

www.universitaria.cl

Dieudo LECLERCQ



Álvaro CABRERA MARAY



UNIVERSIDAD
DE CHILE



Directores de la publicación:

Dieudonné Leclercq
Universidad de Liège (ULg)

Álvaro Cabrera Maray
Universidad de Chile (UCH)

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

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

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

Dieudonné Leclercq

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

Álvaro Cabrera Maray

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

Contenidos del libro IDEAS:

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

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

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

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

La lista de los capítulos y el resumen de cada uno

aparece a continuación después de este capítulo.

CAPÍTULO XVIII

PdP: Pruebas de Progreso

DIEUDONNÉ LECLERCO, ÁLVARO CABRERA Y CEES VAN DER VLEUTEN

“Usted extraerá de la prueba lo que hubiera puesto en ella”¹⁵⁵.

VAN DER VLEUTEN, 1998.

A. El Aprendizaje basado en Problemas en medicina en la Universidad de Maastricht

El sistema de evaluación que se presenta en este capítulo, conocido como las Pruebas de Progreso (PdP), es tan original que antes de describirlo se hace necesario comprender el contexto en el que ha nacido y continúa hasta hoy desarrollándose: el PBL (Problem Based Learning) o ABP (Aprendizaje Basado en Problemas) implementado en la Facultad de Medicina de la Universidad de Maastricht.

A.1. Contexto histórico

Durante el periodo 1965-1970¹⁵⁶, en la Universidad Mc Master en la ciudad de Hamilton, Ontario (Canadá), varios decanos y profesores¹⁵⁷ de la Facultad de Medicina iniciaron la implementación del método ABP – PBL (Aprendizaje Basado en Problemas – Problem Based Learning). Durante el mismo periodo, Howard Barrows junto a S. Abramson, en la USC¹⁵⁸, inventaron el uso de “pacientes simulados”.

Ya como Decano de la formación de Posgrado en la Universidad Mc Master, Barrows, en 1971, desarrolló el método de PBL que se describe a continuación, cuyas ideas y conceptos han sido diseminadas por Barrows y Tamblyn en su libro de 1980.

La Facultad de Medicina de la Universidad de Maastricht fue la segunda escuela en el mundo que implementó el ABP. Como fue construida en 1972, desde su fundación adoptó el enfoque del ABP y su filosofía, que coloca el aprendizaje de los estudiantes al centro del currículum. Así, incluso la arquitectura del “paisaje de estudio”, contigua al hospital, fue concebida en función de esta nueva pedagogía.

¹⁵⁵ Lo que significa que los estudiantes tienen en cuenta el contenido y los procesos mentales evaluados por una prueba al definir sus formas de estudiar (*Test driven curriculum*).

¹⁵⁶ La cronología de este párrafo es una comunicación personal de Geoff Norman a D. Leclercq.

¹⁵⁷ Los profesores J. Evans (cardiólogo, 1er Decano), F. Mustard (cardiólogo, 2do Decano), B. Spaulding, B. Walsh, J. Anderson (muy influyente sobre el concepto de PBL). Una segunda ola la constituyeron profesores como M. Campbell, B. Mueller, D. Sackett, V. Neufeld y H. Barrows.

¹⁵⁸ University of Southern California.

A.2. Los objetivos y sus implicancias metodológicas

Además de perseguir objetivos específicos relativos a que sus estudiantes sean “competentes en medicina”, el movimiento del ABP persigue desarrollar competencias transversales como “ser capaz de autoformarse con el apoyo de sus compañeros, a partir de problemas”, que son cruciales en la perspectiva del aprendizaje continuo durante toda la vida.

Lo anterior tiene como implicancia importantes diferencias metodológicas entre este enfoque pedagógico y aquel de los programas clásicos de formación en medicina. Estos programas tradicionales, con fuerte énfasis teórico, generalmente basan la formación en los primeros años en una enseñanza ex cátedra de contenidos como matemáticas, física, química, biología; a lo que sigue la bioquímica, anatomía, etc. Con frecuencia la responsabilidad de integrar esos conocimientos es dejada a las capacidades de cada estudiante, sin acompañamiento del docente. La carga de enfrentar a los estudiantes con la realidad (pacientes reales) se concentra en los últimos años de la carrera, lo que crea problemas de motivación en los estudiantes. La formación técnica (proceder en una anamnesis, observación, auscultación, manipulación, punciones, inyecciones) intenta ser asegurada por talleres donde se practican métodos como observación, imitación, ejercitación, autoscopia, etc.

El sistema del ABP en Maastricht coincide con enfoques más tradicionales en la forma de asegurar la formación técnica, a través de *skillslabs* que funcionan con los métodos clásicos, pero la formación teórica, la integración de saberes y el desarrollo de competencias genéricas son asegurados con el método innovador de los “7 jumps” (7 saltos), que combina grupos tutoriales (enfaticando la colaboración) y autoformación en el “paisaje de estudio” (enfaticando la formación autónoma), en fases llamadas de “ida” y “vuelta”.

A.3. La fase “ida” de un grupo tutorial (con un paciente simulado): encuentro con un caso clínico

En la Facultad de Medicina de la Universidad de Maastricht todos y cada uno de los más de 1.500 estudiantes (hay 200 y hasta 300 en cada grado, y se requieren seis años o grados para ser médico general) pertenecen a un grupo de 8 estudiantes, que se reúne dos veces por semana en lo que se llama el “grupo tutorial”. Durante una hora, y en presencia de un “tutor” (miembro del cuerpo de profesores) el equipo sostiene un encuentro con un “paciente simulado”, es decir, una persona que no está realmente enferma pero que lo simula según un “rol” (una viñeta clínica) que ha aprendido¹⁵⁹. El objetivo del equipo es “entender mejor lo que le pasa o lo que le puede pasar”, pero sin hacer un diagnóstico ni tampoco prescripciones. Durante esta hora, el grupo de ocho

¹⁵⁹ En algunas facultades de medicina se trata de pacientes estandarizados, debido a que los actores que los representan no pueden improvisar de ninguna forma.

estudiantes, dirigidos por uno de ellos –rol que se cambia en cada encuentro–, pasa por 5 de los 7 saltos (Schmidt, 1993) para resolver un problema de este tipo:

- (1) Clarificación de los términos y conceptos que no son directamente comprensibles
- (2) Definición del problema
- (3) Análisis del problema y tentativas de explicaciones
- (4) Elaboración de la lista de explicaciones
- (5) Formulación de los objetivos de aprendizaje

El paciente no puede contestar más que lo que está escrito en su rol. El tutor no expone contenidos ni enseña medicina, sino que asegura que el grupo plantee bien el problema y recorra los 5 saltos. El uso de la palabra es moderado por uno de los 8 estudiantes, que es el presidente del encuentro (*chairman*). Las decisiones del grupo (del tipo “¿Qué vamos a intentar saber/averiguar?” y “¿Quién va a buscar cada información?”) son escritas por un “secretario” (que también cambia en cada encuentro del grupo) sobre un pizarrón visible por todos. Al final de la hora los miembros del equipo se separan para cumplir con la etapa 6 de los 7 saltos. Esta etapa de indagación se realiza en el “paisaje de estudio” (*Study landscape*¹⁶⁰) o paisaje de aprendizaje: una inmensa biblioteca “de trabajo”.

A.4. La fase 6 de los 7 saltos: en el paisaje de estudio

Durante dos días, en el paisaje de aprendizaje, los miembros del equipo buscan documentos (libros, revistas, periódicos, sitios Web, imágenes y videos) donde puedan comprobar las informaciones que les permitan contestar al problema planteado por el paciente de la etapa “ida”. Los libros importantes están disponibles en decenas de copias. Existen varias salas para el trabajo individual, en su mayoría equipadas con computadores. También hay espacios para reuniones de trabajo y estudio en pequeños grupos.

A.5. La fase “vuelta” del grupo tutorial (etapa 7 de los 7 saltos)

Después de dos días transcurridos principalmente en el paisaje de aprendizaje, el grupo de ocho estudiantes y el tutor se encuentran de nuevo. Los miembros comparten las informaciones que han recolectado sobre el tema, y discuten sus vínculos e implicancias hasta acordar una explicación para el caso.

En la Universidad de Liège, que desde 2002 adoptó el método de ABP, cada miembro del grupo debe llegar a esta reunión con un esquema fisiopatológico, y el grupo construye un esquema común.

Esta etapa se lleva a cabo durante una hora. La hora siguiente sirve para la presentación de un nuevo caso con otro paciente simulado. De este modo, el grupo debe considerar dos casos por semana.

¹⁶⁰ Se pueden ver fotografías de este paisaje de estudio en los complementos multimedia de este libro.

Un grupo funciona durante seis semanas, y luego de ese plazo se constituyen nuevos grupos (al azar).

El contenido de los casos no es elegido al azar, y sigue una evolución creciente en su complejidad. Los casos de los primeros grados tratarán sobre dolores de espalda y no sobre cáncer, que sí será abordado en años superiores.

B. Las Pruebas de Progreso: concepto e historia

Como hemos visto, son los grupos los que deciden el contenido que van a aprender, de acuerdo al problema que les plantea el caso en estudio, en el paso n° 5 de los siete saltos. Los casos son los mismos para los 30 grupos que conforman un grado (primer año, por ejemplo) y funcionan en paralelo. Sin embargo, el contenido estudiado no es exactamente el mismo entre todos los grupos, pues han tomado diferentes decisiones en el salto 5.

Es un postulado de la Facultad que todos los estudiantes progresen en el dominio de la medicina en todas sus dimensiones. Para evaluar los progresos han ideado dos sistemas de evaluaciones, con funciones formativa y sancionadora, que coexisten: (1) los "Bloc tests" (pruebas de un 'bloque'), que tienen lugar cada seis semanas y abordan los contenidos relativos a las enfermedades presentadas en los casos del mismo periodo, y (2) las "Pruebas de Progreso - PdP", que ocurren cuatro veces por año (es decir cada tres meses), y donde el contenido testeado abarca toda la medicina, incluyendo ciencias básicas y conocimientos de varias especialidades¹⁶¹. Exactamente el mismo test es aplicado a todos los estudiantes de la Facultad, desde el 1° al 6° grado, lo que implicó 800 estudiantes en 1996 y cerca de 2.000 en 2012.

Introducidos en 1976 en la Facultad de Medicina de Maastricht¹⁶² por Wijnen (Van der Vleuten y Wijnen, 1990), estas pruebas contienen entre 200 y 250 preguntas "Verdadero-Falso-No sé (o "?")", agrupadas alrededor de microcasos escritos (viñetas clínicas). Tiene una duración de 4 horas máximo y se realiza el mismo día para todos los estudiantes en inmensas salas.

Típicamente, el puntaje total de un estudiante resulta de la fórmula C-I, donde "C" es el número de respuestas correctas e "I" el número de incorrectas. Las respuestas "?" (no sé) reciben 0 punto. Esta fórmula corresponde a la *correction for guessing* clásica.

¹⁶¹ Se puede ver un listado de los temas de la prueba en la Tabla 2, sección E.

¹⁶² Parece ser simultáneo a la introducción del mismo principio en la Universidad de Missouri en Kansas City (UMKC).

C. La construcción de una Prueba de Progreso¹⁶³

Existe un Comité Revisor de las Pruebas de Progreso (CR-PdP)¹⁶⁴, compuesto por ocho personas que se mantienen en este cargo durante cuatro años¹⁶⁵. Tiene en sus manos las misiones que se describen a continuación:

- (1) El CR recolecta las preguntas que han sido concebidas por los Departamentos. Estas deben ser preguntas que médicos recién graduados hayan contestado correctamente. Los Departamentos añaden a cada pregunta (ítem) comentarios (referencias) que serán entregados a los estudiantes después de la PdP, como parte de la retroalimentación.
- (2) Sin hacer ninguna revisión, el CR categoriza estas preguntas incorporándoles códigos, por ejemplo, en una de las 15 categorías de contenido que aborda una PdP. Un mismo Departamento puede enviar preguntas para más de una categoría.
- (3) Ocho meses antes de la fecha de la PdP, el CR extrae (al azar en cada categoría) entre 400 y 500 ítems del banco de preguntas. El número de preguntas disponibles por categoría excede el número necesario a incluir en la prueba. Por supuesto, mientras más preguntas ha enviado un Departamento, más grande es la probabilidad de que algunas de ellas sean elegidas al azar para integrar una o más categorías, pero está reglamentado que la tasa de ítems originados en un mismo Departamento no puede exceder el 6% del total de una PdP.
- (4) El CR aprueba, modifica o rechaza las preguntas seleccionadas al azar. Este proceso se lleva a cabo en 6 reuniones en una amplia sala con muchos libros disponibles para verificar contenidos. Si es necesario, se consulta algún aspecto con el Departamento que ha concebido el ítem. Como un mismo comité está trabajando sobre más de una PdP durante el mismo periodo, usualmente se encuentran lidiando con cerca de mil ítems. Entre 1976 y 1998, sobre 19.000 ítems seleccionados al azar, 13.000 fueron aprobados y se rechazaron 6.000, pero en el 80% de los aprobados se necesitaron adaptaciones. 97% de las adaptaciones fue sobre la formulación y 56% fueron modificaciones relativas al contenido. Además, el CR asegura una distribución casi igual de las preguntas con respuesta correcta "Verdadero" y "Falso" (cerca de 50% - 50%). También respetan la distribución de 40% de ítems sobre contenidos de ciencias de base, 40% de clínicas y 20% de ciencias comportamentales.
- (5) Después de cada PdP (4 veces por año), el CR envía a cada Departamento la lista completa de sus ítems incluidos en el banco de preguntas, indicando el estatuto en que se encuentran: aún no seleccionado (al azar), utilizado sin modificaciones, utilizado modificado, o rechazado.

¹⁶³ Verhoeven *et al.* (1998) en Verhoeven (2003, p. 41).

¹⁶⁴ Progress test Review Committee-PTRC.

¹⁶⁵ Uno/a de los/las profesores/as que integran el comité está encargado/a de los aspectos administrativos y logísticos.

correctas, *Inc*, de incorrectas, y ² de “no sé”. En la zona derecha de la tabla aparecen la Media y la desviación estándar (*Dev*) de los resultados colectivos de la cohorte del estudiante. Por ejemplo, en el tema 1 (Anatomía) la cohorte logró un promedio de 39% de respuestas correctas, con una desviación estándar de 14%. En las 13 preguntas de este tema este estudiante obtuvo 54% de respuestas correctas (es decir, 7 preguntas), lo que está sobre el promedio del grupo (39%), incluso está más que +1 *Dev* por arriba del promedio. Esto se indica por la presencia de los signos “+”. Cuando la tasa de éxito individual es +0,5 *Dev* por sobre la Media del grupo, el símbolo es solo “+”.

Tabla 2: Resultados entregados a un estudiante

Descripción	Resultados individuales					Resultados colectivos (n = 54)							
	NdP	Cor.	Inc.	?	C-I	Cor	Dev	Inc	?	Dev	C-I	Dev	
1. Anatomía/embriología	13	54++	0--	46	54++	39	14	20	16	40	19	19	23
2. Bioquímica	13	77+	8-	15	69++	68	13	15	11	17	11	52	21
3. Farmacología	7	57+	0-	43	57++	38	19	14	17	48	25	24	27
4. Fisiología	11	73	0--	27++	73++	67	16	19	11	14	12	48	24
5. Genética - Biol. Cel.	8	50	25	25	25	57	16	22	10	21	14	36	23
6. Inmunología	4	75+	25+	0-	50	65	17	15	15	19	20	50	25
7. Microbiología médica	6	83	0-	17+	83	85	12	6	8	10	13	79	16
8. Patología	10	60	0-	40	60	59	14	7	10	34	16	52	18
TOTAL CIENCIAS DE BASE	72	65+	6--	29	60++	58	9	15	7	26	11	43	11
9. Medicina del trabajo	1	0	0	100	0	13	34	7	26	80	40	6	45
10. Cirugía general	16	38--	6--	56	31++	33	13	18	11	48	19	15	14
11. Cardiología	10	30--	10-	60++	20	49	17	21	14	29	21	28	22
12. Dermatología	5	40	0	60	40	31	33	7	15	62	34	25	39
13. Ginecología-Obstetricia	6	33	17	50	17	29	21	13	13	57	21	16	29
14. Práctica familiar	9	44	11--	44+	33+	40	14	27	13	33	17	13	21
15. Medicina interna	20	65++	10-	25-	55++	49	13	17	8	35	15	22	16
16. Pediatría	8	38	0--	63+	38+	36	19	16	14	48	16	20	29
17. Otorrinolaringología	5	20	20+	60	0	22	18	8	15	70	22	14	25
18. Neurología	2	50+	0	50-	50++	19	31	3	11	79	34	16	32
19. Oftalmología	3	33	33++	33-	0	37	28	9	17	54	29	28	37
20. Ortopedia	5	20+	0	80	20++	10	15	3	7	87	19	6	13
21. Pneumología	5	40	0-	60+	40	42	22	10	13	48	22	33	28
22. Medicina de rehabilitación	3	0-	33+	67-	33	15	22	12	23	73	26	3	37
23. Urología	1	100+	0	0	100+	74	44	15	36	11	31	59	73
TOTAL CIENCIAS CLÍNICAS	99	40	9-	51	31+	37	10	15	7	48	14	21	10
24. Salud preventiva	2	0	100++	0-	100-	6	16	42	42	53	46	-36	43
25. Epidemiología	12	75++	8--	17	67+	65	18	18	9	17	18	48	22
26. Legislación sanitaria	2	50	50++	0--	0	40	40	13	26	47	42	27	52
27. Psiquiatría clínica	6	0-	0-	100++	0	18	19	10	16	72	28	8	22
28. Psicología médica	3	33+	33+	33-	0	17	25	18	25	65	41	-1	27
29. Sociología médica	2	100++	0	0-	100++	52	37	7	18	41	41	44	42
TOTAL CIENCIAS COMPORTAMENTALES	27	48	19	33	30	42	13	17	8	41	20	26	10
TOTAL FINAL	198	51+	9--	40	41++	45	9	16	6	39	13	30	9

Cada estudiante además recibe otra versión de los mismos resultados, pero presentados con un agrupamiento de preguntas diferente al de la Tabla 2. Este se ordena por sistemas biológicos y sociales, mientras el de la Tabla 2 se estructura por disciplinas:

Tabla 3: Resultados entregados de otra manera al mismo estudiante

DESCRIPCIÓN	RESULTADOS INDIVIDUALES					RESULTADOS COLECTIVOS (N = 54)							
	NdP	Corr	Incor.	?	C-I	Cor	Dev	Inc	Dev	?	Dev	C-I	Dev
1. Sistema respiratorio	18	56++	6--	39	50++	44	11	23	10	33	14	22	16
2. Hematología y sist. Linfático	15	73+	0-	27+	73++	65	12	17	9	18	11	48	19
3. Sistema músculo-esquelético	19	37+	16	47-	21	30	9	12	10	58	16	18	11
4. Salud mental	9	11	11	78	0	18	18	13	15	70	28	5	16
5. Sistema reproductivo	13	46+	15	38-	31+	35	13	16	9	49	15	19	17
6. Sistema cardiovascular	23	43-	4--	52++	39	54	15	19	11	27	17	35	21
7. Sist. endocrino + metabolismo	14	57-	14	29+	43	65	13	14	9	22	11	51	18
8. Piel	9	22	11	67	11	27	21	14	12	59	26	14	21
9. Salud social	7	43+	43++	14--	0	30	20	19	18	52	32	11	19
10. Sistema digestivo	19	58++	5-	37	53++	45	12	15	10	40	17	29	15
11. Sistema urinario	12	50+	0--	50+	50++	44	11	16	13	40	18	28	17
12. Sistema nervioso	21	48++	10	43-	38++	31	12	12	10	57	18	18	14
13. Varios	9	78	0-	22+	78	82	13	6	6	12	13	76	16
14. Metodología y Filosofía	10	80+	10-	10	70+	67	20	19	12	14	19	49	26
Total final	198	51+	9--	40	41++	45	9	16	6	39	13	30	9

F. Comparaciones transversales de los resultados

En la Figura 1⁷⁰ se presentan los 24 resultados promedio (4 cada año) de algunas cohortes. El primer PdP rendido en Maastricht fue en octubre de 1977. Se ve que lo rindieron estudiantes de 4 cohortes: en la cohorte del grado 4 el promedio de éxito fue 51%; para el grado 3 fue 38%; para el grado 2 fue 29%; y para el grado 1 fue 12%. Se puede ver que las pendientes son más o menos equivalentes en todas las cohortes y que al final (en el PdP n° 24) la tasa media de éxito se acerca al 60%.

G. Evolución de los resultados de una cohorte

Al analizar los resultados de la cohorte que rindió su primera PdP en octubre 1978⁷¹ (Figura 1), obtenidos en las 24 PdP a las que se enfrentaron durante su formación, a un ritmo de 4 cada año durante 6 años, se ve que:

⁷⁰ Van der Vleuten *et al.* (1996, p. 104).

⁷¹ Van der Vleuten y Wijnen (1990), p. 31.

- El promedio de las respuestas correctas (menos las incorrectas) era 11% en octubre 1978. Se ha vuelto 18% en diciembre de 1978, 22% en marzo de 1979, 23% en junio de 1979, y 22% en octubre de 1979.
- A menudo, la primera PdP después de las vacaciones (a la derecha de las líneas verticales) obtiene una tasa de éxito más baja que la inmediatamente anterior.

La línea punteada indica los resultados obtenidos por médicos (entre 50 y 100 cada año) egresados hace un año de varias escuelas de medicina, a quienes se les paga por rendir la PdP. Excepto en una PdP de 1982 han obtenido puntajes totales mejores que los estudiantes del sexto año¹⁷².

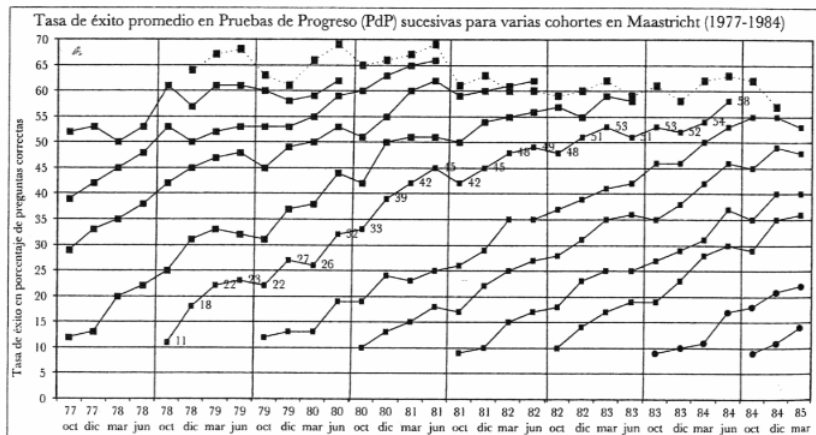


Figura 1: Evolución de los resultados promedios de algunas cohortes entre 1977 y 1985 en la Facultad de Medicina de Maastricht

H. Las ventajas de las Pruebas de Progreso

H.1. Validez teórica de contenido o de cobertura

Preguntas Verdadero-o-Falso (pvf) o Preguntas de Selección Múltiple (psm), si son corregidas por lectura óptica de marcas, permiten utilizar muchas preguntas cubriendo muchos saberes. Así, podemos obtener una buena validez de “cobertura” del contenido.

H.2. Validez consecucional

- a) *Consecuencias sobre la clarificación de los objetivos de aprendizaje.* El estudiante tiene, desde la primera vez que se enfrenta a la PdP, una visión total de los tipos de problemas

que va a encontrar durante el ciclo completo de sus estudios de medicina. Esto le permite contextualizar de mejor forma su estudio, cuando está solo en el paisaje de aprendizaje, o cuando se discute en grupo.

- b) *Consecuencias sobre los métodos de estudio.* El estudiante no se prepara para obtener un éxito momentáneo en un test y luego olvidar, lo más rápido posible, los contenidos testeados, como se observa con demasiada frecuencia en los exámenes en la universidad. Con las PdP nada se puede olvidar. El estudiante no puede decir “ese es un contenido del año pasado” o “eso ya lo aprobé”, subentendiendo que tiene el derecho a haberlo olvidado. No es posible “empollar” o “calentar” el examen durante la última semana o las últimas horas. El mejor método para prepararse es estudiar regularmente.
- c) *Consecuencia en la rapidez de la retroalimentación.* La automatización de la corrección permite entregar retroalimentación rápidamente y de manera muy detallada y sutil. Esto promueve las conductas de búsqueda del estudiante, las orienta, y facilita el autoanálisis y la autorregulación.

H.3. Validez informativa

Partiendo de la base que la institución que aplica la PdP posee la intención y la capacidad de analizar los datos resultantes, una PdP puede entregar información y ayudar a identificar:

- a) *Las diferentes evoluciones según el ámbito o área de contenido (conocimientos, habilidades clínicas, habilidades comportamentales)*

La Figura 2 presenta este tipo de evoluciones¹⁷³ a lo largo de las 24 aplicaciones de la PdP en medicina en la Universidad de Maastricht, en un periodo de 10 años: “evolución ascendente lineal en el puntaje total, resultando de la combinación (y compensación) de evoluciones ascendentes de curva bajando en los conocimientos y las habilidades comportamentales, y de curva subiendo en las habilidades clínicas” (Van der Vleuten *et al.*, 1996, p. 107).

- b) *Las diferentes “quinéticas” de los procesos de adquisición de conocimiento*

Albano *et al.* (1996) describen los resultados de una misma PdP rendida por estudiantes de seis facultades de medicina diferentes, ubicadas en tres países: Holanda (2), Alemania (1) e Italia (3).

Los resultados totales eran poco diferentes en el grado final (6^o año). Sin embargo, existían diferencias en los momentos en los que cada una de las grandes competencias

¹⁷³ La línea central es el promedio de las curvas de predicción (según ecuaciones de regresión) y las dos líneas que la cercan son los intervalos de confianza (con 95% de seguridad). Los datos son acumulados a partir de 10 cohortes (1983-1993).

¹⁷² En Van der Vleuten *et al.* (1996), p. 104.

y habilidades (conocimiento en las ciencias de base, habilidades clínicas, etc.) había progresado, y estas diferencias pudieron explicarse por razones curriculares. Estas variaciones de los “tempos” de las evoluciones fueron bautizadas por estos autores como *knowledge acquisition kinetics*.

Cada PdP sirve como un pre-test para los que siguen, y como un post-test para los anteriores, de modo que los docentes pueden vincular sus intervenciones e innovaciones con los resultados. Lo anterior les permite tomar decisiones, acerca de ajustes y mejoras en su docencia, con base en evidencias. Esta regulación con base en la evidencia constituye el sueño de los militantes de la “evaluación de la clase” (*classroom assessment*) propuesta por Angelo y Cross en 1993.

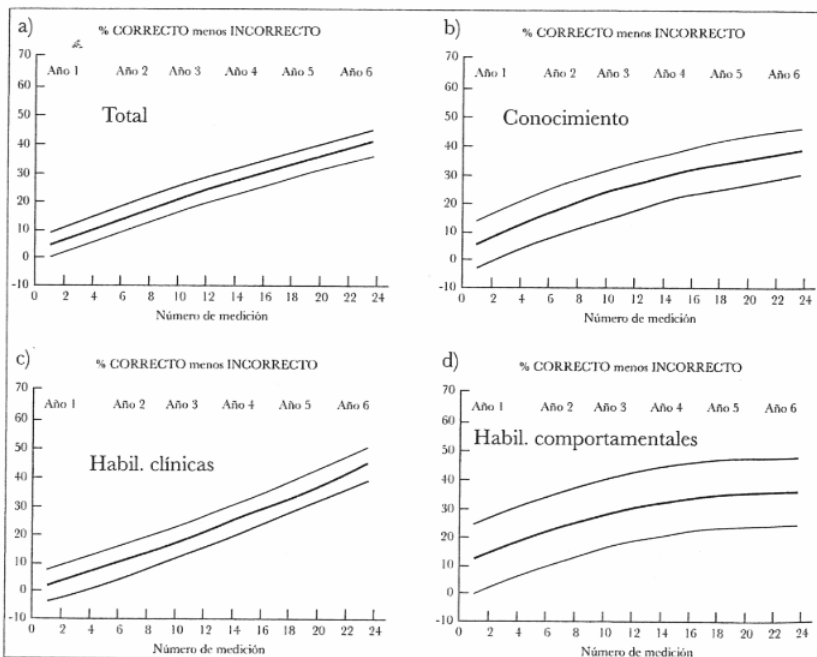


Figura 2: Evoluciones (pendientes) diferentes según el ámbito en 10 cohortes

c) La coherencia entre resultados de aprendizaje y evaluación

Verhoeven (2003, p. 50) describe los resultados de una investigación donde para cada ítem de una PdP se identificó si estaba “conectado” o “no conectado” con el “referencial de competencias”, que fue distribuido a los profesores y estudiantes al inicio del año. Se realizó en 1994 y abarcó 230 ítems rendidos por 863 estudiantes de los seis grados de la carrera. Se determinó que 179 ítems estaban conectados (78%) y 51 ítems estaban “no conectados” (22%). En el grado 6 el

promedio de respuestas correctas para los 179 ítems “conectados” fue 61%, y para los “no conectados” fue 41%.

d) La relación entre los procesos (modos de estudio) y los productos (resultados en los exámenes)

Verhoeven (2003, pp. 111-121) describe las relaciones entre los resultados en las PdP y los modos de estudio de los estudiantes (activo vs pasivo; en superficie vs en profundidad), medidos por cuestionarios:

- Los estudiantes con modos de estudio deseables (en profundidad, activo) logran un puntaje más alto en las PdP.
- Aunque existe una correlación alta entre los resultados positivos y la cantidad de preparación específica en el caso de otras evaluaciones (como los ECOE o las Pruebas de Bloque), esta correlación es muy baja en el caso de las PdP.

I. Las desventajas de las Pruebas de Progreso¹⁷⁴

I.1. Ausencia de validez diagnóstica en primer año

Solo una pequeña fracción de las preguntas de las PdP refleja los contenidos del currículum del primer año (no más del 20% de los ítems), de modo que el sistema es poco sensible a los progresos de los debutantes. Por tanto, no es el más apropiado si se trata de identificar a aquellos estudiantes novatos que están en dificultades.

I.2. Problema de validez de replicabilidad

Aunque la fiabilidad (*reliability*) de las PdP es buena dentro de un año de formación (alfas¹⁷⁵ entre 0,7 y 0,8) y a través de los diferentes años (alfas de 0,95 o más), y aunque la construcción de las PdP es muy estandarizada, las variaciones de desempeño (de performance) entre las PdP de un mismo año son amplias, lo que provoca dificultades en la fijación del umbral de éxito. Por eso, en la U. de Maastricht han desarrollado el principio de los umbrales de éxito relativos, que son referencias normativas.

Para esto se postula que:

- a) los estudiantes de las cohortes sucesivas de un mismo grado tienen la misma calidad promedio
- b) los estudiantes hacen los mismos esfuerzos durante el año
- c) las diferencias de resultados promedio en las PdP resultan de las diferencias entre las dificultades intrínsecas de las preguntas

¹⁷⁴ Según Van der Vleuten *et al.* (1996), p. 106.

¹⁷⁵ El alfa de Cronbach es el índice matemático de la fiabilidad del total de las notas en una prueba. Su rango va de 0 a 1 (ver Capítulo 23).

I.3. Problemas de validez de aceptabilidad

- Es necesaria una organización central* para concebir, administrar y procesar las PdP, y es un método de evaluación *costoso* en términos del tiempo que consumen los procesos de revisión de las preguntas (pero que aseguran la validez teórica). En la Universidad de Maastricht este tiempo se gana al haber suprimido el formato de exámenes creados y administrados por cada docente.
- La selección de los ítems es difícil*. Como no existen vínculos estrechos entre los casos del ABP y las PdP, la identificación de preguntas corresponde a los objetivos fundamentales del currículum, y puede ser un proceso difícil para los creadores de las PdP. Esto resulta en problemas para prevenir la incorporación de ítems demasiado especializados o de detalles.
- Las PdP no funcionan bien en programas (currícula) heterogéneos* con una especialización temprana.

J. La fijación del umbral de éxito relativo (UER)

En Maastricht, históricamente, se han aplicado sucesivamente tres métodos para fijar el umbral:

- El método de Angoff (1971): una comisión de jueces, formada por una combinación de docentes y de “estudiantes *borderline*”¹⁷⁶, estima, para cada ítem, “la probabilidad de que un estudiante *borderline* dé una respuesta correcta”. El umbral de éxito en la PdP es el promedio de las estimaciones de los jueces para todos los ítems.
- Según el método propuesto por Wijnen (Muijtjens *et al.*, 1998), se calculaba el promedio de la cohorte. El valor resultando de “el promedio $-1,96$ desviaciones estándar” era considerado como el *umbral de éxito*.
- Hoy en día, según el método propuesto por Cohen-Schotanus, Van der Vleuten y Bedner (1996), se calcula el promedio de los resultados totales del 10% de los mejores estudiantes¹⁷⁷, y este promedio es considerado como el “máximo relativo”. El umbral de éxito queda en el 60% de este máximo relativo. Siendo la distribución de los resultados en forma “de Gauss”, aproximadamente el 16% de los resultados se ubica bajo el umbral (relativo) de éxito (Verhoeven, 2003, p. 27).

El resultado de estas referencias normativas es que no se sabe, antes de la aplicación de la PdP, cuál será el umbral de éxito.

¹⁷⁶ “Un estudiante ‘borderline’ es un estudiante que pasa una cantidad promedio de tiempo estudiando, logra un conocimiento apenas suficiente para proseguir al grado superior, pero a menudo tiene dificultades de pasar arriba del umbral de éxito en los PdP” (Verhoeven, 2003, p. 86).

¹⁷⁷ Este porcentaje arbitrario puede variar y ser el 5%, por ejemplo.

K. La utilidad (U) de un método de evaluación

Van der Vleuten (1996) propuso un modelo teórico, aún no aplicado en la práctica, en el cual

$$U = R \times V \times E \times A \times C$$

donde

R es Replicabilidad (fiabilidad);

V es Validez conceptual (o Teórica de constructo);

E es impacto Educativo (o validez Consecuencial);

A es Aceptabilidad; y

C es Costos.

Una U de puntaje máximo es imposible, pero ninguno de los cinco factores puede tener un valor 0. Los factores pueden tener un peso (coeficiente de importancia) diferente según los retos y objetivos de la evaluación (¿intención formativa? ¿Función sancionadora?).

Los criterios de calidad de Dispositivos de Evaluación de los Aprendizajes, ETR-CPAD, propuestos por Leclercq (2006)¹⁷⁸, constituyen un sistema similar (pero que genera un “perfil” en vez de un “puntaje total”) para juzgar las ventajas y desventajas del uso de PdP en un programa de formación y en circunstancias particulares.

L. Tipos de preguntas en las Pruebas de Progreso

L.1. Preguntas Verdadero-o-Falso (PVF)

En sus PdP entre 1977 y 2005, y poniendo el enfoque sobre la rapidez de respuesta, la Universidad de Maastricht eligió utilizar entre 200 y 250 Preguntas Verdadero-o-Falso en sus PdP. Desde 2006 utiliza PSM.

L.2. Preguntas de Selección Múltiple (PSM)

Son varias las universidades que han implementado Pruebas de Progreso usando PSMs. La Universidad de Missouri-Kansas City (UMKC) ha desarrollado un QPE (Quarterly Profile Examination) de 400 PSMs. Willoughby *et al.* (1991) observaron que los resultados acumulados de los QPE estaban bien correlacionados con los resultados en el NBME (National Board Medical Examiners), prueba de nivel nacional en EE.UU., necesaria para ejercer la medicina. También la Peninsula Medical School at Plymouth (UK) usa PSMs en la PdP que ha implementado.

¹⁷⁸ Ver Capítulo 4.

La Universidad de Mc Master (Canadá), que adoptó el principio de las PdP en 1992, utiliza 180 PSMs en su PPI (Personal Progress Index). Blake *et al.* (1996) mencionan que la correlación entre la performance en el PPI y en la Prueba de licenciatura del Consejo Médico de Canadá evolucionó desde 0,12 para la primera aplicación (un mes después de haber ingresado a la facultad de medicina) hasta 0,60 para la nota acumulativa de todas los PPI. Además observaron que una importante tasa de los estudiantes que lograron un nivel bajo en los PPI se autoadministraron remediaciones.

En la Facultad de Medicina de la Universidad de Otago, Nueva Zelanda (Schwartz, 2009), desde el año 2003, 900 estudiantes rinden dos veces cada año una PdP en dos partes, cada una de 3 horas de duración. La primera consta de 200 preguntas, y la segunda de 400. Las correlaciones entre los resultados en las dos últimas PdP (las del último año de formación) y el examen final varían entre 0,74 y 0,79.

L.3. Preguntas de Respuesta Breve (PRB)

La Facultad de Medicina de la Universidad de Utrecht, Alemania (Rademakers *et al.*, 2005), introdujo las PdP en los años 4to y 5to del currículum, usando PRBs, tres veces por año, en lo que llaman UPT (Utrecht Progress Tests). El contenido de las preguntas es “lo que debe ser conocido de memoria sin ninguna preparación”, y se refiere a lo que se ha denominado “Test de Dominio” (Gagné *et al.*, 1992). Se esperaba que los resultados alcanzaran al menos el 80% de respuestas correctas, y en ese porcentaje se fijó el umbral de éxito, en lugar del 60% de los tests tradicionales. Cada versión de un UPT utiliza 40 casos clínicos, cada uno seguido por una y hasta tres preguntas clínicas, más una y hasta tres preguntas de ciencia básica (los ítems son construidos por docentes de estas ciencias básicas y clínicas). El tiempo promedio de preparación / utilización de un caso es 4,7 horas: 2 horas para la concepción, 10 minutos de discusión, 1 hora para revisión y 1,5 para corregir las respuestas de los 200 estudiantes. Los resultados fueron decepcionantes: sucesivamente 45% y 48% de éxito en vez del 80% esperado (que ningún estudiante alcanzó). Rademakers *et al.* consideran que 1/3 de las preguntas eran demasiado difíciles, y/o trataban de detalles. También en Utrecht (como en Maastricht) la referencia del máximo absoluto (100%) fue reemplazada por un máximo relativo, que es “el puntaje promedio del 10% de los mejores estudiantes de 5to año”. Los estudiantes alcanzan lo esperado (aprueban) cuando obtienen el 80% de este máximo relativo. Los autores observan que el tiempo de corrección es “razonable cuando los números de estudiantes y de preguntas son limitados”.

M. ¿En papel o computador?

En la Facultad de Medicina de la Universidad de Maastricht, las PdP funcionan con respuestas marcadas sobre hojas pre-imprimidas, diseñadas para ser leídas por un lector óptico de marcas (LOM).

Como consecuencia de las quejas, tanto de los estudiantes como de los docentes, relativas a que el tiempo requerido para la rendición de las PdP “cortaba” el tiempo destinado a las clases, la Universidad de Otago (Nueva Zelanda) comenzó en 2007 a administrar las PdP vía computadores. Los estudiantes pueden rendir la PdP en cualquier momento dentro de un periodo de dos semanas. El número de preguntas por PdP fue reducido a la mitad. Schwartz (2009) menciona que:

- El 15% de los estudiantes confiesa que consulta (en internet y en libros) las respuestas correctas de preguntas de las PdP anteriores antes de rendir una.
- Hay estudiantes que confiesan que no rinden las PdP con la misma seriedad que antes, porque en la actualidad existe demasiada flexibilidad. Esto hace decir a Schwartz: “Hagas lo que hagas, no puedes dejar contentos a todos”.

Las distribuciones de los resultados parecen las mismas que las obtenidas en versiones “en papel”, pero la correlación entre la PdP del fin del año 2007 con el examen final de la carrera bajó desde 0,65 en 2007 hasta 0,53 en 2008. Al parecer este cambio se motiva por las razones *a* y *b* expuestas arriba.

En 2009, semanas antes de las PdP de ese año, Schwartz envió e-mails a los estudiantes del 5° año, explicando lo que ocurrió en 2007 y 2008, y recordándoles que la mejor retroinformación se puede obtener de una PdP cuando el estudiante se presenta sin preparación específica y sin consultar las respuestas correctas antes. Puede ser la razón por la que, en 2009, la correlación alcanzo 0,70, casi el valor de antes del 2006.

N. Programas internacionales que usan las Pruebas de Progreso (en 2010)

Freeman *et al.* (2010) elaboraron la siguiente lista de instituciones, presentes en los cinco continentes, que usan las PdP como parte de sus dispositivos de evaluación de los aprendizajes a nivel de un programa completo.

- El Grupo Holandés, compuesto por cinco facultades de medicina en Holanda: Groningen, Leiden, Maastricht, Nijmegen y VU Amsterdam, y adicionalmente la Ghent University en Bélgica.
- La Universidad de Mc Master (Canadá), incluyendo un programa de pregrado y un programa de asistente médico, y la Canada Limerick University.
- El grupo Charite, con instituciones de Alemania y Austria: Berlin Regel, Berlin reform, Witten, Aachen, Bochum, LMU Munich, Koln, Munster, Hannover, Mannheim, Regensburg (Alemania); Graz, Innsbruck (Austria).
- El grupo NBME¹⁷⁹ 1: Barts, St. George's London, Leeds y Queens University, Belfast (UK).
- El grupo NBME 2: University of South Florida y Case Western Reserve University.

¹⁷⁹ National Board of Medical Examiners.

- Southern Illinois University, Vanderbilt, University of New Mexico, Penn State, Texas Tech, Medical College of Georgia, University of Minnesota, en EE.UU.
- University of Manchester School of Medicine (UK).
- Peninsula College of Medicine and Dentistry (UK).
- Swansea University, College of Medicine (UK).
- University of Tampere, Finlandia
- The College of Medicine at King Saud bin Abdulaziz University for Health Sciences (KSAU-HS), en Arabia Saudita.
- Karaganda State Medical University, en Kazajtan.
- Otago Medical School, en Nueva Zelanda.
- Sao Paulo Escola de Medicina, en Brasil.
- University of Indonesia, Medical School.
- Universidad Católica de Mozambique.
- Universidad de Pretoria, Sudáfrica.

O. Las retroalimentaciones desde 2010¹⁸⁰

O.1. Un banco de preguntas y de datos común a varias universidades

En 1995 se asociaron a Maastricht las facultades de medicina de otras dos universidades holandesas: Groningen y Nijmegen. Estas escuelas adoptaron las PdP y definieron que el proceso, incluyendo la creación del banco de preguntas y el análisis posterior de los resultados, resultara de una cooperación equilibrada y recíproca entre ellas (Van der Vleuten *et al.*, 2004). En 2006 se les unió la Facultad de Medicina de Leiden. Desde ese año las cuatro facultades imparten el mismo día la misma PdP de 200 ítems en formato de PSM, enterando 4 PdP por año. Los resultados de los 7.500 estudiantes son procesados en un banco de datos común.

O.2. Referencias normativas: los percentiles para posicionar los resultados

Los resultados de una PdP son presentados gráficamente según los percentiles en que se distribuyen los puntajes obtenidos por un grupo de estudiantes. Para una cohorte o para el total de la carrera, de una o de varias universidades, se informan siempre los mismos valores:

- Percentil 5: quinto puntaje más débil entre todos los resultados (inicio de la zona coloreada)
- Percentil 95: excelente (fin de la zona coloreada)
- Percentil 15: umbral de éxito

¹⁸⁰ Esta sección está inspirada en Muijtjens *et al.* (2010).

- Percentil 70: umbral de “distinción”
- Percentil 50 (o Media): ayuda a posicionar a cada estudiante sobre o bajo la Media

O.3. Retroalimentación individual

La Figura 3 muestra una hoja de retroalimentación para un estudiante en una PdP. Se observa el resultado de un grupo (cohorte), de acuerdo a los percentiles, para cada dimensión evaluada: salud preventiva, aspectos moleculares y celulares, sistema cardiovascular, etc.

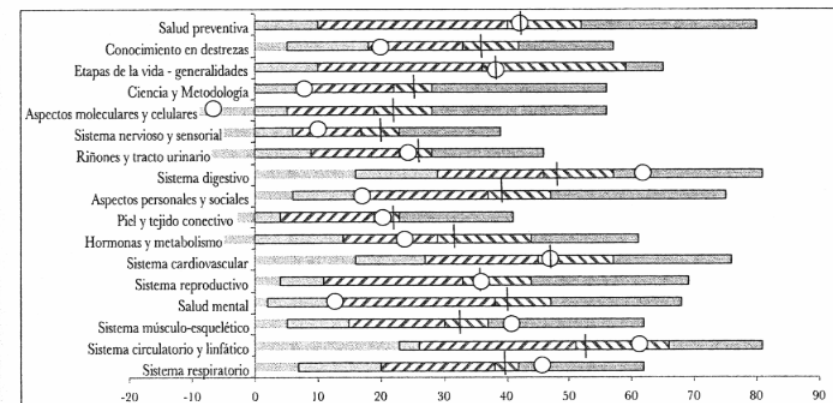


Figura 3: Percentiles para cada grupo de preguntas por un estudiante del año 3 en su doceava PdP

En la representación gráfica los círculos blancos indican la posición del estudiante. Las líneas verticales negras son las posiciones de las Medias (percentil 50), y las barras horizontales muestran la distribución de los resultados del grupo. El inicio de la barra marca el percentil 5; la zona gris se extiende entre el percentil 5 y 15; la zona con fondos de trama va del percentil 15 a 70; la zona gris de la derecha marca del percentil 70 a 95, donde termina la barra. Se ve que los puntajes de este estudiante están en 7 ocasiones bajo la Media; en otras 7 dimensiones están sobre la Media, y en 3 ocasiones están en la Media. En los *aspectos moleculares y celulares* este estudiante obtiene un puntaje por debajo del percentil 5.

O.4. Retroalimentación longitudinal

La Figura 4 indica la evolución del dominio de saberes de “la medicina general”, para el caso de un estudiante y su grupo, durante 3 años de formación. Se muestran los resultados de 12 PdP rendidas (1 a 12), junto a la predicción de los resultados en las

próximas dos PdP (13 y 14); predicciones que el estudiante debe intentar confirmar y mejorar, si son positivas, o contradecir si son negativas.

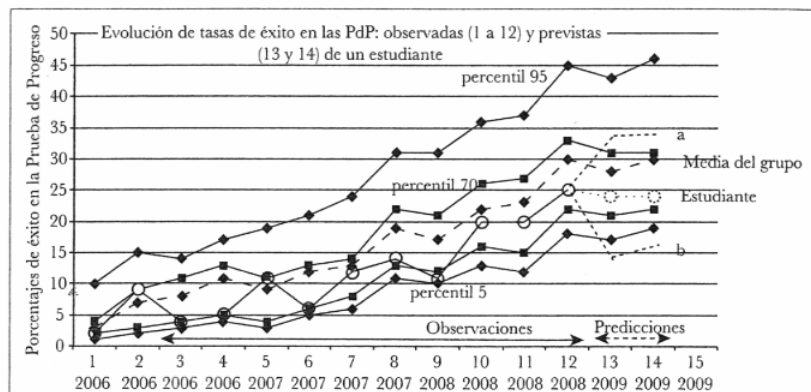


Figura 4: Resultados puntuales (círculos) de un estudiante hasta la PdP 12 (incluida), posicionados en los percentiles de las distribuciones de puntajes de los miembros de su grupo

Las curvas que enmarcan la del estudiante indican los mismos límites que las zonas coloreadas en la Figura 3: la primera curva indica el percentil 5, la segunda el 15, la curva de línea punteada muestra la Media del grupo, las 2 siguientes curvas marcan los percentiles 70 y 95. La sección más a la derecha del gráfico presenta las predicciones para las dos futuras PdP (13 y 14), mostrando:

- La Media estadísticamente más probable del grupo (M)
- El resultado estadísticamente más probable de este Estudiante (E)
- Los límites de confianza (P.05) del valor predicho para el estudiante (E), arriba (a) y abajo (b)

O.5. ProF: una forma de (auto) "Retroalimentación Electrónica Flexible"¹⁸¹ para los docentes

El principio de ProF consiste en permitir a cada docente y a cada estudiante recibir, por internet, las presentaciones que desee a partir de los datos acumulados, durante años, por las universidades de Maastricht, Groningen, Nijmegen y Leiden. Por ejemplo, un docente puede comparar los resultados de sus estudiantes con los de las otras universidades.

En un ejemplo que se muestra en los complementos multimedia de este libro se ve que los resultados de una universidad están por debajo de la Media de las otras 3

¹⁸¹ El título del artículo de Muijtjens *et al.* (2010) es "Flexible Electronic Feedback"; ProF es el programa informático que lo hace posible.

universidades hasta la PdP n° 11, pero estas diferencias desaparecen durante el tercer año (PdP 11 y 12).

El docente puede pedir al software ProF los mismos tipos de resultados para cada grupo de preguntas (para cada sistema, por ejemplo). Mediante "la inspección de varios tipos de puntajes (los "No sé", los errores), un profesor puede verificar si su disciplina tiene una evolución diferente de las otras, indicando, quizás, que algunos ítems son demasiado difíciles" (Muijtjens *et al.*, 2010, p. 494).

O.6. ProF: una forma de (auto) "Retroalimentación Electrónica Flexible" para los estudiantes

Añadimos la palabra "auto" para insistir sobre la contribución del beneficiario en la definición de la información que recibe. Un estudiante, normalmente, empieza "comparando sus resultados con los de su cohorte en su universidad. Cuando detecta una categoría débil, el paso siguiente puede ser inspeccionar la presentación longitudinal para verificar si la debilidad es ocasional o si es persistente longitudinalmente. La evolución puede servir como una señal adicional que indica que hay que mejorar. Via discusiones de estos materiales con otras personas pertinentes (posiblemente pares, mentores, supervisores) se puede diseñar el cómo mejorar (diferentes universidades utilizan diferentes sistemas de apoyo en el estudio; algunas utilizan los portafolios digitales)" (Muijtjens *et al.*, 2010, p. 494).

Conclusiones

La última sección (O. Retroalimentaciones) ilustra que una función sancionadora (certificativa) no excluye funciones formativas. Las PdP constituyen una oportunidad y una herramienta preciosa para investigaciones sobre el desarrollo del conocimiento y del dominio de un ámbito. Esto, además de las ventajas en términos de mejorar las dimensiones de validez de la evaluación, que se presentaron en la sección H.

Referencias

- ALBANO, M.-G., DECCACHE, A., GODIBIULE, A. Y D'IVERNOIS, J.F. (2009). Development of publications on patient education in chronic diseases from 1999 to 2009. *Educ. Ther. Patient*, 1 (2), 101-107.
- ANGELO, T. Y CROSS P. (1993). *From Classroom Assessment Techniques. A Handbook for College Teachers*, 2nd Ed. NY: Jossey Bass.
- ANGOFF, W. (1971). Scales, norms and equivalent scores. In R.L. Thorndike (Ed). *Educational measurement*. Washington DC: American Council on Education. P: 508-600.
- BARROWS, H.S. Y TAMBLYN, R.M. (1980). *Problem-Based Learning: An Approach to Medical Education*. New York: Springer Publishing Company.

- BLAKE, J.M., NORMAN, G.R., KEANE, D.R., MUELLER, C.B., CUNNINGTON, J. Y DIDYK, N. (1996). Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Academic Medicine*. 1996 Sept; 71(9):1002-7.
- COHEN-SCHOTANUS, J., VAN DER VLEUTEN, C. Y BEDNER, W. (1996). En betere cesuur bij testamens: de beste studenten als referentiepunt. In Ten Cate, T. et al; (Eds). *Gezond Onderwijs (Health education)*. Vol 5. Houtem/ Diegem, p. 83-88.
- FREEMAN, A., VAN DER VLEUTEN, C., NOUNS, Z. Y RICKETTS, C. (2010). Progress testing internationally. *Medical Teacher*, 32, 451-455.
- GAGNÉ, R., BRIGGS, L. Y WAGER, W. (1992). *Principles of Instructional Design*. Forth Worth: Harcourt Brace Jovanovich.
- LECLERCQ, D. Y VAN DER VLEUTEN, C. (1998). PBL -Problem Based Learning ou APP- Apprentissage Par Problèmes, chapitre 8 de D. Leclercq (Ed.), *Pour une pédagogie universitaire de qualité*, Sprimont: Mardaga, pp. 187-205.
- LECLERCQ, D. (2006). L'évolution des QCM. In G. Figari et L. Mottier-Lopez. *Recherches sur l'évaluation en Education*. Paris: L'Harmattan, p. 139-146.
- MUIJTJENS A., HOGENBOOM, R., VERWIJNEN, G. Y VAN DER VLEUTEN, C. (1998). Relative or absolute standards in assessing medical knowledge using progress test. *Advances in Health Sciences Education*. 3, p. 81-87.
- MUIJTJENS, A., TIMMERMANS, I., DONKERS, J., PEPEKAZMP, R., MEDEMA, H., COHEN-SCHOTANUS, J., THOBEN, A., WENINK, A. Y VAN DER VLEUTEN, C. (2010). Flexible electronic feedback using the virtues of progress testing. *Medical Teacher*. 32, p. 491-496.
- RADEMAKERS, J., TEN CATE, T. Y BAR, P. (2005). Progress Testing with short answer questions. *Medical Teacher*, 27: p. 578-582.
- SCHMIDT, H.G. (1993). "Foundations of Problem-Based Learning: Some Explanatory Notes," *Medical Education*, 27.
- SCHWARTZ, P. (2009). A most unusual examination: Progress testing at the University of Otago Medical School. <http://akoaooteaaroa.ac.nz/node/3496> (consultado el 8 de enero de 2014).
- VAN DER VLEUTEN, C. Y WYNEN, W. (1990). *Problem-based learning: perspective from the Maastricht experience*. Amsterdam: Thesis.
- VAN DER VLEUTEN, C. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education*. 1, p. 46-47.
- VAN DER VLEUTEN, C., VERWIJNEN G. Y WIJNEN, W. (1996). Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical teacher*. Vol. 8, 2, P. 103-109.
- VAN DER VLEUTEN, C. (1998). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education*. 1, p. 46-67.
- VAN DER VLEUTEN, C. SCHUWIRTH, L., MUIJTJENS, A., THOBEN, A., COHEN-SCHOTANUS, J. Y VAN BOVEN, C. (2004). Cross institutional collaboration in assessment: A case on progress testing. *Medical teacher*. 26, p. 719-725.
- VERHOEVEN, B., VERWIJNEN, G., SCHERPOBIER, A., SCHUWIRTH, L. Y VAN DER VLEUTEN, C. (1998). Quality assurance in test construction. The approach of a multidisciplinary central test committee. In *Education for Health*. 12, pp. 49-60.
- VERHOEVEN, B. (2003). *Progress testing. The utility of an assessment concept*. PhD dissertation. Maastricht University: Faculty of Medicine.
- WILLOUGHBY, L., ARNOLD, L., CALKINS, V. Y NEADERHISER, M. (1991). The Concurrent and Predictive Validity of the Quarterly Profile Examination. *Educational and Psychological Measurement*, 51: p. 691-697.

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>