¿Fueron comunicadas las respuestas correctas en un tiempo breve después del examen?²⁰⁵

²⁰⁵ Lo que constituye un asunto muy importante, como se ve en la demostración de Karraker (Capítulo 13, D.1).

¿Fueron comunicadas a los estudiantes las estadísticas (índices de discriminación) de los ítems?

Después del examen ¿pudo el estudiante recibir explicaciones acerca de sus errores?

La Facultad insistió en que los docentes cambiaran esta situación, y el claustro de profesores lo hizo, de modo que dos años después las respuestas a estas mismas preguntas tuvieron avances notables en el nivel de satisfacción.

B) PROCESO

Ya hemos mencionado que las etapas 1, 2, 3 y 8 del CGCPE no están vinculadas a un examen en particular, sino que a un *ámbito de estudio*. En la etapa 8 se analizan todas las estadísticas de la prueba, para ver si es necesario cambiar algo en la Tabla de Especificación o TdE (ver sección C2).

C) DISCUSIÓN Y MEJORA

Este proceso de revisión (de mejora) para las pruebas futuras puede incluir la adición o la modificación de los aprendizajes a evaluar, o de las proporciones de cada uno de los procesos mentales incluidos en la TdE. Esto puede hacerse después de un intercambio con los especialistas del SMART.

C.9. Utilización de EXAMS-on-line con función formativa

El profesor Hanzen, de la Facultad de Veterinaria de la ULg, propuso a sus 270 estudiantes rendir 10 pruebas optativas con función formativa, una cada semana, y una prueba con función sancionadora (obligatoria), lo que significaba un máximo de 2.970 utilizaciones posibles. Cada prueba contenía 10 Preguntas de Selección Múltiple (PSMs) con 4 Soluciones Generales Implícitas (ver Capítulo 13). Las estadísticas de rendición de las 11 pruebas (Hanzen *et al.*, 2010) muestran 1.699 rendiciones (57%). En promedio, cada estudiante rindió 5,8 pruebas. 1/8 de los estudiantes no rindió ninguna de las pruebas con intención formativa, y la misma tasa las realizó todas. Las estudiantes fueron más participativas, con un promedio de 6,4 pruebas versus un promedio de 4,7 para los varones.

En la encuesta realizada por el docente, un 65% declaró que en más del 75% de las veces compararon sus respuestas con las correctas y leyeron las justificaciones entregadas por el docente como parte de la retroalimentación de las pruebas con intención formativa. Este comportamiento fue adoptado en mayor medida por las estudiantes que por los varones. El 74% de los estudiantes declaró que en más de dos tercios de las ocasiones intentaron explicarse la razón de sus errores.

Hay una correlación positiva entre el número de utilizaciones de las pruebas con intención formativa y la performance en la prueba con intención certificativa. Una

decisión de regulación que siguió a esta experiencia fue disminuir el número de pruebas con función formativa (menos de 10), pero aumentar el número de preguntas (más de 10) en cada una de ellas.

C.10. Conclusiones sobre EVAL ESTU

Las cuatro ventajas que más aprecian los profesores en este sistema son: (1) la posibilidad de asegurar la estandarización de sus exámenes (las mismas preguntas a todos los estudiantes en el mismo tiempo), especialmente cuando el grupo es grande (centenas de estudiantes); (2) el tiempo ganado por cada profesor que usa el sistema, en términos de ahorro de tiempo de corrección; (3) la posibilidad de ser diagnóstico para el caso de cada estudiante, y de entregarle esos datos con detalle en la retroalimentación; (4) la rapidez de la retroalimentación, para el profesor en primer lugar, y después para los estudiantes.

Junto a estas ventajas, añadimos la posibilidad de multiplicar las pruebas de entrenamiento con una función formativa (ver Capítulo 9) y de testear hipótesis relativas al impacto de innovaciones sobre el aprendizaje.

D. Cajas de voto electrónicas (CVE o clickers)

Un modo de animar los cursos y preparar en el aula a los estudiantes para rendir exámenes es la utilización de cajas de voto electrónicas. Esta técnica permite al docente (y a los estudiantes) darse cuenta inmediatamente de los errores, de sus frecuencias, y reaccionar sin demora (Leclercq *et al.*, 1999; Gilles *et al.*, 2000; Noel-Lambot y Detroz, 2003).

D.1. Descripción de las cajas de voto

Son múltiples las expresiones para denominar a esta herramienta: *clickers* (Schackow *et al.*, 2004; Wood, 2004; Duncan, 2005; Knight y Wood, 2005), *Audience Response Systems-ARS* (Banks, 2006; Kay y Lesage, 2009), *Audience response technology* (Mc George *et al.*, 2008; Smrek, 2011), *Classroom response systems* (Bruff, 2009), *Télévotants* (Leger *et al.*, 2010), *Boîtiers de vote* (Leclercq *et al.*, 1999; Gilles *et al.*, 2000).

En los complementos multimedia de este libro pueden verse imágenes de distintos tipos de cajas de voto electrónicas (CVEs), con modelos con diversas funciones y cantidad de teclas: aquellas que permiten seleccionar una alternativa y confirmar o cambiar respuesta, otras que además permiten introducir un grado de certeza (solo 2 niveles (H=High / L=Low), en Ikkal y Sibley, 2006). Hay CVEs que permiten identificar quién ha contestado, pero se requiere introducir previamente parámetros como el nombre o código del utilizador en cada CVE.

En el departamento de Física de la Universidad de Colorado, en Boulder, cada estudiante tiene su propio *clicker*, proporcionado por el departamento contra un pago que recuperará al finalizar el año. Varias aulas están equipadas con un sistema de recepción, y en estas salas el profesor puede improvisar la utilización de los *clickers*, simplemente pidiendo “por favor, tomen sus *clickers*...”.

D.2. Preparación del uso de las cajas de voto

En la ULg, un docente que planea utilizar las cajas de voto no solo debe reservarlas, sino que también debe reservar el tiempo del operador que manejará el sistema y estará presente en el momento de su uso.

Si se quiere que las preguntas aparezcan sobre la pantalla, tanto antes de que los estudiantes contesten como cuando aparezcan los resultados de la votación (tasas de votos para cada alternativa), el docente tiene que introducir estas preguntas y sus alternativas en el sistema, lo que se hace mediante el envío de un fichero al SMART. Sin embargo, este envío previo no es obligatorio. Las preguntas pueden incluso ser improvisadas durante una clase, tanto por el profesor como por los estudiantes. Esto último es lo que pasaría si varios estudiantes proponen hipótesis diferentes, y todos los integrantes del curso deben elegir cuál de estas hipótesis prefieren.

D.3. Uso de las cajas de voto

El profesor debe proponer una pregunta, con opciones de solución numeradas, escritas a mano sobre el pizarrón o proyectadas. Desde un computador el operador da la señal para el comienzo de la votación. En el sistema de la ULg aparece proyectado en la pantalla del aula el número de votos emitidos (confirmados y enviados por los estudiantes), en tiempo real. Por supuesto, este número cambia rápidamente y permite al docente decidir cuándo terminar la recepción de los votos (otra vez desde el computador). A menudo, para no perder demasiado tiempo, los docentes terminan la recepción de votos cuando el 90% de los estudiantes ha contestado. El profesor puede también elegir no proyectar los resultados y ser solo él/ella quien los conoce. El SMART entrega al docente las trazas (huellas), es decir, las estadísticas de las respuestas a cada pregunta.

D.4. Encuestas a estudiantes sobre cómo usan las cajas de voto

Leclercq *et al.* (1999) obtuvieron los siguientes resultados a dos preguntas contestadas por sus estudiantes, quienes habían utilizado de manera repetida las cajas de voto. El curso era masivo (300 estudiantes), de modo que debieron compartir una caja de voto cada tres estudiantes, lo que los forzaba a discutir la respuesta para intentar llegar a un acuerdo:

Tabla 7: Frecuencias con las que el estudiante declara concebir una respuesta personal a una del docente de acuerdo al formato de entrega de la respuesta (mano alzada vs cajas de voto)

¿En cuál porcentaje de las veces intenta Ud. concebir una respuesta personal?	Media
Cuando el profesor invita a levantar la mano para expresar una respuesta	40%
Cuando se utilizan cajas de voto	78%

Un 13% de los estudiantes contestó que, en este último caso (cajas de voto), “Dejo a los otros la tarea de concebir una respuesta”.

D.5. Medición del impacto del uso de las cajas de voto

Eric Mazur, profesor de física en la Universidad de Harvard, ha elaborado (1997) sobre el concepto de “Instrucción por los pares” o *Peer Instruction*. En este contexto, durante una clase, el docente realiza preguntas concebidas para evaluar la comprensión de conceptos (que Mazur llama *Concept tests*), las que proveen a los estudiantes la oportunidad de testear su comprensión, respondiendo vía *clickers*. El/la estudiante puede verse reafirmado/a (en caso de respuesta correcta) o descubrir un error. En el momento de intercambio entre pares, el/la estudiante debe intentar corregir su mala comprensión del contenido con el apoyo de sus compañeros, o ayudarles a otros comunicándoles su forma de entender la solución del problema, según haya respondido incorrecta o correctamente. El cambio posterior al intercambio entre pares es medido a través de una nueva respuesta vía *clickers*. Mazur ha observado que el uso de cajas de voto con sus estudiantes de un curso de física permitió duplicar la ganancia relativa promedio en evaluaciones sobre el concepto de fuerza.

Schackow *et al.* (2004) observaron, con estudiantes de medicina, que los resultados (las ganancias) son mejores con cajas de voto que sin ellas, y que este efecto perdura en el tiempo (un mes después).

Ikkal y Sibley (2006), para enseñar el concepto de fuerza, utilizaron cajas de voto desde 1991, y compararon las ganancias relativas entre un pre y un post-test con las ganancias relativas de 1990 (sin cajas de voto), que constituyeron la “línea de base”. La Tabla 8 presenta sus observaciones promedio.

Tabla 8: Ganancias Relativas sin CVE (1990) y con CVE (1991-1997)

	1990	1991 cve	1993 cve	1995 cve	1997 cve
Ganancia Relativa	25%	49%	55%	64%	74%

Los mismos autores recolectaron las opiniones de sus estudiantes, quienes manifestaron que en un curso con cajas de voto electrónicas, en comparación con un curso sin ellas, “los estudiantes se ven mucho más estimulados a pensar” (Ikkal y Sibley, 2006).

En una revista de investigaciones sobre este tema, Duncan (2005) concluye que el uso de cajas de voto electrónicas en clase:

- aumenta el aprendizaje activo.
- aumenta la probabilidad de retención de conceptos cuando han contestado personalmente a las preguntas (aun cuando su primera respuesta fuera errónea), comparado con la situación en la cual simplemente reciben la información.
- aumenta la probabilidad de retención de información cuando han discutido con sus compañeros antes de contestar.
- las preguntas “con *clickers*” entregan retroalimentación rápida no solo a los estudiantes sino también al docente, de modo que puede detectar rápidamente si hay confusiones.
- favorece la “enseñanza por pares”.
- aumenta el placer de aprender de los estudiantes (es lo que estos expresaron en una encuesta).

D.6. El futuro de las cajas de voto

Muy probablemente, en el futuro cercano las funciones que hoy cumplen las cajas de voto electrónicas serán encargadas a computadores en poder de cada estudiante. Esto presenta varias ventajas; por ejemplo:

- la identificación de quién ha contestado, permitiendo un monitoreo de los avances y un seguimiento personalizado.
- la posibilidad de que los estudiantes contesten a PRBs (respuesta breve), con grados de certeza, y envíen comentarios que el docente puede proyectar para que toda la clase los conozca.
- la posibilidad de que cada estudiante reciba, en tiempo real, retroalimentaciones como la evolución de sus notas y de sus índices metacognitivos.

En una fotografía en los complementos multimedia de este libro se puede ver parte de un aula donde cerca de 300 estudiantes rindieron un Test Espectral Metacognitivo (TEM -ver Capítulo 9) durante el curso ISE de D. Leclercq, en diciembre de 2008. Cerca de 100 estudiantes usaron su computador personal para rendir un TEM informatizado (con la ayuda de J. Sougné y S. Delcomminette), mientras sus compañeros (cerca de 200) rendían la misma prueba en versión papel. Es probable que los tablets del tipo i-Pad sean utilizados para las mismas funciones.

E. Permitir a varios jueces corregir a distancia la misma hoja de respuestas

Una reciente realización (Gilles *et al.*, 2010) del SMART, a petición de la Facultad de Medicina de la ULg, consiste en permitir a varios docentes corregir, a distancia, una misma hoja de respuestas. Esta hoja contiene textos escritos a mano, gráficos y dibujos. Los jueces escriben sus puntajes directamente en la copia informatizada, gracias al programa EXAM-QROL desarrollado por el SMART.

Referencias

- BANKS, D. (Ed.) (2006). Audience response systems in higher education: applications and cases. Hershey, PA: Information Science Publishing.
- BRUFF, D. (2009). Teaching with classroom response systems: creating active learning environments. San Francisco: Jossey Bass.
- CRAHAY, V., GILLES, J.L., HAUSOUL, E., MAINFERME, R. y TINNIRELLO, S. (2006). Vers une qualité docimologique accrue des évaluations des acquis des apprenants à l'aide d'un cycle de construction et de gestion des tests standardisés. XXIIIème Congrès de l'Association Internationale de Pédagogie Universitaire 15 - 18 Mai 2006, Monastir Tunisie. http://www.smart.ulg.ac.be/smartweb/documents/aipu2006/presentation_tv_i_aipu.pdf (consultado el 11 de enero de 2014).
- DETROZ, P. y MAINFERME, R. (2007). Evaluation des enseignements à l'Université de Liège. 24e Congrès de l'Association Internationale de Pédagogie Universitaire (AIPU), Vers un changement de culture en enseignement supérieur. Regards sur l'innovation, la collaboration et la valorisation. Faculté des sciences de l'éducation de Montréal, Canada.
- DUNCAN, D. (2005). Clickers in the classroom. How to enhance science teaching using classroom response systems. San Francisco: Pearson Education.
- GILLES, J.L., BOURGUIGNON, J.-P. y DETROZ, P. (2000). Les questionnaires à choix multiple: utilisation pour l'enseignement en groupe avec boîtiers électroniques. Revue Médicale de Liège, 55: 12, pp. 1047-1050.
- GILLES, J.L., PIETTE, S.-A., DETROZ, P., TINNIRELLO, S., PIRSON, M., DABO, M. y LÉ, H. (2005). The electronic Construction and Quality Control in Standardized Testing platform project (e- Ceqcst, in Demetriou, A., Dochy, F. *et al.* (Eds). European Association For Research on Learning and Instruction (EARLI) - 11th Biennial Conference of EARLI: Book of Abstracts. Nicosia: University of Cyprus, Department of Educational Sciences.
- GILLES, J.L., DETROZ, P., CRAHAY, V., BONNET, P. y TINNIRELLO, S. (2010). La plateforme EXAMS, un « Assessment Management System » pour instrumenter la construction et la gestion qualité des évaluations des apprentissages. In J.-. Blais. Evaluation des apprentissages et TIC. Tome 2. Québec. Les presses de l'université Laval, p. 43-63.
- HANZEN, C., CRAHAY, V., DETROZ, P. y LECLERCQ, D. (2010). Impact de tests formatifs en ligne sur l'implication et la perception de leur utilité par des étudiants en médecine vétérinaire. XXVIe Congrès de l'Association de Pédagogie Universitaire. 17-21 mai. Rabat, Maroc.
- IKBAL, J. y SIBLEY, J. (2006). Interactive teaching. Classroom Response Systems at ubc.
- KAY, R. y LESAGE, A. (2009). Examining the benefits and challenges of using audience responses systems: A review of the literature. Computers and Education. 53. P. 819-827.
- KNIGHT, J. y WOOD, W. (2005). Teaching more by lecturing less. Cell Biol Educ. 4 (4), p. 298-310.
- LECLERCQ, D., WILLAIN, J.CH., DENIS B., POUMAY, M., GILLES, J.L., ORBAN, M. y JANS V. (1999). Votes en amphithéâtre électronique pour animer de grands auditoires universitaires selon six paradigmes d'apprentissage / enseignement, In J.P. Bécharde et D. Grégoire (Eds), Apprendre et Enseigner autrement. Actes du 16° colloque international de l'Association Internationale de pédagogie Universitaire (AIPU), Montréal : HEC, 567-578.
- LEGER, M., BOURQUE, J. y RICHARD, J.-F. (2010). Influence des télévotants sur le résultat scolaire: une méta-analyse. IJTHE-RITPU - Revue Internationale des technologies en pédagogie universitaire. 7 (2), p. 35-47.
- MAZUR, E. Case study: Clickers used to teach physics at Harvard. http://tdu.massey.ac.nz/GmA1_Learning_Technologies_Online/Teaching%20with%20Learning%20Technologies%20e1_7/html/course_files/case_study_1.html (consultado el 27 de marzo de 2013).
- MAZUR, E. (1997). Peer Instruction: A User's Manual Series in Educational Innovation (Prentice Hall, Upper Saddle River, NJ).

- MC GEORGE, E., HOMAN, S., DUNNING, J., ELMORE, D., BODIG, G., EVANS, E., KHICHADIA, S., LICHTI, S., FENG, B. y GEDDES, B. (2008). Student evaluation of audience response technology in large lecture classes *Education Tech Research Dev.* 56:125-145.
- NOEL-LAMBOT, F., DETROZ, P. (2003). Utilisation de boîtiers de vote électroniques lors de travaux pratiques et travaux dirigés de biologie (niveau 1re candidature). Actes du colloque: les technologies de l'information et de la communication: mutations dans la formation scientifique universitaire, CIRUISEF-AUF, Dakar.
- SCHACKOW, T., MILTON, C., LOYA, L. y FRIEDMAN, M. (2004). Audience response system: Effect on learning in family medicine residents. *Family medicine.* 36, 496-504.
- SMREK, R. (2011). Two-Way Learning. <http://ohsonline.com/articles/2011/12/01/two-way-learning.aspx> (consultado el 11 de enero de 2014).
- Wood, W. (2004). Clickers: a teaching gimmick that works. *Dev. Cell.* 7, pp. 796-798.

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>