



Perfiles Educativos

ISSN: 0185-2698

perfiles@unam.mx

Instituto de Investigaciones sobre la  
Universidad y la Educación  
México

Tirado, Felipe; Backhoff, Eduardo; Larrazolo, Norma  
La revolución digital y la evaluación: un nuevo paradigma  
Perfiles Educativos, vol. XXXVIII, núm. 152, abril-junio, 2016, pp. 182-201  
Instituto de Investigaciones sobre la Universidad y la Educación  
Distrito Federal, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=13244824011>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica  
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

# La revolución digital y la evaluación: un nuevo paradigma

FELIPE TIRADO\* | EDUARDO BACKHOFF\*\* | NORMA LARRAZOLO\*\*\*

La revolución digital ha dado lugar a una nueva era en la evaluación educativa a gran escala; ofrece posibilidades que hace una década eran impensables. Junto con los avances de las ciencias cognitivas y la psicometría fue posible lograr avances sustanciales, como la generación automática de reactivos. El propósito de este trabajo es presentar un nuevo paradigma de evaluación, basado en campos semánticos, que se vale de la tecnología digital. Este paradigma permite evaluar competencias cognitivas de mayor complejidad, a través de respuestas construidas o semi-construidas y de crédito parcial, superando así muchas de las limitaciones del formato tradicional de lápiz y papel y de opción múltiple. Se describen las características teóricas, conceptuales, técnicas y computacionales del Examen de Competencias Básicas (EXCOBA) y el Generador Automático de Exámenes (GenerEx), lo que permite generar ítems isomorfos y exámenes equivalentes.

*The digital revolution has given way to a new age in large-scale educational assessment, offering possibilities that were inconceivable a decade ago. Together with progress in cognitive sciences and psychometrics, it has been possible to make significant progress, such as automated question generation. The aim of this article is to present a new assessment paradigm, based on semantic fields, using digital technology. This paradigm allows for evaluating more complex cognitive competencies through constructed or semi-constructed responses and partial credit, thus overcoming many limitations of the traditional paper-and-pencil and multiple-choice formats. The theoretical, conceptual, technical and computational characteristics of the Basic Competencies Exam (EXCOBA) and the Automatic Item Test Generation tool (GenerEx), which permits the generation of isomorphs and equivalency tests, are described.*

## Palabras clave

Evaluación  
computarizada  
Campos semánticos  
Respuesta construida  
Generador de exámenes  
Ítems isomorfos

## Keywords

Computerized  
assessments  
Semantic fields  
Constructed responses  
Item automatic generator  
Isomorphic items

Recepción: 10 de agosto de 2015 | Aceptación: 18 de octubre de 2015

\* Profesor titular en la Facultad de Estudios Superiores (FES) Iztacala de la Universidad Nacional Autónoma de México (UNAM) (México), donde dirige el programa de Investigación PsicoEducativa. Doctor en Educación por la Universidad Autónoma de Aguascalientes. Líneas de investigación: evaluación educativa, diseños educativos y mediación digital. Publicaciones recientes: (2015), "Entornos educativos, educación en contextos", *Revista de la Red Iberoamericana de Pedagogía*, vol. 4, núm. 7, pp. 6-15; (2015, en coautoría con A. Miranda y Z. Delgado), "Una estrategia educativa de colaboración en línea: co-evaluación y meta-evaluación", en *Memoria electrónica. XIII Congreso Nacional de Investigación Educativa*, México, COMIE/Universidad Autónoma de Chihuahua. CE: ftirado@unam.mx

\*\* Consejero de la Junta de Gobierno del Instituto Nacional para la Evaluación de la Educación (INEE) (México), responsable del desarrollo de los Exámenes de la Calidad y el Logro Educativos (EXCALE) para evaluar la calidad del Sistema Educativo de México (educación básica); representante de México en el Programme for International Student Assessment (PISA). Doctor en Educación. Líneas de investigación: desarrollo y validación de instrumentos de evaluación educativa, diseño de instrumentos no cognitivos y validez de pruebas de logro. Publicación reciente: (2016, en coautoría con F. Ferreyra), "Validez del generador automático de ítems del examen de competencias básicas", *Revista Electrónica de Investigación y Evaluación Educativa*, vol. 22, núm. 1, art. 2, s/pp. CE: backhoff@inee.edu.mx

\*\*\* Directora de contenidos de Métrica Educativa, A.C. Doctorado en Educación por la Universidad Iberoamericana (Noroeste). Línea de investigación: evaluación educativa. Publicación reciente: (2015, en coautoría con J.C. Pérez, E. Backhoff y R. Guaner), "Análisis de la estructura cognitiva del área de habilidades cuantitativas del EXHCOBA mediante el modelo LTM de Fisher", *Revista Internacional de Educación y Aprendizaje*, vol. 3, núm. 1, pp. 25-38. CE: nlarrazolo@metrica.edu.mx

## PRESENTACIÓN

La revolución digital está penetrando cada vez con mayor intensidad en todos los ámbitos educativos, y la evaluación no es una excepción. Bennett (2004) afirmó hace más de una década que ningún tema tendría mayor importancia para la innovación y la práctica futura de la evaluación educativa que las computadoras y la Internet. Su predicción fue correcta, ya que la informática y la evaluación se han desarrollado a un ritmo vertiginoso. La revolución digital en el campo de la evaluación educativa ofrece ahora posibilidades que apenas hace una década eran impensables, y ha dado lugar al advenimiento de una nueva era en la evaluación.

Entre las múltiples ventajas de los recursos digitales para la evaluación, hay que destacar que pueden generar preguntas y procesar las respuestas por algoritmos en cuestión de décimas de segundo; establecer enlaces en línea de alta velocidad para transmitir flujos de información en tiempo real y diferido; usar recursos multimedia para hacer presentaciones de textos, imágenes, sonido y video en formatos de alta definición; articular vínculos dinámicos como hipertexto (enlaces internos) e hipervínculo (enlaces web) que permiten relacionar distintas fuentes de información de manera instantánea y evaluar la interacción (respuestas en pantalla). También se puede medir el tiempo de respuesta, generar bases de datos relacionadas que permiten acceder, recuperar, procesar, interpretar e integrar información diversa; y programar con lenguajes de cómputo el diseño, formato, administración y calificación de una amplia variedad de reactivos que superan muchas de las limitaciones de las preguntas tradicionales de opción múltiple. Estas potencialidades han dado lugar a una nueva generación de evaluaciones del aprendizaje, cuyas características principales son: la digitalización de todos sus procesos, la administración ramificada (no lineal), la medición a profundidad de competencias

cognitivas por respuesta construida y la generación automática de ítems (*automatic item generation*).

Los exámenes más acreditados elaborados por el Educational Testing Service, como son el Test of English as a Foreign Language (TOEFL iBT) y el Graduate Record Examination (GRE), ahora trabajan en plataformas digitales. La evaluación del Programme for International Student Assessment (PISA), de la Organización para la Cooperación y Desarrollo Económico (OECD, por sus siglas en inglés), ha señalado que la aplicación correspondiente a 2015 se realizará principalmente en formato digital (OECD, 2015).

En México, desde 1993, el Examen de Habilidades y Conocimientos Básicos (EXHCOBA) se ha aplicado en formato digital a cientos de miles de aspirantes a ingresar a instituciones universitarias de educación media superior y superior (Backhoff *et al.*, 1996).

No obstante lo anterior, a pesar de que ha habido un gran avance en la digitalización de la evaluación de los aprendizajes, muchas de las pruebas computarizadas siguen evaluando conocimientos y habilidades con base en reactivos de opción múltiple, toda vez que ha sido la forma tradicional para hacer evaluaciones válidas y confiables que son rápidas de calificar de manera automatizada. Esto ha sido particularmente importante para las evaluaciones a gran escala, donde miles, o decenas de miles de estudiantes tienen que ser evaluados en tiempos acotados. En el mejor de los casos, algunas evaluaciones de gran escala ya utilizan reactivos de respuesta construida, a partir de respuestas cortas, para después ser calificadas por un grupo de especialistas entrenados a través de rúbricas, tal y como se reporta en el programa PISA for Development (OECD, 2014).

Por la variabilidad que suele haber en las respuestas construidas, así como lo tardado y costoso que puede significar su calificación a través de rúbricas, la mayoría de las instituciones responsables del desarrollo de pruebas

estandarizadas ha seguido utilizando los procedimientos tradicionales para diseñar y usar evaluaciones donde se requiere que el estudiante seleccione una respuesta dada. Por lo general, estos procedimientos demandan que especialistas en contenidos especifiquen los conocimientos, habilidades o competencias que se desean evaluar; que otro grupo de expertos elabore las especificaciones definidas de los reactivos que se utilizarán en las pruebas; y que se hagan estudios piloto de validación de los reactivos para finalmente conformar la versión definitiva de la prueba con aquellos ítems que logren obtener los estándares psicométricos internacionalmente aceptados de validez y confiabilidad. Usualmente se construyen varios reactivos para cada especificación, con el objetivo de tener dos o más versiones de una prueba con reactivos equivalentes. Esto implica que se pueden tener versiones diferentes de un mismo examen mientras sean conceptual y psicométricamente equivalentes, pero hay dos inconvenientes: 1) es difícil construir reactivos de opción múltiple que sean diferentes y a la vez isomorfos (equivalentes conceptual y psicométricamente); y 2) el procedimiento es lento y costoso.

A pesar de que estas pruebas pueden tener varias versiones, sufren el desgaste natural al usarse repetida e intensivamente, como es el caso de las pruebas de ingreso a las universidades, porque se filtran las preguntas, los aspirantes se van aprendiendo las respuestas y en consecuencia pierden su validez. También las respuestas de opción múltiple tienen el inconveniente de ser fáciles de copiar, además de que pueden ser respondidas correctamente al azar. Es por estas razones que los ítems de estas pruebas deben renovarse en forma constante, con el objetivo de mantener vigente la validez de sus resultados.

La mayoría de exámenes estandarizados a gran escala evalúa con el formato de respuesta. La razón principal es que así se pueden calificar objetivamente con gran rapidez; aun si la vía de administración del examen es por

lápiz y papel, se pueden usar lectores ópticos. Adicionalmente, la psicometría ha desarrollado en las últimas décadas análisis y procedimientos que mejoraron la evaluación, como la teoría de respuesta al ítem (*item response theory*), que revolucionó en los años sesenta del siglo pasado muchos de los planteamientos de la teoría clásica de la evaluación. Esto ha ayudado a preservar la opción múltiple.

Es bien sabido que el formato de elección de respuesta presenta limitaciones serias que no se pueden obviar, tal y como lo señalan muchos autores. Entre sus limitaciones más relevantes destaca que no evalúa la comprensión de conocimientos a profundidad (pensamiento complejo), ni las habilidades de ejecución (Rothstein *et al.*, 2010). Asimismo, hay importantes críticas a dos de sus premisas básicas que sostienen que: 1) es posible desagregar los aprendizajes en unidades pequeñas de respuesta, en tanto el todo se puede descomponer en sus partes; y 2) el aprendizaje se puede estudiar e interpretar con base en elementos aislados de su contexto. Resnick y Resnick (1992) comentan que estos supuestos no corresponden con lo que ocurre en la vida real, donde todo está relacionado y los significados se configuran en contextos, los cuales hacen que se interpreten las propiedades del entorno desde un referente cultural. Las evaluaciones suelen ser útiles para la escuela, pero no para la vida o la vida profesional.

Ng y Chan (2009) analizaron los llamados exámenes objetivos y concluyeron que éstos se limitan a dos clases de respuestas correctas: las que el examinado conoce, y las que acierta por azar. Estas últimas, dadas por correctas, son una fuente de error que contamina y aminora la validez de estos exámenes.

El modelo de opción múltiple suele presentar entre tres a cinco opciones de respuestas posibles, donde las respuestas incorrectas se conciben técnicamente como “distractores” (distráen), incorporan “ruido”, y aunque son plausibles, solamente una opción debe ser la correcta. Este formato atomiza el saber y promueve

la concepción del aprendizaje fragmentado, circunscrito a precisiones, lo que suele ser aprendido por memorización. De este modo se restringe al sustentante a elegir entre las limitadas opciones disponibles y no se le permite que desarrolle y construya sus propias respuestas.

También puede ocurrir que el examinado conozca parte de la solución, que tenga ciertos conocimientos del área, porque el conocimiento se da en un continuo, y no de manera atomizada en respuestas puntuales, como se concibe en las respuestas que son dicotómicas (correcto-incorrecto). En este tipo de exámenes no se reconoce que hay distintos niveles de conocimiento, no se valora el conocimiento parcial ni se registra; no se le otorga crédito parcial, simplemente se ignora y se pierde esa información.

Teniendo presentes las limitaciones observadas en las evaluaciones tradicionales disponibles en formato de selección de respuesta aplicadas en lápiz y papel, y dadas las amplias potencialidades que ofrecen los medios digitales, en las últimas décadas se han explorado nuevos planteamientos para mejorar los procesos de evaluación. En este trabajo se presenta un nuevo paradigma de evaluación de competencias académicas, sustentado en tres grandes características: 1) evaluar habilidades cognitivas complejas en contexto, a través de tareas que requieran del sustentante que construya sus respuestas, y no simplemente las seleccione; 2) formular las tareas de evaluación con base en reactivos equivalentes que son técnicamente isomorfos; 3) generar de manera automática, en un sistema de cómputo, las tareas de evaluación, permitiendo el control de niveles de dificultad; y 4) calificar las respuestas con un esquema de crédito parcial. Este modelo está integrado por dos componentes: el Examen de Competencias Básicas (EXCOBA) y un editor de reactivos computarizado integrado por el programa GenerEx (generador de exámenes).<sup>1</sup>

## MARCO CONCEPTUAL

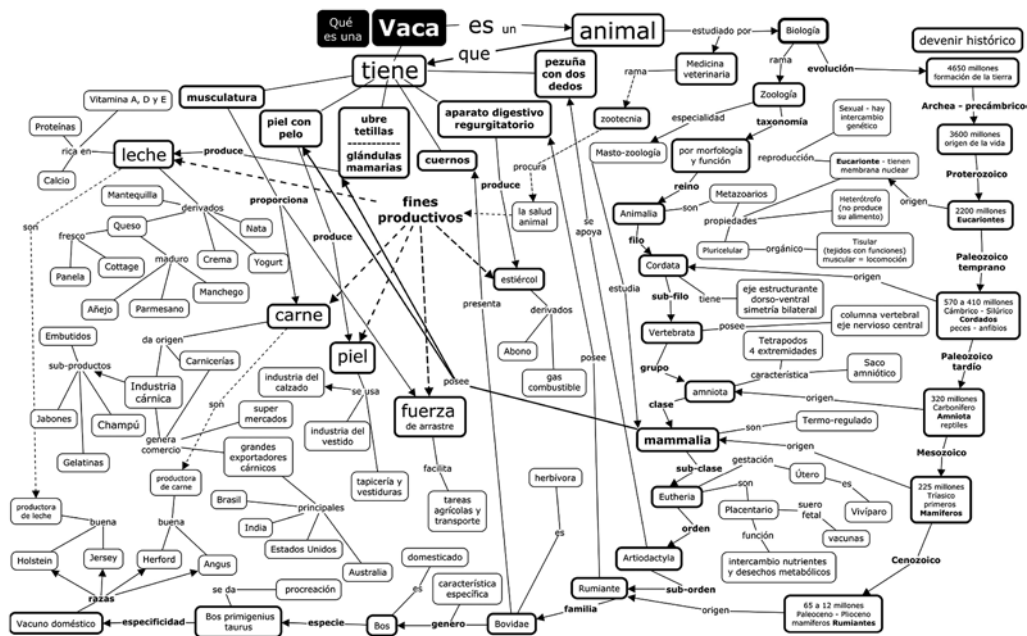
La evaluación de los aprendizajes requiere partir de un paradigma que tenga una concepción de la educación como un sistema complejo, bajo una visión del aprendizaje entendido como las resultantes de procesos multifactoriales, en los que se originan efectos interactivos no lineales entre los factores, lo cual ocurre en procesos longitudinales que forman trayectorias a lo largo del tiempo, ante diversas circunstancias, distintas situaciones, en múltiples instituciones y organizaciones; por todo ello se hace necesario concebir análisis sistémicos (Lemke y Sabelli, 2008).

Los primeros planteamientos para explicar los procesos de la apropiación del conocimiento desde el punto de vista psicológico aparecen con las primeras teorías de la cognición, las cuales tuvieron su origen en los años veinte del siglo pasado, con los trabajos experimentales de los psicólogos alemanes de la escuela de la Gestalt. Posteriormente siguió un gran desarrollo de las teorías cognoscitivas a mediados del siglo XX, particularmente en Europa, con los planteamientos psicogenéticos de Jean Piaget y psicosociales de Lev Vygotsky, y en los Estados Unidos con los de Jerome Bruner y David Ausubel.

Actualmente el constructivismo es el paradigma dominante en la psicología educativa; concibe al desarrollo del conocimiento como un proceso de construcción propia del educando, en tanto no se da una reproducción fiel de los estímulos provenientes del ambiente externo, ni tampoco la codificación apegada a un texto o discurso del profesor (Serrano y Pons, 2011). En el paradigma constructivista hay una controversia epistemológica que surge de la atribución del carácter interno o externo en el proceso de construcción del conocimiento: unos enfatizan más el carácter individual (interno) y otros el carácter social de la construcción bajo una concepción socio-cultural (Coll, 2001).

<sup>1</sup> Para mayor información sobre ambos desarrollos consúltese la página de Métrica Educativa: [www.metrica.edu.mx](http://www.metrica.edu.mx) (consulta: 21 de mayo de 2015).

Figura 1. Campo semántico representado en un mapa conceptual



Fuente: elaboración propia.

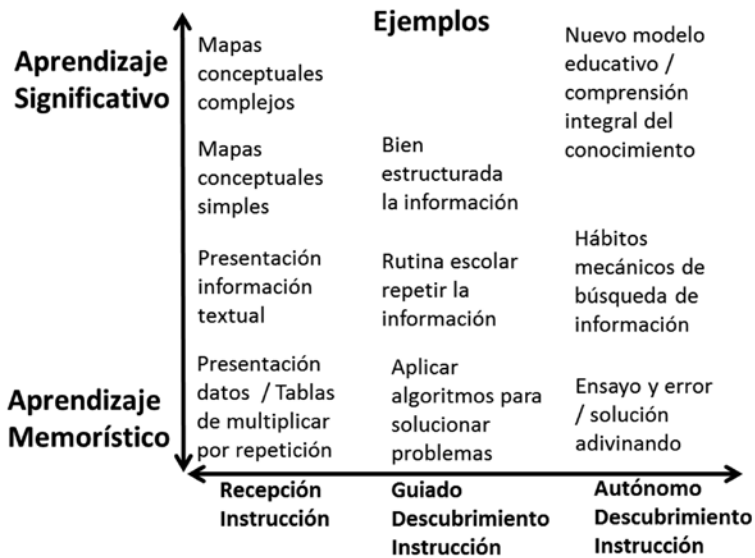
La tesis central del paradigma constructivista establece que el conocimiento y el aprendizaje no son resultado de la experiencia directa, sino el producto de una actividad cognitiva constructiva, donde las personas interpretan las experiencias de vida y construyen representaciones que se codifican en una estructura o malla semántica dotada de significado y significación. Esta posición es compartida por las distintas tendencias de la investigación psicológica y educativa contemporáneas. Sin embargo, se pueden identificar *grosso modo* tres posiciones: el constructivismo cognitivo, con raíces en la psicología y epistemología genéticas de Jean Piaget, donde se enfatiza la construcción de la estructura de conocimientos por parte del aprendiz, bajo un proceso de equilibrio entre la asimilación y la acomodación (Piaget, 1970); el constructivismo de orientación social, inspirado en los planteamientos de Lev Vygotsky, que concibe el proceso del desarrollo del pensamiento como un fenómeno social que está mediado básicamente por el lenguaje (Vygotsky, 1986);

y el construccionismo socio-cultural que sustenta Mikel Cole, entre otros, con un enfoque en el que se pone mayor énfasis en la elaboración del conocimiento a partir de las mediaciones y el contexto socio histórico cultural del aprendiz. En este tercer enfoque se destacan también las prácticas lingüísticas, pero concebidas como artefactos que median la actividad cognitiva, lo que plantea una visión de una psicología cultural con una orientación más antropológica (Cole, 1998).

El Examen de Competencias Básicas está basado, particularmente, en algunos aspectos del constructivismo cognitivo socio-cultural, que está sustentado en que el conocimiento de las personas se codifica en estructuras que forman mallas o campos semánticos (Luria, 1981). La *teoría de campo* no sólo establece las relaciones conceptuales de una estructura de cognición jerarquizada, sino que asume que la evocación conceptual tiene relaciones con distintas valencias (fuerzas de enlace). El EXCOBA está diseñado para evaluar competencias académicas básicas que adquieren las



Figura 2. Representación ortogonal del aprendizaje



Fuente: Novak y Cañas, 2008.

personas durante su escolarización, las cuales constituyen elementos que permiten tener éxito escolar en niveles de estudio superiores y en la vida.

El conocimiento se codifica en categorías que irradian un campo semántico de significación, donde hay relaciones jerárquicas (supra-ordinadas / ordinadas / sub-ordinadas), en tanto unos conceptos incluyen a otros, por ejemplo, el concepto *vaca*, está incluido en conceptos generales (supra-ordinados) como ser un animal mamífero. Hay conceptos específicos que la definen (sub-ordinados), como tener cuernos y dar leche, que incluye a sus derivados, como mantequilla y quesos; y éstos a los diferentes tipos, etcétera. La estructura de cognición ha dado lugar a concebir los *mapas conceptuales*, que corresponden a una visualización de un campo semántico. En el diagrama de la Fig. 1 se presenta un ejemplo de irradiación conceptual de un *campo semántico* a partir del concepto *vaca*, representado en un mapa conceptual.

Novak y Cañas (2008) plantean que hay dos dimensiones ortogonales en los procesos de aprendizaje: la primera va del aprendizaje

de datos que se logra por simple *recepción*, con muy limitado significado, y transita hacia el aprendizaje *significativo* hasta llegar a campos complejos del conocimiento (representado en el eje Y de la Fig. 2); la otra dimensión se refiere a las dinámicas del aprendizaje, que inicia *por recepción* (pasivo y memorístico) y evoluciona hacia el aprendizaje *por descubrimiento* (activo y por comprensión - *insight*), representados en el eje X.

El planteamiento es que en el aprendizaje *por recepción* el contenido a aprender se presenta al alumno en su forma final y el estudiante debe asimilar o incorporar los conceptos a su estructura de conocimientos de modo arbitrario, por un procedimiento memorístico basado en la simple *repetición*, de lo que resulta un conocimiento estático que tiene poca o nula operatividad cognitiva. Por el contrario, en el aprendizaje *por descubrimiento*, el alumno debe reorganizar la información, integrarla o ajustar su estructura cognoscitiva existente al transformarla para obtener el producto final deseado. En la segunda dimensión, el aprendizaje *significativo* ofrece operatividad cognitiva; esto implica

que el objeto de aprendizaje se relaciona de modo sustancial y no arbitrario con la estructura de conocimientos del alumno, con lo que ya sabe, y por lo mismo adquiere operatividad cognitiva que estimula una actitud a favor del aprendizaje (Ausubel *et al.*, 1978).

La comprensión se da cuando hay construcción de significado que permite la operatividad cognitiva, la reflexión que posibilita entender las relaciones de pertinencia que subyacen a una idea o proposición. Entre más profunda es la reflexión, mayor es la significación, las relaciones de pertinencia son más extensas (como en el ejemplo del concepto *vaca*), y por lo mismo, la fuerza de la huella mnémica es más resistente al olvido. En el proceso de evaluación se busca evaluar competencias ubicadas en los extremos de los ejes ortogonales (presentados en el cuadro de Novak y Cañas, 2008), relaciones conceptuales extensas y desarrollo autónomo, que equivale a reflexionar a profundidad.

La mayor eficacia del aprendizaje significativo reside en cuatro grandes ventajas: facilita nuevos aprendizajes, produce efectos de memoria más resistentes al olvido, se superan detalles concretos (precisiones) y, lo más importante, genera la comprensión que habilita la reflexión para construir nuevas respuestas, tanto críticas como creativas. Esta orientación evoca distintas actitudes hacia el aprendizaje y la evaluación, ya que es muy diferente a la evaluación tradicional, que suele requerir el

recuerdo literal del aprendizaje original, así como el reconocimiento de respuestas dadas para evaluar la comprensión. Los dos tipos de aprendizaje planteados por Ausubel *et al.* (1978) no son dicotómicos ni excluyentes; coexisten y se conforman en un continuo, como se planteó en la Fig. 2. En consecuencia, en las nuevas estrategias de evaluación educativa se prioriza el aprendizaje significativo, y se tiende a dejar relegada la estrategia de opción múltiple, por estar más relacionada con procesos de identificación puntual de conocimiento memorizado.

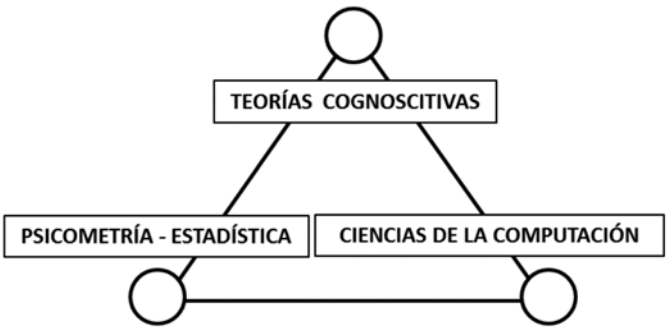
De acuerdo con lo expuesto en este apartado, el EXCOBA pretende recuperar los aprendizajes significativos que tuvieron lugar tanto en la educación básica como en la media superior, ocurridos por comprensión o por descubrimiento (*insight*). El examen evalúa aquellos conceptos y procedimientos expresados en el currículo que el estudiante logró relacionar sustancialmente, y que permanecieron de manera organizada y no arbitraria en su estructura cognoscitiva.

### EL MODELO DEL EXCOBA

El modelo del Examen de Competencias Básicas se sustenta en tres soportes teóricos de gran fortaleza: las teorías de la cognición, la psicometría y las ciencias de la computación, como se ilustra en la Fig. 3.

En el diseño conceptual, el tipo de reactivos utilizados en el examen corresponde a

Figura 3. Soportes teóricos



Fuente: elaboración propia.



lo que denominamos *reactivos estructurales constructivos*. El concepto de *reactivo estructural* es porque la actividad a evaluar apela a un campo semántico del conocimiento, y no a un contenido puntual. Por lo tanto, el sustentante debe mostrar la capacidad para responder a un sistema de conceptos coherentes, entrelazados, organizados, estructurados, como es en química la tabla periódica de los elementos, y no a un concepto único y desarticulado, como sería mencionar o reconocer el número o peso atómico de un elemento dado. El nombre de *reactivo constructivo* refiere al constructivismo cognitivo, ya que el sustentante deberá construir sus propias respuestas, como comprender que la fracción de  $\frac{2}{5}$  de un área dada equivale a 0.4, o que corresponde al 40 por ciento; en este ejemplo, se requiere que construya la respuesta marcando la parte del área que corresponda a la fracción indicada en cualquiera de sus tres expresiones ( $\frac{2}{5}$ , 0.4, 40 por ciento).

Otra particularidad de los reactivos estructurales es la cantidad de respuestas solicitadas por reactivo: hay ítems donde el resultado es único y sólo puede ser calificado como correcto o incorrecto (ítems dicotómicos); pero en otros ítems se requiere de múltiples respuestas, dos o más (ítems politómicos), de manera que se puede calificar en función del número de respuestas que el sustentante pudo contestar correctamente con un esquema de *crédito parcial*. Un ejemplo puede ser identificar seis estados y seis capitales en un mapa de la república mexicana, de manera que hay 12 respuestas, de las cuales unas pueden estar bien y otras no; en este caso el crédito se otorga de acuerdo al número de respuestas correctas, de manera que si contesta seis bien se otorga 0.5 de punto; si fueran nueve se daría 0.75 de punto, es decir, proporcional al número de aciertos.<sup>2</sup>

Los reactivos estructurales no necesariamente descartan el formato de selección de

respuesta, pero se dan dentro de las exigencias de diferentes contextos a evaluar por medio de planteamientos estructurales, como se verá en un ejemplo en el que se requiere seleccionar *has* (de haber) o *haz* (de hacer). Los procesos de evaluación no se pueden reducir a un único formato, sino que se configuran y distribuyen en diferentes formatos, de acuerdo con los campos semánticos. Los reactivos estructurales se pueden plantear en cualquier campo del saber: en muchos casos, como en la aritmética, la física, la química, la geometría y el álgebra, el sustentante debe construir la respuesta y anotar la solución numérica o literal, según se requiera; pero en otros casos los ítems presentan campos estructurados y se solicita al sustentante construir su respuesta organizando los componentes de la tarea evaluativa.

Como se indicó, otro de los componentes teóricos del modelo se sustenta en la *psicometría*, que también tiene sus orígenes a principios del siglo XX, con el primer test de inteligencia para medir la edad mental de los estudiantes en una escala, elaborada en 1908 por Alfred Binet y Théodore Simon. La psicometría constituye un cuerpo disciplinario desarrollado por más de un siglo para hacer mediciones rigurosas de los fenómenos psicológicos, como son la inteligencia, el aprendizaje y la personalidad, a partir de elementos que permitan medir, con sustento empírico, dichos atributos psicológicos; es así como se ha pretendido obtener indicadores objetivos, precisos, estables, válidos y confiables, lo que ha dado lugar a la construcción de test y pruebas, entre otras, las relacionadas con el rendimiento académico o escolar. Para fortalecer sus análisis, la psicometría se ha valido del uso de las matemáticas y, en particular, de la teoría de la probabilidad, que cristalizó en la *teoría clásica de la medida*. Hay otros planteamientos, como la *teoría de respuestas al ítem*, que a diferencia de la primera no parte

2 Si fuera necesario, los reactivos podrían ser ponderados, de tal manera que se les den pesos diferenciados de acuerdo con su importancia curricular o nivel cognitivo.

de la distribución normal de la dificultad de los ítems dentro de un continuo, sino que asume que el nivel de dificultad es propio de cada ítem, es decir, es independiente; se trata de estimar la probabilidad de que una persona con determinado nivel de competencias pueda contestar correctamente una pregunta dentro de un rango de dificultad dado en una escala. La psicometría ha logrado análisis tan complejos y sustentados que muchos especialistas los han reconocido como aportaciones relevantes al desarrollo del conocimiento matemático.

El tercer componente está dado por las *ciencias de la computación*, las cuales se han desarrollado vertiginosamente desde finales del siglo pasado, transformando todos los campos del saber, las tecnologías, aplicaciones industriales y telemáticas. En el EXCOBA el modelo está sustentado en procesos de ingeniería de tests que hacen posible aplicar algoritmos automatizados, como es el caso de GenerEx, que permite evaluar campos semánticos a través de generar *reactivos isomorfos de respuesta construida o semi-construida*.

El ambiente computarizado facilita la mediación para poder aprovechar diversos códigos de conocimiento, como la representación visual, dada la facilidad de incorporar y recrear imágenes dinámicas que son factibles de ser transformadas por el sustentante; además, permite medir el tiempo de reacción y respuesta de cada sustentante en cada uno de los reactivos de la prueba, lo cual ofrece ventajas que son impensables en el formato tradicional de lápiz y papel. La aplicación computarizada constituye una nueva dimensión.

## PROPIEDADES DEL EXCOBA

Después de 20 años de investigación e innovación, se logró concebir un nuevo paradigma y consolidar el Examen de Competencias Básicas, el cual aprovecha las ventajas antes enunciadas de la cognición, la psicometría y la tecnología digital. Este examen

computarizado de aplicación a gran escala fue diseñado para seleccionar y diagnosticar a los estudiantes que aspiran ingresar a los niveles de educación media superior (EMS) y educación superior (ES). El examen evalúa competencias a partir de la operatividad cognitiva de los aprendizajes básicos y estructurales que los estudiantes debieron adquirir durante su trayectoria escolar. Su originalidad consiste en la concepción, diseño e instrumentación de un modelo de evaluación de competencias educativas (para la escuela y para la vida), concebido con una visión estructuralista y constructivista, que es más cercana a los problemas que se presentan en la vida cotidiana; lo que es referido como *evaluación auténtica*.

La selección del contenido se realiza mediante la identificación, en el currículo, de aprendizajes esperados que se consideren básicos y estructurales, es decir, que sean el sustento conceptual que da andamiaje para poder estructurar y comprender otros conocimientos que son más complejos; también se considera la relevancia, en tanto permiten integrar otros conocimientos que dan competencias para la vida dado un contexto socio-cultural. Por ejemplo, al definir la relevancia de países o ciudades algunos resultan de mayor importancia para ciertos contextos: países como España o ciudades como Washington tienen mayor relevancia para los mexicanos dadas las múltiples relaciones políticas, económicas, históricas y culturales que se tienen con ellos, en comparación con países como Kenia o ciudades como Asunción, donde las relaciones son menos relevantes y frecuentes.

La estructura del examen se basa en el análisis conceptual de los planes y programas de estudio, lo que requiere un riguroso análisis longitudinal de las articulaciones de los contenidos curriculares para identificar las habilidades y conocimientos sustantivos implicados en las diversas disciplinas (Backhoff *et al.*, 2006). Se trata de reconocer los conocimientos básicos estructurales; observar cómo

Figura 4. Estructura del EXCOBA por niveles de dificultad



Fuente: elaboración propia.

a lo largo del currículo éstos se amplían y complejizan de manera articulada. De aquí la estructura de complejidad multinivel que define los niveles de dificultad (Fig. 4). Por ejemplo, se identifica cuándo es la primera vez que se presenta en el currículo el concepto de átomo, y se da seguimiento para observar el tratamiento que se le da a esta noción en niveles superiores. El concepto de átomo es básico y estructural para poder comprender conceptos más complejos como el de isótopo, porque se requiere comprender la estructura atómica (núcleo) y sus componentes (protones y neutrones) para poder entender que el desbalance forma un isótopo. Se trata, entonces, de evaluar conocimientos básicos que se precisan en los planes de estudio de los programas de educación obligatoria, y que se definen como aprendizajes esperados.

Así mismo, se presenta una propuesta innovadora con relación a la forma de evaluar dichas competencias curriculares, en el sentido de alejarse del formato tradicional de selección de respuesta. En el nuevo modelo se utilizan formatos de respuesta construida y semi-construida, que implican la ejecución de respuestas en contexto. Así resultan los reactivos más naturales o auténticos para evaluar el aprendizaje, dado que en la vida cotidiana se

responde en contextos, de acuerdo a la “evaluación auténtica” (Chittenden, 1991; Darling-Hammond y Snyder, 2000).

El modelo del EXCOBA consiste en presentar tareas evaluativas complejas, que requieren del sustentante construir las respuestas o integrarlas dentro de un esquema o contexto. Esto permite otorgar crédito parcial a las respuestas, con base en sus aciertos, errores y omisiones, con lo que se mejora la precisión de la evaluación.

## PROTOTIPOS DE REACTIVOS

El GeneEx es capaz de proveer una infinidad de versiones de reactivos isomorfos estadísticamente equivalentes en diferentes campos del conocimiento que son definidos en una *tabla de especificaciones*.

En el campo del conocimiento matemático, tanto los indicadores nacionales obtenidos por Excale (Examen para la calidad y el logro educativo), como los internacionales de PISA (Programme for International Student Assessment), permiten apreciar que una de las áreas de mayor dificultad para ser comprendida por los estudiantes son las relaciones numéricas básicas, como porcentajes, decimales o fracciones.


Figura 5. Reactivo de matemáticas

**A**

Selecciona las partes de la figura que indica la fracción.  
Haz clic sobre las partes que elijas, y si deseas desmarcar haz clic nuevamente sobre ellas.

La figura contiene 25 fracciones

Fracción:  $\frac{2}{5}$

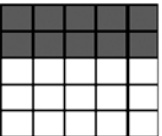


**B**

Selecciona las partes de la figura que indica la fracción.  
Haz clic sobre las partes que elijas, y si deseas desmarcar haz clic nuevamente sobre ellas.

La figura contiene 25 fracciones

Fracción:  $\frac{2}{5}$




**C**

Selecciona las partes de la figura que indica la fracción.  
Haz clic sobre las partes que elijas, y si deseas desmarcar haz clic nuevamente sobre ellas.

La figura contiene 25 fracciones

Fracción:  $\frac{2}{5}$



Fuente: elaboración propia.

Una manera de evaluar esta habilidad es pidiéndole al estudiante que marque el número de partes que son necesarias para igualar la fracción que se solicita, en este caso, dos quintos ( $\frac{2}{5}$ ), como se ve en la Fig. 5-A.

Aquí el sustentante debe construir su respuesta marcando el número de partes que correspondan a dos quintos de la figura (el cuadrado). No hay un orden predefinido y puede haber muchas respuestas para responder este reactivo correctamente. Con esto se rompe la idea común de que ante una pregunta hay, y debe haber, solamente una respuesta correcta, más cuando se trata de un problemas de matemáticas. En la Fig. 5 se observan dos opciones distintas de respuestas (B y C) entre muchas otras posibles, que son igualmente correctas y equivalentes, aunque son distintas; lo que muestra que es un reactivo de comprensión.

En la Fig. 6 se puede ver un reactivo isomorfo de la pregunta anterior, en tanto que

explora la misma área de conocimiento (corresponde al mismo contenido), y tiene un nivel de dificultad equivalente; de igual manera, es de respuesta construida, y puede haber múltiples formas de contestarlo correctamente.

Para entender el funcionamiento del GenerEx en este tipo de reactivos, de una plantilla de ítems isomórficos la computadora selecciona una fracción ( $\frac{2}{5}$ ,  $\frac{3}{5}$ ,  $\frac{4}{5}$ ,  $\frac{2}{6}$ ,  $\frac{3}{6}$ ,  $\frac{4}{6}$ , etc.), de manera que la figura contenga múltiplos de la fracción correspondiente. Así se puede presentar un número indefinido de reactivos equivalentes en dificultad y contenido. Es importante resaltar que puede haber muchos reactivos isomórficos cambiando las figuras (octágonos, rombos, cuadrados, círculos, etc.), así como su color y tamaño, por lo que nuevamente se puede decir que es muy difícil que dos sustentantes tengan un ítem con la misma fracción, la misma figura, del mismo color y en el mismo momento; por lo

Figura 6. Reactivo isomorfo

Selecciona las partes de la figura que indica la fracción.  
Haz clic sobre las partes que elijas, y si deseas desmarcar haz clic nuevamente sobre ellas.

La figura contiene 10 fracciones

Fracción:  $\frac{2}{5}$



Fuente: elaboración propia.

tanto, es casi imposible copiar en este tipo de exámenes, pero aún más importante, también es muy difícil obtener aciertos por respuestas dadas al azar.

En la Fig. 7 se presenta un reactivo correspondiente al área de lenguaje. Se evalúa la comprensión ortográfica en el contexto de una narrativa que constituye un campo semántico que otorga significado específico a las opciones de respuesta. No se trata de reglas ortográficas puntuales de memorización, sino que se refiere a la comprensión del uso apropiado de acentos diacríticos como: *esta* cuando es adjetivo demostrativo, *ésta* cuando es pronombre y *está* cuando es verbo, o una letra que transforma la función de un verbo, como: *has* de haber y *haz* de hacer. Si una persona no

reconoce estas diferencias, sus competencias para expresarse apropiadamente por escrito estarán más limitadas en comparación con quien sí las domina. Lo relevante es que la competencia tanto de escritura como de lectura están vinculadas, de aquí que es posible obtener alguna valoración por medio de estos indicadores, en tanto habrá una correlación entre quien comprende estas reglas ortográficas, sabe escribir con mayor propiedad y es más asiduo/a a la lectura, es decir, hay correlatos que ofrecen indicadores indirectos del dominio en ciertos campos del conocimiento.

En el primer recuadro (A) de la Fig. 7 se da un ejemplo de un reactivo de comprensión de reglas ortográficas dentro de un contexto de escritura. En el segundo recuadro (B) se puede

Figura 7. Reactivo de competencias verbales

Lee con atención el siguiente fragmento y escoge (da clic) la palabra que corresponde al texto de acuerdo a las reglas ortográficas.

has - haz / está - esta / qué - que / sí - si / cómo - como / tu - tú

A María \_\_\_\_ comiendo. Su papá le dijo, primero \_\_\_\_ tarea, y vamos al cine, \_\_\_\_ no la \_\_\_\_ terminado \_\_\_\_ a las 4, ya no vamos. María preguntó ¿a \_\_\_\_ hora empieza la película? Su papá le dijo: \_\_\_\_, ayer \_\_\_\_ me dijiste que a las 5, ella dijo \_\_\_\_ es cierto.

B María esta comiendo. Su papá le tu primero tu tarea, y vamos al cine, si no la tu terminado \_\_\_\_ a las 4, ya no vamos. María preguntó ¿a \_\_\_\_ hora empieza la película? Su papá le dijo: \_\_\_\_, ayer \_\_\_\_ me dijiste que a las 5, ella dijo si es cierto.

C María está comiendo. Su papá le dijo, primero haz tu tarea, y vamos al cine, si no la has terminado como a las 4, ya no vamos. María preguntó ¿a qué hora empieza la película? Su papá le dijo: cómo, ayer tú me dijiste que a las 5, ella dijo sí es cierto.

Fuente: elaboración propia.

Figura 8. Reactivo comprensión tabla periódica

Coloca (arrastra) los núcleos atómicos de los elementos que se presentan en la casilla correspondiente en la **Tabla Periódica**, y escribe su nombre y su símbolo:

Contiene 6 protones      Contiene 8 protones      Contiene 3 protones      Contiene 30 protones

Núcleos atómicos con protones

1 1.0079 <b>H</b> HIDRÓGENO	2 4.0026 <b>He</b> HELIO
3 6.941 <b>Li</b> LITIO	4 9.0122 <b>Be</b> BERILIO
11 22.990 <b>Na</b> SODIO	12 24.305 <b>Mg</b> MAGNESIO
19 39.098 <b>K</b> POTASIO	20 39.098 <b>Ca</b> CALCIO
37 85.468 <b>Rb</b> RUBIDIO	38 87.62 <b>Sr</b> ESTRONCIO
55 132.91 <b>Cs</b> CESIO	56 137.33 <b>Ba</b> BARIO
87 223.019 <b>Fr</b> FRANCIO	88 226.025 <b>Ra</b> RADIO
21 44.956 <b>Sc</b> ESCANDIO	22 47.867 <b>Ti</b> TITANIO
23 50.942 <b>V</b> VANADIO	24 51.086 <b>Cr</b> CROMO
25 54.938 <b>Mn</b> MANGANESES	26 55.845 <b>Fe</b> HIERRO
27 58.933 <b>Co</b> COBALTO	28 58.933 <b>Ni</b> NIOBIO
29 63.546 <b>Cu</b> COBRE	30 65.38 <b>Zn</b> ZINC
31 69.723 <b>Ga</b> GALIO	32 72.64 <b>Ge</b> GERMANIO
33 74.922 <b>As</b> ARSENICO	34 75.94 <b>Se</b> SELENIO
35 79.904 <b>Br</b> BROMO	36 83.798 <b>Kr</b> KRIPTON
39 88.906 <b>Y</b> ITRIO	40 90.907 <b>Zr</b> ZIRCONIO
41 91.224 <b>Nb</b> NIOBIO	42 92.906 <b>Mo</b> MOLEBDENO
43 95.94 <b>Tc</b> TECNICIO	44 97.907 <b>Ru</b> RUTENIO
45 101.07 <b>Rh</b> RODIO	46 106.42 <b>Pd</b> PALADIO
47 107.87 <b>Ag</b> PLATA	48 112.41 <b>Cd</b> CADMIO
49 114.82 <b>In</b> ESTADIO	50 118.71 <b>Sn</b> ESTADIO
51 127.60 <b>Sb</b> ANTIMONIO	52 127.60 <b>Te</b> TELURO
53 126.91 <b>I</b> YODO	54 126.91 <b>Xe</b> XENON
57 138.91 <b>La</b> LANTANIO	58 140.91 <b>Ce</b> CELESTIO
59 140.91 <b>Pr</b> PRASEODIMIO	60 144.24 <b>Nd</b> NEODIMIO
61 144.24 <b>Pm</b> PROMETIO	62 150.36 <b>Sm</b> SAMARIO
63 151.96 <b>Eu</b> EUROPIO	64 157.25 <b>Gd</b> GADOLINIO
65 158.93 <b>Tb</b> TERBIO	66 162.50 <b>Dy</b> DYSMADIO
67 164.93 <b>Ho</b> HOLMIO	68 167.26 <b>Er</b> ERBIO
69 168.93 <b>Tm</b> TERCIO	70 173.05 <b>Yb</b> YTERBIO
71 174.97 <b>Lu</b> LUTECIO	
89 223.019 <b>Ac</b> ACTINIO	90 226.025 <b>Th</b> TORMIO
91 227.03 <b>Pa</b> PASTORIO	92 238.03 <b>U</b> URANIO
93 238.03 <b>Np</b> NEPTUNIO	94 244.04 <b>Pu</b> PLUTONIO
95 244.04 <b>Am</b> AMERICIO	96 247.07 <b>Cm</b> CURCIO
97 247.07 <b>Bk</b> BERKELIO	98 251.08 <b>Cf</b> CALIFORNIO
99 251.08 <b>Es</b> EINSTEINIO	100 252.08 <b>Fm</b> FERMIUM
101 252.08 <b>Md</b> MENDELIVIO	102 258.10 <b>No</b> NOBELIO
103 259.10 <b>Lr</b> LAWRENCIO	

**LANTANIDOS**

**ACTINIDOS**

Fuente: elaboración propia.

apreciar la manera en que se comporta el sistema cuando el sustentante contesta. Obsérvese que al colocar el cursor en la línea de la palabra demandada, automáticamente se despliegan varias opciones de respuesta, de las que se requiere seleccionar por comprensión, no por memorización, la que es apropiada para el contexto específico. En el último recuadro (C) se ofrece el reactivo cuando se ha respondido apropiadamente.

Es importante resaltar que el reactivo no se contesta con una respuesta, sino que requiere de 10 respuestas, de tal forma que se puede calificar con crédito parcial; de esta manera se hace posible evaluar con mayores niveles de precisión. El conocimiento se da en un continuo: en un campo semántico se pueden tener nociones, se puede saber más o menos sobre cierto tema. El conocimiento no es dicotómico: sí se sabe o no se sabe; es politémico, en el sentido de que se tienen ciertos conocimientos y otros se ignoran.

En la Fig. 8 se presenta un prototipo de reactivo para evaluar conocimientos de química. Un conocimiento fundamental básico estructural en química es la tabla periódica, dado que organiza, clasifica y distribuye a los elementos químicos conforme a sus propiedades y características; constituye una estructuración de todos los elementos a partir de su número o peso atómico, y conforma un ordenamiento en periodos y grupos que se distinguen por sus propiedades y características físico-químicas. La tabla periódica constituye un campo semántico, un mapa conceptual cuya comprensión permite la operatividad cognitiva. Solicitar al sustentante que escriba el nombre y símbolo del elemento a partir del número atómico no trae consigo una respuesta de memorización, ya que lo que se requiere es comprender la organización de la tabla y leer para transcribir las respuestas puntuales. De igual manera se podría solicitar en otro reactivo isomórfico que transcriba



el valor del peso atómico de los elementos que se cuestionan, ya que igualmente están dados en la tabla.

Este reactivo permite una evaluación auténtica porque en la vida cotidiana todos podemos fácilmente consultar la tabla periódica. Es absurdo pretender memorizarla, a pesar de que todavía hay profesores que así lo requieren y que lo que evalúan es su memorización.

Ahora se presentan algunos prototipos de reactivos en el terreno de los conocimientos sociales relacionados con el área de geografía e historia. Igualmente en estas áreas se presentan campos de conocimiento estructurados donde el sustentante tiene que integrar (construir) su respuesta. Para ejemplificar

hemos considerado la evaluación de los conocimientos de geografía política que tiene una persona. Para evaluar este tipo de conocimiento, en un reactivo del EXCOBA se le presenta al sustentante un mapa de la geografía política de México con los 32 estados y sus capitales, sin sus nombres, y a un lado se ofrece un listado con el nombre de seis estados y seis capitales, los cuales son seleccionados al azar por el sistema computarizado atendiendo un criterio de distribución: dos del norte, dos del centro y dos del sureste. Se solicita que seleccione cada uno de los nombres que están en el listado (haga clic sobre el nombre) y los “arrastre” (*drag*) hasta el lugar correspondiente en el mapa, tal y como se ve en la Fig. 9.

Figura 9. Reactivo de geografía política



Fuente: elaboración propia.

De esta manera se explora el conocimiento geográfico que tiene el sustentante sobre un contexto general estructurado, lo que resulta en un proceso de evaluación más auténtica, ya que se apega a las circunstancias en que se suele presentar en la vida cotidiana el conocimiento de la geografía política del país. Esto contrasta mucho con lo que sería una pregunta similar formulada en opción múltiple, tal como: “la capital del Estado de México es: a) Chetumal, b) Campeche, c)

Toluca, d) Mexicali”. En este reactivo de opción múltiple sólo hay una opción correcta, mientras que el reactivo estructural se construye con 12 posibles respuestas, se exploran más conocimientos y se requiere saber más. El reactivo estructural podría contener más elementos o menos, y se puede calificar con crédito parcial, es decir, tener una evaluación ponderada en función del número de elementos que se hayan colocado de manera correcta.

La decisión de cuántos elementos debe contener cada reactivo, y de si se debe dar crédito parcial ponderado o redondear, se tiene que tomar con base en los análisis empíricos que permitan obtener los mejores indicadores psicométricos de validez predictiva. Esto es así porque en los procedimientos digitales (la analítica de datos, *big data*) es posible registrar todas las respuestas y que éstas vayan alimentando una base de datos; de esta manera, el sistema puede procesar, seleccionar y organizar las preguntas de acuerdo al registro histórico por orden de dificultad, definida de manera operativa: muy fáciles (preguntas en que aciertan los sustentantes en más de 80 por ciento en promedio), fáciles (donde acierta entre 60 y 79 por ciento), de dificultad media (acierta entre 40 y 59 por ciento), difícil (acierta entre 20 y 39 por ciento), y muy difícil (acierta no más de 19 por ciento). Para el caso del ejemplo de los estados y capitales de México, se puede programar para que el sistema presente estados del norte, del centro y del sur, en cada caso uno de ellos difícil y el otro fácil, lo que permitirá estructurar reactivos de doble criterio: distribución geográfica y dificultad.

En el reactivo de opción múltiple presentado puede haber un error de sesgo de orden fonético (Mexicali - Estado de México), lo cual queda anulado en el formato estructural, porque se responde a una representación integral en un campo estructurado (un mapa). Otro sesgo en la pregunta de opción múltiple es que quienes son originarios del Estado de México o entidades federativas vecinas se verán beneficiados porque les será más fácil saber la respuesta correcta que a una persona que es oriunda de un estado lejano, es decir, la referencia a Toluca será más frecuente para quien vive en ella o cerca de ella. Este sesgo no se da en el reactivo estructural constructivo, ya que, como se ha dicho, el examinado requiere colocar el nombre de seis estados y seis capitales, dos del norte, dos del centro y dos del sur, lo que neutraliza los sesgos favorables o desfavorables por la región de procedencia;

todos tendrán equiparadas, de alguna manera, las ventajas de la cercanía y las desventajas de la lejanía.

Lo más relevante de este tipo de reactivos es que evalúa conocimientos estructurales que son consustanciales para la comprensión y asimilación de conocimientos vinculados a esta área semántica. El conocimiento no se da atomizado. Por ejemplo, cuando se conoce una noticia y se tiene conocimiento del lugar de su ocurrencia, la nota cobra otro significado y significación, porque hay mayor irradiación semántica. Si se analiza el conocimiento de geografía política con otros conocimientos relacionados, se observa que hay correlación: no es común saber el nombre del gobernador de un estado y no saber de qué estado se trata y dónde se ubica. Si se ignora esto es claro que el significado y la significación semántica son pobres. Si se explora el conocimiento que se tiene de las estructuras de gobierno (municipio - estado - federación), de la división de poderes (legislativo - ejecutivo - judicial), y de los gobernantes, se observa que hay una correlación que es indicativa de niveles de dominio de esta área de conocimiento. No será lo mismo quien asiste a votar y no logra reconocer la diferencia entre candidatos a la diputación estatal y la federal, respecto de quien sí la distingue. Esto es lo que da sustento profundo a este procedimiento de evaluación por reactivos estructurales, en tanto explora campos semánticos.

Esta misma estrategia se puede seguir para obtener indicadores de niveles de dominio de áreas de conocimientos en geografía internacional, al solicitar la ubicación en un mapamundi de ocho países y ocho capitales, dos del continente americano, dos del europeo, dos del africano y dos del asiático; uno fácil y otro difícil, definidos operativamente. El estructurar reactivos con preguntas fáciles y difíciles es importante porque así se construyen escalas: por ejemplo, identificar la capital de Estados Unidos (Washington) será más fácil que la de Paraguay (Asunción), dadas

las marcadas diferencias en las implicaciones políticas, económicas, históricas y culturales que se tienen. El poder de discriminar se da al contar con escalas (preguntas muy fáciles, fáciles, medio fáciles, medio difíciles, difíciles, muy difíciles), cuya congruencia entre el poder de discriminación y confiabilidad se puede medir con la prueba Alpha de Cronbach. Una prueba no es un agregado de preguntas, es un diseño estructurado y balanceado de preguntas que componen escalas de media; para juzgar si una persona es alta o baja de estatura, se requiere tener una escala de referencia definida dentro de un contexto.

De igual manera, en un mapamundi se puede explorar la ubicación de sitios

históricos, aportaciones culturales, personajes históricos y políticos; o bien de otra naturaleza, como ríos, montañas, fauna, puntos de coordenadas, etcétera, en tanto constituyen indicadores del dominio de ciertos campos semánticos de conocimiento.

Para finalizar se ofrece en la Fig. 10 el prototipo de un reactivo en formato digital apoyado por representación visual para valorar conocimientos de historia de México. En contraste a un planteamiento de opción múltiple, el ítem consiste en solicitar a los sustentantes que acomoden (arrastrando) las imágenes en la línea del tiempo, atendiendo el orden en que gobernaron los personajes históricos que se presentan.

Figura 10. Reactivo de historia de México



Fuente: elaboración propia.

En la pregunta subyace una estructura cronológica fundamental para poder organizar y comprender la revolución mexicana, ya que Díaz fue derrocado por el movimiento revolucionario que emplazó Madero, Madero fue derrocado por un golpe de Estado que encabezó Huerta, Huerta fue derrotado por las fuerzas revolucionarias encabezadas por Carranza, y Carranza fue asesinado por orden de Obregón. La respuesta implica un problema estructural, porque se requiere

concatenar los acontecimientos, acomodarlos en el orden histórico, lo que traza una visión básica estructural de la historia de México de la segunda década del siglo XX que fue particularmente crítica y trascendental.

En el formato de opción múltiple esta pregunta sería: “De los siguientes personajes, cuál es el orden en que gobernaron: a) F. Madero - P. Díaz - V. Carranza - V. Huerta. - A. Obregón; b) V. Huerta - P. Díaz - V. Carranza - F. Madero - A. Obregón; c) P. Díaz - F. Madero - V. Huerta

- V. Carranza - A. Obregón; d) F. Madero - V. Carranza - P. Díaz - V. Huerta - A. Obregón.

Es importante observar que cuando esta pregunta se formula en formato digital, a diferencia de la opción múltiple, se plantea mediante la representación visual dada por la imagen, así como la representación conceptual por los nombres de los personajes y el eje cronológico. Las imágenes contribuyen a facilitar la elaboración de la respuesta porque en ellas hay conocimiento representado en códigos visuales, lo cual ayuda si se tiene la representación icónica de los personajes; este conocimiento es importante para el logro de las competencias. Además, la estrategia de ir arrastrando las imágenes para colocarlas en la línea del tiempo hace que la actividad sea más creativa, rápida y fácil en tanto se va estructurando la integración de la respuesta; esta manera de formular el reactivo cambia la actitud del sustentante respecto al de opción múltiple, donde sólo se reconocen los nombres y el ordenamiento de los personajes. De acuerdo con la taxonomía de Bloom y Krathwohl (1966), el reconocimiento es el nivel más bajo en la elaboración de una respuesta, lo cual permite apreciar lo limitado que resulta un formato de selección de respuesta comparado con los reactivos estructurales constructivos.

Otro parámetro que se puede ponderar en los reactivos es la dimensión del error. No todos los errores son iguales; unos demuestran mayor desconocimiento porque son más graves que otros. Por ejemplo, en la secuencia histórica de los presidentes enunciados, no es un error equivalente el colocar a Obregón antes de Carranza, a antes de Díaz, ya que el primero es un error de años y el segundo de décadas. De aquí que en las propiedades de un error también hay potencialidades para afinar los procesos de evaluación, al ponderar el nivel de los errores cometidos.

En la base de datos que configura esta pregunta se puede poner el listado de 20 presidentes mexicanos, los más importantes por su relevancia histórica, y seleccionar, de acuerdo

al registro histórico, seis de ellos al azar, por orden de dificultad: dos fáciles, dos de dificultad media y dos difíciles.

## DISCUSIÓN

Es importante considerar algunas de las limitaciones que ahora presentan la evaluación computarizada y la generación automática de exámenes. Por un lado se sabe que en México la infraestructura computarizada es aún limitada en las escuelas públicas, especialmente en las del nivel básico y particularmente en las que se ubican en zonas rurales. Así mismo, muchos docentes están poco familiarizados y carecen de entrenamiento para utilizar recursos digitales con propósitos educativos. Las condiciones escolares del país hacen que la evaluación por computadora sea aún un tema ajeno para muchos profesores, e inaccesible para la realidad cotidiana del salón de clases en la mayoría de las escuelas públicas del nivel básico, por lo que muchos estudiantes pueden enfrentar dificultades cuando se presenten por primera vez a una evaluación que les requiera utilizar una computadora. Claro que esto puede ser obviado si previamente se les proporciona el entrenamiento requerido.

Es importante considerar que el uso de las tecnologías digitales presenta crecimientos exponenciales, particularmente los dispositivos móviles de interconexión (teléfonos inteligentes y tabletas) que son cada vez más económicos y más utilizados, aún en sectores de ingreso limitado. Las competencias digitales están en constante expansión, lo que parece ser una transformación cultural progresiva irreversible, característica de nuestro tiempo.

También es importante tener presente que aunque los generadores automáticos de ítems pueden resolver el problema de la construcción de ítems isomorfos y exámenes paralelos, representan tan sólo una manera más de acercarse al complejo proceso de la evaluación de las competencias ganadas por los escolares y del logro educativo alcanzado. Los docentes

sin duda deben seguir cultivando los muy diversos procedimientos de evaluación que se pueden realizar en el aula a partir de los productos generados por los alumnos, así como aprovechar otros referentes, como pueden ser procesos de evaluación auténtica, co-evaluación, auto-evaluación y meta-evaluación, todos estos aprovechados en una dinámica longitudinal de evaluación formativa.

Hay que comprender que los diversos procesos de evaluación son complementarios y no excluyentes; unos no cancelan a los otros, por el contrario, agregan, al crear una visión desde diferentes perspectivas, y conforman representaciones que se deben integrar en una coherencia y congruencia explicativa de la complejidad educativa.

## CONCLUSIÓN

Los sistemas de cómputo están transformando de manera muy profunda los procesos de evaluación en todos los campos; en el campo de la medicina, por ejemplo, la revolución de los datos masivos (*big data*) y la telemática permiten ahora levantar registros y monitorear el ritmo cardíaco de decenas de miles de personas con dispositivos electrónicos móviles, como un sensor de pulsera, y así evaluar los efectos de un fármaco en contextos y tiempo real, es decir, se puede hacer una *evaluación auténtica* del comportamiento del sistema cardiovascular.

Dado el crecimiento exponencial en la capacidad de almacenar, transferir, investigar, depurar (desagregar y robustecer), visualizar, distribuir y analizar enormes cantidades de información, gracias al formato digital, se ha desarrollado la *minería de datos*, la cual, en el campo de la investigación educativa, ha dado lugar al *learning analytics* (Ari y White, 2014), que permite hacer mediciones, análisis y reportes para comprender los procesos del aprendizaje y su evaluación.

Con el GenerEx y el EXCOBA se obtienen registros de cuántos segundos tardó cada

sustentante en contestar cada pregunta de la prueba, lo que permite hacer análisis estadísticos integrados bajo la teoría clásica de la medida y la teoría de respuesta al ítem (TRI), para así definir nuevos valores psicométricos; esto ofrece una aproximación de vanguardia en el campo de la evaluación educativa.

En los sistemas de cómputo un estudiante, al ingresar, genera una *identidad electrónica*, que permite obtener trayectorias longitudinales de su desempeño; esto transforma los procesos de evaluación. El estudiante puede responder en línea a la hora y lugar que sea de su conveniencia; incluso si se interrumpe la energía en el momento de contestar un examen, es posible continuar posteriormente ya que, como todo está soportado en bases de datos, se puede reconfigurar la situación. Al crear la identidad electrónica los sistemas de seguridad permiten obtener registros que no son vulnerables a la suplantación, tales como fotografías e información biométrica a partir de tomas de huellas digitales o registro del iris ocular.

La identidad electrónica permite llevar registro del uso de la guía del examen en línea (demo), su frecuencia, tiempo de entrenamiento, velocidad de respuesta considerando el tiempo de reacción (entre que aparece una pregunta y se emite una respuesta al dar un clic) y el tiempo de respuesta (entre que aparece una pregunta y se da por terminada al pasar a otra pregunta); con ello se hace posible analizar indicadores que permitan evaluar los reactivos en contextos diferentes al de la aplicación de un examen.

Además, dado el soporte digital, hay plena transparencia al imprimir el resultado obtenido en el momento de concluir el examen; además, se puede revisar detenidamente, en otro momento, la ejecución de cualquier sustentante para despejar todas las dudas que pudiera tener, lo que también puede generar aprendizaje. Con los resultados en el examen, las trayectorias de desempeño y los cuestionarios de contexto se pueden realizar análisis

multinivel de los factores socioculturales asociados al desempeño y crear políticas institucionales de atención diferenciada.

Pero a nuestro juicio la ventaja más relevante de la revolución digital en el campo de la evaluación está en la transformación operativa que permite utilizar procedimientos más acordes a los paradigmas cognoscitivos. En el EXCOBA ha sido posible aplicar los principios del construccionismo, considerando que el conocimiento está integrado en estructuras cognitivas, en campos semánticos que tienen una irradiación de aprendizaje expansivo, como se ilustró en el ejemplo de *la vaca* de la Fig. 1. Las preguntas no se responden por reconocimiento, sino que hay una elaboración de la respuesta, dado un contexto, donde es muy difícil obtener aciertos por azar o por copia. Tampoco tiene que ser dicotómica (acierto o error); puede ser politómica, ya que requiere varias respuestas de las cuales algunas pueden ser correctas y otras no. Esto posibilita

la evaluación con crédito parcial, lo que contribuye a mejorar los niveles de precisión de los procesos de evaluación.

La mediación digital permite evaluar el conocimiento representado en diferentes códigos, como el visual, a partir de imágenes dinámicas, o la ubicación de eventos dentro de un mapa, donde se configura una *gestalt*. En un mapa están implícitas todas las distancias entre dos puntos (ciudades o pueblos); no es conocimiento puntual, es heurístico, de manera que el sustentante responde ante contextos y construye su respuesta. Este proceso es mucho más cercano a lo que ocurre en la vida cotidiana, y por lo mismo más próximo a lo que se conoce como *evaluación auténtica*, que evalúa competencias para la vida, no sólo para la escuela.

Con el nuevo paradigma que se ha planteado en este trabajo, se pueden apreciar algunas de las múltiples ventajas de la revolución digital en el campo de la evaluación educativa.

## REFERENCIAS

- ARI Larusson, Johann y Brandon White (2014), *Learning Analytics. From research to practice*, Nueva York, Springer.
- AUSUBEL, David, Joseph Novak y Helen Hanesian (1978), *Educational Psychology. A Cognitive View*, Nueva York, Holt, Rinehart & Winston.
- BACKHOFF, Eduardo, Miguel Ibarra y Martín Rosas (1996), "Desarrollo y validación del sistema computarizado de exámenes (SICODEX)", *Revista de la Educación Superior*, vol. XXXV, núm. 1 (97), pp. 1-8.
- BACKHOFF, Eduardo, Andrés Sánchez, Margarita Peón, Lucía Monroy y María Tanamachi (2006), "Diseño y desarrollo de los exámenes de la calidad y el logro educativos", *Revista Mexicana de Investigación Educativa*, vol. 11, núm. 29, abril-junio, pp. 617-638.
- BENNETT, Randy (2004), "How the Internet will Help Large-Scale Assessment Reinvent Itself", en Mitchell Rabinowitz, Fran Blumberg y Howard Everson (eds.), *The Design of Instruction and Evaluation: Affordances of using media and technology*, Nueva York, Routledge, pp. 101-128.
- BLOOM, Benjamin y David Krathwohl (1966), *Taxonomy of Educational Objectives*, Handbook I, Cognitive domain, Nueva York, McKay.
- CHITTENDEN, Edward (1991), "Authentic Assessment, Evaluation, and Documentation of Student Performance", en Vito Perrone (ed.), *Expanding Student Assessment*, Alexandria, VA, Association for Supervision and Curriculum Development, pp. 22-31, en: <http://calteach.ucsc.edu/courses/documents/Hein%20assessment.pdf#page=33> (consulta: 15 de febrero de 2005).
- COLE, Michel (1998), *Cultural Psychology, a Once and Future Discipline*, Cambridge, MA, Harvard University Press.
- COLL, César (2001), "Constructivismo y educación, la concepción constructivista de la enseñanza y el aprendizaje", en César Coll, Jesús Palacios y Álvaro Marchesi (comps.), *Desarrollo psicológico y educación 2. Psicología de la educación escolar*, Madrid, Alianza Editorial, pp. 157-186.
- DARLING-Hammond, Linda y Jon Snyder (2000), "Authentic Assessment of Teaching in Context", *Teaching and Teacher Education*, vol. 16, núm. 5-6, pp. 523-545.



- LEMKE, Jay y Nora Sabelli (2008), "Complex Systems and Educational Change. Towards a new research agenda", *Educational Philosophy and Theory*, vol. 40, núm. 1, pp. 118-129.
- LURIA, Alexander (1981), *Language and Cognition*, Nueva York, Wiley & Sons.
- Ng, Annie y Alan Chan (2009), "Different Methods of Multiple-Choice Test, Implications and Design for Further Research", *Proceedings of the International MultiConference of Engineers and Computer Scientists 2009*, vol. II, Hong Kong, IMECS 2009, marzo, pp. 1-6, en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.159.379&rep=rep1&type=pdf> (consulta: 11 de abril de 2015).
- NOVAK, Joseph y Alberto Cañas (2008), *The Theory Underlying Concept Maps and How to Construct Them*, Technical Report IHMC CmapTools, Florida, Institute for Human and Machine Cognition, pp. 1-36, en: <http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf> (consulta: 10 de julio de 2015).
- Organisation for Economic Co-operation and Development (OECD) (2014), *PISA for Development. Capacity Needs Analysis Report: Zambia*, en: [http://www.oecd.org/pisa/aboutpisa/Pisa%20for%20Development\\_Zambia\\_FINAL%20with%20cover.pdf](http://www.oecd.org/pisa/aboutpisa/Pisa%20for%20Development_Zambia_FINAL%20with%20cover.pdf) (consulta: 10 de julio de 2015).
- Organisation for Economic Co-operation and Development (OECD) (2015), "PISA 2015 'Computer Platform Development'", en: <http://www.oecd.org/pisa/pisaproducts/PISA-2015-computer-platform.pdf> (consulta: 10 de julio de 2015).
- PIAGET, Jean (1970), "Piaget's Theory", en Paul H. Mussen (ed.), *Carmichael's Manual of Child Psychology*, vol. I, Nueva York, Wiley & Sons. pp. 703-732.
- RESNICK, Daniel y Lauren Resnick (1992), "Assessing the Thinking Curriculum", en Bernard Gifford y Mary Catherine O'Connor (eds.), *Changing Assessments. Alternative views of aptitude, achievement and instruction*, Boston, Kluwer, pp. 37-73.
- ROTHSTEIN, Richard, Helen Ladd, Diane Ravitch, Eva Baker, Paul Barton, Linda Darling-Hammond, Edward Haertel, Robert Linn, Richard Shavelson y Lorrie Shepard (2010), "Problems with the Use of Student Test Scores to Evaluate Teachers", The Economic Policy Institute, Research and Ideas for Shared Prosperity, en: <http://www.epi.org/publication/bp278/> (consulta: 20 de abril de 2014).
- SERRANO, Juan Manuel y Rosa María Pons (2011), "El constructivismo hoy, enfoques constructivistas en educación", *Revista Electrónica de Investigación Educativa*, vol. 13, núm. 1, en: <http://redie.uabc.mx/vol13no1/contenido-serranopons.html> (consulta: 13 de junio de 2014), pp. 1-27.
- TIRADO, Felipe, Eduardo Backhoff, Norma Larrazolo y Martín Rosas (1997), "Validez predictiva del examen de habilidades y conocimientos básicos (EXHCOBA)", *Revista Mexicana de Investigación Educativa*, vol. II, núm. 3. pp. 67-84.
- VYGOTSKY Lev (1986), *Thought and Language*, Cambridge (MA), Massachusetts Institute of Technology Press.