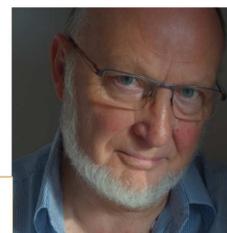


www.universitaria.cl

Dieudo LECLERCQ



Álvaro CABRERA MARAY



UNIVERSIDAD
DE CHILE



Directores de la publicación:

Dieudonné Leclercq
Universidad de Liège (ULg)

Álvaro Cabrera Maray
Universidad de Chile (UCH)

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

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

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

Dieudonné Leclercq

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

Álvaro Cabrera Maray

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

Contenidos del libro IDEAS:

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

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

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

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

La lista de los capítulos y el resumen de cada uno

aparece a continuación después de este capítulo.

P A R T E

3

Evaluación
de saberes

o

recursos
aislados

CAPÍTULO XIII

Las Preguntas de Selección Múltiple (PSMs): del currículum oculto a la vigilancia cognitiva

DIEUDONNÉ LECLERCO Y ÁLVARO CABRERA

Introducción a los capítulos 13 y 14

En el año 2001 Case y Swanson dedicaron un libro a la redacción de preguntas para el famoso test USMLE⁷⁶, gestionado por el NBME⁷⁷ estadounidense, y cuyo objetivo es evaluar a los médicos en los EE.UU. Los autores explicitan lo que los motivó en primer lugar a realizar este trabajo: "...las pruebas tienen una influencia importante sobre el aprendizaje de los estudiantes". Case y Swanson postulan (p. 9) que "Una primera meta de un examen es comunicar lo que usted considera como importante" y precisan (p. 13) que "la pregunta debe tratar de un contenido importante". Este problema de contenido no será abordado en los dos capítulos siguientes, pero al respecto existen fórmulas para elegir contenidos de preguntas o problemas para la evaluación, según su importancia en la vida real (ver Capítulo 8, sección B.5).

Asumir la fuerte influencia que ejercen las consecuencias en el aprendizaje de los estudiantes implica preocuparse de los mensajes que se vehiculan a través de los métodos de evaluación y los tipos de preguntas que se formulan, en términos de *epistemología* y de *hábitos mentales*. No querer ver estos retos es continuar ignorando el concepto de *currículum oculto* y dejar escapar la posibilidad de transformar y mejorar (a veces sustancialmente) las preguntas que utilizamos en nuestros instrumentos de evaluación.

Uno de los tipos de pregunta que con mayor frecuencia se usa en las pruebas desarrolladas y aplicadas en nuestras universidades son las llamadas PSM (*MCQ = multiple choice questions*). Utilizaremos la sigla PSM pues es la que con mayor frecuencia se encuentra en la literatura en español, aun cuando este tipo de preguntas aparezca referido en más de una forma en publicaciones y sobre todo en textos de trabajo de las universidades. Esta dificultad deriva del doble significado de la palabra "choice", que puede implicar tanto la opción que se le presenta al estudiante como la selección de una(s) de ellas. Así, la expresión que más se encuentra es "preguntas de selección múltiple", aunque no sea del todo exacta ya que muchas veces el estudiante realiza una selección "única" y no "múltiple". "Preguntas con soluciones múltiples" es una expresión que mantiene la sigla, pero tampoco es exacta pues muchas veces (en PSM

⁷⁶ The United States Medical Licensing Examination® (USMLE®)

⁷⁷ National Board of Medical Examiners, cuya misión es proteger la salud del público mediante la evaluación actualizada de los profesionales de la salud, e incluye investigación en evaluación y el desarrollo de instrumentos de evaluación.

clásicas) hay solo una solución correcta. Otro grupo de documentos usa la expresión “preguntas de opción múltiple” –POM.

Las PSM potencialmente pueden ser un peligro para el aprendizaje o una gran herramienta, y es posible para el docente realizar esta transformación. El Capítulo 13 estará dedicado a esto. El resto del libro de Case y Swanson (2001) está dedicado a que “los ítems estén bien estructurados, de modo de evitar defectos que beneficien a los estudiantes familiarizados con los tests” (p. 13). Este asunto se desarrollará en el Capítulo 14: “Reglas de redacción de las PSMs”.

En la siguiente sección se hablará del “currículum oculto”, pero hay otras denominaciones como currículum “escondido”, o “latente”, o “encubierto”. Se puede ver una perspectiva comparada en Kentli (2009).

También se usará la palabra “nocosomial” para describir los efectos, en los estudiantes, de algunos tipos de evaluaciones. Se llama “nocosomiales” a las enfermedades generadas dentro del y por el sistema de salud (el hospital en particular), como por ejemplo una infección intrahospitalaria.

A. El currículum oculto

A.1. ¿Qué es el currículum oculto?

Se puede definir el currículum oculto como *lo que nadie enseña, pero que todos aprenden*. El sociólogo francés Emile Durkheim (1925) insistió en que la escuela enseña más que aquello que es oficial, aún no conscientemente, porque aparece como un *consensus*. Ph. Jackson (1968) enumera estas enseñanzas o condicionamientos: “esperar en silencio, gesticular poco, esforzarse por terminar el trabajo, cooperar, respetar a sus docentes y a sus pares, ser puntual, comportarse de forma civilizada”. Sobre estas reglas se sostienen las actividades rutinarias en la escuela (Giroux, 2001; Torres, 1998). Díaz Barriga (2006) hace notar que estas reglas *tienen una estrecha relación con cómo se actúa, más que con la información de lo que se dice*. Dreeben (1967) argumenta que las estructuras (mentales) de la escuela enseñan sobre la autoridad y sirven a los intereses de los poderosos. Se puede decir también, como Bourdieu, que sirven a la “reproducción social”. Y al mismo tiempo es innegable que algunas de las características enunciadas por Jackson permiten el funcionamiento de la escuela. Como Lynch (1989), se debe constatar que el currículum oculto tiene una serie de limitaciones, por sus consecuencias para la institución y los estudiantes:

- (1) El currículum oculto encarna la teoría del consenso que refuerza la estabilidad y rechaza cambios que pueden resultar de conflictos.
- (2) La escuela transmite las normas y las representaciones como si existieran sin problemas y fueran indiscutibles, de modo que trata a los estudiantes como receptores pasivos de la socialización, sin la habilidad de construir sentidos por sí mismos.

A.2. ¿Qué se aprende de modo oculto en la escuela? ¿Cuáles normas y representaciones?

Al hablar específicamente de evaluación de los aprendizajes, el currículum oculto nos enseña:

- Que se deben contestar todas las preguntas (aunque existen preguntas que no deben ser contestadas, o porque son absurdas, o porque son demasiado invasivas, o porque no tenemos suficiente información para dar una respuesta).
- Que las preguntas están bien formuladas, siempre.
- Que siempre hay una respuesta correcta y solo UNA (a pesar de que en la realidad existen preguntas con varias respuestas correctas, aunque diferentes).

A.3. ¿Cómo participan las preguntas en el currículum oculto? ¿Cuáles son sus efectos nocosomiales?

Las pruebas y los tipos de preguntas utilizadas mayoritariamente en el sistema plantean dos problemas de validez consecucional (ver Capítulo 4):

- (1) el riesgo de adormecer la espontaneidad (o vigilancia cognitiva) de los estudiantes, habituándoles a esperar que *otros* planteen los problemas, que los planteen bien, que imaginen y propongan soluciones, que simplifiquen a tal nivel el problema que sea suficiente responder Sí o No, aun cuando en muchas situaciones de la vida primero debamos darnos cuenta de que hay un problema, o notar que está mal planteado, o se deban imaginar soluciones nuevas.
- (2) el riesgo de habituar a los estudiantes a adoptar modos de pensamiento y de razonamiento deformados, torcidos por el modo de preguntar. Por ejemplo, las preguntas Verdadero-Falso pueden vehicular la idea de que en el conocimiento todo es blanco o negro, lo que Perry (1985) llama “el nivel 1 de desarrollo epistemológico: el dualismo”. Otro ejemplo: las PSMs del tipo clásico (1 respuesta, 1 de las *k* soluciones explícitas es correcta, consigna A, ver sección F2) pueden inducir en los estudiantes estrategias como (a) leer las soluciones antes que la pregunta, (b) no preocuparse de verificar si la pregunta tiene sentido (porque han recibido garantía de que existe una respuesta correcta entre las soluciones propuestas), (c) no preocuparse de determinar si existen suficientes datos (información) como para responder la pregunta (por la misma razón).

A.4. A mayor Requerimiento (de la pregunta) menor Espontaneidad (del estudiante)

La Figura 1 representa gráficamente varios conceptos:

- (1) Un “ *eje bipolar Requerimiento – Espontaneidad*”, donde el primero es el nivel de descripción de la pregunta que se hace al estudiante, siendo el máximo las PVF (Pre-

guntas Verdadero-Falso) que simplifican las variables hasta un Sí o No, en el extremo izquierdo del eje. En la posición 0 (cero) se ubica la ausencia de pregunta, pero con un problema presente del cual el estudiante debe darse cuenta. En el plano derecho se ubican posiciones positivas de la espontaneidad o iniciativa, avanzando hacia la incerteza que demanda espontaneidad y vigilancia de parte del estudiante, porque el problema no está presente en la actualidad (se planteará en el futuro) o porque el requerimiento consiste en atraer al estudiante en la dirección de malas decisiones.

- (2) Un *punto de aplicación* de cada tipo de pregunta sobre este eje.
- (3) las *intensidades* de los efectos que estas preguntas producen en términos de (a) esfuerzo para producir una respuesta, simbolizado con las amplitudes de las flechas enteras \Rightarrow (b) las amplitudes del peligro de adormecer la vigilancia cognitiva (flechas punteadas) \leftarrow
- (4) las significaciones de estos retos negativos (flechas punteadas) o positivos (flechas enteras) en cada tipo de pregunta, comentados en la columna de la derecha.

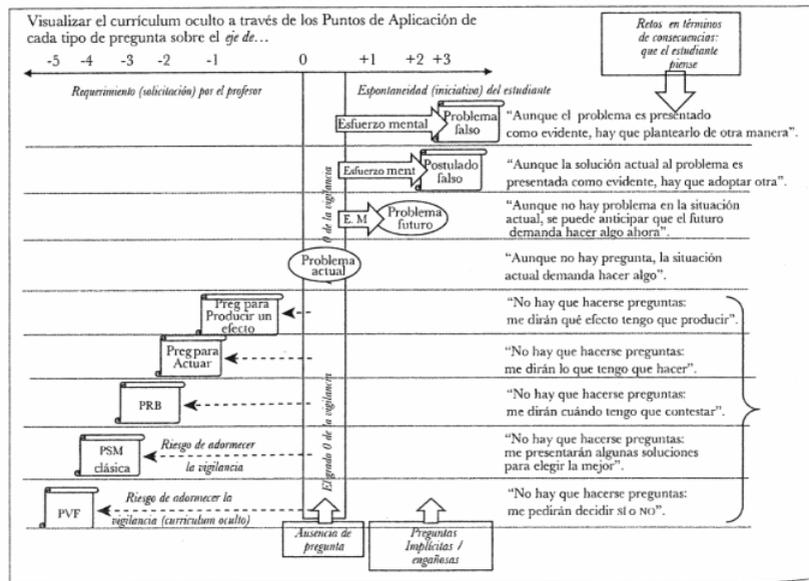


Figura 1: Taxonomía de los vectores opuestos "Espontaneidad/Vigilancia" y "Requerimiento"

Los efectos opuestos del requerimiento y de la espontaneidad son presentados como *vectores* que, como en física, tienen dirección, punto de aplicación y amplitud o intensidad.

A.5. ¿Cuándo es importante la diferencia entre vigilancia y requerimiento?

En algunas profesiones, por ejemplo los expertos médicos que deben juzgar el daño sufrido por un paciente, se hace la distinción entre (a) una respuesta espontánea ("Doctor, me duele cuando...") y (b) una respuesta como reacción a un requerimiento ("Ya que usted me lo pregunta, doctor, le contesto que me duele").

En la vida común y corriente se producen a menudo situaciones de los niveles 0, -1 y -2 de requerimiento (o de los niveles 0, +1 y +2 de la vigilancia / espontaneidad): es decir, cuando se debe detectar un problema en una situación actual; o anticipar un problema futuro; o detectar un postulado falso. Esto último ocurre cuando (a) el problema es planteado como si todo fuera normal, aunque en realidad no lo es, o (b) cuando la presentación del problema contiene un error, ya sea (b.1) con la intención de hacer daño (pues quien formula el problema sabe que abusa, que manipula, que engaña, que es pernicioso), o (b.2) por ignorancia.

La falta de previsión (es decir, de capacidad para detectar problemas futuros) nos hace hoy sufrir numerosas consecuencias de situaciones que aún no han atraído la atención que merecen. Campesinos de Bangladesh sufren el hundimiento de sus tierras porque otros pueblos del Himalaya cortaron demasiados árboles miles de kilómetros más al norte. El pulmón natural de la selva de Brasil se debilita a razón de segundos porque se ha sobreexplotado la tala de árboles. Accidentes nucleares ocurren porque las precauciones han resultado insuficientes.

A.6. ¿Y durante las evaluaciones?

Durante la formación, estas situaciones de niveles 0, -1 y -2 del requerimiento (o 0, +1 y +2 de la espontaneidad / vigilancia) pueden ser simuladas en juegos de roles, como en ECOES⁷⁸. Es lo que con frecuencia los estudiantes llaman "trampas", y los evaluadores llaman "falsos requerimientos" o "preguntas implícitas o tácitas" (Leclercq, 1986) o *No Warning questions*.

- Ej1: En un ECOE de *farmacia*: los examinadores han introducido un error en la tasa (proporción) de un medicamento en la receta de un médico que el cliente entrega al farmacéutico. Esta situación se ubica en -2 (requerimiento) / +2 (vigilancia), sobre el eje, pues una receta (prescripción médica), en principio, está correcta y el farmacéutico debe ejecutarla. Además esta situación coloca al farmacéutico en una posición socialmente difícil: deber telefonar al médico para verificar si se ha equivocado.
- Ej2: En un ECOE de ciencias de *hospital*: la sonda del catéter está doblada, y la enfermera debe darse cuenta de eso cuando visita a la persona enferma *por otro motivo* (ubicación sobre el eje: -1 (requerimiento) / +1 (vigilancia), porque el foco de la atención está puesto en otro asunto).

⁷⁸ Exámenes Clínicos Objetivos Estructurados (ver Capítulo 8).

Ej3: En un ECOE de *psicología*, en el papel del padre que consulta quejándose del comportamiento de su niña, los examinadores han introducido comportamientos narrados de este padre de los cuales el estudiante-psicólogo debe notar que son tóxicos para la niña y que existe hostigamiento (ubicación sobre el eje: -2 (req.) / +2 (vig.), porque el caso es presentado como si el comportamiento de la niña fuera el inadecuado). Además este caso pone al psicólogo en una situación difícil, porque el padre es su cliente y, a pesar de esto, debe denunciarlo a la policía.

A.7. Las PSM: potencialmente una gran herramienta o un peligro

Cuando los estudiantes entran a la universidad ya han vivido el currículum oculto durante al menos 12 años⁷⁹. Si queremos compensar sus efectos debemos actuar rápido. Tenemos solo algunos semestres para despertar la vigilancia cognitiva y promover la espontaneidad, de modo que es necesario *sobre-aprender* (en inglés *overlearning*; en francés: *sur-apprentissage-sur-entraînement*). Leclercq (1986, p. 127-144) propuso una técnica que permite entrenar la vigilancia sistemáticamente (es decir, numerosas veces), y es aplicable tanto en cursos pequeños como en grupos masivos: el uso de Soluciones Generales Implícitas (SGI) en las Preguntas de Respuesta Breve (PRB) y Preguntas de Selección Múltiple (PSM).

Por ejemplo:

- | |
|--|
| 6. Ninguna (de las soluciones propuestas) |
| 7. Todas (las soluciones propuestas) |
| 8. Falta(n) datos en el enunciado (o tronco o introducción) como para contestar (elegir una sola solución) conforme a la consigna. |
| 9. Absurdo en el enunciado (o tronco o introducción) ¡Prioridad! porque no hay que contestar. |

Figura 2: Las 4 SGI propuestas por Leclercq (1986, p. 127-145)

Cuando la Pregunta es de Respuesta Breve (PRB), las SGI “Ninguna” y “Todas” no están en uso. Con Preguntas Verdadero-Falso (PVR) tampoco la SGI “Absurdo” está en uso.

Estas Soluciones son *Generales* porque se aplican a todas las preguntas de un test. Son *Implícitas* porque no se repiten en cada pregunta (quien rinde el test debe recordar que existen como posibilidades de respuesta).

A continuación se muestran 5 ejemplos de PSM + SGI, siguiendo la consigna de la Figura 2:

⁷⁹ En Chile la educación básica y media suman 12 años obligatorios para todos los niños y jóvenes; dos niveles anteriores (educación pre-escolar: kinder y pre-kinder) tienen amplia cobertura, pero es discutible que refuercen los mismos aspectos (negativos) del currículum oculto mencionados más arriba. En Bélgica la escolaridad obligatoria previa a la educación superior alcanza los 15 años.

Tabla 1: Ejemplos de preguntas con SGI y sus respuestas correctas

P1. ¿Cuál es la capital de Francia? 1. Lyon 2. Niza 3. Marsella
P2. ¿Qué edad tenía Neruda? 1. 23 años 2. 41 años 3. 75 años
P3. ¿Cuál de estas regiones es Parte de Gran Bretaña? 1. Inglaterra 2. País de Gales 3. Escocia
P4. ¿En qué año se encontraron O'Higgins y Julio César? 1. En 52 aC 2. En 1800 3. En 1815
P5. ¿Cuál es la capital de Perú? 1. Tokio 2. Lima 3. Roma

Las respuestas correctas son:

Pregunta 1	6. Ninguna, porque la respuesta correcta es París.
Pregunta 2	8. Falta(n) datos, porque no se precisa en cuál fecha. Si es cuando inicia su carrera como cónsul la respuesta es 23 años; cuando fue electo senador tenía 41 años; a los 67 años recibió el Premio Nobel de Literatura. No se puede contestar “Todas”, pues murió a los 69 años, de modo que la solución ‘75 años’ es falsa.
Pregunta 3	7. Todas (las tres soluciones son correctas sin excluirse, lo que no es el caso en P2).
Pregunta 4	9. Absurdo, porque los periodos de vida de los dos personajes no se encuentran.
Pregunta 5	3. Lima, porque preguntas con una respuesta correcta, presentada entre las opciones clásicas, deben seguir existiendo en un test con SGI.

Han existido y perduran muchos debates y muchas experiencias sobre las Soluciones Generales Implícitas “Ninguna” (en inglés *NOTA*⁸⁰) y “Todas” (*AOTA*⁸¹). Huang *et al.* (2007) mencionan que autores como Osterlind (1989), Ebel y Frisbie (1991), y Frary (1989) observaron que *AOTA* es una consigna que discrimina mejor que la consigna simple.

Rodríguez (1997, p. 18), refiriéndose a investigaciones sobre la SGI *NOTA*, señala que (1) entre 19 investigaciones, en 16 de ellas la SGI *NOTA* aumenta la dificultad (nunca la baja), y que (2) entre 13 investigaciones, en 7 de ellas no modifica la discriminación, en 5 la baja y en una la aumenta.

La crítica más importante y que con mayor frecuencia se escucha es que “no sabemos si el estudiante conoce la respuesta correcta cuando elige la solución ‘Ninguna’”. Podemos responder que si no usamos los grados de certeza (ver Capítulo 16), tampoco sabemos si un estudiante conoce la respuesta correcta cuando elige la respuesta correcta explícita.

Una fórmula para asegurarse consiste en pedir al estudiante que escriba *una justificación* cada vez que contesta “Ninguna” o “Faltan datos” o “Absurdo”⁸², justificación que el docente tendrá que leer, aumentando su trabajo. El debate (ver Capítulo 9) es otra forma de resolver este problema.

⁸⁰ None of the above (*NOTA*).

⁸¹ All of the above (*AOTA*).

⁸² Lo que sirve también al desarrollo del espíritu crítico.

B. Definir las Preguntas de Selección Múltiple (PSM)

Puede parecer extraño haber evocado tipos poco comunes de PSMs antes de dar una definición del concepto. La intención era preparar esta(s) definición(es), porque *existen muchos tipos de PSM: ¡Es toda una familia!*

B.1. Definiciones demasiado vagas

En la literatura⁸³, la definición de Preguntas de Selección Múltiple (PSM) a menudo se limita a un modo de evaluación en el cual los respondientes deben elegir la mejor de entre varias soluciones listadas. Ya hemos visto que hay otras posibilidades, y veremos otras más y con mayor detalle. Los componentes de una PSM (nótese el singular) comúnmente son definidos como (1) un cuerpo o tronco, y (2) soluciones entre las cuales hay una correcta, siendo las demás distractores.

Aunque correctas, estas definiciones están incompletas, lo que resulta en ambigüedades. Por ejemplo, una PSM puede estar mal formulada si debe respetar una consigna A, pero bien formulada si debe respetar una consigna B. Por esta razón, la consigna debe ser parte de la definición de una PSM. Lo mismo se puede decir del sistema de cotejo.

B.2. Una definición a geometría variable⁸⁴ con ocho componentes de las PSMs⁸⁵:

Tabla 2: Definición a geometría variable de los ocho componentes de una PSM comparada con un árbol invertido

Las PSM se componen de ...		Comparación con un árbol invertido
... los objetivos de la prueba completa	¿POR QUÉ evaluar? ¿QUÉ evaluar (cuáles dimensiones)? ¿Con qué nivel de capacidad diagnóstica?	El bosque
... los objetivos de la PSM	¿Cuáles partes del (los) objetivo(s) de la prueba contribuye a evaluar?	Su contribución al bosque (consistencia interna de la prueba)
... una consigna	... que informa la forma de responder el tipo de preguntas que siguen (como el número de soluciones correctas, de respuestas esperadas y de consecuencias en términos de lista de cotejo).	Las raíces del árbol
... un cuerpo o tronco ⁸⁶	... el que debe presentar una pregunta (no bastan afirmaciones que no formulan una pregunta).	Tronco
... algunas soluciones propuestas	Que pueden ser apropiadas a una sola pregunta o a varias preguntas de un test, de modo explícito o implícito (por ejemplo, las Soluciones Generales Implícitas)	Ramas
... la(s) respuesta(s) correctas	... que no deben depender de la persona que contesta. Y los grados de error de las opciones incorrectas (si la consigna anuncia que hay errores de varios tipos -como fatales vs benignos, por ejemplo)	Rama(s) fecunda(s) Ramas infecundas
... las consecuencias en términos de lista de cotejo	Esta lista, que puede llamarse tabla o escala, al menos debe precisar las TC (Tarifas en caso de Respuesta Correcta), TI (Tarifas en caso de Respuesta Incorrecta) y TOM (Tarifa en caso de omisión)	Frutos comestibles y frutos no comestibles (punibles)
... las características edumétricas.	... como las popularidades de las soluciones, o sus índices de discriminación (conocidos luego de experimentación).	Validación por nutricionistas

⁸³ Las primeras PSM fueron aplicadas a gran escala en 1916 en los "Otis Alpha Army Tests" para evaluar la inteligencia de los reclutas y militares de EE.UU.

⁸⁴ Se puede limitar a la columna de la izquierda; se inspira en la definición de Leclercq (1986, p. 13)

⁸⁵ Nótese el plural: hay muchos tipos de PSMs.

⁸⁶ En inglés, "Lead in" o "stem".

C. Modelos de procesos mentales y formas de asignar puntaje en las Preguntas de Selección Múltiple (PSM)

Bruce Choppin (1970) describió tres modelos de la actividad mental de una persona contestando una PSM simple o clásica, es decir, con una solución correcta entre k presentadas, y vinculó estos modelos con las instrucciones (consignas), con los tipos de PSM, y con las listas de cotejo.

C.1. El modelo 1 postula que

- | |
|--|
| (1) cuando sabe la respuesta, la persona la elige, y |
| (2) cuando no la sabe, elige al azar entre las k soluciones. |

De este modelo sigue el método de asignar puntaje que se llama "la corrección por adivinación clásica" (*the classical correction for guessing*), que consiste en la siguiente escala de cotejo:

Por cada respuesta correcta, $TC = +1$ punto.

Por cada omisión, $TOM = 0$ punto.

Por cada respuesta incorrecta, $TI = -1 / (k-1)$ punto, siendo k el número de soluciones, de modo que:

Tabla 3: Las tarifas de la corrección *for guessing* clásica

Si $k =$	2	3	4	5	6	7	8	9
$TI =$	-1	-0,5	-0,33	-0,25	-0,20	-0,16	-0,14	-0,125

Se puede verificar que un estudiante que contestara completamente al azar obtendría un *puntaje esperado* (el más probable) de 0, como ilustraremos en un ejemplo con una prueba de 20 PSM clásicas con 4 soluciones (una correcta). Si un estudiante elige 20 veces al azar, como tiene 1/4 de chances de elegir la solución correcta, obtendrá 5 correctas y 15 incorrectas. Ganará 5 veces 1 punto y 15 veces $-0,33$ puntos. Su nota matemáticamente esperada (NME) para el total de la prueba equivale a: $NME = 5 + (15 \times -0,33) = 5 - 5 = 0$.

La fórmula genérica es $NME = (TC \times p) + (TI \times q)$

donde p es la probabilidad de éxito y q la probabilidad de fracaso ($q = 1-p$).

No contestar resulta en la misma nota (0) que (de una manera probabilística⁸⁷) contestar al azar.

⁸⁷ "De una manera probabilística" significa que a veces tendrá una nota superior al azar y en otras ocasiones inferior, pero el promedio de esos casos (la esperanza matemática) es 0.

C.2. El modelo 2 postula que

- (1) cuando sabe, el/la estudiante elige la solución correcta, y
- (2) cuando no la sabe, él/ella empieza eliminando las soluciones incorrectas, y
- (3) solo después, elige entre las soluciones que quedan.

De este modelo surgen consignas (instrucciones) que piden al estudiante identificar las soluciones incorrectas (y no las correctas) y eliminarlas. Coombs *et al.* (1956) propusieron esta consigna y la escala de cotejo que deriva: +1 punto por cada solución errónea eliminada, y +1-k puntos por la eliminación de la solución correcta. De este modo, si hay 4 soluciones en una PSM clásica, la nota a esta pregunta puede variar entre +3 (las 3 soluciones falsas han sido identificadas) y -3 (ninguna de las 3 soluciones erróneas ha sido identificada, pero la solución correcta ha sido eliminada). Entre esos dos extremos las notas pueden ser -2, -1, 0, +1, +2. Es decir, 7 posibilidades en lugar de 3. Otra vez se puede verificar que un estudiante que contesta al azar o que elimina todas las soluciones o que no contesta recibe la misma nota: 0. *Esta consigna reconoce que el conocimiento parcial existe.*

C.3. El modelo 3

Este modelo no hace la distinción “cuando el estudiante conoce/sabe” y “cuando no conoce/no sabe”. Este modelo es el de los investigadores, entre los cuales nos contamos, que piensan que hay un *continuum* de conocimiento, y que, entre los dos casos extremos (el del conocimiento total y el de la ignorancia total) existen niveles de conocimiento parcial (De Finetti, 1965). Este modelo postula que,

- cuando debe contestar a una pregunta (ya sea una PSM o una Pregunta de Respuesta Breve - PRB), una persona
- (1) empieza evocando varias soluciones posibles
 - (2) las ordena de la más probable a la menos probable (subjectivamente)
 - (3) si es forzada (por la consigna) a entregar una sola respuesta, entrega la que tiene la probabilidad subjetiva más alta⁸⁷.

Este modelo 3 postula que hay casos en los cuales la respuesta entregada tiene una probabilidad (subjetiva) alta y otros en que esta probabilidad es baja. Se desprende la idea de pedir al estudiante que entregue, además de su respuesta, un grado de certeza en la misma (ver Capítulo 16).

C.4. Un cuarto modelo⁸⁹

Si le es permitido ser más sutil, el estudiante es capaz de producir respuestas más diagnósticas y más fiables. Es lo que observan Evans y Misfeldt (1974) que permi-

⁸⁸ Esto se puede verificar fácilmente de forma experimental.

⁸⁹ Añadido por Leclercq.

ten “Contestar Hasta la Correcta” (CHC)⁹⁰: cuando es informado/a que su primera respuesta es incorrecta, el/la estudiante puede dar una segunda (*idem* con una tercera, etc.), y recibir un puntaje parcial. Hanna (1974) llama a este proceso CHC *total feedback*, oponiéndolo al proceso de *partial feedback* que se limita a informar al estudiante si su primera respuesta es correcta o no. Muchos autores comparten la idea de un cuarto modelo. Por ejemplo, Sabers y Feldt (1968, p. 256) cuando escriben: “*Mala información y conocimiento parcial pueden ser importantes factores en la respuesta del estudiante*”, como una de las conclusiones de su estudio experimental que recomienda insistir a los estudiantes para que contesten a cada una de las preguntas (al contrario de la consigna de *correction for guessing clásica*, que promueve la omisión pues penaliza el error). Leclercq (2013) –ver capítulo 16– recomienda contestar todas las preguntas, pero utilizando grados de certeza: “Cuando usted no sepa, entregue la certeza 0% y conteste algo”.

C.5. Discusión sobre la asignación de puntajes

Aunque la CFG (*correction for guessing clásica*) es con frecuencia presentada como LA fórmula para asignar puntajes con las PSMs⁹¹, consideramos que no es la más apropiada porque:

- (1) está basada en un modelo teórico débil (ver arriba),
- (2) no tiene valor formativo alguno para el/la estudiante,
- (3) no informa al docente sobre las capacidades del estudiante,
- (4) funciona de forma ciega, postulando que TODAS las respuestas correctas son dadas al azar.

Por todas estas razones debe ser remplazada por (o combinada con) un sistema que utilice grados de certeza y asigne puntaje en forma acorde (ver capítulos 16 y 17).

D. Peligros y desventajas de todas la Preguntas de Selección Múltiple (PSM)

D.1. El peligro de memorizar los distractores como si fueran las respuestas correctas

Este peligro es relativo a la *calidad (validez) Consecuencial* de las PSMs. Ya en 1926 Remmers y Remmers habían observado que la rendición de pruebas verdadero-falso comporta el riesgo de que los estudiantes reconozcan como correctas proposiciones falsas

⁹⁰ Answer Until Correct (AUC).

⁹¹ De hecho se usa en la Prueba de Selección Universitaria (PSU) de nivel nacional en Chile.

que habían aparecido en una prueba VF anterior⁹². Llamaron a esto el *Efecto de sugestión negativa*. Como lo denunció Skinner (1958) “un día, para contestar a una pregunta, el estudiante extraerá de su memoria errónea una solución falsa que ha encontrado en una PSM”. En un provocador artículo (*Las PSM como instrumento para perpetuar falsos conceptos*), Preston (1965) verificó experimentalmente esta hipótesis, con el siguiente dispositivo experimental compuesto por dos grupos:

Tabla 4: La experiencia de Preston (1965)

	Grupo 1 (control)	Grupo 2 (experimental)
Fase 1	Preguntas de Respuestas Breves	Preguntas de Respuestas Breves
Fase 2		PSM sobre el mismo contenido
Fase 3	Mismas Preguntas de Respuestas Breves	Mismas Preguntas de Respuestas Breves

Preston observó que, en el grupo 2, en la fase 3 aparecieron errores que no estaban presentes en la fase 1 y que provinieron de la fase 2 (pues estos errores eran soluciones falsas –distractores– de las PSMs). Estos errores no aparecieron en el grupo 1. Karraker (1967) repitió esta experiencia con una diferencia crucial: inmediatamente después de la fase 2 entregó las respuestas correctas a los estudiantes, lo que resultó en la ausencia (en el grupo 2) de las respuestas erróneas originadas por los distractores de las PSMs.

Esta demostración⁹³ tiene una consecuencia extremadamente importante:

los docentes que utilizan PSMs deben entregar las soluciones correctas a los estudiantes después de la prueba.

(de no hacerlo estarán directamente causando un daño a algunos estudiantes y contribuyendo a perpetuar falsos conceptos⁹⁴).

Hasta este momento, muchos docentes no conocían estas investigaciones ni sus conclusiones e implicancias, pero ahora han *perdido la inocencia* en lo relativo al tema de las PSMs y su debida retroalimentación. Esta expresión (perder la inocencia) la formulaba Benjamin Bloom (1972) para indicar que quien se ha formado en pedagogía no puede seguir justificándose en que “no sabía”.

⁹² Jacoby y Hollingshead (1990) observaron el mismo efecto con docentes y errores de ortografía presentes en las hojas de respuesta de sus estudiantes.

⁹³ Repetida en Butler y Roediger (2007).

⁹⁴ Lo que violaría el principio de que (para lograr resultados destacados) “el profesor puede desplegar cualquier método de enseñanza que le parezca apropiado, siempre y cuando no provoque un daño a los estudiantes (o a cualquier otro) en el proceso” (Bain, 2004, p. 5).

D.2. Las otras desventajas de las Preguntas de Selección Múltiple (PSM)

- (1) *El trabajo de redacción* de los ítems y las opciones de respuesta son demandantes en tiempo, como se muestra en este capítulo y el siguiente. Para los distractores, el ideal es disponer de una lista de errores frecuentemente cometidos por los estudiantes y que sean reveladores de procesos mentales erróneos, de modo que estos errores sirvan como distractores diagnósticos. Una forma de hacerlo es presentar, como una evaluación con intención formativa, preguntas que demanden una respuesta breve (abierta); estas preguntas serán luego convertidas en PSMs usando como distractores algunos de los errores cometidos por los estudiantes. En esta tarea los docentes que tienen una larga experiencia en la enseñanza de un contenido pueden ayudar a los debutantes, compartiendo con ellos su banco de preguntas.
- (2) Es relevante recabar la *opinión de los pares* antes de presentar las PSMs en un examen, lo que también demanda tiempo. Los colegas pueden aconsejar sobre el nivel de dificultad, la forma de redacción, etc., pues puede ser que lo que es evidente para la persona que ha redactado la pregunta no es tan evidente para otros lectores. En la Universidad de Maastricht (ver Capítulo 18), en la creación de las PVF de las Pruebas de Progreso, en un 15% las preguntas sometidas al equipo de evaluación son rehechas o rechazadas. Por supuesto, estas revisiones requieren tiempo (proceso incompatible con la improvisación) y confidencialidad, especialmente cuando se trata de exámenes. El análisis de los resultados por ítems de los años anteriores también contribuye a mejorar las PSMs.
- (3) Una *presentación oral* de las PSM es imposible, a no ser que
 - se trate de soluciones fijas y permanentemente a la vista (sobre la pared de la clase o en una pantalla). Por ejemplo, declinaciones en Latín “1. Nominativo 2. Vocativo 3. Genitivo 4. Acusativo 5. Dativo 6. Ablativo”. Esas soluciones pueden llamarse Generales (porque son válidas para todas las preguntas) y, en este caso, Explícitas (porque no hay otras soluciones posibles y los estudiantes deben considerar únicamente estas Generales).
 - se limite a Preguntas Verdadero-Falso-No sé (o PVFNS), como se ilustra en el Capítulo 16 con pacientes diabéticos.
- (4) *La tentación de concebir preguntas “de detalles”*. Noizet y Caverni (1978, p. 155) recuerdan que “la(s) respuesta(s) correcta(s) a una PSM debe(n) ser “consensuable(s)”, es decir, deben obtener el acuerdo de todos los especialistas. Por ejemplo es “consensuable” que se atribuya cierta obra a un autor, pero es “menos consensuable” que este mismo autor pueda ser clasificado como “romántico”. Por esta razón, muchos docentes se limitan a utilizar preguntas cuyo contenido trata solo de hechos factuales, no midiendo los procesos superiores de la taxonomía de Bloom (ver Capítulo 15), sino solo la memorización de hechos sobre los que existe acuerdo entre los expertos. Una de las consecuencias de ello podría ser una “parcelación” (atomización) del conocimiento, donde el estudiante sabe muchas cosas pero des-

vinculadas, desarticuladas, sin integración. Preguntas con libros abiertos y Soluciones Generales Implícitas ayudan a evitar este peligro.

- (5) *La facilidad de copiar* las respuestas del vecino. Un código (1, 2 o 3) se transmite (con gestos) más fácilmente que una respuesta breve o de desarrollo. Para evitar eso el docente debe preparar varias versiones de la prueba. Cada versión contiene las mismas preguntas pero no en el mismo orden sobre las hojas. Si las respuestas son leídas con un Lector Óptico de Marcas es el computador el que “desembrolla” el asunto.
- (6) *La posibilidad de dar “Respuestas correctas gracias al azar”*, lo que D’Hainaut (1973) llama “elecciones felices por ignorancia”. Simétricamente, los errores en las PSM clásicas no pueden ser interpretados sin ambigüedad: el/la estudiante ¿dio esta respuesta incorrecta (a) de buena fe (por una concepción errónea) o (b) por azar? Existen varias formas de prevenir este riesgo:
 - aumentar el número de soluciones para reducir la probabilidad de una respuesta correcta al azar. En algunos casos es imposible. Por ejemplo, en alemán, si preguntamos el género de la palabra *Stein* hay solo 3 posibilidades (Masculino, Femenino, Neutral).
 - penalizar los errores, por ejemplo con la aplicación de la “corrección por adivinar” –*correction for guessing*– (ver sección A de este capítulo).
 - utilizar los grados de certeza (ver Capítulo 16), que no solo resuelven este problema sino que añaden muchas otras ventajas, especialmente mejorando la validez teórica o epistemológica, como lo enseña el modelo 3 de Choppin (ver antes, sección C). Por ejemplo, en el caso del género (masculino, femenino, neutral) de la palabra “*Stein*”, un grado de certeza de 33% significa ignorancia total.
- (7) *La imposibilidad de evaluar el nivel “Síntesis – Formulación – Expresión”* de la taxonomía de Bloom. A aquellos que se rehúsan a utilizar las PSM por esta única razón se les puede contestar (como Wood, 1977) que hacerlo sería como rehusarse a utilizar un estetoscopio porque no sirve de nada en el examen de los ojos y de los oídos. Sin embargo, esta es una razón importante para no limitar la evaluación a *solo* la utilización de PSMs. Un DEA –Dispositivo de Evaluación de los Aprendizajes– utiliza por definición varios instrumentos, que evalúan una gama de competencias y recursos. Una utilización excesiva de PSMs podría tener una consecuencia negativa: no dar a los estudiantes la posibilidad de ejercitar y evaluar la formulación de sus ideas.
- (8) *No se debe presentar soluciones erróneas en contenidos perceptivos*, por ejemplo aquellos en que la percepción visual (como en la ortografía de uso⁹⁵) o auditiva (como en la pronunciación) importa y constituye la única manera de memorizar la respuesta correcta.
- (9) *El problema de la uniformización de las personas y de las sociedades*. Este peligro parece serio si se usa (abusa) de PSMs simples de consigna tipo A (“Hay una de las k soluciones que es correcta y se puede dar una respuesta”; ver sección F2). La situación se modifica si usamos PSMs con SGI y si se organiza un debate posterior (ver Capítulo 9) en el cual se aceptan soluciones correctas propuestas (y bien argumentadas)

⁹⁵ No es lo mismo con la ortografía gramatical porque está basada en un razonamiento.

por los estudiantes, diferentes a la propuesta original del profesor. Jenkins *et al.* (1958) utilizaron una prueba (Kent-Rosanoff) de asociación de (100) palabras: los testeados deben producir una palabra asociada a una palabra dada. Estudiantes de la Universidad de Minnesota fueron examinados en dos ocasiones, con 23 años de diferencia. Para cada palabra (pregunta) se calcula la palabra (respuesta) más popular y su porcentaje entre todas las respuestas dadas a la pregunta. En 1929 este índice era (en promedio) 29% y en 1952 era 38%. Los mismos autores no han observado el mismo cambio en estudiantes de instituciones que hablan francés o alemán (países donde se utilizan menos PSM). Estos resultados parecen indicar una restricción de la diversidad (aumento de la homogeneidad) del campo mental entre los estudiantes de esta universidad (¿y de Estados Unidos?). Tal conclusión debe ser considerada con precaución, porque en 32 años la tasa de los estudiantes que ingresan a la universidad aumentó mucho. Sin embargo, hay que permanecer atentos y vigilantes ante este riesgo.

E. Ventajas de las Preguntas de Selección Múltiple (PSM)

Como lo muestra su historia y evolución durante un siglo (ver Capítulo 4), las preguntas de selección múltiple han sido concebidas y perfeccionadas para contribuir, de mejor forma que otros instrumentos, a mejorar las dimensiones de validez (calidades) de la evaluación. A continuación entregamos una enumeración de estas mejoras de calidad, organizadas según el modelo ETICPRAD de tipos de validez de la evaluación (ver Capítulo 4).

E.1. Contribuir a la validez Teórica conceptual

El uso de PSM cuenta con evidencias que refuerzan algunas de sus ventajas sobre otros instrumentos de evaluación que se usan aproximadamente en los mismos niveles, para evaluar el dominio de recursos similares y que demandan los mismos procesos mentales (como las preguntas de respuesta breve -PRB). Sax y Collet (1968, p.172) concluyen así su comparación experimental: “*Exámenes con PSMs motivan a los estudiantes a desempeñarse igual de bien que las preguntas de evocación, cuando los ítems son difíciles y ejercitar a los estudiantes con PSM mejora el rendimiento en los exámenes de evocación*”. Los autores piensan que puede ocurrir lo contrario si la ejercitación es con PRB. Concluyen que “*Es más difícil engañar al docente (respecto del conocimiento del estudiante) enfrentado a alternativas (de respuesta) muy cercanas que requieren sutiles discriminaciones, que engañarlo con ítems de tipo ‘Respuesta breve’*”.

Más recientemente, Schuwirth (1998, p. 49) concluye en la misma dirección: “*No se debe presupuestar que todas las PSMs son más fáciles que las (paralelas) PRBs si existe un sesgo, su efecto no es dramático y debe ser ‘balanceado’ contra las otras ventajas de las PSMs (tiempo de administración para asegurar la validez de ‘cobertura’, tiempo de corrección)*”.

Roediger y Marsh (2005) destacaron la importancia del número de proposiciones; de la cantidad de soluciones posibles. En una experiencia hicieron que sus estudiantes rindieran una prueba de 144 preguntas (PSM): 72 indagaban sobre un texto que los estudiantes habían leído, y 72 eran sobre uno que no habían leído. Además, en cada grupo de 72 PSM, 24 tenían 2 soluciones (Verdadero-Falso), 24 tenían 4 soluciones y 24 tenían 6 soluciones. Sus resultados (Figura 3) indican que mientras más grande el número de soluciones, más difícil es la pregunta (su tasa de éxito es más baja) y mayor es la diferencia medida entre los contenidos leídos y no leídos: discrimina mejor el conocimiento resultando de la lectura.

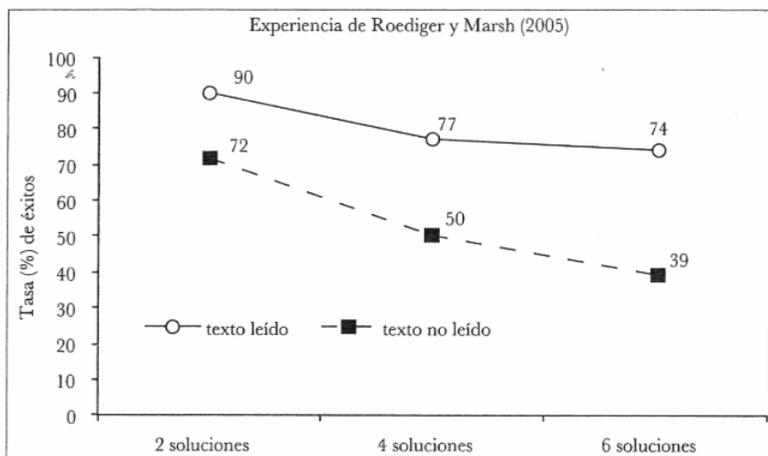


Figura 3: Relación entre el número de soluciones y: (1) la tasa de éxito; (2) la diferenciación entre el hecho de haber o no leído un texto informativo

E.2. Contribuir a la *validez Teórica de contenido* o de cobertura

Un alto número de PSMs permite testear un gran número de contenidos, capítulos de libro, temas... o procesos mentales. La siguiente es una tabla de Contenidos x Procesos:

Tabla 5: Tabla de cobertura "Contenidos-procesos mentales" de una prueba ficticia

	Total	Capítulo 1	Capítulo 2	Capítulo 3	Capítulo 4	Capítulo 5
Memorización	11	3	1	2	3	2
Comprensión	13	4	3	2	3	1
Aplicación	6	1	2	1	1	1
Análisis	10	2	1	2	2	3
Total	40	10	7	7	9	7

E.3. Contribuir a la *validez de Replicabilidad*

A continuación se muestra la fiabilidad (que va de 0 a 1 en el eje horizontal) de la nota final en términos de *reducción del error de medición*, según el número de preguntas de la prueba (eje vertical) y el número (k) de soluciones de la PSM. Las relaciones son expresadas por una fórmula matemática⁹⁶ y por el siguiente gráfico:

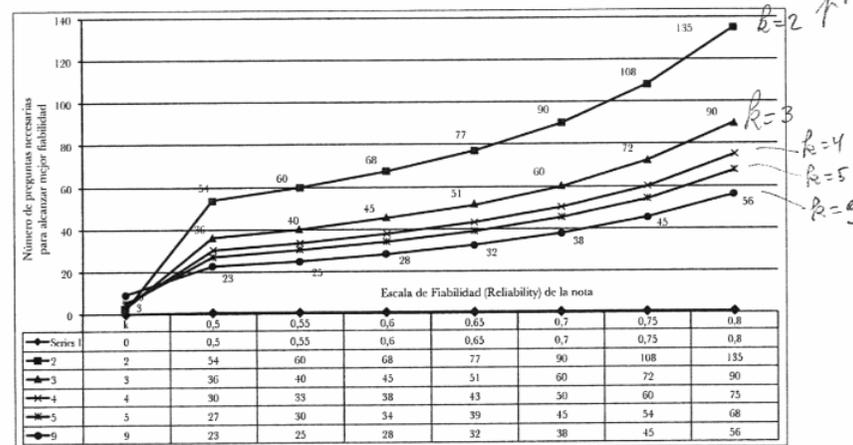


Figura 4: Número de PSM y de soluciones vinculado a la fiabilidad de la nota de un test

La fórmula es $NP = (9 / (1 - F)) \times ((k + 1) / (k - 1))$.

donde NP = Número de preguntas necesarias
y F = la Fiabilidad (de 0,5 hasta 0,8) que queremos alcanzar.

Se ve que para alcanzar una fiabilidad (*reliability*) de la calificación de 0,8 se necesitan 135 PSM, aunque lo mismo puede ser alcanzado con solo 68 PSM con 5 soluciones (dos veces menos). La misma fiabilidad puede lograrse con 56 ítems de PSM con 9 opciones de respuesta, es decir, 5 soluciones clásicas + 4 Soluciones Generales Implícitas (SGI).

E.4. Contribuir a la *validez de Aceptabilidad* para los estudiantes

Esta aceptabilidad se ve reforzada por los beneficios en términos de rapidez de respuesta: no tienen que verificar la ortografía ni la redacción, tampoco el estilo o el vocabulario, de modo que pueden contestar a 100 PSMs en 3 horas. Pueden contestar a menos preguntas si son PSM con SGI y libro abierto, pues consultar documentos demanda tiempo. Además, el estudiante tiene la garantía de que todos los contenidos

⁹⁶ De Ebel, 1969, p. 450, citada por Ledercq, 1986, p. 153.

importantes serán cubiertos y que no será una víctima del azar en la pregunta que le corresponda contestar.

E.5. Contribuir a la *validez de Aceptabilidad* para los docentes

Esta aceptabilidad se ve reforzada por los beneficios en términos de rapidez de corrección. La buena gestión de las PSMs tiene como resultado que un docente dispone, en un día o dos, de todos los resultados de un test (por grupo, por cada estudiante, por cada respuesta). El hecho de que cada solución sea codificada mediante un número o una letra facilita el proceso de corrección, ya sea manual o por computador (vía el teclado o la lectura óptica de marcas -LOM), que permite retroalimentaciones detalladas.

E.6. Contribuir a la *validez Deontológica* (ética)

Esto se logra a través de la objetividad de la corrección; tan objetiva que se puede automatizar. De este modo se evitan los efectos de corrección descritos en el Capítulo 6. Sin embargo, un debate inmediatamente después de la prueba, comunicando las respuestas correctas, es importante no solo para evitar daños nosocomiales (ver sección D) sino también para asegurar las calidades psicométricas de cada pregunta, como los índices de discriminación.

E.7. Contribuir a la *validez Consecuencial* de autorregulación

Porque los estudiantes pueden autocorregirse, siempre y cuando hayan recibido las respuestas correctas después del test. Esto calibra el juicio sobre los propios conocimientos / desempeños, y promueve la autoplanificación para la mejora (ver Capítulo 9).

E.8. Contribuir a la *validez Consecuencial* epistemológica

Porque favorece la comunicación colectiva de las respuestas correctas y el *debate colectivo* “en caliente”. En este espacio cada persona es invitada a justificar públicamente su solución (la 4, por ejemplo) que es por todos conocida; todos los participantes toman partido y entregan argumentos para validar como correcta la misma u otras soluciones, también conocidas por todos. Los beneficios en términos epistemológicos (Perry, 1970 y 1985), de representaciones de lo que es el conocimiento, de lo que significa “ser experto”, “comprender”, “argumentar”, ya han sido ilustrados en el Capítulo 9.

E.9. Contribuir a la *validez Consecuencial* formativa

En la experiencia descrita en la sección E1, Roediger y Marsh (2005) pidieron a sus estudiantes rendir una prueba de PSMs con 2, 4 y 6 soluciones. Luego de un tiempo destinado a una tarea sobre otro tema, pidieron a sus estudiantes rendir otra prueba con

preguntas abiertas (de respuesta construida) de 4 tipos: preguntas que trataban sobre textos leídos o no leídos, preguntas que habían sido presentadas anteriormente en la prueba de PSM (pero sin revelar las respuestas correctas después de la prueba), y otras que no habían sido presentadas anteriormente, con sus posibles combinaciones. Los autores calcularon las tasas de éxito cuando las preguntas trataban de textos que fueron leídos o no leídos, y testeados o no testeados con PSMs. También calcularon las tasas de errores correspondientes a soluciones falsas (distractores) que habían aparecido en las PSMs. Los resultados se muestran en la Tabla 6.

Tabla 6: % de éxito y de respuestas falsas inspiradas por los distractores de la prueba anterior con PSMs en la experiencia de Roediger y Marsh (2005)

		Número de soluciones en la PSM durante la prueba anterior (0 = no fueron testeados con PSM)			
		0	2	4	6
% éxito	Textos leídos	40%	67%	61%	61%
	Textos no leídos	16%	34%	28%	26%
% distractores	Textos leídos	4%	6%	8%	9%
	Textos no leídos	6%	9%	13%	15%

La Tabla 6 muestra que:

- (1) Cuando los estudiantes han sido expuestos a PSMs la tasa de éxito es mayor.
- (2) El porcentaje de errores inspirados en distractores crece de acuerdo al aumento en el número de soluciones de las PSMs en la fase anterior, como se anticipó que ocurriría.
- (3) Este fenómeno (2) es menos amplio cuando los estudiantes han leído el texto vinculado a las preguntas, en comparación a cuando no lo han leído.
- (4) La presencia de porcentajes de errores inspirados por distractores superiores a 4%, cuando los estudiantes han sido expuestos a PSM, puede aparecer como una contraindicación del uso de PSMs. Mala noticia. Sin embargo, se debe recordar que las respuestas correctas no han sido comunicadas a los estudiantes. En una clase debe ocurrir lo contrario: se debe retroalimentar a los estudiantes con las respuestas correctas, lo que asegura que no ocurra el fenómeno del error inspirado en el distractor (ver sección D1).

Los estudiantes pueden ser invitados a crear ítems de PSMs ellos mismos, como lo hicieron Mc Leod y Snell (1996) en el Departamento de Medicina de la Universidad de Mc Gill. Al contrario de lo que Mc Keachie (1986) había observado (“solo una tasa de 10% de los ítems recibidos era utilizable”), observaron que, aunque no todos los estudiantes habían entregado preguntas, 90% de las que los estudiantes habían creado eran aceptables. Sin embargo, algunos contenidos estaban sub representados. En sus 3 exámenes (cada uno de 70 PSMs de tipo A “clásico”) Mc Leod y Snell introdujeron

62 (sobre los 210) ítems construidos por los estudiantes. La tasa de preguntas que presentaba un “caso clínico” era 46% en las PSMs construidas por los docentes (y con un éxito medio de 66%) y 24% en las construidas por los estudiantes (y con éxito medio de 69%). Además, jueces (expertos médicos) juzgaron que las preguntas de los estudiantes tenían una mejor autenticidad (validez Ecológica; “*face validity*”) comparadas con las de los docentes. Una posible explicación es que *Las preguntas de los estudiantes reflejan mejor la población de pacientes que estos estudiantes encuentran en los hospitales donde hacen sus prácticas de terreno.*

F. Formas de presentar Preguntas de Selección Múltiple (PSM) y sus consignas

F.1. Formulaciones del tronco (cuerpo)

A) PRESENTAR EL TRONCO CON O SIN PUNTOS DE INTERROGACIÓN

(Fa) ¿Cuál es el nombre del río central de Costa de Marfil? 1. 2. 3. 4.
 (Fb) Completar: “..... es el río central de Costa de Marfil”. 1. 2. 3. 4.
 (Fc) El río central de Costa de Marfil se llama..... 1. 2. 3. 4.
 (Fd) Aquí siguen nombres de ríos.
 ¿Cuál de ellos es el central de Costa de Marfil? 1. 2. 3. 4.

El hecho de no haber signos de interrogación en las Formulaciones Fb y Fc justifica que varios autores no utilicen la palabra “pregunta” sino la palabra *ítem*⁹⁷ para designarlos. Las diferencias entre las formulaciones Fa, Fb, Fc, Fd son superficiales. Sin embargo, la formulación Fd (Aquí siguen) es más apropiada cuando el tronco es largo, presentándolo en varias frases (ver regla 11 en el Capítulo 14).

B) PREGUNTAS DIRECTAS, ÍTEMS INDIRECTOS Y PSEUDOPREGUNTAS

Las preguntas de Formulación Fa, Fb, Fc, Fd son *directas*. Permiten contestar antes de haber leído las soluciones porque en realidad (y no solo en las PSM) hay una sola respuesta correcta. Los ítems siguientes (de formulación Fe y Ff) son *indirectos* porque fuerzan a leer las soluciones antes de responder:

(Fe) Entre las ciudades que siguen, ¿cuál está ubicada en Perú? 1. 2. 3. 4.
 (Ff) Entre los ríos que siguen, ¿cuál *no* desemboca en el Pacífico? 1. 2. 3. 4.

Los ítems (Fe) y (Ff) evalúan la capacidad de *reconocer*, aunque las preguntas de

⁹⁷ Haladyna y Downing (1989); Case y Swanson (2001).

formulación Fa, Fb, Fc, Fd pueden evaluar la *evocación*, si la respuesta correcta es “Ninguna” (+ justificación).

Los siguientes ítems son *pseudopreguntas*

(Fg) Entre las afirmaciones siguientes, ¿Cuáles son correctas? 1. 2. 3. 4.
 (Fh) ¿Cuándo el hidrógeno combustiona en el aire? 1..... 2..... 3..... 4.....

No son PSMs, tampoco son preguntas, pero introducciones comunes a varias PVF tendrían que ser tratadas en esta forma. Hemos visto (en sección A7 antes) el interés de utilizar Soluciones Generales Implícitas (SGI). Estas se aplican más cómodamente con preguntas que con ítems.

C) PREGUNTAS DE FORMA POSITIVA Y DE FORMA NEGATIVA

La forma negativa de una pregunta aumenta la dificultad (Terranova, 1969; Dudycha y Carpenter, 1973).

Ahora que estas precisiones de vocabulario están hechas, podemos considerar el concepto de *consignas* (o instrucciones).

F.2. Las consignas de las PSMs

A) LA CONSIGNA A (SIMPLE O CLÁSICA)

“Hay una de las *k* soluciones que es correcta y se puede dar una respuesta”.

Ya hemos denunciado el peligro de desarrollar en el estudiante malos procesos de respuesta: empezar considerando las soluciones antes que haber leído bien la pregunta, no formular su propia respuesta, no preguntarse si la pregunta tiene sentido o si contiene un absurdo, o si el tronco presenta datos suficientes para contestar. Por todas estas razones la condenamos como un instrumento que tiene más efectos perniciosos que beneficios. Las Soluciones Generales Implícitas son una mejora capital.

Se puede distinguir entre:

- las PSMs A+ (como “positivas”) que piden al estudiante elegir la respuesta correcta
- de las Preguntas A- (como “negativas”) que piden al estudiante elegir la solución incorrecta entre las otras (que son correctas).

B) LA CONSIGNA B (LA MEJOR SOLUCIÓN – THE BEST SOLUTION)

“Entre estas soluciones elige la mejor”.

Este tipo de consigna conviene con preguntas del tipo:

“¿Cuál es la mejor traducción de *To be or not to be*?”

1. Ser o no ser 2. Estar o no estar 3. Ser o ser no 4. No ser sino ser”

El riesgo de este tipo de consigna es que “cuál es la respuesta más correcta” puede ser discutible si es que existen dudas respecto a “cuál punto de vista es el mejor”.

Ejemplo de Case y Swanson (2001, pp.16-17):

VINETA: Un señor de 32 años presenta una historia de 4 días de debilidad progresiva en las extremidades. Ha tenido buena salud excepto por una infección respiratoria hace 10 días. Su temperatura es 37,8°C. Su presión arterial es 130/80 mm Hg, su pulso 94 p/m y sus respiraciones son 42 p/m y superficial. Tiene debilidades simétricas en ambos lados de la cara y los músculos proximales y distales de las extremidades. Su sensibilidad está intacta. No se encuentran reflejos tendinosos profundos, y hay flexión como respuesta plantar.

PREGUNTA: Entre los diagnósticos siguientes, ¿cuál es el diagnóstico más probable?

1. Encefalomieltis diseminada aguda
2. Síndrome Guillain-Barré
3. Miasthenia gravis
4. Poliomieltis
5. Polimiositis

Se debe notar que las opciones incorrectas no son totalmente falsas; se pueden ordenar en el siguiente eje de 'menos probable' a 'más probable':

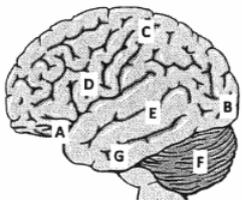
4	3	1	5	2
Menos probable				Más probable

Figura 5: Pregunta con consigna tipo B: La mejor Solución (The Best Solution). Case y Swanson (2001)

Este tipo de situación (consigna B) está vinculado con la técnica de los Tests de Concordancia de Scripts o TCS (ver Capítulo 19).

C) LA CONSIGNA C (DE CORRESPONDENCIA)⁹⁸

Esta forma constituye una manera económica de preguntar *muchas PSM* con un tronco común y soluciones comunes, lo que permite ganar mucho espacio. En el ejemplo, el tronco común es la sentencia “¿Qué zona designa el área?” y la figura del cerebro, y las soluciones comunes son A, B, C, D, E, F, G.



Preguntas: ¿Qué zona designa el área... ?

P1. de la visión

P2 del olfato

P3 del oído

P4 de la lectura

P5. de la decisión

P6. de los movimientos

P7. del equilibrio

P8. del gusto

Respuestas correctas: P1 = B, P2 = A, P3 = E, P4 = C, P5 = Ninguna, P6 = C, P7 = F, P8 = A.

Figura 6: Pregunta con consigna de correspondencia

⁹⁸ En inglés es llamado “Extended matching items” o “R type” (Case y Swanson, 2001, p. 71).

Por supuesto, no debe existir el mismo número de “soluciones” que de preguntas, pues una vez que 7 de las 8 preguntas han sido contestadas la respuesta a la pregunta que queda se favorecería demasiado. Debe haber una disimetría, con preguntas cuya solución es ninguna (como P5), soluciones que no son correctas para ninguna pregunta (como solución G), y soluciones que son la solución correcta de varias preguntas (por ejemplo, la solución A para P2 y P8).

D) LA CONSIGNA PSN (PREGUNTAS CON SOLUCIONES NUMEROSAS)

Este tipo de consigna ha sido descrita en detalle en el Capítulo 15 (sección B2). Consiste en proveer de una hoja con decenas de soluciones posibles (varias veces más que las requeridas por las preguntas) ordenadas alfabéticamente. Se fomenta la evocación antes que el reconocimiento.

E) LA CONSIGNA LM (LONG MENU)

Cuando una persona introduce las primeras letras de una palabra vía el teclado de un computador, hay programas que completan la palabra. Schuwirth (1998) utilizó este principio en tests (por computador) de diagnóstico en medicina. Por ejemplo, si el/la estudiante teclea “diab”, el programa completa “diabetes”. Esto fuerza a el/la estudiante a evocar la solución (no reconocerla de entre un set de soluciones), porque hay centenas de soluciones en la lista que conoce el programa.

F) LA CONSIGNA A-R (ASEVERACIÓN – RAZÓN)

También llamadas “de tipo E”, esta forma de test consiste en dos columnas, una de aseveraciones (PA) y la otra de razones (PR). Para cada PA y cada PR el/la estudiante debe contestar si es correcta o no (Verdadero - Falso), y además debe designar cuál aseveración está vinculada con cuál razón.

En su libro sobre los ítems en el USMLE del NBME, Case y Swanson (2001, p. 124) envían este tipo de consigna al cementerio, pues comporta una carga mental resultante en excesivas ambigüedades cuando se trata de interpretar errores.

G) LA CONSIGNA DE TIPO K

Ejemplo: Después de que el tronco ha descrito lo que son A, B, C y D, las soluciones propuestas son

$$1 = A \quad 2 = B \quad 3 = C \quad 4 = D \quad 5 = A y B \quad 6 = B y C \quad 7 = A, B y D \quad 8 = B y D.$$

Esta consigna también ha sido enviada al cementerio de las consignas por Case y Sawnsom (2001, p. 122), porque se produce demasiada confusión y demasiadas solu-

ciones que son parcialmente correctas. En el ejemplo anterior, si la Respuesta correcta (RC) es 6, entonces las soluciones 2, 3, 5, 7 y 8 son parcialmente correctas. Por esta razón Leclercq limitó su consigna SGI a solo 4 SGI, entre las cuales está "Todas", pero no utiliza combinaciones de varias soluciones, lo que evita la situación de la consigna K.

H) LA CONSIGNA "SOLUCIONES GENERALES IMPLÍCITAS" (SGI)

Esta consigna ha sido descrita en la sección A7 de este capítulo. Por supuesto, las distintas SGI pueden ser introducidas progresivamente, por ejemplo en la escuela primaria, pero en la universidad consideramos que todas deben ser introducidas para no seguir perdiendo tiempo (ya es una lástima que los estudiantes no las hayan encontrado durante su educación secundaria).

G. Listas de cotejo para las Preguntas de Selección Múltiple (PSM)

G.1. ¿Corrección por adivinación? (para desincentivar la respuesta al azar)

En la sección C se presentó la corrección por adivinación (*correction for guessing*), aunque no recomendamos este tipo de puntaje sin utilizar grados de certeza (ver Capítulo 16).

G.2. Gravedad de los errores

Los niveles de gravedad de los errores pueden ser definidos y ordenados por expertos, sobre un eje (como en el ejemplo de Case y Swanson en la sección F2). Un error fatal puede ser penalizado de forma mucho más severa que uno benigno.

Referencias

- BAIN, K. (2004) What the Best College Teachers Do. Harvard University Press.
- BLOOM, B.S. (1972) L'innocence en pédagogie, Education - Tribune Libre, 135, 14-20.
- BUTLER, A.C., y ROEDIGER, H.L., III (2007). *Testing improves long-term retention in a simulated classroom setting*. European Journal of Cognitive Psychology, 19, 514-527.
- CASE, S. y SWANSON, D. (2001). *Constructing Written Test Questions for the basic and clinical sciences*. Philadelphia: National Board of Medical Examiners (NBME).
- CHOPPIN, B. (1970). *An IEA Study of Guessing. A Proposal*, Stockholm: International Association for the Evaluation of Educational Achievement, unpublished memorandum, IEA/TR/9.
- COOMBS, C.H., MILHOLLAND, J.E. y WOMER, F.B. (1956). The assessment of partial knowledge. Educational Psychological Measurement, 16, 13-37.
- DE FINETTI, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item. British Journal of Mathematical and Statistical Psychology, 18, 87-123.
- D'HAINAUT, L. (1973). Etude d'une nouvelle variable pour l'analyse statistique des expériences pédagogiques. Bulletin de Psychologie, 305, 26, 622-630.

- DÍAZ BARRIGA, A. (2006). UNAM. La educación en valores: Avatares del currículum formal, oculto y los temas transversales. Revista Electrónica de Investigación Educativa, 8 (1). Consultado en: <http://redie.uabc.mx/vol8no1/contenido-diazbarriga2.html>, Pág. 7 y 8 (consultado el 5 de enero de 2014).
- DREBEN, R. (1967). *On What is Learned in School*. London: Addison-Wesley.
- DUDYCHA, A.L. y CARPENTER, J.B. (1973). Effects of item format on item discrimination and difficulty. Journ. Appl. Psychol., 58, 11-121.
- DURKHEIM, E. (1925; trad.1961). *Moral Education*. New York: Free Press.
- EBEL, R.L. (1969). Expected reliability as a function of choices per item. Educational and Psychological Measurement, 29, 565-570.
- EBEL, R.L. y FRISBIE, D. (1991). *Essentials of educational measurement*. Prentice Hall (Englewood Cliffs, N.J.).
- EVANS, R. y MISFELDT, K. (1974). Effects of self-scoring procedures on test reliability. Perceptual and motor skills. 38, 12-48.
- FRARY R.B. (1989). Partial-credit scoring methods for multiple-choice tests. Applied Measurement in Education, 2(1), 79-96.
- GIROUX, H. A. (2001). *Theory and Resistance in Education*. London: Bergin y Garvey.
- HALADYNA, T. y DOWNING, S. (1989). A taxonomy of Multiple-Choice Item-Writing rules. Applied Measurement in Education, 2 (1), 37-50.
- HANNA, G. (1977). A study of reliability and validity effects of total and partial immediate feedback in multiple-choice testing.
- HUANG, Y., TREVISAN, M. S., y STORFER, A. (2007). The impact of the "all-of-the-above" option and student ability on multiple choice tests. International Journal of the Scholarship of Teaching and Learning, 1(2),
- JACKSON, Ph. (1968). *Life in Classrooms*. New York: Holt, Reinhart y Winston.
- JACOBY, L.L. y HOLLINGSHEAD, A. (1990). Reading student essays may be hazardous to your spelling: Effects of reading incorrectly and correctly spelled words. Canadian Journal of Psychology, 44, 345-358.
- JENKINS, J., MINK, W. y RUSSEL, W. (1958). Associative clustering as a function of verbal association strength. Psychol. Rep., 4, 127-136.
- KARRAKER, R. J. (1967). Knowledge of results and incorrect recall of plausible of multiple choice alternatives, Journal of educational Psychology, 58, 11-14.
- KENTLI, F.D. (2009). Comparison of hidden curriculum theories. European journal of Educational Studies, 1 (2), 83-88.
- LECLERCQ, D. (1986). *La conception des QCM*. Bruxelles: Labor y Paris: Nathan.
- LECLERCQ (2013). Autoevaluación y grados de certeza. Capítulo 16 de Leclercq y Cabrera. *IDEAS. Innovaciones en Dispositivos de Evaluación de los Aprendizajes en la enseñanza Superior*. UCH.
- LYNCH, K. (1989). *The Hidden Curriculum: Reproduction in Education, A Reappraisal*. London: The Flamer Press.
- MC KEACHIE, W. J. (1986). *Teaching tips: A guide-book for the beginning college teacher* (4th ed.). Ann Arbor, MI: G. Wahr.
- MC LEOD, P. y SNELL, L. (1996). Student-generated MCQs. Medical teacher. 18, n°1, 23-25.
- NOIZET, G. y CAVERNI, J.P. (1978). *Psychologie de l'évaluation scolaire*, Paris: PUF.
- OSTERLIND, S. J. (1989). *Examinations: Design and construction*. Boston: Kluwer Academic Publishers
- PERRY, W.G., Jr. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart and Winston.
- PERRY, W.G. (1985). Different worlds in the same classroom: Students' evolution in their vision of knowledge and their expectations of teachers. In Gullette, M.M. (Ed.), *On teaching and learning*. Volume 1, 1-17. Cambridge, MA: Harvard-Danforth Center for Teaching and Learning.
- PRESTON, R.C. (1965). Multiple choice test as an instrument of perpetuating false concepts, Educational and Psychological measurement, 25, 111-116.

- REMMERS, H. y REMMERS, E. (1926). The negative suggestion effect on true-false examination questions. *Journal of Educational Psychology*, 17. P. 52-56.
- RODRIGUEZ, M. (1997). The art y science of item-writing. A meta-analysis of Multiple-Choice item Format Effects. Paper presented at the AERA. 79p.
- ROEDIGER, M. y MARSH E. (2005). The Positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Vol. 31, No. 5, 1155-1159.
- SABERS, D. y FELDT, L. (1968). An empirical study of the effect of the correction for chance success on the reliability and validity of an aptitude test. *Jour. Educ. Measmt.*, 5, n°3, 251-258.
- SAX, G. y COLLET, L. (1968). An empirical comparison of the effects of recall and multiple-choice test on student achievement. *Jour; Educ. Measmt.* 5, N° 2, 169-173.
- SCHUWIRTH, L.W.T. (1998). Computerized Case-based Testing: an approach to the assessment of medical problem solving. Ph.D. in Education, Maastricht: University of Maastricht.
- SKINNER, B.F. (1958). Reinforcement Today. *American Psychologist*, 13, 3.
- TERRANOVA, C. (1969). The effects of negative items in multiple-choice test items. Doctoral Dissertation. SUN 4 Buffalo. DAI-Univ. Microfilms, n° 69-20, 512.
- TORRES, J. (1998). *El currículum oculto*. Ediciones Morata, Madrid, España.
- WOOD, R. (1977). Multiple-choice: A state of the art report. In Choppin y Postlethwaite (Eds). *Evaluation in Education*. International Progress. Oxford: Pergamon.

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

Dieudonné LECLERCQ y Álvaro CABRERA MARAY 2014

Resumen de cada capítulo

Los editores y autores principales del libro

p. 11-13

Prologo

Álvaro Cabrera &
Dieudonné
Leclercq

Parte 1: Conceptos clave en educación

p. 15-20

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

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

13	<p>Las Preguntas de Selección Múltiples (PSM): del currículo escondido a la vigilancia cognitiva Presenta los retos del currículo oculto y de la espontaneidad vs la limitación a respuestas sobre 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>