

Evaluación estudiantil de la actuación del profesor y su efecto sobre la efectividad docente

CARLOS E. VARELA-RUEDA*

Se realizó un estudio prospectivo de investigación educativa de tipo observacional, comparativo y longitudinal con el objetivo de determinar si la evaluación estudiantil de la actuación docente y la retroinformación inmediata que los alumnos proporcionan a los profesores son de utilidad para mejorar la efectividad docente.

Se diseñó un instrumento de evaluación por medio del cual los alumnos otorgaron una calificación acerca de su desempeño docente a once médicos que participaron como profesores de tres ó cuatro cursos de formación de profesores de medicina. Se compararon las calificaciones de cada profesor con las que él mismo obtuvo en cursos anteriores y se observó diferencia estadísticamente significativa en prácticamente todos los casos ($p < 0.001$), lo cual se interpretó como mejoría en el desempeño después de recibir retroinformación inmediata al término del curso previo. Existió correlación positiva fuerte ($r = 0.86$) entre diferentes evaluaciones practicadas a los alumnos y las calificaciones que los estudiantes otorgaron a los profesores. El instrumento de evaluación resultó ser confiable y razonablemente válido.

Se concluye que la evaluación que los estudiantes hacen acerca de la actuación docente y la retroinformación inmediata que el profesor recibe son de gran utilidad para mejorar la efectividad docente.

CLAVES: Evaluación educativa, Evaluación de profesores, Profesores de medicina, Efectividad docente

Summary

A prospective study of educational research was carried out. Its type was observational, comparative and longitudinal being its main objective to determine whether the student evaluation of the teachers' performance and the immediate feedback that the students gave to the teachers are useful to improve teaching effectiveness.

An instrument of evaluation was outlined with the students give a certain grade on the teaching performance of eleven physicians who participated as teachers in three or four courses for training of medical teachers. Each teacher's grades were compared with the ones they got in preceding courses, and a significant statistical difference was observed in almost all the cases ($p < 0.001$) which was interpreted as an improvement in performance after receiving the immediate feedback at the end of the previous course. There was a strong positive correlation ($r = 0.86$) between different tests applied to the students and the grades that the students give to the teachers. The instrument of evaluation turned out to be reliable and reasonably valid.

In conclusion, the evaluation that the students make of the teaching performance and the immediate feedback that the teacher receives are of great help to improve teaching effectiveness.

Presentado en sesión ordinaria de la Academia Nacional de Medicina, el 9 de noviembre de 1988.

*Académico numerario, Departamento de Investigación Educativa y Formación de Profesores, Instituto Mexicano del Seguro Social.

Casi siempre, en el ámbito educativo la evaluación se ha relacionado solamente con la determinación del grado de aprovechamiento escolar de los educandos. Sin embargo, en la actualidad se sabe que la evaluación educativa se debería realizar a partir de diversos objetos de estudio como son los estudiantes, el aprendizaje, los maestros, los planes y programas, los recursos didácticos, las instituciones y los sistemas educativos.¹ En particular, a la evaluación de la efectividad docente se le debe dar tanta importancia en la educación superior como la que se concede a la valoración de los estudiantes. Para realizarla se dispone de métodos generales que se valen de los logros de los alumnos, de las opiniones de los estudiantes, de otros maestros o del propio profesor que se autoevalúa, o de instrumentos estandarizados que miden la eficiencia docente.²

Uno de los procedimientos más útiles es la evaluación que el alumno hace, desde su particular punto de vista, acerca de lo que le ha parecido el trabajo docente de sus profesores. Dicha evaluación se ha usado desde hace más de cincuenta años³ en niveles educativos y campos de enseñanza muy distintos, tales como son el cálculo en la preparatoria⁴ o en escuelas de ingeniería,⁵ la contaduría y administración en estudiantes que cursan la licenciatura⁶ y la anatomía,⁷ las ciencias básicas,⁸ la metodología clínica⁹ y muchas otras áreas de la medicina.^{10,11}

En Latinoamérica parece utilizarse poco la evaluación estudiantil de la efectividad docente en medicina o, por lo menos, no se ha estudiado y publicado mucho sobre ella, como lo demuestra el hecho de que en la revista *Educación médica y salud*, que es la principal publicación en español dedicada a estos aspectos, no haya aparecido un solo artículo sobre evaluación de la efectividad docente en los 21 años transcurridos desde la aparición de su primer número en 1966.

Algunos investigadores se han mostrado escépticos acerca de la utilidad de las evaluaciones practicadas por los estudiantes para evaluar la enseñanza,¹² ya que podrían tener sesgos por distintas causas como son el llamado efecto doctor Fox,^{13,14} el momento en que se practican,¹⁵ el nivel de los estudiantes,¹⁶ los procedimientos que se utilizan¹⁷ o porque pueden llegar a tener problemas de confiabilidad y validez, razón por la que estos factores han sido los que más se han investigado.¹⁸

Por otra parte, un aspecto que se ha estudiado poco y acerca del cual no existe acuerdo todavía, es el que se refiere al valor real de la retroinformación estudiantil inmediata. En particular, en el escenario médico dicho valor no ha sido establecido plenamente, ya que Bandaranayake ha informado de resultados contradictorios¹⁹ y sólo Irby y col. encontraron que la retroinformación inmediata es suficientemente sensible para identificar diferencias en la enseñanza entre los profesores, las cuales no se pueden apreciar tan bien cuando dicha retroinformación se proporciona tardíamente.²⁰

El objetivo de este estudio es determinar si la evaluación estudiantil de la actuación docente y la retroin-

formación inmediata que los alumnos proporcionan a los profesores son de utilidad para mejorar la efectividad docente.

Material y métodos

Se diseñó un protocolo de investigación educativa de tipo observacional, comparativo y longitudinal en el cual se siguió prospectivamente el desempeño docente de once médicos que tuvieron participación como profesores en los diferentes cursos que se impartieron en el Centro de Formación de Profesores del Instituto Mexicano del Seguro Social (IMSS) en el periodo comprendido entre el mes de mayo de 1985 y el mes de septiembre de 1986. En el diseño del estudio se tomaron en consideración los aspectos siguientes:

a. Descripción de los cursos

Se realizaron cuatro cursos de formación de profesores de medicina (CFPM) que se identificaron con los números romanos del I al IV. Los CFPM tuvieron en promedio 525 horas de duración, mismas que fueron distribuidas a lo largo de quince semanas. La estructura curricular fue de tipo modular integrativo²¹ y en cada CFPM se programaron nueve unidades didácticas, las que se codificaron con números arábigos consecutivos del 1 al 9. El promedio de tiempo dedicado a las unidades didácticas fue de 57 horas que se impartieron mediante técnicas didácticas grupales. La temática general fue la siguiente: teorías de grupos y comunicación humana (unidad 1), administración (2), metodología cde la investigación científica (3), metodología estadística (4), implicaciones sociales de la educación (5), psicología del aprendizaje (6), planificación didáctica (7), metodología de la enseñanza (8) y evaluación educativa (9).

b. Características de los alumnos

Todos los participantes en los CFPM fueron médicos que tenían especialidades distintas y que antes de ingresar al CFPM fungían como profesores titulares, adjuntos o auxiliares de los cursos de especialización en medicina (CEM) que bajo la forma de una residencia se realizan en las unidades de atención médica del IMSS localizadas en las ciudades principales del país. A los diferentes CFPM asistieron 72 médicos: 17 al I, 19 al II, 19 al III y 17 al IV.

c. Características de los profesores

Los CFPM fueron impartidos en forma conjunta por dos tipos de profesores que se denominaron invitados (PI) y de tiempo completo (PTC). Los PI fueron médicos egresados de alguno de los CFPM realizados previamente; trabajaban regularmente como profesores de un

CEM y acudieron al Centro de Formación de Profesores para participar en los CFPM como PI. 19 PI asistieron al curso I, 16 al II, 23 al III y 12 al IV. Sin embargo, en el estudio sólo se incluyeron nueve PI (identificados con letras de la A a la I), porque fueron los que cumplieron el criterio de inclusión de haber participado en la docencia de tres CFPM como mínimo durante el tiempo dedicado a la investigación, aunque no fuera en forma consecutiva. Siete de los PI fueron especialistas en medicina familiar y, de los otros dos, uno era pediatra y el otro ginecólogo. Todos los PI provinieron de algún CEM impartido en el interior del país; seis fueron profesores titulares y tres adjuntos. Ninguno de los PI recibió estímulos económicos por participar en los CFPM.

Los dos PTC (identificados como J y K) fueron especialistas en medicina familiar que cumplieron su tiempo completo en el Centro de Formación de Profesores del IMSS. En cada CFPM los PTC participaron en varias unidades didácticas. En el estudio sólo se tomaron en cuenta aquellas unidades en las que los PTC hubieran intervenido por lo menos tres veces aunque no fueran consecutivas; tal fue el caso en cuatro unidades distintas para cada PTC.

d. Evaluación a los alumnos

En cada unidad didáctica la evaluación del área cognoscitiva se hizo mediante pruebas escritas de tipo opción múltiple. El desempeño en los trabajos prácticos se valoró por medio de guías de observación y trabajos escritos realizados por los alumnos, mientras que el área afectiva sólo se evaluó con guías de observación. Debido a que en los CFPM se utilizó la evaluación por criterio,²² se requirió que los alumnos lograran un mínimo de setenta por ciento de los objetivos del curso para poder acreditarlo.

Para evaluar integralmente los conocimientos adquiridos durante todo el CFPM, se practicó un examen con cien reactivos de opción múltiple que correspondieron a las nueve unidades didácticas y que se distribuyeron en una proporción semejante a la importancia relativa de cada unidad del curso. Este examen se practicó tanto al inicio como al final del CFPM para poder valorar las ganancias absoluta y relativa de conocimientos, tanto de cada alumno como del grupo como un todo.

e. Evaluación a los profesores

Se realizó por medio de una escala evaluativa en la que se describe en forma sucinta la conducta a evaluar, misma que se coteja contra un número comprendido en la escala que va del cero al diez. Con este instrumento cada profesor fue evaluado por los alumnos acerca de los trece aspectos siguientes: precisión de los objetivos, orden en la presentación de los contenidos, grado de dominio de los temas, claridad en la entrega de los con-

tenidos, amabilidad en la presentación de los temas, utilización de la metodología didáctica, utilidad de los ejercicios y laboratorios realizados, colaboración para el desarrollo de ejercicios y laboratorios, disposición para aclarar dudas y precisar conceptos, ambiente social, actitud hacia el grupo, capacidad para despertar interés y reconocimiento de las propias limitaciones.

Con anticipación suficiente, todos los PI y PTC fueron informados de que la evaluación que se les practicaría no tendría fines promocionales, ni sería utilizada para definir si podrían conservar el cargo o para otorgarles estímulos económicos, ya que el propósito explícito era contar con un elemento adicional que les permitiera mejorar su efectividad docente y que la evaluación y la retroinformación sólo se harían a los profesores que las aceptaran en forma voluntaria, lo cual ocurrió en todos los casos.

La evaluación a los PI y PTC fue practicada en forma anónima por todos los alumnos al concluir las actividades académicas de cada unidad didáctica, sin que los profesores estuvieran presentes, y antes de que éstos evaluaran a los alumnos. En los cuatro CFPM se utilizó el mismo instrumento; antes de aplicarlo, los estudiantes recibieron una explicación acerca de su propósito y contenido así como un adiestramiento sobre la manera de utilizarlo.

Una vez que se resolvió el examen de conocimientos y sólo después de que fueron evaluados los trabajos prácticos y el área afectiva de los alumnos, los profesores fueron informados inmediatamente acerca de las calificaciones que les otorgaron los estudiantes. Acto seguido el grupo de alumnos proporcionó a todos los profesores de la unidad didáctica una retroinformación verbal complementaria de las calificaciones.

En la investigación se compararon los promedios de calificaciones obtenidas secuencialmente por los profesores en los diferentes CFPM. Adicionalmente, en el caso de los PTC se compararon las calificaciones obtenidas en las diferentes unidades didácticas de cada CFPM en que participaron. El estudio fue ciego simple ya que ni los alumnos ni los profesores supieron que los resultados de las calificaciones y de la retroinformación iban a ser utilizados en la investigación.

f. Análisis estadístico

Los valores promedio de las calificaciones se calcularon con los métodos de estadística descriptiva habituales y se expresaron como la media aritmética (\bar{x}) \pm la desviación estándar (DS). También se empleó el coeficiente de correlación lineal r de Pearson.²³

La comparación de las mediciones sucesionales de las calificaciones de los cuatro CFPM se hizo con análisis de varianza de una vía.²⁴ Para hacer la comparación múltiple de medias se utilizaron el método de la diferencia significativa honesta de Tukey²⁵ o el método de Scheffé,²⁶ según que existiera o no el mismo número

de casos en cada grupo. Para efectuar la comparación de medias entre dos muestras independientes se empleó la prueba *t* de Student de dos colas.²⁷

La confiabilidad de las calificaciones otorgadas por los alumnos se calculó por el método de Ebel,²⁸ mientras que la confiabilidad de las calificaciones asignadas a los exámenes de conocimientos se determinó por el método de las mitades.²⁹

La significancia estadística fue definida como un valor de *p* menor de 0.05.

Resultados

El rango de las calificaciones que los alumnos otorgaron a los profesores en los trece criterios evaluados mostró diferencias. La calificación mínima recibida por un profesor fue de 77.65 en el curso I, 81.05 en el II, 86.84 en el III y 89.41 en el IV, mientras que la calificación máxima fue de 100.00 en cada uno de los cuatro CFPM.

En el Cuadro I se presentan los promedios de las calificaciones que los educandos asignaron a cada uno de los PI y PTC. Al practicar la comparación múltiple de medias secuenciales obtenidas por cada profesor en la unidad didáctica en que participó se observaron diferencias estadísticamente significativas: $p < 0.05$ para el PTC K (unidad 7), $p < 0.01$ para el PI H y $p < 0.001$ para los otros quince casos.

La confiabilidad de las calificaciones que los alumnos proporcionaron a cada profesor en los tres o cuatro CFPM (Cuadro I) varió desde 0.03 en el caso del PI B hasta 0.93 en el PTC K (unidad 5), en tanto que la confiabilidad de las calificaciones estudiantiles dadas en un curso fluctuó entre 0.01 para el PI B y 0.81 para el PTC K (unidad 5).

En el Cuadro I también aparecen los promedios de calificaciones alcanzados por todos los profesores (PI + PTC). La confiabilidad de las calificaciones en los cursos fue de 0.47 y en un curso de 0.28. Se observó un aumento progresivo de los promedios obtenidos en los diferentes cursos y el análisis de varianza reveló diferencia significativa (52 grados de libertad; $F = 5.84$; $p < 0.01$).

Después de calcular los promedios de calificaciones obtenidos en las nueve unidades didácticas sólo para el grupo de PI (Cuadro II, renglón 1), se encontró que la confiabilidad de las calificaciones en todos los cursos fue de 0.48 y en un solo curso de 0.27. Al hacer el análisis de varianza se encontró diferencia significativa (27 grados de libertad; $F = 6.88$; $p < 0.01$). Al efectuar la comparación posterior entre los pares de las medias no se encontró diferencia significativa ($p > 0.05$) en los casos siguientes: profesores B y H (entre los cursos I y II), profesores A y H (entre los cursos II y III) y profesores D y F (entre los cursos III y IV), mientras que en las otras 24 comparaciones la diferencia observada fue significativa ($p < 0.001$).

Los promedios de las calificaciones otorgados al PTC J en las unidades 1, 4, 6 y 9 (Cuadro 2I, renglón 2) tuvieron confiabilidad de 0.44 en todos los cursos y de 0.26 en un curso. En el análisis de varianza no existió diferencia significativa (12 grados de libertad; $F = 0.82$; $p > 0.05$), por lo que al hacer la comparación entre los diferentes pares de medias tampoco se encontró diferencia significativa ($p > 0.05$). Al practicar la comparación longitudinal por pares de las medias de las calificaciones alcanzadas en las unidades didácticas de un mismo curso (Cuadro III, no se observó diferencia significativa ($p > 0.05$) entre los promedios de las unidades 1 y 4 (curso III), en tanto que en todas las otras catorce comparaciones la diferencia fue significativa (desde $p < 0.025$ hasta $p < 0.0005$).

El PTC K alcanzó en las unidades didácticas 2, 3, 5 y 7 los promedios de calificaciones que aparecen en el Cuadro II (renglón 3). La confiabilidad de las calificaciones en los cursos fue de 0.47 y en un curso de 0.32. En el análisis de varianza no se encontró diferencia significativa (11 grados de libertad; $F = 1.35$; $p > 0.05$) y, por ello, tampoco la hubo entre las medias de los diferentes cursos ($p > 0.05$). Al comparar longitudinalmente y por pares los promedios de las calificaciones obtenidos en las diferentes unidades didácticas de cada curso (Cuadro IV), no se observó diferencia significativa entre la unidad 2 y la 3 en el curso I ($p > 0.50$), ni entre la unidad 5 y la 7 en el curso III ($p > 0.10$), mientras que en los otros nueve pares de medias la diferencia fue significativa (desde $p < 0.05$ hasta $p < 0.0005$).

En el examen integral de conocimientos efectuados a los alumnos al finalizar el CFPM se logró un promedio grupal superior a setenta por ciento en cada curso (Cuadro V). En los cuatro exámenes la confiabilidad fue alta (0.70 en el curso I, 0.89 en el II, 0.84 en el III y 0.62 en el IV) y el error estándar al determinar la confiabilidad fue relativamente bajo (3.24, 3.29, 3.70 y 2.88 respectivamente). En los cuatro CFPM existió diferencia significativa entre el promedio grupal de calificaciones en la evaluación diagnóstica y el de la final ($p < 0.0005$), mientras que en el valor del coeficiente de correlación lineal *r* de Pearson entre ambos exámenes no hubo diferencia significativa en caso alguno ($p > 0.05$).

Se encontró una correlación positiva fuerte ($r = 0.86$) entre los promedios de las calificaciones grupales obtenidos por los alumnos en la evaluación integral final y los promedios de las calificaciones que los estudiantes otorgaron a los profesores. (Cuadro VI).

Todos los alumnos aprobaron el CFPM en que participaron; el rango de las calificaciones globales estuvo entre 76.64 y 92.31 en el curso I, entre 77.82 y 89.14 en el II, entre 73.99 y 90.84 en el III y entre 76.97 y 90.23 en el IV. Se observó una correlación positiva fuerte ($r = 0.86$) entre los promedios de las calificaciones obtenidas por los profesores y los promedios en la calificación global (que incluye las evaluaciones

Cuadro I

Promedios de las calificaciones otorgadas por los alumnos a cada uno de los profesores (PI más PTC)

Unidad didáctica	Profesor	Calificaciones ($X \pm DS$)				Confiabilidad	
		Curso I	Curso II	Curso III	Curso IV	En los cursos	En un curso
1	A	87.88 \pm 3.41	95.06 \pm 1.63	96.29 \pm 2.00	-	0.71	0.45
	J	96.79 \pm 1.59	95.42 \pm 1.74	98.06 \pm 0.81	96.83 \pm 1.73	0.86	0.60
2	B	89.77 \pm 2.82	91.21 \pm 1.23	94.33 \pm 1.69	-	0.03	0.01
	K	89.68 \pm 2.84	-	93.89 \pm 1.93	94.75 \pm 2.36	0.28	0.12
3	C	81.90 \pm 3.16	85.83 \pm 2.91	90.65 \pm 2.01	-	0.46	0.22
	K	90.18 \pm 0.81	91.05 \pm 1.43	-	95.56 \pm 1.91	0.58	0.32
4	D	93.98 \pm 1.27	96.19 \pm 1.29	97.74 \pm 0.39	98.10 \pm 0.57	0.46	0.17
	J	99.01 \pm 0.78	96.75 \pm 0.61	98.54 \pm 0.72	-	0.05	0.02
5	E	90.59 \pm 1.62	92.63 \pm 1.18	96.67 \pm 1.05	-	0.62	0.35
	K	-	94.41 \pm 0.87	92.28 \pm 0.60	98.28 \pm 0.60	0.93	0.81
6	F	-	88.74 \pm 1.77	95.73 \pm 1.35	96.54 \pm 1.41	0.75	0.50
	J	99.63 \pm 0.37	98.26 \pm 0.60	-	99.66 \pm 0.35	0.26	0.10
7	G	88.96 \pm 1.71	91.82 \pm 2.43	94.98 \pm 1.30	-	0.32	0.21
	K	97.15 \pm 1.77	98.10 \pm 0.57	98.68 \pm 0.73	-	0.09	0.03
8	H	94.39 \pm 1.66	95.63 \pm 1.31	96.52 \pm 0.53	-	0.64	0.38
9	I	-	95.00 \pm 0.87	97.01 \pm 1.52	98.01 \pm 0.54	0.32	0.14
	J	-	98.77 \pm 0.39	99.23 \pm 0.33	99.95 \pm 0.16	0.59	0.32
$x \pm DS$	Todos	92.30 \pm 4.88	94.05 \pm 3.49	96.44 \pm 2.20	97.74 \pm 1.60	-	-
Confiabilidad		-	-	-	-	0.47	0.28

PI: Profesores invitados.

PTC: Profesores de tiempo completo.

Cuadro II

**Promedios de las calificaciones otorgadas por los alumnos
a los profesores participantes en los CFPM**

Profesores	Calificaciones ($X \pm DS$)				Confiabilidad	
	Curso I	Curso II	Curso III	Curso IV	En los cursos	En un curso
Todos los PI (en las 9 unidades didácticas)	89.64 \pm 3.89	92.46 \pm 3.28	95.55 \pm 1.99	97.84 \pm 0.97	0.48	0.27
PTC J (En las unidades 1, 4, 6 y 9)	98.48 \pm 1.22	97.30 \pm 1.32	98.61 \pm 0.48	98.81 \pm 1.41	0.44	0.26
PTC K (En las unidades 2, 3, 5 y 7)	92.34 \pm 3.41	94.51 \pm 2.88	96.95 \pm 2.17	96.56 \pm 1.48	0.47	0.32

Cuadro III

**Valores de t para las diferencias entre los promedios de las calificaciones
obtenidas por el profesor J en las diferentes unidades didácticas de cada CFPM**

Unidades didácticas comparadas	Valores de t para la diferencia de medias			
	Curso I	Curso II	Curso III	Curso IV
1 y 4	4.2423 (e)	2.4987 (c)	1.5343 (a)	-
1 y 6	6.0477 (e)	5.3452 (e)	-	5.5542 (e)
1 y 9	-	6.5079 (e)	4.6339 (e)	6.2208 (e)
4 y 6	2.5279 (c)	6.1134 (e)	-	-
4 y 9	-	9.6648 (e)	3.0179 (d)	2.6104 (c)
6 y 9	-	2.4688 (b)	-	-

(a) $p > 0.10$ (b) $p < 0.025$ (c) $p < 0.02$ (d) $p < 0.01$ (e) $p < 0.0005$

Cuadro IV

**Valores de t para las diferencias entre los promedios de las calificaciones
obtenidas por el profesor K en las diferentes unidades didácticas de cada CFPM**

Unidades didácticas comparadas	Valores de t para la diferencia de medias			
	Curso I	Curso II	Curso III	Curso IV
2 y 3	0.5865 (a)	-	-	2.0652 (c)
2 y 5	-	-	7.4437 (e)	5.1640 (e)
2 y 7	7.7327 (e)	-	8.0414 (e)	-
3 y 5	-	6.9536 (e)	-	3.1491 (d)
3 y 7	12.4040 (e)	15.8644 (e)	-	-
5 y 7	-	12.2898 (e)	1.3984 (b)	-

(a) $p > 0.50$ (b) $p > 0.10$ (c) $p < 0.05$ (d) $p < 0.005$ (e) $p < 0.0005$

de conocimientos, trabajos prácticos y actitudes realizadas en todas las unidades didácticas) que lograron los alumnos. (Cuadro VI).

Discusión

Todos los profesores observados en este estudio mejoraron progresivamente su actuación en las unidades didácticas en que participaron, como lo muestran las diferencias significativas que se encontraron entre las calificaciones que les otorgaron los alumnos de los diferentes cursos. Esto es relevante porque en cada curso los alumnos fueron diferentes y no supieron de las calificaciones anteriores obtenidas por el profesor; también lo es porque los estudiantes pro-

porcionaron las calificaciones en forma anónima, hecho que les permitió libertad de acción y de pensamiento.¹¹ Es pertinente señalar que los alumnos calificaron a los profesores antes de que éstos los evaluaran, lo cual evitó cierta presión sobre ellos y dificultó que los resultados de la evaluación que les hicieron los maestros influenciaron su opinión acerca de la actuación de los profesores, aunque hay que tomar en consideración que otros autores han demostrado que tal efecto no es importante.¹⁵

La mejora que los profesores pueden lograr tiene un límite que depende de la disposición y capacidad de cada profesor. Tal límite fue alcanzado por los PI A, D, F y H, ya que no mejoraron más después de dos o tres cursos, lo que se evidenció cuando la diferencia entre

Cuadro V

Resultado de los exámenes integrales de conocimientos practicados a los alumnos de los CFPM

Concepto		Curso I	Curso II	Curso III	Curso IV
Alumnos (n)		17	19	19	17
Evaluación diagnóstica	(X ± DS)	29.59 ± 6.13	29.50 ± 8.33	28.89 ± 6.75	29.53 ± 6.97
Evaluación final	X ± DS	76.76 ± 6.46	75.58 ± 9.93	79.26 ± 9.15	79.53 ± 4.68
	Confiabilidad	0.70	0.89	0.84	0.62
	Error estándar	3.24	3.29	3.70	2.88
t		22.0851 (*)	15.1490(*)	18.7946 (*)	23.8225 (*)
r		0.20 (**)	0.37 (**)	0.23 (**)	0.14 (**)

(*) p < 0.0005 (**) p > 0.05

Cuadro VI

Correlación entre los promedios de las calificaciones obtenidas por los alumnos en las diferentes evaluaciones practicadas en los CFPM y los promedios de las calificaciones que los estudiantes otorgaron a los profesores (PI más PTC)

Evaluación practicada	Calificaciones (X ± DS)				r
	Curso I	Curso II	Curso III	Curso IV	
A los alumnos (final integral de conocimientos)	76.76 ± 6.46	75.78 ± 9.93	79.26 ± 9.15	79.53 ± 4.68	
A los profesores (por los estudiantes)	92.30 ± 4.88	94.05 ± 3.49	96.44 ± 2.20	97.74 ± 1.60	0.86
A los alumnos (global) (*)	84.60 ± 3.65	84.38 ± 3.23	85.00 ± 3.73	85.67 ± 3.40	0.86

(*) Incluye conocimientos, trabajos prácticos y actitudes en todas las unidades didácticas del curso.

los promedios de calificaciones dejó de ser significativa. Sin embargo, los otros profesores todavía tuvieron diferencias significativas entre las calificaciones recibidas después de participar en varios cursos, hallazgo que como era de esperarse muestra que el desempeño docente está sujeto a las diferencias individuales existentes entre los distintos profesores.

El grupo de PI mejoró significativamente de curso a curso, lo que contrasta con el caso de los dos PTC que no mostraron diferencias significativas en los promedios obtenidos en los diferentes cursos. Probablemente ello sea debido a que los PTC ya habían alcanzado su nivel estándar de desempeño y a que la retroinformación sólo les permitió mantenerlo, en tanto que como los PI tienen menos experiencia, todavía están por lograr su estándar personal. Tal pudo ser la situación de los PI B, C, E, G e I que progresaron mucho, pero que parecen tener posibilidades de hacerlo más. En el caso de los otros PI el tiempo destinado a este estudio no permitió observar si el estándar de desempeño alcanzado se mantuvo posteriormente.

El seguimiento longitudinal de los PTC en las diferentes unidades de un mismo curso mostró que sí mejoraron significativamente a lo largo de dicho curso y que sólo ocasionalmente llegaron al límite de su desempeño, lo cual sugiere que, aunque es difícil, el profesor puede mejorar algo cada vez que trabaja con un grupo de estudiantes, sobre todo si se toma en cuenta que con el instrumento de evaluación se valoraron trece criterios generales que se refieren a una serie de conductas complejas, mismas que sólo excepcionalmente se pueden hacer con perfección y, sobre todo, en forma permanente. Al observar esta parte de los resultados pudiera pensarse que pudo haber estado presente un efecto halo semejante al conocido como doctor Fox,¹³ ya que como los PTC participaron en varias unidades didácticas, su personalidad pudo haber seducido a los estudiantes. Sin embargo, parece que no fue así, ya que de haberse presentado tal efecto, se hubiera logrado pronto el estándar de desempeño y los promedios hubieran sido de diez por ser esta la calificación máxima posible. De cualquier forma, estos resultados concuerdan con los de Feinstein y Levine, quienes observaron que cuando se hace la retroinformación estudiantil inmediata sobre la calidad de la enseñanza, las opiniones de un mismo grupo de alumnos tienden a mejorar ante nuevas actuaciones del profesor.⁸

Las calificaciones que los alumnos otorgaron a los profesores tuvieron una confiabilidad media. En particular, la consistencia interna de las calificaciones estudiantiles fue más clara entre los diferentes cursos. La confiabilidad observada significa que los alumnos interpretaron el instrumento de evaluación y el desempeño de los profesores en forma consistente, es decir, que los estudiantes fueron capaces de diferenciar con regularidad el trabajo de los distintos profesores así como las diversas conductas que les evaluaron, aun

cuando varios profesores participaron a un mismo tiempo.²⁰

Por sí mismos, los logros de los alumnos sirven para evaluar la efectividad docente, aunque para la determinación de tales logros el profesor no sea el único factor que participa, ya que también intervienen otros, entre los que se encuentran las características del plan de estudios y de la institución educativa así como la propia capacidad y dedicación de los alumnos, por lo que no sería correcto atribuirle sólo al profesor toda la responsabilidad del éxito o fracaso del alumno.² Sin embargo, los logros de los alumnos sí son de gran utilidad para validar las calificaciones que ellos proporcionan a los profesores.^{7,17} Es así que la correlación positiva fuerte que se encontró tanto con la evaluación integral como con la evaluación global que se practicaron a los alumnos permite inferir que el instrumento utilizado para evaluar a los profesores es razonablemente válido.

El seguimiento del desempeño docente que se hizo en este estudio ha permitido demostrar que cuando el profesor recibe retroinformación inmediata sobre su actuación, cuenta con algunos elementos que, aunque tienen un fuerte contenido de subjetividad, le permiten reflexionar y planear mejor sus actividades futuras en busca de una actuación tal que le permita contribuir más eficazmente a que los alumnos logren resultados cada vez mejores en los cursos que imparte. Por ello parece ser claro que si el profesor es inteligente, muestra interés por perfeccionarse y aprecia el valor del señalamiento de sus errores y aciertos, mejorará su desempeño después de conocer los resultados de la apreciación estudiantil.²⁰

La evaluación del profesor es uno de los problemas más importantes e interesantes que enfrenta la evaluación educativa, pero a pesar de ello no se practica con la frecuencia que se debiera, probablemente debido a que ha llegado a considerarse como un tabú o a que pudiera creerse que deteriora la imagen del profesor.² Esto podría ser particularmente cierto cuando quien evalúa es el propio alumno, ya que se pudiera pensar que no está capacitado para evaluar en forma eficiente el desempeño docente. Sin embargo, los estudiantes son las personas más calificadas para evaluar la efectividad docente,³⁰ ya que su opinión proporciona una muestra del efecto que el trabajo del profesor ha tenido en los propios alumnos, así sea nada más el grado de satisfacción que han experimentado a propósito del curso en que han participado.¹² Por otra parte, los educandos son las personas que se afectan en forma inmediata por la actuación del profesor, son quienes tienen la mejor oportunidad de presenciar más directamente dicha actuación y, aunque los profesores no lo quisieran, de cualquier manera los estudiantes los evalúan, por lo que es mejor organizar esa evaluación en forma sistemática y tomar en cuenta tanto sus resultados como los de la retroinformación que completa el proceso de comunicación que el propio profesor ha iniciado con su labor docente.³¹

Parece ser que la clave para lograr resultados positi-

vos con el uso de la evaluación y retroinformación estudiantil está en la formulación de preguntas adecuadas,¹⁸ dirigidas a las personas indicadas,¹⁶ en la manera apropiada¹¹ y en el momento pertinente,¹⁵ lo cual pudiera parecer obvio, pero hay que reconocer que no siempre se cumplen estos preceptos.³² Debido a que todos estos aspectos fueron atendidos con cuidado en este estudio y a juzgar por los resultados observados, se puede concluir que la evaluación estudiantil acerca del desempeño docente así como la retroinformación inmediata que el profesor recibe como consecuencia, son de gran utilidad para mejorar la efectividad docente, y que cualquier utilización que se hiciera de la retroinformación estudiantil debiera iniciarse con la consideración de su relación con el proceso enseñanza-aprendizaje como un todo.³²

Referencias

1. DE ALBA, A.; DIAZ, B.A. y VIESCA, A.M.: *Evaluación: análisis de una noción*. Rev. Mex. Sociol. 1984; 46: 175.
2. VILLARROEL, C.: *Evaluación de los aprendizajes en la enseñanza superior*. Caracas. Ediciones Paulinas, Universidad Central de Venezuela, 1974, pág. 117.
3. REMMERS, I.H.: *The relationship between student marks and attitude toward instructors*. School. Soc. 1928; 28: 759.
4. RODIN, M.; y RODIN, B.: *Student evaluations of teachers*. Science. 1972; 177: 1164.
5. ZELBY, L.W.: *Student-faculty evaluation*. Science. 1974; 183: 1267.
6. ARIAS, G.F.: *El inventario de comportamientos docentes (ICD): un instrumento para evaluar la calidad de la enseñanza*. Perfiles Educativos. 1984; No. 23 (Mar.-Abr.): 14.
7. MENDELSON, M.A.; CANADAY, S.D. y HARDIN, J.H.: *The relationship between student ratings of course effectiveness and student achievement*. Med. Educ. 1978; 12: 199.
8. FEINSTEIN, E. y LEVINE H.G.: *Impact of student ratings on basic science portion of the medical school curriculum*. J. Med. Educ. 1980; 55: 502.
9. THOMPSON, W.G.: *Students' ratings of instructors in a multidisciplinary course*. J. Med. Educ. 1985; 60: 120.
10. LANCASTER, C.J.; MENDELSON, M.A. y ROOS, G.R.: *The utilization of student instructional ratings in medical colleges*. J. Med. Educ. 1979; 54: 657.
11. SMITH, S.R. y PAULEN, L.J.: *Use of anonymous student evaluations of faculty members in U.S. medical schools*. J. Med. Educ. 1984; 59: 196.
12. GROVER, P.L.: *Evaluation of instructional skills of medical teachers: the participant observer in the medical school*. Med. Educ. 1980; 14: 12.
13. NAFTULIN, D.H.; WARE, J.E. y DONNELLY, F.A.: *The doctor Fox lecture: a paradigm of educational seduction*. J. Med. Educ. 1973; 48: 630.
14. WARE, J.E. y WILLIAMS, R.G.: *The Dr. Fox effect: a study of lecturer effectiveness and ratings of instructors*. J. Med. Educ. 1975; 50: 149.
15. CANADAY, S.D.; MENDELSON, M.A. y HARDIN, J.H.: *The effect of timing on the validity of student ratings*. J. Med. Educ. 1978; 53: 958.
16. SLOTNICK, H.B. y DURKOVIC, R.G.: *Dimensions of medical students' perceptions of instruction*. J. Med. Educ. 1975; 50: 662.
17. FREY, P.W.: *Student ratings of teaching: validity of several factors*. Science. 1973; 182: 83.
18. RIPPBY, R.M.: *Student evaluations of professors: are they of value?* J. Med. Educ. 1975; 50: 951.
19. BANDARANAYAKE, R.: *Utilization of feed-back from student evaluation of teachers*. Med. Educ. 1978; 12: 314.
20. IRBY, D.M.; SHANNON, N.F.; SCHER, M.; PECKHAM, P.; KO, G. y DAVIS, E.: *The use of students ratings in multi-instructor courses*. J. Med. Educ. 1977; 52: 668.
21. PANZSA, M.: *Enseñanza modular*. Perfiles Educativos. 1981; No. 11 (Ene.-Mar.): 30.
22. MORAN, O.P.: *La evaluación de los aprendizajes y sus implicaciones educativas y sociales*. Perfiles Educativos. 1981; No. 13 (Jul.-Sep.): 21.
23. AITKEN, A.C.: *Statistical mathematics*. 8a. ed. Edimburgo: Oliver and Boyd, 1957, pág. 86.
24. HARDYCK, C.S. y PETRINOVICH, L.F.: *Introduction to statistics for the behavioral sciences*. Filadelfia, WB. Saunders, 1969, pág. 135.
25. LEVIN, J.: *Fundamentos de estadística en la investigación social*. 2a. edición. México. HARLA, 1979, pág. 164.
26. DOWNIE, N.M. y HEATH, R.W.: *Métodos estadísticos aplicados*. 3a. edición. México. HARLA, 1973, pág. 238.
27. SPIEGEL, M.R.: *Teoría y problemas de probabilidad y estadística*. México. McGraw-Hill, 1976, pág. 215.
28. EBEL, R.L.: *Estimation of the reliability of ratings*. Psychometrika. 1951; 16: 407.
29. GRONLUND, N.E.: *Medición y evaluación de la enseñanza*. México. Editorial Pax, 1973, pág. 121.
30. ROUS, S.N.; BAMFORD, J.C.; GROMISCH, D.; RUBIN, S. y SALL, S.: *The improvement of faculty teaching through evaluations: a follow-up study*. J. Surg. Res. 1972; 13: 262.
31. GUTIERREZ, S.R.: *Introducción a la didáctica*. 2a. edición. México. Editorial Esfinge, 1980, pág. 222.
32. MORRIS, V.D.: *A positive approach to the utilization of student feedback in medical education*. J. Med. Educ. 1976; 51: 541.

COMENTARIO OFICIAL

JOSE DE JESUS VILLALPANDO-CASAS*

El trabajo que tengo ocasión de comentar invita a reflexionar sobre uno de los aspectos fundamentales de la educación médica: la efectividad docente. El tema enfocado por el académico Varela Rueda permite conocer más acerca del desempeño de los docentes y la manera de mejorarlo, según se desprende de los resultados de la investigación realizada con sus propios alumnos, a quienes ha formado concienzudamente como profesores, observando su cometido bajo una acertada metodología.

La efectividad docente de los individuos sujetos al estudio se infiere del resultado de los cursos impartidos por ellos, valorado a través de las calificaciones obtenidas por los alumnos, las cuales se correlacionan con la conclusión de la apreciación sistematizada que éstos hacen sobre la actuación de los profesores y que les comunican a manera de retroinformación cuando terminan su participación, dando lugar a un proceso cuyo desenlace es la mejoría significativa de la efectividad docente en el siguiente curso, hasta el límite de la capacidad y disposición individual.

La investigación ubica al profesor como uno de los elementos más importantes del proceso enseñanza-aprendizaje, tan controvertido por algunas corrientes de educación superior, y confiere a los alumnos un papel trascendente para el desempeño docente, como fenómeno inherente del proceso de comunicación que tiene lugar en la interacción profesor-alumno.

Asimismo, incursiona en un campo que tradicionalmente se considera vedado, como es la evaluación estudiantil de la actuación del profesor, por no concederse a los alumnos la capacidad o madurez necesaria para ello, siendo como son los directamente

afectados. Por cierto, los alumnos siempre externalizan juicios sobre sus profesores, aunque de manera informal y sin su conocimiento o aceptación; así, éstos se ven privados de opiniones que pudieran ser benéficas para todos: profesores y alumnos.

El trabajo que se presenta ofrece una alternativa viable para incrementar la eficiencia de la educación y coadyuvar a resolver el problema de la defectuosa actuación de ciertos profesores de medicina y su repercusión en la preparación de los estudiantes. Para ello es necesario contar con la disposición de ambas partes de intervenir para mejorar y que se acepte la retroinformación. Además, como bien lo señala el autor, también se requiere hacer coincidir las preguntas adecuadas, sobre las personas indicadas, formuladas de la manera apropiada y aplicadas en el momento pertinente, para que las respuestas puedan ser aprovechadas.

Una de las tareas más enaltecedoras de los seres humanos es la de contribuir al desarrollo de los demás individuos. Esta ha sido la principal de las actividades profesionales de Carlos Ernesto Varela Rueda; baste mencionar entre sus importantes logros en el campo de la educación médica la sustancial participación en la creación de la especialidad de medicina familiar en nuestro país y en la preparación de numerosos profesores de medicina en la institución donde presta sus servicios. Ahora, la Academia Nacional de Medicina abre su recinto y lo acoge como uno de sus miembros más distinguidos en reconocimiento a su trascendente labor formadora de recursos humanos y méritos personales. Es por ello que tengo la satisfacción de ofrecerle la bienvenida a nuestra corporación.

* Académico numerario. Jefatura de Enseñanza e Investigación. Instituto Mexicano del Seguro Social.