

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](http://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](mailto: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](mailto: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 XIX

### TCS: Test de Concordancia de Script

VALÉRIE MASSART, ANNE COLLARD Y DIDIER GIET

#### A. Nota histórica

En el año 2000 Charlin, Roy, Brailovsky, Goulet y Van der Vleuten publicaron un artículo describiendo el TCS: Test de Concordancia de Script, una nueva técnica para evaluar la capacidad clínica de los médicos. En 2002 Charlin, Van der Vleuten y otros dos autores publicaron una versión en francés de estas ideas.

Esta técnica se basa en el uso de la información para la revisión de hipótesis. Por lo tanto, pensamos que describir el TCS en este libro es de potencial interés para todas las profesiones donde es necesario plantear hipótesis y revisarlas.

#### B. Objetivos de un Test de Concordancia de Script

En su práctica cotidiana los profesionales de la salud se ven enfrentados a problemas variados, de mayor o menor complejidad, con frecuencia mal estructurados o mal planteados (ver Capítulo 7), y con información incompleta. Uno de los resultados es que en muchos casos no existe unanimidad entre los expertos sobre la solución a implementar. Es característico del razonamiento clínico trabajar en situación de incerteza, donde los médicos que tienen muchos años de experiencia logran mayor éxito que los que tienen menos experiencia. Según Charlin *et al.* (2002, p. 136),

La investigación sobre el razonamiento clínico ha demostrado que esos profesionales divergen en su modo de caminar hacia la solución de problemas mal definidos, aunque llegan en general a los mismos diagnósticos y decisiones terapéuticas. Cuando se intenta introducir Preguntas de Selección Múltiple (PSMs) en los exámenes que involucran problemas complejos, los expertos no alcanzan acuerdo sobre la solución correcta. El TCS es innovador desde tres puntos de vista: la pregunta presentada al estudiante, el modo de recolectar su respuesta, y el proceso para establecer su puntaje.

Así, el TCS ha sido concebido como una forma de evaluar, en contexto escolar, la capacidad de razonamiento clínico de los médicos en formación. Esta herramienta debe sumarse a otras, y complementarse en un Dispositivo de Evaluación de los Aprendi-

zajes (DEA) que en conjunto evalúe todos los resultados de aprendizaje comprometidos en el curso o programa.

En el sexto año de formación de la carrera de medicina, en la Universidad de Liège (Bélgica), se realiza un examen de selección para ingresar a la especialización de medicina general. En ese momento de evaluación se practicaban (1) Pruebas escritas con PVF (ver Capítulo 13) sobre saberes aislados, (2) Exámenes Clínicos Objetivos y Estructurados –ECOES– (ver Capítulo 8), y (3) Entrevistas Médicas Simuladas –EMS– (ver Capítulo 8.D.5). Una debilidad de este DEA era que en todas esas formas de evaluación existen pocas dudas acerca de cuáles respuestas podrán ser consideradas correctas, al menos en la opinión de quienes conciben las situaciones y las preguntas. Lo anterior no se orienta a la evaluación de la capacidad de analizar y deducir en la incerteza y con información incompleta, como se hace en el razonamiento clínico. Buscando re-equilibrar su Dispositivo de Evaluación de los Aprendizajes, el Departamento de Medicina General (DMG) de la Universidad de Liège, así como han hecho otras escuelas de medicina en el mundo, decidió añadir el TCS a uno de sus DEAs –en este caso al que se aplica al momento de la selección para la especialidad de Medicina General.

### C. La noción de Script<sup>182</sup> y la forma de la prueba

El postulado en el que se basa el TCS es que, en situación clínica, los médicos movilizan las redes de conocimientos que han construido, en las cuales han establecido relaciones entre una patología y sus características clínicas (como síntomas y señales). Esto les permite tomar decisiones relativas a la fortaleza de cada hipótesis, hasta el nivel de rechazar algunas y profundizar otras. Se denomina *script* tanto a las redes que existen como a la forma de su movilización: en inglés es un guion, un texto dramático (que existe previamente pero debe ser interpretado para “cobrar vida”), y una secuencia de comandos en un computador (con un programa escrito previamente que debe ser ejecutado).

Una prueba TCS se estructura en una serie de *grupos* de preguntas. La Tabla 1 presenta un grupo de 3 preguntas, basadas en una viñeta clínica común, llamada “Situación inicial” (en letras gruesas en la Tabla 1). Cada *pregunta* tiene 3 componentes, presentados en 3 columnas: una hipótesis (columna izquierda), una información suplementaria (columna central), y una escala de 5 posiciones entre las cuales el estudiante debe elegir una opción (columna derecha). En la Tabla 1 hemos indicado la distribución de las *respuestas de 16 expertos*. Por supuesto, esta distribución no es comunicada al estudiante antes de que conteste.

Tabla 1: Ejemplo de un grupo de 3 preguntas (sobre la misma viñeta) de un TCS en medicina general (medicina familiar) utilizada en el curso del Prof. D. Giet (ULg)

Viñeta 1:

Javier, 36 años, chofer de camión de carga pesada, presenta lumbalgias desde hace 3 semanas.		...su intención inicial se vuelve...				
Si Ud. piensa <i>prescribir</i>	Y Ud. averigua que...	-2	-1	0	+1	+2
un relajante muscular de tipo tetrazepam (de la clase de las benzodiazepinas)	...ya ha presentado somnolencia al usar loratadina (clase de los antiestamínicos)	9 exp.	5 exp.	2 exp.		
un antiinflamatorio no esteroideo (piroxicam)	sufre de hipertensión arterial desde hace 5 años, actualmente estabilizada sobre lisinopril (clase de los inhibidores de la enzima de conversión de la angiotensina)	2 exp.	14 exp.			
18 sesiones de fisioterapia y re-educación	...las radiografías únicamente muestran señales de artrosis lumbar			3 exp.	9 exp.	4 exp.

Significación de los niveles de la escala:

-2	-1	0	+1	+2
inútil o no pertinente	menos útil, menos pertinente	no influenciada por los nuevos datos	útil y deseable	Imprescindible

El grupo de preguntas que se muestra en la Tabla 1 son preguntas acerca de tratamiento. En los TCS se intercalan *tres tipos* de grupos de preguntas. Estos son:

- (1) Preguntas acerca del *diagnóstico* (si su hipótesis es...)
- (2) Preguntas de *investigación* (si Ud. pensaba pedir un examen complementario de...)
- (3) Preguntas de *tratamiento* (si Ud. pensaba prescribir la siguiente terapia...) como el caso del ejemplo.

### D. El sistema de cotejo

Se pide a un panel de expertos que conteste a cada una de las preguntas, y se establece la distribución de sus respuestas para cada caso. En la Tabla 1 la distribución de las respuestas de los 16 jueces (expertos) a la primera pregunta muestra que no todos están de acuerdo, aunque se ve una cima (*peak*) en la opción *inútil o no pertinente*.

El principio de cotejo para asignar puntaje es que el estudiante recibirá el valor que en esta distribución tenga la celda que ha elegido, y que por supuesto no conoce de antemano. En el ejemplo de la Tabla 1, en la pregunta 1, el máximo puntaje que un estudiante puede recibir es 9 (si elige *inútil o no pertinente*) y el mínimo es 0 (si elige cualquiera de los dos niveles del extremo derecho). Con la respuesta *menos útil, menos pertinente*, el estudiante recibirá 5 puntos, y con la respuesta *no influenciada por los nuevos datos* recibirá 2 puntos.

Si se quiere que cada pregunta de un TCS tenga el mismo peso que las otras, es posible atribuir puntos tomando en cuenta la moda (el *peak*) de la distribución. En este escenario:

<sup>182</sup> Charlin, Tardif y Boshuizen (2000).

- Elegir *inútil* o *no pertinente* recibirá  $9/16 = 0,56$  puntos  
*menos útil, menos pertinente* recibirá  $5/16 = 0,31$  puntos  
*no influenciada por los nuevos datos* recibirá  $2/16 = 0,125$  punto.

El puntaje también puede ser calculado por la correlación entre las respuestas del estudiante y la del grupo con la fórmula del Kappa de Cohen (que varía de  $-1$  a  $+1$ ).

¿Cuál sería el máximo de puntaje que se puede obtener en una pregunta de rcs?

Es variable de acuerdo al *peak* de la distribución, pero como se ve en la sección E más abajo, sobre la construcción de un rcs, nunca puede ser 16 sobre 16. Si hubiese una sola respuesta elegida por todos los jueces no se trataría de un problema apropiado para evaluar la capacidad de razonar clínicamente, a través de un rcs, sino más bien de una pregunta sobre un contenido apropiado para ser evaluado con Preguntas de Selección Múltiple (psm) o Preguntas Verdadero-o-Falso (pvf).

### E. La construcción de un Test de Concordancia de Script

En 2009, en el Departamento de Medicina General (DMG) de la Facultad de Medicina de la Universidad de Liège, siete profesores del DMG escribieron viñetas clínicas, cada una compuesta de tres preguntas. Resultaron 83 viñetas que fueron presentadas a 13 expertos en medicina general que ejercen en terreno. El proceso de revisión de las respuestas de estos jueces se organizó en los siguientes pasos:

- (1) Verificación de la homogeneidad de los expertos. Este paso resultó en la eliminación de un experto, cuyas respuestas estuvieron sistemáticamente lejos de los otros jueces.
- (2) Eliminación de las preguntas que presentaron una distribución bimodal de las respuestas de los expertos. Es decir, las preferencias se agruparon de forma equivalente a ambos lados de la escala.
- (3) Eliminación de las preguntas sin distribución. Es decir, todos los expertos eligieron la misma respuesta.

Finalmente, este proceso resultó en 50 viñetas, de tres preguntas cada una, para construir un rcs: 24 viñetas de diagnóstico, 13 de investigación y 13 de terapia.

El rcs así construido fue rendido por 31 estudiantes. Después de eliminar 5 viñetas, las 45 que quedaron presentaron un coeficiente de coherencia interna (Alfa de Cronbach) de 0,81. Lo anterior significa que el valor (numérico) de esta forma de cotejo (y de medición) tiene una buena fiabilidad (*reliability*).

### F. Las dimensiones de validez ETICPRAD<sup>183</sup> de un Test de Concordancia de Script

Charlin *et al.* (2002) mencionan las investigaciones edumétricas que demuestran lo siguiente:

- *La validez Teórica* (de *constructo*). En radiología, Charlin *et al.* (1998a y 1998b) observaron que los médicos más experimentados logran notas más altas que los “internos en hospital”, y estos más altas que los “externos al hospital”. Lo anterior indica que el rcs está construido de forma que logra reflejar los desarrollos de competencias en el tiempo durante el ejercicio de la profesión.
- *La validez Informativa*. Sibert *et al.* (2002) construyeron un rcs de urología en francés y lo tradujeron al inglés. Lo aplicaron a expertos y estudiantes de Francia y Canadá, y observaron que los estudiantes de Francia lograron mejores notas cuando son evaluados con las pautas de los jueces de Francia, y lo mismo ocurre para los estudiantes de Canadá. Lo anterior indica una sensibilidad del rcs al contexto.
- *La validez Predictiva*. Brailovsky *et al.* (2001) aplicaron un rcs sobre conocimientos clínicos en cirugía a todos los estudiantes, dos años antes del fin de la carrera. Al mismo tiempo, los estudiantes de la especialidad de medicina general además participaron de un ECOE. Los resultados confirmaron la hipótesis de que las notas en el rcs predicen mejor los resultados en el razonamiento clínico que los resultados respecto a las habilidades clínicas.
- *La validez de Replicabilidad* (fiabilidad-*reliability*). Los resultados de las tres investigaciones mencionadas arriba muestran que una replicabilidad de 0,8 se logra con un rcs de 60 ítems, que se puede rendir en una hora. Charlin *et al.* (2002, p. 142) observaron que cuando los expertos se reúnen para dar las respuestas que consideran correctas en forma colectiva (método de *consensus*), en el 59% de los casos no entregan la misma respuesta que dan individualmente, sobre la escala de 5 niveles.
- *La validez de Aceptabilidad*. Charlin *et al.* (2002, p. 142) observaron que el principio del rcs es bien aceptado tanto por los estudiantes como por los docentes.

### G. Resultados de la utilización de Test de Concordancia de Script

En 2011 se desarrollaron 50 viñetas en el Departamento de Medicina General de la Universidad de Liège. 104 estudiantes entre tercero y sexto año contestaron un rcs compuesto por 20 viñetas o casos, cada uno formado por un grupo de tres preguntas

<sup>183</sup> Ver Capítulo 4.

(60 preguntas en total;  $NP=60$ ). Los mismos estudiantes rindieron una Prueba Verdadero-o-Falso (PVF) de 120 preguntas ( $NP=120$ ), con grados de certeza. Los valores de las notas promedio (y las desviaciones estándar) de las distribuciones se presentan en la Tabla 2.

Tabla 2: Valores estadísticos de la experiencia en el DMG de la U. de Liège, donde NE es el número de estudiantes en cada grupo<sup>184</sup>

Grupo (NE)	Puntos en el TCS (NP=60) (Test de Concordancia de "Script")		Puntos en la Prueba de PVF (NP=120) (Verdadero-Falso)	
	Media	Dev. Estándar	Media	Dev. Estándar
Año 3 (35)	0.57	0.16	75.8 c	8.22
Año 4 (20)	0.55	0.15	65.6	6.16
Año 5 (25)	0.67 b	0.10	68.6	3.90
Año 6 (24)	0.64 b	0.11	66.3	6.29
Tutores (9)	0.79 a	0.05	72.6	7.46

Las correlaciones entre las notas totales en el TCS y en las pruebas de PVF se representan en la Figura 1, para dos grupos de estudiantes: los de los años 3 y 4 (a la izquierda) y de los años 5 y 6 (a la derecha).

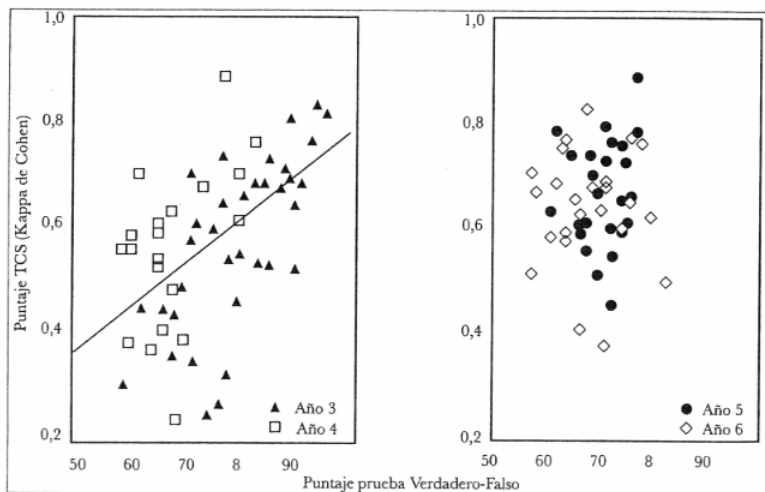


Figura 1: Correlaciones entre las pruebas PVF y TCS en 2 grupos de estudiantes de medicina en 2011 en la Universidad de Liège (DMG)

<sup>184</sup> \*  $p < 0.05$  comparando año 3 con año 6; <sup>b</sup>  $p < 0.05$  comparando año 3 con año 4; <sup>c</sup>  $p < 0.05$  comparando año 4 con año 6.

A partir de la Figura 1 se pueden hacer las siguientes observaciones:

- (1) El promedio de las notas en la prueba de PVF no mejora al comparar los dos grupos: años 3-4 (grupo 1) y años 5-6 (grupo 2). Esto indica que no es en los conocimientos donde se produce una mejora durante los últimos años de formación.
- (2) El promedio de notas en el TCS del grupo de estudiantes más avanzados (años 5 y 6) es más alto que la nota promedio del grupo de los años 3 y 4. Esto indica que la competencia de razonamiento clínico va mejorando a medida que se avanza en los años de estudio.
- (3) La correlación entre los resultados en las PVF y el TCS es mejor en los años 3 y 4 ( $r = 0.53$ ) que en los años 5 y 6 ( $r = 0.07$ ). Parece que en los dos últimos años las diferencias interestudiantes en el dominio de conocimientos juegan un rol menos importante. Se ve que entre los estudiantes hay menores diferencias en los resultados de la prueba con PVF: el ancho de la nube de puntos (o desviación estándar) es más estrecho.

## H. Investigaciones en desarrollo sobre los Test de Concordancia de Script

El Departamento de Medicina General de la ULg ha iniciado una serie de investigaciones acerca de la evaluación de los aprendizajes logrados por los estudiantes. Algunas preguntas de investigación centradas en el TCS son:

- ¿Cuáles son las correlaciones entre las notas en el TCS y otros tipos de pruebas?
- ¿Cómo determinar el umbral de éxito en el TCS?
- ¿Cuáles son las diferencias entre varios paneles de expertos?
- ¿Qué se debe cambiar cuando los casos y preguntas son analizados según los criterios de la Medicina Basada en Evidencias? (ver Capítulo 7).
- ¿Cuál es la validez consecuencial (ver Capítulo 4) del TCS, es decir, su impacto sobre los estudiantes?, ¿modifican sus comportamientos de aprendizaje?
- ¿Cómo evolucionan (considerando varios años sucesivos) los resultados comparados de los estudiantes en medicina en Pruebas de tipo PVF y de tipo TCS? Ver Collard *et al.* (2009).

Charlin *et al.* (2002) han puesto en su agenda de investigaciones:

- ¿Cuál es el número necesario de expertos en un panel de TCS?
- ¿Cuáles son las diferencias entre paneles constituidos por expertos en MG, de acuerdo con el lugar de trabajo (en hospital, en práctica rural, urbana, etc.)?
- ¿Cuáles son los ítems (preguntas) con mejor discriminación? (es decir, capaces de distinguir entre los estudiantes más competentes y los menos competentes en un

grupo) ¿Los ítems con un gran consenso entre los expertos? ¿O los con divergencias entre ellos? La hipótesis de los autores es que son los del segundo tipo.

## I. Conclusiones

El rcs permite evaluar de modo estandarizado (con las mismas preguntas para todos los estudiantes) un gran grupo de estudiantes en poco tiempo, y con una corrección objetiva. Charlin *et al.* (2002, p. 143) piensan que “no hay un instrumento que, por sí solo, pueda evaluar la capacidad clínica”, de modo que el rcs debe ser parte de un Dispositivo de Evaluación de los Aprendizajes (DEA) que utilice varios instrumentos de manera concomitante. Por ejemplo las Preguntas de Selección Múltiple (psm, ver Capítulos 13 y 14), los ECOES (ver capítulo 8), los AFC (análisis fraccionado de casos; ver Capítulo 7), y los rcs, todos ellos con formatos de exámenes estandarizados y calificados mediante corrección objetiva, junto a los portafolios (Giet *et al.*, 2012), como evidencia de la integración de aprendizajes (ver Capítulo 12). Aunque, hasta hoy, el rcs ha sido esencialmente desarrollado en medicina, puede inspirar aplicaciones en otros ámbitos donde se trabaje en la incerteza y con información incompleta.

## Referencias

- BRAILOWSKY, C., CHARLIN, B., BEAUSOLEIL, S., COTÉ, S. Y VAN DER VLEUTEN, C. (2001). Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: An exploratory study on the Script Concordance Test. *Medical Education*, 35, pp. 430-436.
- CHARLIN, B., BRAILOWSKY, C., BRAZEAU-LAMONTAGNE, L., SAMSON, L. Y LEDUC, C. (1998a). Script Questionnaires: Their use for assessment of diagnostic knowledge in radiology. *Medical Teacher*, 20, pp. 567-571.
- CHARLIN, B., BRAILOWSKY, C., LEDUC, C., Y BLOUIN, D. (1998b). The Diagnostic Script Questionnaire: A new tool to assess a specific dimension of clinical competence. *Advances in Health Sciences Education*, 3, p. 51-58.
- CHARLIN, B., TARDIF, J. Y BOSHUIZEN, H. (2000a). Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Academic Medicine*, 75, pp. 182-190.
- CHARLIN, B., ROY, L., BRAILOWSKY, C., GOULET, F. Y VAN DER VLEUTEN, C. (2000b). The Script Concordance Test, a Tool to Assess the Reflective Clinician. *Teaching and Learning in Medicine*, 12: 189-195.
- CHARLIN, B., GAGNON, R., SIBERT, L. Y VAN DER VLEUTEN, C. (2002). Le test de concordance de script, un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale*, 3 : 135-144.
- COLLARD, A., GELAES, S., BREDART, S., DEFRAIGNE, J.-O., VANBELLE, S., BONIVER, J. Y BOURGUIGNON, J.-P. (2009). Reasoning versus knowledge retention and ascertainment throughout a PBL curriculum. *Medical Education*, 43 (9), pp. 854-865.
- GIET, D., MASSART, V., DEUM, M. Y BRICHANT, J.-F. (2012). Dispositif pédagogique en médecine palliative visant le développement de la réflexivité chez des étudiants. Utilisation d'un portfolio électronique. *Pédagogie Médicale*, 13 (1), pp. 61-64.
- SIBERT, L., CHARLIN, B., CORCOS, J., GAGNON, R., GRISE, P. Y VAN DER VLEUTEN, C. (2002). Stability of clinical reasoning assessment results with the script concordance test across two different linguistic, cultural and learning environments. *Medical Teacher*, 24, pp. 537-542.

# 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	<b>ATOME (Alineamiento en un Tablero de Objetivos, Métodos y Evaluaciones.</b> 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.	<b>D.Leclercq &amp; Álvaro Cabrera</b> p. 23-34
2	<b>Los componentes de un dispositivo de evaluación de los aprendizajes (DEA)</b> 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.	<b>D. Leclercq</b> p. 35-50
3	<b>El prisma de las características de un Dispositivo de Evaluación de los Aprendizajes (DEA)</b> 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.	<b>D. Leclercq</b> p. 51-82
4	<b>ETIC PRAD: Ocho criterios de validez de un Dispositivo de Evaluación de los Aprendizajes (DEA)</b> 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).	<b>D. Leclercq</b> p. 83-92
5	<b>Autodescribir y evaluar el Dispositivo de Evaluación de los Aprendizajes (DEA) de un curso</b> 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.	<b>D. Leclercq &amp; Álvaro Cabrera</b> p. 93-102



6	<p><b>La calificación subjetiva de los desempeños complejos: Criterios y rubricas</b>                  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><b>D. Leclercq &amp; Álvaro Cabrera</b>  p. 103-128</p>
7	<p><b>Evaluar la capacidad de resolver problemas</b>                  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><b>D. Leclercq, S. Delcomminette</b>  (HERS) &amp; A. Cabrera p. 129-152</p>
8	<p><b>ECO: Exámenes Clínicos Objetivos y Estructurados</b>                  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><b>G. Philippe (ULg), D. Leclercq &amp; J-P. Bourguignon (ULg)</b> p. 153-170</p>
9	<p><b>Meta cognición y Tests Espectrales Metacognitivos (TEMs)</b>                  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><b>D. Leclercq &amp; Álvaro Cabrera</b> p. 171-196</p>
10	<p><b>Evaluar los Aprendizajes en la Pedagogía Por Proyectos (PPP)</b>                  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><b>Álvaro Cabrera</b> p. 197-220</p>
11	<p><b>Evaluar la contribución de cada participante a un trabajo grupal</b>                  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><b>D. Leclercq, P. Gillet (ULg), M. Erpicum (ULg) &amp; A. Cabrera</b> p. 221-242</p>
12	<p><b>Los Portfolios: Hacia una evaluación más integrada y coherente con el concepto de desempeño complejo</b>                  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><b>M. Poumay (ULg) &amp; Chr. Maillard (ULg)</b> p. 243-260</p>

13	<p><b>Las Preguntas de Selección Múltiples (PSM): del currículo escondido a la vigilancia cognitiva</b>                  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><b>D. Leclercq &amp; Álvaro Cabrera</b>                   p. 261-286</p>
14	<p><b>Reglas de redacción de las Preguntas de Selección Múltiples y la habilidad para responder pruebas</b>                  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><b>D. Leclercq</b>                   p. 287-300</p>
15	<p><b>Evaluar procesos cognitivos según la Taxonomía de Bloom</b>                  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><b>D. Leclercq</b>                   p. 301-328</p>
16	<p><b>Auto-evaluación con grados de certeza: un microscopio para la evaluación de los aprendizajes</b>                  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><b>D. Leclercq</b>                   p. 329-356</p>
17	<p><b>Grados de certeza y docimología: como calificar</b>                  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><b>D. Leclercq</b>                   p. 357-386</p>
18	<p><b>PdP: Pruebas de Progreso</b>                  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><b>D. Leclercq, A. Cabrera &amp; C. Van der Vleuten (U. Maastricht)</b>                   p. 387-408</p>
19	<p><b>TCS : El Test de concordancia de Script</b>                  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><b>V. Massart (ULg), A. Collard (ULg) D. Giet (ULg)</b>                   p. 409-418</p>

20	<p><b>Concebir Dispositivos de Evaluación de los Aprendizajes (DEA) al nivel de un programa</b>                  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><b>D. Leclercq, C. Van der Vleuten &amp; A. Cabrera</b> p. 419-430</p>
21	<p><b>Retroinformaciones (Feedbacks )</b>                  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><b>D. Leclercq, M. de la Fuente (UCH) &amp; A. Cabrera</b> p. 431-454</p>
22	<p><b>Los roles de un SMART: Servicio Metodológico de Apoyo a la Realización de Tests</b>                  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><b>D. Leclercq &amp; P. Detroz (ULg)</b> p. 455-476</p>
23	<p><b>Índices cuantitativos en Docimología</b>                  Consiste en un catálogo de conceptos útiles para tratar cuantitativamente los datos resultando de evaluaciones estandarizadas como</p> <ul style="list-style-type: none"> <li>-los tipos de categorías (nominales, ordinales, métricas).</li> <li>-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>).</li> <li>-las presentaciones gráficas de distribuciones.</li> <li>-índices de comparación o de progreso: la amplitud del efecto (AE), la ganancia relativa (GR).</li> <li>-la fiabilidad de la nota (<i>reliability</i>) al total de la prueba y el alfa de Cronbach.</li> <li>-el umbral de éxito, fijado a priori o a posteriori.</li> <li>-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</li> <li>-el análisis automática de una prueba</li> <li>-el valor heurístico de los nubes de puntos.</li> </ul>	<p><b>D. Leclercq, R. Roco (Chile) &amp; A. Cabrera</b> p. 477-543</p>
24	<p><b>Index de los autores</b> 426 autores citados.</p>	<p><b>D. Leclercq &amp; A. Cabrera</b> p. 545-549</p>
25	<p><b>Index de los conceptos</b>                  Se puede bajar gratuitamente via <a href="http://hdl.handle.net/2268/180060">http://hdl.handle.net/2268/180060</a></p>	<p><b>D. Leclercq &amp; A. Cabrera</b></p>