

Evaluación del paquete de enseñanza asistida por computadora: "Los aminoácidos"

Leonor Fernández Rivera-Río, Klaus-Dieter Gorenc¹

Departamento de Bioquímica, y ¹Secretaría de Educación Médica, Facultad de Medicina

(Recibido, octubre 19, 1994, aceptado, marzo 16, 1995)

Resumen

Con el paquete de autoría Authorware Professional versión 2.0 Macromedia se desarrolló un software amigable para la plataforma PC con cinco conceptos básicos acerca de los aminoácidos: isomería D y L, grupo alfa amino, grupos radicales y predicción de la carga eléctrica de los aminoácidos sometidos a diferentes pH, donde se incluyeron los ejercicios correspondientes. La participación de los siete grupos de este estudio exploratorio evaluativo fue voluntaria. El paquete fue probado con 185 alumnos del primer año de la licenciatura de Médico Cirujano de la Facultad de Medicina de la UNAM que resolvieron un pre-test, usaron el paquete durante 90 min y resolvieron un post-test inmediatamente después. Los resultados muestran que el uso incrementó significativamente en el número de respuestas correctas en el post-test comparado con el pre-test (valor de $t = 14.27$; $gl = 184$ $p = 0.0000$). Tanto el pre-test como el post-test fueron validados y confiabilizados: pre-test con validez por construcción = 81.5% y α de Cronbach = 0.9355 y el post-test con validez por construcción = 82.1% y α de Cronbach = 0.9382, valores que marcan el control de calidad del producto.

Palabras clave: Aminoácidos - Enseñanza - Software - Evaluación - Pre y Post-test

Summary:

A friendly software package was developed with Authorware Professional Version 2.0 Macromedia for PC platform with five basic concepts; i.e. D and L isomerism, alpha amino group, radicals and electrical charge prediction under different pH. Exercises were included. Participation in this explorative-evaluative study was voluntary. This package was tested by 185 students of the Biochemistry course at the Faculty of Medicine of the National Autonomous University of Mexico. Students solved a pre-test, utilized the package during 90 minutes and solved a post-test immediately afterwards. Results obtained reveal a significant increase in correct answers in the post-test, compared with the pre-test (t value = 14.27; $df = 184$; $p = 0.0000$). Construction validity and reliability of pre and post-test were performed: pre-test construction validity = 81.5% and α Cronbach = 0.9355; post-test construction validity = 82.1% and α Cronbach = 0.9382; these values are related to the control of the product quality.

Key words: Aminoacids - Teaching - Software - Evaluation - Pre and Post-test

Introducción

El facilitar la enseñanza de los conceptos básicos de la bioquímica es útil para los alumnos de varias carreras, ya que ésta es una ciencia básica que forma parte de los planes de estudio de las carreras de Médico Cirujano, Médico Veterinario Zootecnista, Cirujano Dentista, Químico Farmacobiólogo, Psicología, así como en los de la Tecnología de alimentos, Nutrición y Biomedicina, entre otras.

En los últimos años, la bioquímica se ha desarrollado enormemente y, en la actualidad los conocimientos de esta disciplina han permitido el avance de otras ciencias como son la biología molecular, la fisiología,

la farmacología, etcétera, lo que ha dado como consecuencia el desarrollo de nuevas tecnologías. Este avance ha provocado que la cantidad de información que se maneja sea demasiado grande para que un maestro pueda estar actualizado en todos los temas, además de que los libros de los que se toma la información muchas veces son traducciones que tienen varios años de atraso. La práctica profesional de cualquier carrera del área médico-biológica requiere un nivel de conocimientos suficiente para que los conceptos básicos de bioquímica puedan ser utilizados para resolver problemas clínicos.

La bioquímica es una ciencia abstracta¹ en la que los conceptos teóricos son enseñados a través del uso de modelos, debido a que no se pueden ver ni las moléculas ni los átomos. Hasta ahora en México no se ha usado a la computadora como herramienta auxiliar en la docencia, mientras que en Australia, desde la década de los ochenta, se ha observado un importante incremento^{2,3}; sin embargo, es posible que ésta sea una ayuda con gran potencial para la enseñanza en el futuro, debido a que permite la producción de modelos tridimensionales de moléculas sencillas y de moléculas complejas como las proteínas y los ácidos nucleicos, lo que facilita la comprensión de algunos conceptos bioquímicos como el de complementaridad que se observa entre el substrato y la enzima o entre el antígeno y el anticuerpo. La simulación de cambios en el pH del medio podría permitir la comprensión de cómo una proteína mantiene su función ya que tiene una conformación determinada por las interacciones iónicas, hidrofílicas e hidrofóbicas entre los grupos radicales de los aminoácidos que la forman. Otras ventajas de este tipo de herramienta son que el usuario podrá hacer uso del paquete las veces que lo desee, además de que el *software* puede ser actualizado en forma más rápida y económica que los libros.

La enseñanza de la bioquímica plantea algunas dificultades atribuibles al alumno y otras que son propias de la materia. En el alumno las diferencias en estrategias y estilos de aprendizaje y en la capacidad para manejar la información; los prejuicios acerca de la materia, los conocimientos previos insuficientes, la incapacidad para comprender información abstracta, los grupos numerosos, son, entre otros, los factores obstaculizantes. Para saber si una reacción se ha llevado a cabo se usan métodos indirectos que detectan cambios fisicoquímicos que ocurren en las moléculas. Otro elemento a considerar es un aumento explosivo en la información derivada de las investigaciones en esta materia.

En el Departamento de Bioquímica de la Facultad de Medicina de la Universidad Nacional Autónoma de México (UNAM), se ha iniciado el Proyecto de Desarrollo de *Software* Para la Enseñanza de Bioquímica Asistida por Computadora (EBAC)¹. Este proyecto tiene como objetivo el mejorar la comprensión de algunos conceptos básicos de esta materia proveyendo a los alumnos con una herramienta adicional cuya función es la de complementar a la clase tradicional y a la consulta del material impreso. El objetivo del proyecto es la elaboración de paquetes de *software* para la EBAC¹. Para lograr éste, y como un primer paso, se elaboró el

programa tutorial "Los aminoácidos"¹ en el que se desarrollaron los conceptos que se refieren a las características fisicoquímicas de los aminoácidos y que están contenidos dentro del "Plan de Objetivos del Curso de Bioquímica"⁴ que se imparte durante el primer año del Plan Unico de la Carrera de Médico Cirujano de la Facultad de Medicina de la UNAM⁵. Para la evaluación del *software* se postuló que no existen diferencias significativas entre las mediciones del conocimiento efectuadas antes y después de someter a un grupo de alumnos bajo este tipo de enseñanza, y en lo que respecta a los instrumentos de medición, se postuló que efectivamente miden lo que pretenden medir y que además lo efectúan con precisión.

Material y métodos

Características del equipo de cómputo. Para el desarrollo y evaluación de este paquete de EBAC¹ se seleccionó la plataforma PC y se usó una computadora 486 con 8.0 Mb de memoria RAM, disco duro de 240 Mb y Monitor a color SVGA y el paquete de autoría *Authorware Professional* Versión 2.0 de Macromedia. El paquete se probó en el Centro de Cómputo de la Facultad de Medicina de la UNAM, usando computadoras PC 386 y 486 con 2 Mb de RAM, disco duro de 125 Mb y monitor a color SVGA.

Método para el diseño del paquete de software "Los aminoácidos". Para el diseño del paquete se siguió el siguiente procedimiento:

1. Después de la revisión de la literatura referente al tema⁶⁻⁹ y de seleccionar los tópicos relevantes de dicha información se eligió el método de enseñanza para este paquete. Este consiste en la exposición de la teoría por medio de texto, acompañada de animaciones en las que se representan diferentes fenómenos como son la ionización de los grupos amino y carboxilo y la modificación de gráficos de titulación de α -aminoácidos. Se incorporaron también las estrategias que un experto usa para la aplicación de los conocimientos en la resolución de problemas específicos; por ejemplo, para diferenciar a un aminoácido hidrofílico de un hidrofóbico un experto será capaz de reconocer en la fórmula la presencia de ciertos radicales químicos (-SH, -OH, -COOH, etcétera) asociados al radical de los aminoácidos hidrofílicos y ausentes en los aminoácidos hidrofóbicos⁶⁻⁹. Cada segmento de exposición teórica está precedido de la información acerca de los objetivos de la sección y termina con un resumen. Se optó porque el usuario pueda iniciar la revisión del tema por la

parte que mejor le parezca; no obstante, se recomienda que la primera vez siga el orden establecido en el menú.

2. Los objetivos instruccionales del paquete son que el alumno identifique:

- ¿Cuáles son las funciones que desempeñan los aminoácidos en los seres vivos?
- ¿Cuáles son los átomos y los grupos químicos que los forman?
- ¿Qué es un isómero?
- ¿Qué es un estereoisómero?
- ¿En qué consiste la actividad óptica?
- ¿Cuáles son las características de los grupos radicales de los D y de los L α -aminoácidos?
- ¿Cuál es el comportamiento de los diferentes aminoácidos cuando son sometidos a diferentes pH?

3. Se estableció que el usuario del paquete debe tener conocimientos previos acerca de la estructura del agua, de pH y de sistemas amortiguadores de pH.

4. La interfase usuario-computadora se diseñó de manera que las interacciones paquete-usuario se lleven a cabo únicamente por medio del *ratón*, esto con el objeto de hacer que el paquete sea lo más *amigable* posible.

5. El contenido teórico del paquete fue evaluado por cinco profesores pertenecientes a la planta docente del Departamento de Bioquímica de la Facultad de Medicina de la UNAM, quienes sugirieron algunas modificaciones que fueron consideradas en el producto final.

Evaluación exploratoria del paquete. La evaluación exploratoria de este paquete se hizo con los alumnos de los grupos 1102, 1105, 1109, 1111, 1133 y 1136 ($n = 185$ alumnos) de la generación 1993-1997, que cursaron Bioquímica durante el primer año en la Carrera de Médico Cirujano de la UNAM, dentro del primer "bloque" del curso. La selección de los grupos se hizo con base en la cooperación voluntaria de los maestros.

En todos los grupos en los que se probó el paquete, los alumnos pudieron usarlo sin necesidad de capacitación especial y la única información que se les

proporcionó fue cómo entrar al paquete que se instaló previamente en un directorio del disco duro y se desplegó en Windows en una ventana especial por medio de un ícono.

El instrumento de evaluación (examen) se diseñó para responder a las siguientes preguntas:

1. ¿Es este paquete útil para facilitar el aprendizaje de las propiedades fisicoquímicas de los aminoácidos?
2. ¿Puede el alumno reconocer el grupo amino, el grupo carboxilo, el grupo radical y el carbono en una fórmula?
3. ¿Puede el alumno predecir qué grupos se ionizan al cambiar el pH del medio?
4. ¿Puede el alumno predecir qué carga eléctrica neta tendrán los aminoácidos cuando son sometidos a diferentes pH?
5. ¿Puede el alumno identificar a los isómeros D y L de los aminoácidos al observar una fórmula?

El examen consta de 15 fórmulas de α -aminoácidos diferentes y se pidió al usuario identificar:

- Cuáles aminoácidos pertenecen a la serie D y cuáles a la serie L (15 ítems).
- Identificar a los grupos radicales (3 reactivos)
- Los aminoácidos con radicales ácidos (2 preguntas).
- Los grupos α -amino (3 ítems).
- Predecir la carga de seis de los aminoácidos a pH de 1.5, 7.0 y 10.0 (18 reactivos).

Para operacionalizar la evaluación, se utilizó un diseño de un solo grupo conocido como de pre y post-test¹⁰, cuya estructura indica que se requiere desarrollar una variable endógena emanada del mismo *software* para evaluar tanto a nivel general como individual la capacidad de enseñanza de esta herramienta. Tanto el pre-test como el post-test fueron construidos con 41 reactivos cuya respuestas son dicotómicas (cierto o falso). El pre-test fue aplicado a 185 alumnos antes del uso formal del paquete de *software* "Los aminoácidos"¹ durante 90 minutos. El post-test tiene una correspondencia exacta con la primera medición y se aplicó inmediatamente después del uso del paquete. Para determinar si el pre-test y el post-test miden lo que pretenden medir se

empleó la combinación de dos técnicas de validez, la de criterio exterior en su modalidad de concurrente y la de construcción¹¹; para la primera se extrajo el coeficiente de Cramer y para la segunda se empleó el análisis factorial como procedimiento estadístico confirmatorio^{12,13}. Para completar el binomio y establecer los errores en la medición se emplearon los dos métodos de confiabilidad que requieren una sola medición: división por mitades y la consistencia interna^{11, 13}. Para establecer la validez tanto del pre como del post-test, se utilizó una matriz de correlaciones tetracóricas¹⁴, donde se combinaron cuatro algoritmos factoriales con cinco rotaciones con base en un esquema de $(41 \times 40) / 2 = 820$ correlaciones.

A esta combinación algorítmica con rotaciones se ofreció una matriz de asociaciones lineales entre los 41 reactivos del pre-test y del post-test denominada tetracórica. Mediante este procedimiento se calcularon los coeficientes de correlación tetracórica entre las 41 respuestas dicotómicas, bajo el supuesto de que las variables continuas --de razón, según Friedrichs¹³ -- simulen una distribución normal, que fue recomendado por Richter, Klemm y Zahn en 1990¹⁴.

A partir de este modelo heurístico reductivo escalonado y siguiendo las sugerencias de Johnson y Wichern¹⁵, se definió al reactivo como:

- elemento integrante de factores cuyo valor propio --eigenvalue-- ≥ 1 ;
- con carga ≥ 0.50 en un cierto factor; además
- con carga negativa o $\leq 50\%$ de la carga registrada en los restantes factores;
- consistencia de los reactivos a través de ≥ 2 algoritmos factoriales,
- consistencia de los reactivos a través de ≥ 3 rotaciones.

Por lo tanto, la combinación de estos criterios está dirigida a obtener reactivos con una elevada especificidad; es decir, mantener restringidos los reactivos --conocimientos-- falsos-positivos.

Realizada la evaluación tanto del pre como del post-test, la toma de decisión para seleccionar los mejores instrumentos para medir la efectividad del *software* para proporcionar el conocimiento sobre aminoácidos se llevó a cabo mediante el

procedimiento meta-analítico propuesto por Neiss en 1992¹⁶, orientada a efectuar generalizaciones a través de los resultados arrojados por los análisis particulares dentro de este mismo estudio¹⁶⁻²⁰. Considerando que este estudio exploratorio se orientó únicamente a evaluar tanto el *software* como los instrumentos de medición aplicados antes y después de emplear esta herramienta educativa, no se incluyó un grupo control constituido por educandos sometidos a la forma tradicional empleada en la Facultad de Medicina, aunado a que hubiese sido necesario extraer una muestra aleatoria de docentes y dicotomizarlos para entrenarlos a impartir este tema de aminoácidos, a nivel tradicional y no tradicional siguiendo los parámetros establecidos en la estructura del *software*, además de ejercer un control; por ejemplo, con videograbaciones para establecer que efectivamente no existen diferencias significativas dentro de los dos grupos de estos docentes y, para ello, se hubiera requerido evaluar las grabaciones a través de la técnica de confiabilidad denominada inter - rater - reliability; siendo, por lo menos, necesaria la participación de tres jueces independientes que deberían calificar las grabaciones mediante instrumentos de medición previamente validados y confiabilizados. Con este procedimiento se hubieran elevado considerablemente los costos de operación.

Resultados

Cuantificación de los errores en la medición. Al no confirmarse la estructura teórica subyacente tanto del pre-test como del post-test, se procedió a utilizar el estadístico multivariado factorial como un procedimiento para reducir los 41 reactivos que conforman los dos instrumentos de medición¹².

Pre-test. El concepto teórico del *software* se operacionalizó²¹ mediante el pre-test --y post-test-- a través de cinco dimensiones o conceptos derivados¹³. La primera dimensión corresponde al de la serie D y L medido por 15 indicadores --variables, items, reactivos--; la siguiente, al grupo α -amino con tres indicadores; posteriormente, la del grupo radical también con tres indicadores; la penúltima, al radical ácido con dos indicadores y el quinto concepto derivado se refirió a la predicción del comportamiento ácido-básico, que está constituido por 18 indicadores. Esto corresponde al "deber ser", que al contrastarse con el "ser" expuesto en la primera columna del primer cuadro, se observó, por un lado, fuera de lo esperado, un incremento de cinco a nueve en el número de las dimensiones --factores--. Esto se debió a que la dimensión de la serie D y L se tricotomizó al igual que

el concepto derivado de la predicción del comportamiento ácido-básico y, por el otro, una disminución de 41 a 35 en el número de los indicadores debido a que no cubrió el tercer criterio definitorio de reactivo expuesto en el apartado anterior, correspondiendo estos seis reactivos al concepto derivado de la serie D y L, ubicados en el primer --último item-- y quinto factores, con lo que el pre-test se redujo a ocho factores o dimensiones. Sin embargo, se halló que este instrumento fue confiable, más no válido, permitiendo generalizar los resultados únicamente a la muestra examinada, pero sin la posibilidad de efectuar la generalización externa²², por no integrarse el binomio conformado por la validez y la confiabilidad, debido a que un instrumento, en este caso el pre-test, puede ser confiable, mas no válido, pero un instrumento válido también tiene que mostrar la precisión de la medición²³.

Ante este resultado, los autores citados¹², sugieren que se debe de efectuar el reanálisis de los datos, ya que la exclusión obligada de estos seis items modifica los niveles de correlación tetracórica y, por ende, se debe proceder a determinar el efecto, que al caracterizarse por la desaparición de un factor se halló una modificación en el contenido de los otros factores; verbigracia, la descrita para los reactivos de la serie D y L, que se encontraban en el primer factor antes de efectuar la reducción, que al perder su peso por excluirse el 40% de sus reactivos, producto de un desbalance, ocuparon, por ello, el primer factor los reactivos dirigidos a establecer la predicción del comportamiento ácido-básico, ya que mostraron una estabilidad en su cohesión como factor. A pesar de estos cambios, no se observó un efecto importante entre los porcentajes que explican la varianza de los factores, que es extensible a los diversos coeficientes dirigidos a medir la precisión de la medición, que es factible demostrar cuando se contrastan las dos primeras columnas del Cuadro 1. Este resultado señala, que con esta reducción los beneficios pueden considerarse como limitados, cualitativamente hablando.

Post-test. En contraste con el pre-test (primera columna Cuadro 1), los 41 reactivos del post-test se dispersaron en 10 factores, excluyéndose un item debido a que no cubrió el criterio de la carga factorial ≥ 0.50 . Este reactivo pertenece a los orientados a explorar el conocimiento acerca del comportamiento ácido-básico y esto explica el alto porcentaje de la explicación total de varianza así como el de la precisión que se encuentran esquematizados en la quinta columna del Cuadro 1.

Fuera de lo esperado, al examinar el producto de esta primera reducción, se observó la violación del cuarto criterio que integra la definición de reactivo propuesta¹⁵ por otros dos reactivos que también corresponden a la predicción del comportamiento ácido-básico. Por tal circunstancia, se sometieron los 38 reactivos restantes a un nuevo reanálisis¹² que excluyó por la misma situación apuntada previamente, un reactivo más que se encuentra dentro del cúmulo de reactivos dirigidos a examinar el comportamiento ácido-básico. Este proceso reductivo escalonado se debió a que con la exclusión paulatina de items, se observó, tras dos momentos de inestabilidad de los pesos de los reactivos dentro de los factores, que los 37 reactivos complementarios se ubicaron en nueve factores, así como una articulación entre tres items del comportamiento ácido-básico en un solo factor, que durante el primer análisis se encontraba fragmentado en los dos últimos factores y con esta recuperación, incrementándose su peso, desplazó a los dos reactivos destinados a identificar el grupo radical, como se muestra en la sexta columna del Cuadro 1.

En forma semejante a lo hallado con el pre-test, la reducción de los reactivos afectó mínimamente los porcentajes que establecen la explicación total de la varianza de los factores y los coeficientes de confiabilidad; esto es, los cuatro items espurios afectan en forma mínima la validez y la confiabilidad del post-test. (cfr. columnas 5 y 6 del Cuadro 1).

La heterogeneidad de la pérdida de reactivos se debió a que en el pre-test se encontraron dificultades en seis reactivos de la serie D y L para ubicarlos dentro de cierto factor, mientras que éstas se establecieron en los items que miden el comportamiento ácido-básico en el post-test. Por lo tanto, y con fines comparativos, ambos instrumentos fueron homogeneizados con 24 reactivos, que se encuentran localizados en las columnas 3 y 7 del Cuadro 1.

Pretest y post-test ajustados. Al someter ambos instrumentos a los procedimientos para establecer si éstos efectivamente miden lo que pretenden medir; es decir, el conocimiento sobre siete aspectos abstractos de la bioquímica operacionalizados mediante 24 items y además de que la medición sea precisa, se estructuraron, salvo ligeras dislocaciones dentro y entre los factores, dos instrumentos con errores en la medición semejantes e inclusive con pesos más elevados que las versiones originales y reducidas de las variables endógenas. Esto se observa en el Cuadro 1 cuando se contrastan las columnas 3 y 7 con las

Cuadro 1. Comparación meta-analítica de los errores de la medición para la toma de decisión.

Columna	1	2	3	4	5	6	7	8	9
	PRE - TEST				POST - TEST				META
	SIN MOD	REDUCIDO	AJUSTADO	META ANALISIS	SIN MOD	REDUCIDO	AJUSTADO	META ANALISIS	META ANALISIS
Validez por criterio exterior simulado ³⁸ :									
Coefficiente de Cramer	.9616	.9576	.9534	$\chi^2=0.01$ gl= 2 p= 20%	.9539	.9514	.9534	$\chi^2=0.01$ gl= 2 p= 20%	$\chi^2=0.03$ gl= 5 p= 20%
Validez por construcción									
Número de factores	9	8	7		10	9	7		
Promedio de rango de los coeficientes factoriales	41.12	49.06	68.63	Prueba Kruskal-Wallis $\chi^2=12.74$ gl=2 p=.0010	43.88	47.49	62.21	Prueba Kruskal-Wallis $\chi^2=9.55$ gl=2 P=.0084	Prueba Kruskal-Wallis $\chi^2=24.83$ gl=5 P=.0001
Promedio de rango de h^2 (comunalidades)	51.07	46.23	55.75	$\chi^2=1.56$ gl=2 p=.4583	48.95	50.09	53.75	$\chi^2=.42$ gl=2 p=.8085	$\chi^2=4.98$ gl=5 p=.4180
Porcentaje de la varianza total	81.5	80.1	83.2		82.1	81.5	83.3		
Promedio de rango	10.28	12.56	15.29	$\chi^2=1.97$ gl=2 p=.3722	10.85	13.06	17.86	$\chi^2=3.50$ gl=2 p=.1735	$\chi^2=5.35$ gl=5 p=.3741
Método de validación dividida en 4 partes (split-sample-validation modificado) para determinar la benignidad de la tetratomización de la puntuación de corte a número de reactivos de los test	41	35	24	$\chi^2=4.46$ gl=2 p=10%	41	37	24	$\chi^2=4.65$ gl=2 p=10%	$\chi^2=9.13$ gl=5 p=10%
b número de reactivos discriminantes	14	12	10	$\chi^2=0.67$ gl=2 p=20%	10	8	13	$\chi^2=1.23$ gl=2 p=20%	$\chi^2=2.22$ gl=5 p=20%
c (b-a)	-27	-23	-14	$\chi^2=4.16$ gl=5 p=10%	-31	-29	-11	$\chi^2=10.25$ gl=2 p=1%	$\chi^2=15.09$ gl=5 p=1%
Puntuación de corte hipotética de nivel de conocimiento:									
muy bajo	0-10	0-9	0-6		0-10	0-9	0-6		
bajo	11-20	10-18	7-12		11-20	10-18	7-12		
regular	21-30	19-27	13-18		21-30	19-27	13-18		
alto	31-41	28-35	19-24		31-41	28-37	19-24		
Tamaño de la muestra en niveles *:									
muy bajo	54	51	40		15	11	13		
porcentaje muy bajo clasificados como:									
muy bajo	94.4	96.1	95.0		100.0	90.9	92.3		
bajos	5.6	3.9	5.0		0.0	9.1	7.7		
regulares	0.0	0.0	0.0		0.0	0.0	0.0		
altos	0.0	0.0	0.0		0.0	0.0	----		

Cuadro 1. (cont.).

Columna	1	2	3	4	5	6	7	8	9
	PRE - TEST				POST - TEST				META
	SIN MOD	REDUCIDO	AJUSTADO	META ANALISIS	SIN MOD	REDUCIDO	AJUSTADO	META ANALISIS	META ANALISIS
Tamaño de la muestra en niveles *:									
bajo	55	57	80		35	43	73		
porcentaje bajo clasificados como:									
muy bajo	7.3	3.5	11.3		17.1	9.3	4.1		
bajos	85.5	91.2	85.0		82.9	86.0	89.0		
regulares	7.3	5.3	3.8		0.0	4.7	6.8		
altos	0.0	0.0	0.0		0.0	0.0	-----		
Tamaño de la muestra en niveles *:									
regular	53	56	47		68	65	99		
porcentaje regular clasificados como:									
regular	0.0	0.0	0.0		0.0	0.0	0.0		
bajos	5.7	3.6	4.3		4.4	7.7	3.0		
regulares	90.6	87.5	87.2		92.6	90.8	97.0		
altos	3.8	8.9	8.5		2.9	1.5	-----		
Tamaño de la muestra en niveles *:									
alto	23	21	18		67	66	-----		
porcentaje alto clasificados como:									
muy bajos	0.0	0.0	0.0		0.0	0.0	-----		
bajos	0.0	0.0	0.0		0.0	0.0	-----		
regulares	13.0	4.8	0.0		3.0	12.1	-----		
altos	87.0	95.2	100.0		97.0	87.9	-----		
Porcentaje de clasificación correcta	89.70	91.89	89.19	$\chi^2=0.05$ gl=2 p=20%	92.97	88.65	93.97	$\chi^2=.15$ gl=2 p=20%	$\chi^2=.24$ gl=5 p=20%
L de Wilks	.03603	.03436	.04347		.04719	.05439	.02708		
Prueba de significancia χ^2	583.22	593.27	555.01		541.12	518.23	633.34		
grados de libertad	39	36	30		30	24	39		
significancia	.0000	.0000	.0000		.0000	.0000	.0000		
L de Wilks	.44874	.56772	.48034		.59356	.55251	.38764		
Prueba de significancia χ^2	140.62	99.64	129.79		92.32	105.60	166.31		
grados de libertad	24	22	18		18	14	24		
significancia	.0000	.0000	.0000		.0000	.0000	.0000		
L de Wilks	.71069	.79161	.8980		.96257	.91666	.8592		
Prueba de significancia χ^2	59.936	41.129	27.090		6.751	15.489	26.619		
grados de libertad	11	10	8		8	6	11		
significancia	.0000	.0000	.0007		.5636	.0168	.0052		
Prueba de significancia									
Box SM	282.17	-----	253.85		159.87	133.47	-----		
F aproximada	2.3997	-----	4.1846		2.5495	3.3917	-----		
grados de libertad	91	-----	55		55	36	-----		
significancia	.0000	-----	.0000		.0000	.000	-----		
Consistencia interna (a de Cronbach)	.9355	.9255	.8748	$\chi^2=.83$ gl=2 p=20%	.9382	.9294	.8777	$\chi^2=.84$ gl=2 p=20%	$\chi^2=1.68$ gl=5 p=20%

Cuadro 1. (cont.).

Columna	1	2	3	4	5	6	7	8	9
	PRE - TEST				POST - TEST				META
	SIN	REDU	AJUS	META	SIN	REDU	AJUS	META	META
	MOD	CIDO	TADO	ANALISIS	MOD	CIDO	TADO	ANALISIS	ANALISIS
Técnica de confiabilidad de división por mitades: correlación entre las formas	.2501	.3505	.2663	$\chi^2=2.49$ gl=2 p=20%	.3759	.3358	.2929	$\chi^2=1.37$ gl=2 p=20%	$\chi^2=4.87$ gl=5 p=20%
Spearman - Brown colas (des)iguales	.4002	.5618	.4206	$\chi^2=6.70$ gl=2 p=10%	.5465	.5028	.4531	$\chi^2=1.73$ gl=2 p=20%	$\chi^2=8.83$ gl=5 p=10%
a de Cronbach primera cola	.9149	.8699	.8442	$\chi^2=1.03$ gl=2 p=20%	.9349	.9419	.8963	$\chi^2=.48$ gl=2 p=20%	$\chi^2=2.85$ gl=5 p=20%
a de Cronbach segunda cola	.9524	.9540	.8799	$\chi^2=1.39$ gl=2 p=20%	.9346	.9133	.8650	$\chi^2=1.00$ gl=2 p=20%	$\chi^2=2.77$ gl=5 p=20%
Promedio de rango de los coeficientes a de Cronbach	80.00	42.00	12.50	Prueba Kruskal-Wallis $\chi^2=83.63$ gl=2 p=.0000	80.00	42.00	12.50	Prueba Kruskal-Wallis $\chi^2=86.60$ gl=2 p=.0000	Prueba Kruskal-Wallis $\chi^2=187.83$ gl=5 p=.0000

(*) No fueron sometidos al meta-análisis debido a que invariablemente se violó la segunda regla implícita en la prueba de significancia de chi - cuadrado; es decir cuando $gl > 2$ entonces $k > 2$; por lo tanto, no más del 20 % de las celdillas podrán tener una frecuencia esperada menor de 5²⁹, preferentemente 10³⁰.

columnas 1 y 2, 5 y 6 respectivamente.

Evaluación descriptiva. Para contrastar los siete grupos, se calculó la ganancia media mínima, que se define como el porcentaje de la ganancia mínima en el 50% de la población examinada y se obtiene mediante la siguiente fórmula, que corresponde a una variante de imputación proporcional²⁴:

$$\%Gmm = [(aPre - aPost) * 100] / 41$$

en donde: %Gmm = por ciento de ganancia media mínima; aPre = aciertos obtenidos con el pretest; aPost = aciertos registrados en el post-test; 41 = número de posibles respuestas correctas en ambas mediciones.

Para efectuar los cálculos, se ordenaron por rango a los alumnos de cada uno de los siete grupos expuestos en el Cuadro 2 y la proporción de ganancia se obtuvo por cada alumno en relación al 50% de ganancia

mínima que correspondió a la población de cada grupo.

Dicho Cuadro 2 indica que la ganancia media mínima se comporta como una variable que depende del conocimiento de base de cada educando; esto, se ejemplifica con el grupo 1102, que muestra una imagen en espejo con el grupo 1136.

Ho: Pretest = Post-test. El promedio de ganancia media mínima global expuesto al final del Cuadro 2 que por cada promedio de conocimiento que mostraban los alumnos antes de ser sometidos al uso del paquete de *software* se incrementó en una razón de 1 : 1.53. Esto no permitió establecer qué tanto este resultado se encuentra influido por el azar; es decir, qué se hubiera obtenido sin someter a la clientela al efecto del *software*; por lo tanto y considerando la estructura del diseño empleado, se procedió a examinar la diferencia entre los promedios registrados

Cuadro 2. Porcentaje de aciertos en el pre-test y en el post-test y ganancia mínima media

Grupo	%Ac Pre-test	%Ac Post-test	Garancia mínima media
1102	45.61	63.17	12.20
1105	64.60	79.75	17.07
1106	45.08	64.14	17.07
1109	49.52	66.16	19.51
1111	33.12	60.42	21.95
1133	23.05	46.81	24.39
1136	20.41	49.80	31.71
Pro medio	40.20	61.47	20.56
Desviación estándar	15.66	10.94	6.28
Error estándar	2.24	1.56	0.90

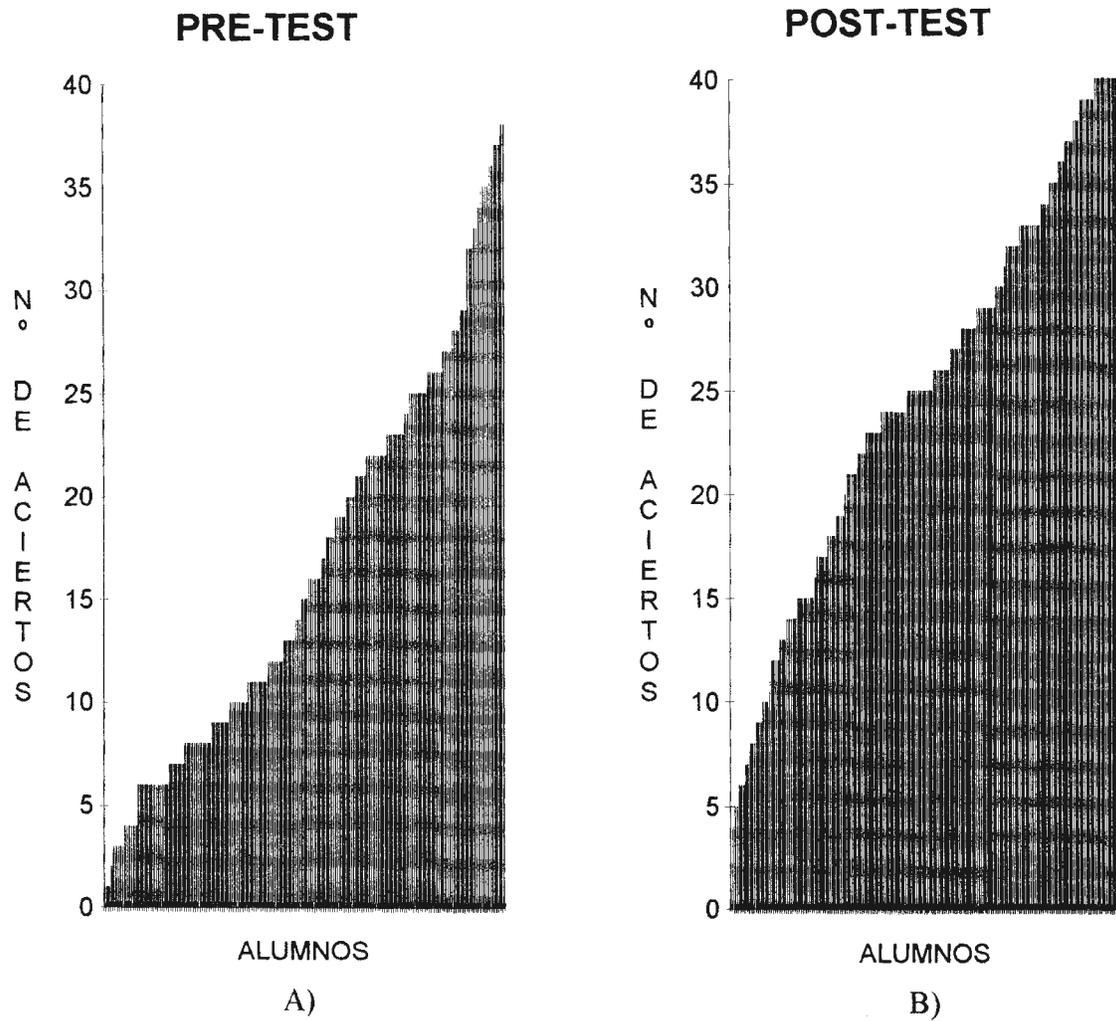
antes y después de aplicar el *software* mediante la prueba de significancia t-Student para muestras dependientes; esto, debido a que el porcentaje de varianza se mantuvo por debajo del 50% en cada uno de los 41 reactivos tanto del pre-test como del post-test.

La Fig. 1 muestra los resultados obtenidos de la aplicación del pre-test y del post-test en todos los grupos estudiados. En el gráfico de la izquierda aparecen los aciertos logrados en el pre-test y en el gráfico de la derecha se ofrecen los aciertos obtenidos en el post-test, después de que los alumnos usaron el paquete durante 90 min. Se observó un incremento en el número de aciertos obtenidos en el post-test. En esta figura se puede establecer el porcentaje de aumento en el número de aciertos de los alumnos que usaron el paquete, observándose que aproximadamente el 50 % de los alumnos exhibió un incremento en el porcentaje de aciertos en el rango situado en 21%.

El pie de figura informa que la hipótesis nula fue refutada a favor del post-test, que invariablemente presentó un mayor promedio y éste se encuentra mínimamente influido por el azar; esto es, independientemente de la estratificación empleada: pre-test y post-test sin reducción, ajustados y por cada uno de los ítems en forma particular.

El procedimiento meta-analítico¹⁶⁻²⁰ se utilizó para efectuar una exposición sumaria de un estudio de seguimiento del comportamiento de los coeficientes obtenidos mediante el de Cramer para establecer la validez a nivel de su criterio exterior en su modalidad concurrente, análisis factorial tanto de su función confirmatoria como de su función reductiva²⁵ en la que se refiere a la validez por construcción¹¹, análisis discriminante²⁶ para determinar la *split-sample-validation* modificada²⁷ y de la confiabilidad medida a través de la consistencia interna y división por mitades^{11, 22, 23} tanto a nivel intra como inter-pre-post-tests.

Los coeficientes sometidos a la prueba de chi-cuadrado de una sola entrada, fueron transformados en proporciones con base en la teoría de los conjuntos para establecer los niveles de sobreposición entre las variables; es decir, elevando al cuadrado a los coeficientes, con lo que se logró estandarizar la medición, cubriendo además el criterio expuesto por Dixon y Massey²⁸ para determinar la distancia que existe entre ellos con base en la distribución de chi-cuadrado. Debido a que no se supuso una distribución normal de los restantes coeficientes expuestos en el Cuadro 1, aunado a que se encontró un porcentaje de la varianza 50%, éstos fueron examinados mediante la prueba de significancia no paramétrica Kruskal-Wallis²⁹.



Prueba de t

Evento	n	$\Sigma (fx / n)$	ϕ	$\sqrt{\frac{\sigma^2}{n}}$	Valor de t	gl	p
Versión original: 41 reactivos							
Pre		17.80	9.82	0.72			
Post	185	26.47	9.89	0.72	14.27	184	0.000
Versión ajustada: 24 reactivos							
Pre		10.98	5.27	0.38			
Post	185	15.60	5.34	0.39	12.82	184	0.000

Fig. 1. Evaluación empírica del software. Resultados obtenidos de la aplicación del pre-test y del post-test. A) = aciertos logrados en el pre-test. B) = Aciertos obtenidos en el post-test

En la síntesis lograda en el Cuadro 1 se encontraba subyacente la toma de decisión con respecto a cuál de las seis modalidades de tests deberá utilizarse en el siguiente ejercicio de evaluación y ésta se basó en 14 contrastaciones para el pre, post y pre-post-tests; verbigracia, en el meta-análisis efectuado para el pretest se halló que en el 85% de las pruebas efectuadas no se registraron diferencias significativas, señalando que el docente puede emplear cualquiera de las tres versiones del pretest; esto, bajo el supuesto de que cada una de las pruebas empleadas tienen el mismo valor; en caso contrario, se tendría que estructurar una escala cuyos valores estarían en relación directa con la capacidad que tenga cada una de las pruebas para explicar el comportamiento del valor total de cada una de las versiones obtenidas a través del proceso reductivo a que fueron sometidos los instrumentos al no confirmarse la teoría subyacente^{11-13, 22, 23, 25}. Lo encontrado en el primer meta-análisis (columna 4, Cuadro 1) se puede generalizar a las versiones del post-test (columna 8, Cuadro 1) donde únicamente se presentó un ligero decremento en el porcentaje de la consistencia interna de 78 a 85%, esquema que prevaleció al meta-analizar los dos meta-análisis previos.

Discusión

Según Tyler^{30,31}, la práctica de evaluar la consecución educativa tiene una larga historia, pero el estudio sistemático de medir con instrumentos y otras formas de la valoración educacional tiene menos de 100 años en el mercado correspondiente. Los cambios profundos registrados en este terreno educativo actualmente se encuentran influenciados por la conceptualización de la evaluación; sin embargo, subraya el autor citado^{30,31} que los procedimientos y los instrumentos de evaluación únicamente han respondido en forma parcial a estos cambios, permaneciendo los resultados en un nivel descriptivo a ultranza. Por tal circunstancia, el reciente incremento en número y posibilidades de dispositivos tecnológicos en la educación, como la televisión, video y computadoras con sus respectivos *softwares*, han requerido proponer estrategias para evaluar la efectividad de éstos en diferentes situaciones educativas. Aunado a que ningún paradigma de la investigación está capacitado para responder a todas las preguntas que surgen en la investigación educativa³.

La primera evaluación, fuera del campo de los procedimientos inferenciales y, por ende, basada en la evidencia empírica³², fue realizada bajo el amparo

de la teoría constructivista piagetiana^{3, 33} ligeramente modificada en el sentido de que originalmente los objetivos educativos cruciales tienen que ser definidos con los educandos, mientras que en este ejercicio los objetivos inmersos en este *software* fueron dispuestos no únicamente a los educandos sino también a los docentes que aceptaron confrontarse con esta herramienta educativa. Bajo este postulado teórico, en 1972 Papert³⁴ también evaluó su sistema modular educativo denominado Ambiente Logo, mientras otros autores permanecen en el terreno discursivo^{35, 36}, y Castro³⁷, al no hallar resultados empíricos en el campo educativo, para ejemplificar la dinámica de los diseños de evaluación, emplea la simulación³⁸. Esta carencia no fue refutada doce años más tarde por Alarcón, Aréchiga y De la Fuente³⁹, que tampoco registraron el comportamiento de la investigación educativa médica en México entre los dos decenios contrastados: 1968-1977 y 1978-1987.

Lo referido por Keeves^{2, 32} con respecto a evaluar la efectividad de los dispositivos tecnológicos, deberá permanecer como un objetivo de índole empírico⁴⁰, que deberá extenderse hacia los *softwares* educativos en general y médicos en particular, debido a que la oferta en el mercado es amplia⁴¹, pero la efectividad educativa es evaluada usualmente con apoyo de constructivismo piagetiano³³. Considerando esta deficiencia, este ejercicio exploratorio se basó en las recomendaciones señaladas por Campbell y Stanley⁴⁰, Castro³⁷ y Keeves^{2, 32} y el *software* fue sometido, en un segundo momento, a la evaluación empírica²¹ para examinar tanto la efectividad de los instrumentos de medición cuyos reactivos fueron derivados de los objetivos educativos cruciales del *software* como de la herramienta educativa, esto para establecer las medidas de seguridad y determinar la capacidad de enseñanza respectivamente con el propósito de ofrecer al mercado nacional un producto educativo con control de calidad.

Las reducciones de los instrumentos de medición exhibidas en el Cuadro 1, podrían fungir como apoyo empírico a lo externado por Bortz en 1984²¹ y Tyler seis años después^{30, 31}, con respecto a que la teoría clásica o tradicional de la medición presenta una declinación de su influencia en el mercado correspondiente, por el peso que muestra la teoría probabilística de los tests; esto es, sobre la magnitud y estructura de los instrumentos. La primera teoría se apoya en una construcción heurística de los tests, mientras que la segunda exige que la variable endógena contenga pocos reactivos derivados estrictamente de la teoría subyacente al

instrumento y que, además, se sostenga a nivel empírico con las técnicas robustas de la validez y de la confiabilidad empleadas en este ejercicio^{11, 22, 23}; es decir, pocos reactivos pero efectivos, con lo que también se reduce el tamaño de la muestra y con ello, el costo para cuantificar los errores en la medición de una variable dependiente⁴². Ante el resultado tan cerrado hallado en la exposición sumaria de los estudios individuales a nivel meta-analítico (columnas 4, 8 y 9 del Cuadro 1)¹⁶⁻²⁰, la decisión se deberá desempatar con base en la propuesta implícita en la teoría probabilística de los tests, además de que las versiones ajustadas del pre y post-test, al ser liberadas de los reactivos espurios, mejoraron invariablemente los niveles de seguridad, aunque no fuera de lo esperado a nivel probabilístico, en contraste con las versiones originales y reducidas. Asimismo, combinando el coeficiente de Cramer^{11, 29, 42} con la capacidad clasificatoria de sujetos del análisis discriminante²⁶ y con la adaptación realizada por Holtzman²⁷ para validar la capacidad divisoria de un instrumento de medición, fue posible pasar de instrumentos cuantitativos a cualitativos, ya que al validar las puntuaciones de corte, tanto del pre como del post-test, se halló que son capaces de discriminar en forma elevada estudiantes de medicina con conocimientos que varían desde muy bajos hasta altos acerca de los conceptos cruciales de los α aminoácidos. No obstante, por la estructura de los programas de evaluación, persiste la posibilidad de ofrecer a los usuarios las deficiencias tanto a nivel grupal como individual y, por ende, los educandos sometidos a este *software* podrían seleccionar a través del menú tutorial los conceptos y donde se registraron las fallas y realizar nuevamente los ejercicios respectivos.

En este ejercicio exploratorio para evaluar tanto los instrumentos de medición como el paquete de *software* se utilizó, de los modelos de evaluación que ofrece el mercado educativo^{43, 44}, el propuesto en 1984 por Miles y Huberman⁴⁵ por su bajo costo de operación, ya que con los otros^{43, 44} se requiere contar o estructurar una red de monitoreo de diversas variables evaluativas bajo condiciones cuasiexperimentales^{2, 3}, incrementándose, como se mencionó en el último párrafo del apartado correspondiente a la descripción del material usado y los métodos empleados, el costo de operación.

Bajo el auspicio de un capital austero y con base en la experiencia previa⁴⁶⁻⁴⁸, el modelo de evaluación empleado⁴⁵ se operacionalizó mediante un diseño no experimental, que comprende establecer a

través de un instrumento válido y confiable^{11, 13, 23, 27, 49} el nivel de conocimiento acerca de los conceptos cruciales de los α -aminoácidos previo y posterior al sometimiento de siete grupos de estudiantes de medicina (Cuadro 2) al *software* Los Aminoácidos, se utilizó un diseño de pre y post-test^{10, 21, 22} y por su estructura, éste se encuentra afectado por la circularidad marcada por Castro³⁷, que puede interrumpirse a través de combinar un diseño 2 por 2 con un pre y post-test de grupos^{10, 21, 22}, además de medir el efecto que tiene la primera medición sobre las diversas formas de proporcionar conocimiento acerca de los aminoácidos y sobre el post-test, con lo que se incrementa tanto el control como el costo. Esto indica, según Castro³⁷, que el incremento significativo registrado en el post-test de la Fig. 1 en comparación con el pretest, deberá permanecer como un argumento falaz, aunque no contradictorio, ya que no es posible afirmar que el cambio observado se deba exclusivamente a la efectividad del *software*. Por lo tanto, siguiendo la recomendación de Castro³⁷, para evaluar el *software* de pH se empleará un diseño de pre y post-test ampliado para contrastar simultáneamente tres diferentes modalidades de enseñanza, implicando un mayor número de docentes voluntarios y con ello, ejercer, al menos, el control sobre la distribución aleatoria respectiva, permaneciendo abiertas las preguntas acerca de la homogeneidad de los docentes en su exposición dentro de cierta modalidad de ofrecer el conocimiento sobre los conceptos cruciales del pH y el efecto del pretest^{10, 21-23, 47}.

La dispersión del apoyo financiero implícito en los ejercicios científicos aislados, se podría controlar con la implantación de la investigación sistemática para evaluar la efectividad de las diversas técnicas educativas empleadas en proporcionar conocimiento acerca de la bioquímica y otras asignaturas incluidas en el Plan Unico de Estudios de la Carrera de Médico Cirujano⁵ de la Facultad de Medicina y el ahorro podría invertirse en el empleo de modelos de evaluación más rigurosos^{43, 44}, que el utilizado en esta exploración empírica con las respectivas limitaciones⁴⁹, donde tampoco se logró establecer el costo del beneficio.

Considerando el principio que en toda medición existe un error implícito¹⁰, las variables operacionalizadas como endógenas únicamente pueden estar constituidas por ítems que previamente fueron sometidos a la evaluación correspondiente con cualquiera de las técnicas de validez y confiabilidad que se encuentran ampliamente difundidas en el

mercado de la instrumentación; es decir, siempre serán *a posteriori*, implicando que la usual colección de reactivos destinados para estructurar los instrumentos de medición únicamente podrán incluirse después de haberse demostrado empíricamente que efectivamente miden lo que pretenden medir y además lo hacen con precisión, subrayando que cada reactivo deberá estar marcado por ambos valores, que podrán evaluarse a través del tiempo, siguiendo las fluctuaciones de este tipo de coeficientes, que al perder esta cualidad, por medio de un sistema de filtros de control de calidad, podrán ser suspendidos en forma temporal para realizar las modificaciones pertinentes o

definitivamente y los reactivos que muestren la mayor estabilidad podrían ser incorporados en la Fase Teórica del Examen Profesional emanado del Plan Unico de Estudios de la Carrera de Médico Cirujano⁵.

Nota. Los lectores interesados pueden solicitar tanto la base de datos y los programas empleados, como las tablas que reflejan el comportamiento de cada uno de los reactivos antes y después de la aplicación del software.

Reconocimientos. El desarrollo del paquete de software para la enseñanza "Los aminoácidos" fue apoyado por la Dirección General de Asuntos del Personal Académico en el Programa de Apoyo a Proyectos de Investigación y de Innovación Docente con la clave No. CD 702592 durante el año de 1993.

Referencias

1. Fernández R-R L. Elaboración de prototipos para el uso creativo de la computadora en la docencia de la bioquímica. BEB 1994; 13:52-57.
2. Keeves JP. The methods of educational inquiry. En: Keeves JP, ed. Educational research, methodology and measurement: An international handbook. Oxford: Pergamon Press, 1990:3-8.
3. De Landsheere G. Research perspectives. En: Keeves JP, ed. Educational research, methodology and measurement: An international handbook. Oxford: Pergamon Press, 1990:9-16.
4. Cea BA, Del Arenal P, Fernández R-R L, Gutiérrez R. Manual de objetivos y prácticas de laboratorio del curso de bioquímica. México: Facultad de Medicina, Universidad Nacional Autónoma de México, 1993.
5. Plan Unico de Estudios de la Carrera de Médico Cirujano. Rev Fac Med UNAM 1993.36:4 Suplemento N° 1, 11-55.
6. Lehninger A L. Biochemistry. New York: Worth Publishers, 1975.
7. Lehninger A L, Nelson DL., Cox M M. Principles of Biochemistry. New York: Worth Publishers, 1993.
8. Boshinski R.C. Bioquímica. Wilmington: Addison-Wesley Iberoamericana, 1991.
9. Voet D. Voet J.G. Biochemistry, New York: John Wiley, 1990.
10. Spector PE. Research designs. Sage University Paper Series on Quantitative Applications in the Social Sciences series 07-023. Beverly Hills: Sage Publications, 1982.
11. Carmines GC, Zeller RA. Reliability and validity assessment. Sage University paper series on Quantitative Applications in Social Sciences, 07-017 Beverly Hills: Sage Publications, 1982.
12. Kim J-O, Mueller ChW. Factor Analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences series 07-014. Beverly Hills: Sage Publications, 1982.
13. Friedrichs J. Methoden empirischer. Sozialforschung. Opladen: Westdeutscher Verlag, 1981.
14. Richter G, Klemm PG, Zahn M. Beitrag zur Lösung des Dreigruppen-Trennproblems in der Alkoholismus-Diagnostik --Vorschlag für einen 9- Item-Screening-Test. Zschr Klin Med 1990;1:79-83.
15. Johnson RA, Wichern DW. Applied multivariate statistical analysis. New Jersey: Prentice Hall, 1982.
16. Neiss A. Erkenntnisgewinn durch Meta-Analyse? Deut Arzbt 1990 20:999-1000.
17. McGaw B. Meta-analysis. En: J P Keeves JP, ed. Educational research, methodology and measurement: An international handbook Oxford: Pergamon Press, 1990:678-85-
18. Johnstone BM, Leino EV, Motoyoshi MM, Temple MT, Fillmore KM, Hartka E. An integral approach to meta-analysis in alcohol studies. Brit J Addic 1991; 86:1211-20.
19. Fillmore KM, Hartka E, Johnstone BM, Leino EV, Motoyoshi MM, Temple MT. A meta-analysis of live course variation in drinking. Brit J Addic 1991; 86:1221-68.
20. Temple MT, Fillmore KM, Hartka E, Johnstone BM, Leino EV, Motoyoshi MM. A meta-analysis of change in marital and employment status as predictors of alcohol consumption on a typical occasion. Brit J Addic 1991; 86:1269-81.
21. Bortz J. Lehrbuch der empirischen Forschung. Berlin: Springer Verlag 1984.
22. Gorenc K-D, Herrera ME, Reynaga J. Manual de diseños de investigación: Texto autoevaluativo. México: Instituto Nacional de Ciencias Penales, 1986.
23. Nadelstischer A. Técnicas para la construcción de cuestionarios de actitudes y opción múltiple. México: Instituto Nacional de Ciencias Penales, 1983.
24. Gorenc K-D. Tengo derecho a tener un último secreto... Acta Médica 1990;101-104:61-89.

25. Kim J-O, Mueller ChW. Introduction of factor analysis. Sage University paper series on Quantitative Applications in Social Sciences, series no 07-013. Beverly Hills: Sage Publications, 1982.
26. Klecka WR. Discriminant analysis. Sage University Paper series on Quantitative Applications in Social Sciences, series no 07-019. Beverly Hills: Sage Publications, 1982.
27. Holtzman WH. Projective techniques. En: Triandis HC, Berry JW, eds. Handbook of cross-cultural psychology. Vol. 2 Methodology. Boston: Allyn & Bacon, 1980:245-78.
28. Dixon WJ, Massey FJ. Introduction to statistical analysis. New York: McGraw-Hill, 1965.
29. Siegel S. Nichtparametrische statistische Methoden. Frankfurt/Main: Fachbuchhandlung für Psychologie Verlagsabteilung, 1976.
30. Tyler RW. Changing concepts of educational evaluation. Int J Educ Res 1986; 10:91-6.
31. Tyler RW. Evaluation for utilization. En: Keeves JP, ed. Educational research, methodology and measurement: An international handbook. Oxford: Pergamon Press, 1990: 155-64.
32. Keeves JP. Research techniques and statistical analysis. En: Keeves JP, ed. Educational research, methodology and measurement: An international handbook. Oxford: Pergamon Press, 1990:501-5.
33. Piaget J, Grize JB, Szeminska A, Bang V. Epistemologie und Psychologie der Funktion. Stuttgart: Klett-Cota, 1977.
34. Papert S. Teaching children thinking. Program Learn Educ Technol 1972; 9:245-55.
35. Novikov L. Probleme der Planung und Organization der pädagogischen Forschung in der Sowjetunion. En: Mitter W, NovikovL, eds. Pädagogische Forschung und Bildungspolitik in der Sowjetunion: Organization, Gegenstand, Methoden. Frankfurt/Main: Deutsches Institut für Internationale Pädagogische Forschung, 1977.
36. Wall WD. The work of the National Foundation for Educational Research in England and Wales. En: Butcher HJ, ed. Educational research in Britain. London: University of London Press, 1968: 15-32.
37. Castro L. Diseño experimental sin estadística. México: Editorial Trillas 1984.
38. Baudrillard J. Cultura y simulacro. Barcelona: Editorial Kairós, 1987.
39. Alarcón D, Aréchiga H, De la Fuente, JR. El estado actual de la investigación médica en México. Ciencia y Desarrollo 1990; 93:55-67.
40. Campbell DT, Stanley JC. Experimental and quasiexperimental designs for research on teaching. En: Gage NL, ed. Handbook of research on teaching. Illinois: Rand McNally, 1963: 171-246.
41. Strassmich M. Media Magic: Computers in science and art. Nicasio: Media Magic, 1994
42. Gorenc K-D, Reynaga J, Beltrán JE, Martínez V. Cálculo de la muestra mínima para evaluar la validez y confiabilidad de pruebas diagnósticas: un estudio preliminar. Rev Intercont Psic Educ 1988; 1:49-68.
43. Sowden S, Keeves JP. Analysis of evidence in humanistic studies. En: Keeves JP, ed. Educational research, methodology and measurement: An international handbook. Oxford: Pergamon Press, 1990:513-26.
44. Sánchez-Guerrero G. Un marco teórico para la evaluación. Cuadernos de Planeación y Sistemas. Facultad de Ingeniería. México, D.F.: Universidad Nacional Autónoma de México, 1990.
45. Miles MB, Huberman AM. Qualitative data analysis: A sourcebook of new methods. Beverly Hills: Sage Publications, 1984.
46. Gorenc K-D, Romero-Tequextle G, Zurita-Campos MA, Cadenas-Sosa DA. Evaluación de seis talleres de diseños de investigación auspiciados por la Procuraduría General de Justicia del estado de Tabasco, México. Universidad y Ciencia 1988; 10:31-54.
47. Gorenc K-D, Alday E, Gorenc J. Medición del efecto de la reacción al instrumento mediante el diseño de dos grupos de Solomon: un ejercicio científico. Rev Intercont Psic Educ 1989; 1-2:259-277.
48. Gorenc, K-D. La utilización del pre y post-test para evaluar un taller de diseños de investigación aplicado en América Latina: reporte de los resultados. Rev Neuro Psiquiat 1991; 54:29-40.
49. Einhorn HJ, Hogart RM. Confidence judgment: Persistence of the illusion of validity. Psychol Rev 1978;5 : 395-416.
50. Nesbit J E. Chi-square. Manchester: Manchester University Press. 1966



REPOSICIÓN DE LA BIBLIOTECA
 DE LA UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
 INSTITUTO DE INVESTIGACIONES PSICOLÓGICAS