

Backhoff, E., Ibarra, M. A. y Rosas, M. (1996). Desarrollo y validación del sistema computarizado de exámenes (SICODEX). *Revista de la Educación Superior*, Vol. XXXV, No. 1 (97). México: Asociación Nacional de Universidades e Instituciones de Educación Superior.

**Desarrollo y validación del sistema computarizado
de exámenes**

Dr. Eduardo Backhoff Escudero, Mtro. Miguel Ángel Ibarra y
Mtro. Martín Rosas Morales

Revista de la Educación Superior
Vol. XXXV, No. 1 (97)

DESARROLLO Y VALIDACIÓN DEL SISTEMA COMPUTARIZADO DE EXÁMENES (SICODEX)

EDUARDO BACKHOFF, MIGUEL ANGEL IBARRA Y MARTÍN ROSAS**

*** Universidad Autonoma de Baja California*

Contenido del Artículo:

[INTRODUCCIÓN](#)

[ANTECEDENTES INSTITUCIONALES](#)

[DESCRIPCIÓN DEL SISTEMA](#)

[RESULTADOS](#)

[VALIDEZ DE LOS RESULTADOS OBTENIDOS](#)

[DISCUSIÓN](#)

[BIBLIOGRAFÍA](#)

INTRODUCCIÓN [Contenido](#)

No es sino hasta la década de los ochenta, con el advenimiento de las computadoras personales, que la computación impacta a las disciplinas sociales. Las primeras aplicaciones empezaron en ambientes profesionales distintos al educativo. Tal es el caso de las pruebas de personalidad computarizadas que se usan en ambientes médicos y psicológicos. Dado que estas evaluaciones se realizan individualmente y no en grupos numerosos, como es el caso típico de los exámenes educativos, el costo por persona examinada no es un factor determinante ya que las necesidades de equipo son reducidas.

El uso de la evaluación psicológica por computadora está muy bien documentada (véase: Ancill Rogers y Carr, 1985; Buttaer, 1987, Carr y Ghosh, 1983; Greist, 1989; Koback, Reynolds y Greist, 1990) y ofrece muchas ventajas cuando las pruebas son válidas y Confiables y se realizan considerando lineamientos profesionales como los de la Asociación Americana de Psicología** (1986) (Koback y, Reynolds y Greist, 1993). Algunas de las ventajas reportadas son:

1. Estandariza la administración de la prueba (Erdman, Klein y Greist, 1985)
2. Hace que las personas se sientan más confortables cuando tienen que responder a preguntas confidenciales (Lucas, Mullins, Luna y McInroy, 1977; Greist y col, 1974; Greist y Klein, 1980)
3. Reducen el tiempo que los evaluadores ocupan en realizar entrevistas y aplicar pruebas

Por su parte, el uso de las computadoras para la administración de pruebas en las instituciones educativas ha empezado a ser una realidad en los últimos años. Poco tiempo atrás, había muy pocas computadoras en las escuelas y su costo era muy

elevado como para dedicarlas exclusivamente a la evaluación.

Sin embargo, esta situación está cambiando rápidamente, ya que:

1. El costo de las máquinas ha decrecido considerablemente
2. El número de computadoras en las escuelas está aumentando día a día, lo que se refleja en más cursos de computación y en la tendencia a automatizar los sistemas administrativos y
3. Los programas (software) son más fáciles de utilizar por personas inexpertas.

Actualmente, podemos afirmar que las ciencias computacionales han ayudado a mejorar las evaluaciones tradicionalmente diseñadas para lápiz y papel. Especialmente, han permitido obtener una mayor información del estudiante y evaluar aspectos adicionales de su ejecución académica. Siguiendo la lógica de Ward (1991) podemos resumir las ventajas del uso de las computadoras para administrar pruebas convencionales en los siguientes apartados.

Individualización de la evaluación A diferencia de las pruebas de lápiz y papel las cuales se administran en grandes grupos y requieren todos los examinados se evalúen a una misma hora con tiempos límites fijos la evaluación computarizada permite: a) una programación flexible y la eliminación de los tiempos límites para contestar las pruebas, a menos que ésta sea una necesidad, b) registrar los tiempos que tarda el estudiante en responder una pregunta o toda la prueba, c) que se califique al estudiante por precisión y/o rapidez, d) que el examinador emplee distintas estrategias en la evaluación tal como dar retroalimentación contingente a cada respuesta y que el estudiante no avance hasta que responda en forma correcta.

Eficiencia y economía en el manejo de la información. En algunas circunstancias escolares como las evaluaciones diagnósticas o de ubicación (;placement), es crucial la generación inmediata de reportes de resultados, ya sea para el estudiante, el administrador y/o el generador del examen. Aun más, a estos reportes se le pueden incorporar reglas para la toma de decisiones, tal como si el alumno ingresa a un curso o a un grupo de acuerdo a su calificación. La eficiencia puede aumentar si los procedimientos de evaluación ligan electrónicamente a otros registros de la institución, como la historia académica, los perfiles socioeconómico y psicológico del estudiante, etcétera.

Economía. La flexibilidad y rapidez así como la generación automática de los reportes de resultados hacen que las evaluaciones electrónicas sean muy económicas. Por ejemplo, en el caso de las evaluaciones para la certificación profesional donde el tiempo del examinado es limitado y crítico, la flexibilidad de este sistema es de gran beneficio. Asimismo, cuando los resultados de una prueba se requieren de una gran rapidez para la toma de decisiones, como en las decisiones para ubicar a un estudiante en un curso, la inmediatez de los resultados es invaluable. En otras situaciones es de

gran ayuda da Posibilidad de manejar grandes volúmenes de información de los estudiantes con el mínimo de tiempo y de personal. Los bancos son un buen ejemplo de cómo se incrementan los beneficios económicos como consecuencia de invertir en la automatización de la información. Podemos esperar que estos cambios ocurran eventualmente instituciones educativas.

Evaluación adaptativa. Como ya se vio, la administración computarizada de las pruebas convencionales mejoran la eficiencia con que se maneja la información del examinado. Sin embargo, esto no hace nada para mejorar la calidad de esta información. El uso, la computadora para administrar las pruebas adaptativas es una forma en que esta tecnología mejora la información que se obtiene de un examen. Una prueba adaptativa es aquella en que al examinado se le presentan las preguntas de acuerdo a su ejecución en el examen y que son más apropiadas para su nivel de conocimientos y habilidades. Hay muchos esquemas para lograr este objetivo, pero esencialmente la prueba procede de la siguiente forma: De un banco de reactivos, se seleccionan unas cuantas preguntas parecida examinado. Después de que se responden, la computadora genera una calificación con la cual estima el nivel académico del estudiante. Con base en esta estimación, el sistema selecciona la siguiente pregunta de los reactivos restantes. Una vez contestada calificada, la computadora reestima el nivel del estudiante y selecciona la siguiente pregunta. Este procedimiento continúa hasta que se agotan las preguntas del examen.

Finalmente, podemos decir que es una realidad que esta revolución electrónica empieza a repercutir fuertemente en el desarrollo de las técnicas de evaluación que utilizan las instituciones educativas, en especial las de educación superior. Por ejemplo, en el Estados Unidos el examen de ingreso para posgrado Graduate Record Examination (GRE) se aplica tanto en su versión de lápiz y papel como en su versión computarizada; y en los próximos años se aplicará esta última. Igualmente, se espera que para fines de esta década el examen de ingreso a la licenciatura Scholastics 'Aptitud Test (SAT) se administre en forma computarizada.

ANTECEDENTES INSTITUCIONALES [Contenido](#)

Ante la ausencia de un examen de ingreso para la educación superior validado, confiabilizado y estandarizado para la población estudiantil mexicana, en 1992 se inició el desarrollo del Examen 'de habilidades y Conocimientos Básicos - EXHCOBA- (Backhoff y Tirado, 1992).

Un año después, se terminó su versión para lápiz y se validó una gran población de estudiantes (15000 aproximadamente de la Universidad Autónoma de Baja California (UABC) y un muestra de aproximadamente 500 estudiantes de la Escuela Nacional .le Estudios Profesionales Iztacala de la UNAM. Posteriormente (se volvió a validar este examen en las universidades estatales de Guanajuato y Zacatecas, con aproximadamente 9000 y 4000 estudiantes, respectivamente.

Recientemente para ubicarse a la vanguardia de la tecnología educativa en México, se desarrolló el Sistema Computarizado de Examen (SICODEX) (Backhoff, Ibarra y Rosas, 1994; 1995) el cual sirvió de base para elaborar la versión computarizada del EXHCOBA. Esta versión (EXTACOBAC) se validó por primera vez en el año en curso con 537 aspirantes que presentaron su examen de ingreso a la UABC.

Los resultados fueron tan alentadores que la UABC decidió crear tres Centros de Evaluación Computarizada en las unidades de Méxicali, Tijuana y Ensenada, para poder aplicar el examen en forma continua al total de aspirantes que desean ingresar anualmente a la universidad. De mayo a julio del año en curso implementamos este sistema en la UABC y se evaluó a cerca de 8200 estudiantes que se inscribieron como candidatos de la UABC. Los resultados obtenidos confirmaron nuevamente la validez de sistema.

Así mismo, el pasado mes de septiembre se validó la versión computarizada con estudiantes de posgrado que iniciaban sus estudios en el Centro de Investigación Científica y de Educación Superior Ensenada (CICESE).

Finalmente en el mes de octubre, como parte de una investigación internacional comparativa, se aplicó el EXTACOBAC en sus dos versiones una muestra representativa y estratificada de estudiantes de recién ingreso de la Universidad Autónoma de Coahuila (UAC), validándose ambas versiones.

DESCRIPCION DEL SISTEMA [Contenido](#)

La interfaz del SICODEX se ha descrito con detalle en otros trabajos (Ibarra y Rosas, 1994; Backhoff, 1994), por lo que en esta ocasión nos referimos a sus generalidades.

Esta interfaz es muy parecida a la de un examen de lápiz y papel, con la diferencia que las preguntas y posibles respuestas se presentan en la pantalla de un monitor. Se contesta fácilmente aunque se desconozca la forma de cómo utilizar una computadora. Para ello, únicamente es necesario estar familiarizado con el teclado normal de una máquina de escribir.

Además del texto de la pregunta y las opciones de respuestas los reactivos pueden contener: 1) Imágenes fijas a color como mapas, diagramas, gráficas y ecuaciones matemáticas, 2) animaciones o figuras en movimiento y 3) textos adicionales que por su extensión no quepan en el área de la pregunta.

El sistema realiza una serie de funciones encadenadas para operar. Estas se pueden agrupar en cinco momentos: 1) cuando el estudiante inicia la sesión, 2) en el momento en que el sistema genera el examen, 3) cuando se prepara al interfaz para presentar el examen y guardar las respuestas del alumno, 4) en el momento en que el estudiante interactúa con el sistema contestando el examen, 5) cuando el examen califica al estudiante y lo saca del sistema.

Es importante destacar que la interfaz le permite al estudiante:

1. "Navegar" o "transitar" libremente en el examen.
2. Corregir, borrar y/o modificar sus respuestas.
3. Contestar el examen total o parcialmente
4. Conocer sus resultados inmediatamente al terminar de responder el examen.

RESULTADOS [Contenido](#)

Como ya se mencionó el SICODEX se ha utilizado en dos ocasiones como examen de admisión de la UABC con aproximadamente 9000 aspirantes. Asimismo, se ha utilizado en otras instituciones como examen diagnóstico (como en el CICESE) y como instrumento de investigación educativa (como en el caso de la UAC).

Con base en estas experiencias, la operatividad del sistema se validó al observar su alta eficiencia, confiabilidad, facilidad de uso y ausencia de problemas para operarlo. Podemos afirmar que el 100 % de los jóvenes han podido contestar el examen sin ningún tipo de problemas.

Aún más, en una encuesta realizada al final del examen, todos los estudiantes encuestados opinaron que preferían la versión computarizada a la tradicional de lápiz y papel por su dinamismo y versatilidad; y aunque algunos de ellos opinaron que se pusieron nerviosos al principio de la sesión, por desconocer la forma de contestar la prueba, todos coincidieron en que su nerviosismo sólo dura el breve tiempo en que aprendieron a utilizar el programa de tres a cinco minutos, aproximadamente).

Los resultados de la operatividad del SICODEX se pueden resumir en los siguientes puntos:

1. Es de fácil uso, ya que el estudiante logra familiarizarse con la interfaz en un lapso breve de tiempo
2. Mantiene la motivación e interés del alumno, debido a su atractiva presentación de gráficos e imágenes animadas.
3. Permite que el estudiante trabaje eficientemente, al llevar un control sobre el tiempo transcurrido y número de preguntas contestadas.
4. Posibilita realizar una cantidad muy grande de exámenes paralelos y evitar que los estudiantes se pasen información durante el examen.
5. Mantiene la seguridad del examen almacenando sus contenidos en forma electrónica y no escrita.
6. Permite la calificación de los resultados en forma inmediata, en el momento en que se termina el examen.
7. Hace eficiente el tiempo y los recursos humanos, ya que se ahorra el proceso de calificación por lector óptico y disminuye la posibilidad de error humano

8. Facilita la investigación educativa, ya que se pueden evaluar diversas respuestas del estudiante, tales como: los tiempos de reacción, las veces que se modifican las respuestas, la duración del examen, etcétera.
9. Permite validar, confiabilizar y actualizar los contenidos del examen al incorporar en el sistema los algoritmos para realizar los análisis estadísticos

VALIDEZ DE LOS RESULTADOS OBTENIDOS [Contenido](#)

Para validar los resultados del SICODEX, fue necesario comparar los resultados de los estudiantes en las dos versiones del examen; lápiz y papel y computadora.

Con este propósito, contrastamos los resultados de dos poblaciones de estudiantes de la UABC que contestaron el EXTACOBAs como parte regular de su proceso de admisión en dos periodos distintos 8208 estudiantes que contestaron la versión de lápiz y papel (verano de 1993) y 537 estudiantes que contestaron la versión computarizada (invierno de 1994).

A manera de réplica, también se compararon otras dos poblaciones de estudiantes del primer año de licenciatura, que participaron en forma voluntaria en un estudio comparativo y que fueron seleccionados en forma aleatoria de la UAC, en septiembre de 1994. En este estudio los estudiantes sólo respondieron las primeras 130 preguntas del examen: 232 utilizaron la versión computarizada y 218 la versión de lápiz y papel.

La tabla 1 muestra los porcentajes de aciertos, por áreas de conocimiento, que obtuvieron los estudiantes de la UABC y la UAC. Como se podrá observar en la gráfica los resultados de ambas versiones son muy similares, replicándose los porcentajes de aciertos en cada una de las áreas evaluadas.

Area	Reactivos	UABC		UAC	
		Lápiz y papel	Computadora	Lápiz y papel	Computadora
H. Verbales	30	17.76	17.46	19.07	17.82
H. Cuantitativas	30	16.42	15.39	18.35	17.66
Lengua Española	15	8.43	8.42	9.80	8.81
Metemáticas	15	6.39	6.03	7.35	7.05
Ciencias Naturales	20	10.80	10.47	10.86	10.38
Ciencias Sociales	20	10.66	10.94	9.96	10.18

Es importante señalar que aunque ambas presentaciones se aplicaron en el proceso regular de admisión de la UABC, no se realizaron en el mismo periodo de ingreso: la de lápiz y papel se respondió en verano, mientras que la computarizada en invierno.

Esta diferencia es importante de mencionar, debido a que en el primer periodo ingresan los mejores estudiantes, dejando la admisión de enero a los jóvenes que no pudieron ingresar en el periodo de verano, o que por alguna razón retrasaron su ingreso a la universidad. Esto explica que el nivel de aciertos en la versión computarizada fuera ligeramente inferior que la de lápiz y papel.

Lo contrario ocurre en el caso de los estudiantes de Coahuila. Los que contestaron la versión computarizada obtuvieron puntuaciones más altas. Esto se puede explicar por la diferencia en el interés que mostraron los estudiantes por contestar voluntariamente un examen computarizado en lugar de uno de lápiz y papel.

Haciendo este análisis un poco más fino de las respuestas de los estudiantes comparamos los resultados de las diez primeras preguntas en ambas versiones (ver la tabla 2). Aquí también observamos que los porcentajes de aciertos fueron muy parecidos, conservando las mismas tendencias antes descritas.

Area	UABC		UAC	
	Lápiz y papel	Computadora	Lápiz y papel	Computadora
Antónimos	52.40	52.10	72.5	73.7
Extensión	72.00	70.70	67.0	67.2
Inferencias	71.30	71.80	70.2	75.4
Significado	64.00	63.80	70.2	75.4
Significado	47.40	46.10	42.7	47.8
Significado	68.70	70.90	74.8	77.2
Refrán	48.90	46.30	46.8	50.4
Comprensión	74.20	70.00	78.0	75.0
Comprensión	47.10	46.30	46.3	45.3
Comprensión	71.00	70.00	64.2	65.1

Finalmente en la tabla 3 se muestran los análisis de confiabilidad para ambas versiones, con la población de la UAC, siendo los Alfa de Grombach generales de la prueba de 0.91 y 0.89 para la versión computarizada y la de lápiz y papel respectivamente.

Area	Reactivos	UABC		UAC	
		Lápiz y papel (N=8208)	Computadora (N=536)	Lápiz y papel (N=218)	Computadora (N=232)
H. Verbales	30	0.66	0.59	0.53	0.57
H. Cuantitativas	30	0.86	0.84	0.83	0.84
Lengua Española	15	0.62	0.50	0.49	0.47
Metemáticas	15	0.75	0.69	0.72	0.72
Ciencias Naturales	20	0.57	0.43	0.51	0.63
Ciencias Sociales	20	0.71	0.62	0.71	0.82

Es interesante notar en la tabla 3 las diferencias en estos índices cuando se compararon las secciones del examen. Esto no debe de sorprender, dado que las partes del examen varían no sólo en su temática y consistencia interna, sino también en el número de reactivos, lo cual hace muy sensible los cambios en la confiabilidad. Así como en los niveles de dificultad de los reactivos, también observamos mejores puntuaciones en los índices de confiabilidad para la versión de lápiz y papel con los estudiantes de la UABC, así como para la versión computarizada con los estudiantes de la UAC.

De estos resultados concluimos que, aunque hubo algunas diferencias en los índices de confiabilidad, éstas no son significativas cuando se comparó el examen en su totalidad (0.91 y 0.89).

DISCUSIÓN [Contenido](#)

Sin lugar a dudas, la computación es una ciencia moderna que posee muchas bondades para utilizarse como herramienta en la investigación y desarrollo tecnológico de la psicología y la educación. Especialmente esto es cierto en el campo de la elaboración y validación de instrumentos de evaluación.

Los desarrollos recientes de la evaluación por computadora representan una verdadera revolución en el campo de la evaluación psicológica y educativa, ya que nos permiten, entre otras cosas :

1. Editar textos, dibujos, fotografías, sonido y vídeo, de tal manera que una persona puede leer, oír, ver, seguir instrucciones y responder a los contenidos temáticos que se le presenten en pantalla.
2. Presentar las preguntas del examen en forma interactiva, o adaptativa, ajustando la

dificultad de las preguntas a la ejecución del estudiante.

3. Retroalimentar la ejecución del individuo en forma contingente, señalándole sus aciertos y errores.
4. Obtener medidas, adicionales al total de aciertos, como es la latencia, duración del examen, número de selecciones hechas en cada reactivo, etcétera.
5. Realizar el análisis de resultados en forma inmediata.

Los resultados obtenidos con el SICODEX confirman muchas de estas ventajas, así como otras que también se señalan en la literatura.

1. No encontramos diferencias significativas entre ambas presentaciones; las correlaciones son muy altas entre reactivos equivalentes y los índices de confiabilidad son muy similares para ambas versiones (Heppner y colaboradores, 1985; Katzy Dalby, 1985; Backhoff, Ibarra y Rosas, 1995).
2. Hay una marcada preferencia de las personas por responder a las versiones computarizadas, independientemente de la experiencia previa del sujeto (Moe y Jotanson, 1988; Backhoff, Ibarra y Rosas, 1995).
3. Las versiones de los tests computarizados que presentan un reactivo a la vez, hace que el individuo ponga mayor atención a la pregunta y responda con mayor cuidado (Vansickle y colaboradores, 1989).
4. La confiabilidad de un instrumento aumenta con la versión computarizada (Cates, 1993).
- 5.- La incorporación de esta tecnología apoya la atención, el entendimiento y la exploración del alumno, ya que representa una comunicación multisensorial (Yanger, 1991 y Hooper 1991).

Podemos concluir diciendo que, aunque las evaluaciones psicológicas y educativas en formatos de lápiz y papel se han utilizado eficientemente por años para obtener información psicológica las personas y evaluar los conocimientos y habilidades de los estudiantes, definitivamente no poseen las capacidades de presentación obtención de información, ni de análisis de resultados si los comparamos, con las nuevas formas de evaluación computarizada.

Sin lugar a dudas, el campo de la evaluación educativa que por mucho tiempo permaneció estancada, se está desarrollando en los últimos años gracias a esta tecnología

BIBLIOGRAFÍA [Contenido](#)

American Psychological Association, Guidelines for computer based tests

interpretations Washington, D.C., 1986.

ANCILL R., ROGERS, D Y CARR, A.C. "Comparison of computerized self-rating scales for depression with conventional observer ratings" *Ach Psychitrica Scandinavia*, 71, 1985, 315-317.

BACKHOFF, E. IBARRA, M.A. Y ROSAS, M. Sistema Computarizado de exámenes (SICODEX), México, SEP, Registro de Derechos de Autor (). 230, 1994.

BACKHOFF, E. IBARRA, M.A. Y ROSAS, M. Sistema Computarizado del examen (SICODEX), Trabajo presentado en el congreso "La Universidad Latinoamericana ante los nuevos escenarios de la región". México, D.F., 1995.

BACKHOFF, E. Y TIRADO, F. "Desarrollo del Examen de Habilidades y Conocimientos Básicos", *Revista de la Educación Superior*, jul-sep.1992, No. 83, p. 95-117.

BACKHOFF, E., IBARRA, M.A. Y ROSAS, M. Automatización del examen de admisión de la UABC. Trabajo presentado en el 29 Simposio Internacional de Ingeniería Mecánica y Sistemas Computacionales. Mexicali B.C., noviembre 1993

BUTCHERJ N. "The use of computers in psychological assessment: An overview of practices and issues", en J.N. Butcher (Ed.), *Computerized psychological assessment* (pp. 3-14), New York, Basic Books, 1987.

CARR, A. C. Y GHOSH, A. "Accuracy of behavioural assessment by Computer", *British Journal of Psychiatry*, 142, 1983, 66-70.

CATES, W.M. A. Small-scale comparison of the equivalence of paper and pencil and computerized versions of student end-of-course evaluations, *Computers in Human Behavior*, 9, 1993, 401-409

KLEIN, M.H. Y GREIST, J.H. "Direct patient-computer interviewing, *Journal of Consulting and Clinical Psychology*, 53, 1985, 760-773

F., ANDERSON, J. FARSTRUP, A., Y WEINDERMAN, N. "Reading performance on a standardized test is better from print than from computer display", *Journal of Reading*, 28, 1985, 321-325.

HOOPER, W.K. "Multimedia Scouting", *IEEE Computer Graphics and Applications*,

11(4), 1991, 52-57.

KATZ, L. y Dalby, J. Computer and manual administration of the Eysenk Personality Inventory", *Journal of Clinical Psychology*, 37, 1985, 586-588.

KOBAK, KA., REYNOLDS, W.M. Y GREIST, J.H. UDevelopment and validation of a computer administered version of the Hamilton Anxiety Scale", *Psychological Assessment*, 5(4), 1990, 487-492.

KOBAK, KA., REYNOLDS, W.M. Y GREIST, J.H. "Development and validation of a computer-administered version of the Hamilton Anxiety Scale", *Psychological Assessment*, 5(4), 1993, 487-492

LUCAS, R.W., MULLINS, P.J., LUNA, C.B. Y MCINROY, D.C. "Psychiatrists and a computer as interrogators of patients with alcohol-related illnesses; A comparison", *British Journal of Psychiatry*, 131, 1977, 160-167.

MOE, K. Y JOHNSON, M. "Participants' reactions to computer testing", *Journal of Educational Computing Research*, 4, 1988, 49-86

VANSICKLE, T., KIMMEL, C. Y KAPES, J. "Test-retest equivalency of computer-based and paper-pencil versions of the Strong-Camp Interest Inventory", *Measurement and Evaluation in Counseling Development*, 22(2), 1989, 88-93.

WARD, W.C. "Using Microcomputers To Administer Tests", College Board y Educational Testing Service, Computerized Placement Tests: Background Readings, New Jersey, 1991.