

Estudio de concordancia clínica en educandos de pre y postítulo en pediatría

Gastón Duffau T.¹, Luisana Navarrete R.¹, Carlos Fernández C.¹

Resumen

La evaluación de destrezas clínicas en educandos debiera medir el nivel de concordancia del grupo, según los objetivos propuestos. Esta aumentaría al progresar el programa docente. Hipótesis: Las destrezas clínicas en educandos de pretítulo muestran baja concordancia entre examinadores ($\kappa < 0,41$) que mejora en el postítulo. Material y método. Con un estudio piloto en 50 casos se estableció un tamaño muestral de 104 a 156 casos para detectar, con error estándar de κ de 0,08, diferencias intra e intergrupo de 0,20, con error alfa de 0,05. Los examinadores fueron parejas posibles de internos ($n = 11$) y médicos alumnos del primer año del Programa de Especialización en Pediatría (APE) ($n = 10$). Se evaluó: gravedad, tiraje, crepitaciones, dificultad respiratoria y compromiso de conciencia. Adicionalmente se obtuvo: edad del paciente, género, estado nutricional, diagnósticos de ingreso, pareja participante y fecha. Estudio estadístico: comparación de proporciones por Ji cuadrado y de índices kappa. Resultados: Ingresaron 358 niños, en dos grupos: 206 casos valorados por becados y 152 por internos, comparables en edad, género y estado nutricional. La evaluación por internos mostró kappa entre $-0,03$ y $0,17$ y por APE entre $0,13$ y $0,68$ con valor $< 0,20$ en 3 de los 5 elementos clínicos estudiados. Al comparar ambos grupos solo se encontró mayor concordancia en médicos en compromiso de conciencia y gravedad. Con solo dos categorías de evaluación por elemento clínico en internos no hubo cambios importantes, pero en APE hubo ascenso de los índices, excepto en compromiso de conciencia. Al comparar los grupos en estas condiciones, se encontró que había diferencia a favor de médicos para todos los elementos clínicos, excepto dificultad respiratoria. Conclusión: del estudio surge la necesidad de revisión de los objetivos y métodos por los cuales se imparte la docencia conducente a la adquisición de destrezas clínicas, puesto que parece poco aceptable el escaso nivel de concordancia alcanzado entre internos y entre médicos APE. Los primeros porque vienen de un curso con énfasis en la práctica semiológica y los segundos porque adicionalmente han contado con tiempo de actividad supervisada en sala.

(Palabras clave: educación médica, concordancia, evaluación.)

Evaluation of clinical skills in pre and postgraduate medical students

Evaluation of clinical skills in pre and postgraduate medical students should establish the level of concordance in each group according to defined objectives. Hypothesis: clinical skills in pregraduate students shows poor interobserver concordance levels (κ index < 0.41) and improves in postgraduate students. Patients and methods: a pilot study of 50 cases permitted sample size estimates of 104 to 156 patients given a standard error of the κ index of 0.08, and alfa error of 0.05 and the power to detect a difference of 0.20. Observers were random couples of interns ($n = 11$) and first year residents ($n = 10$). The clinical findings evaluated were severity of illness, respiratory retractions, rales, respiratory distress and impairment of consciousness. Each record also included age, gender, nutritional status, diagnosis on admission, evaluation couple and date. Statistical analysis of the data included comparisons of proportions (Ji square) and of kappa indexes.

1. Departamento de Pediatría, Cirugía Infantil y Ortopedia, Campus Norte, Facultad de Medicina, Universidad de Chile.

Trabajo recibido el 25 de mayo de 2000, devuelto para corregir el 24 de julio de 2000, segunda versión el 2 de agosto de 2000, aceptado para publicación el 8 de agosto de 2000.

Results: 358 children were admitted to the study, 206 were evaluated by residents and 152 by interns, both groups being of similar age, gender and nutritional status. Kappa indexes generated by intern couples were from -0.03 up to 0.17 and by residents 0.13 to 0.68. These figures were significantly better for residents only when evaluating impairment of consciousness and severity of illness. Reduction of evaluation categories to 2 levels showed significant improvement of kappa indexes only in the group of residents. The latter had better results than interns for all indexes except for respiratory distress. Conclusions: It is imperative to make a critical review of educational objectives and methods directed to the acquisition of clinical skills in pre and postgraduate medical students. The findings in this study show poor level of concordance in both groups and are hardly acceptable because interns have just had semiological training and residents in addition controlled practice.

(Key words: medical education, evaluation, concordance.)

La evaluación del nivel de destreza clínica alcanzado por los educandos se puede realizar por métodos apropiados de observación y verificación de tareas^{1, 2}. De esta manera se obtiene información individual aunque no de un aspecto que pareciera indispensable, que se refiere a determinar hasta qué punto el programa docente ha logrado, en el grupo, un determinado nivel de concordancia en el logro de los objetivos propuestos.

En el desarrollo de un programa docente clínico lo esperable es que el conjunto de educandos adquiera las destrezas que son objetivos, de acuerdo con un plan que conduzca a alcanzar un criterio uniforme en el grupo. Esta uniformidad de criterio clínico debería ir en aumento al progresar el programa docente.

El disponer de un instrumento que permita cuantificar el progreso en destrezas clínicas de un grupo, será muy útil en la medida que incluya características como las siguientes:

a. Que no sea susceptible de preparación por el educando, en el último minuto, como ocurre frecuentemente con las pruebas escritas y exámenes orales, donde se evalúan aspectos cognitivos comúnmente reproductivos más que transferenciales y aspectos psicomotores, b. Que no sea transmisible como una prueba de conocimientos, c. Que utilice los mismos elementos de juicio al inicio y al final y sin embargo que tal conocimiento no afecte el desempeño del educando, como sería repetir el mismo conjunto de preguntas.

El estudio de concordancia es una herramienta bioestadística que permite comparar dos observadores o el desempeño de dos

exámenes en la capacidad que tendría uno de subrogar o reemplazar al otro³. Para el caso que nos interesa, las parejas de observadores de un mayor nivel del programa docente mostrarían un mejor grado de acuerdo en la valoración clínica de los enfermos y ello se reflejaría en el estadístico kappa cuyo valor sería mayor.

Hipótesis: La evaluación de destrezas clínicas en educandos de pretítulo en Pediatría muestra un nivel de concordancia entre observadores regular o inferior (índice kappa < 0,41) y mejora en el postítulo de la misma especialidad.

MATERIAL Y MÉTODO

Lugar: El estudio se desarrolló en el Departamento de Pediatría, Campus Norte con internos de la carrera de Medicina y médicos APE, cuya duración es de tres años.

Campo clínico: Hospital Clínico de Niños "Roberto del Río".

Tamaño de la muestra: Por no disponer de la información pertinente para la estimación del tamaño muestral, se efectuó un estudio piloto con los primeros 50 casos analizados por médicos APE de primer año, que ingresaron en la investigación. Con las combinaciones de prevalencia del hallazgo clínico en análisis (entre los examinadores) y los índices kappa detectados, se determinó que una diferencia razonable de encontrar del estadístico kappa intragrupo e intergrupo, sería de aproximadamente un grado, es decir, 0,20 y al mismo tiempo un valor digno de ser identificado como diferente de kappa = 0. El error estándar del kappa encontrado varió alrededor de 0,08, según elemento clínico revisado, cifra que se estimó pudiera man-

tenerse en las comparaciones intergrupo. Con un error alfa de 5% unilateral, el número de casos necesario, si se mantienen los supuestos, fluctuaría, según manifestación clínica, entre 104 y 156 casos. En este cálculo se excluyó el compromiso de conciencia, que presentó una situación muy diferente de los demás elementos clínicos, con una muy reducida prevalencia. El estadístico kappa se ha revisado en detalle en Rev Chil Pediatr 1999; 70: 314-24.

Los grupos a estudiar se integraron en las combinaciones posibles para conformar parejas de observadores en el nivel de internos de Pediatría ($n = 11$) y médicos APE en su primer año ($n = 10$). Cada pareja estuvo representada en la muestra por un grupo similar de observaciones.

Diariamente uno de los investigadores propuso a una o mas parejas de examinadores el evaluar pacientes que constituyeran ingresos y no sus enfermos habituales. Los elementos clínicos a evaluar fueron del examen físico y correspondieron a presencia e intensidad de:

1. Gravedad inmediata
2. Tiraje
3. Crepitaciones
4. Dificultad respiratoria
5. Compromiso de conciencia

Los 4 primeros fueron valorados en una de las siguientes opciones: (-) (+) (++) y (+++). El último, como (-), obnubilación, somnolencia, sopor y coma.

La información se vació a un formulario *ad hoc* en forma independiente para cada examinador de modo que cada uno ignorara la evaluación del otro.

El formulario mencionado contenía, además, la edad del paciente, su género y estado nutricional por criterio peso/talla según curvas del NCHS y procedimiento sugerido por MINSAL⁴. Asimismo, se consignaron los diagnósticos de ingreso, la pareja participante y la fecha.

El estudio estuvo centrado en manifestaciones relacionadas con patología respiratoria, considerando su elevada frecuencia en el niño hospitalizado.

La hipótesis en verificación no fue conocida por los examinadores al efectuar el estudio de los casos.

El lapso entre el examen efectuado por un educando y el siguiente fue el más breve

posible y en todo caso inferior a 15 minutos. Completado el procedimiento, se indagó con ellos sobre definición de las manifestaciones clínicas estudiadas.

Los examinadores fueron:

- Internos de la Carrera de Medicina (6° año), durante su pasada por la Unidad de Lactantes.
- Médicos en su primer año del Programa de Especialización en Pediatría, después de al menos 3 meses de ingresados al Programa.

Con el fin de disponer de una aproximación al grado de concordancia que pudieran mostrar evaluadores experimentados entre sí, se propuso el estudio de una submuestra de niños a un grupo de 4 de estos.

Para el estudio estadístico de los resultados se recurrió al programa Win Episcope 1.0⁵ que ofrece la ventaja sobre otros de uso común (EpiInfo 6.04 y anteriores o MedCalc), en el sentido de entregar no solo el error estándar de kappa = 0 sino también el del índice encontrado. La comparación de proporciones se realizó por Ji cuadrado y de los índices kappa por el procedimiento sugerido por Cohen⁶. El cálculo de kappa entrega el valor de este corregido por azar aunque no ponderado. En este estudio se aceptó un error alfa de 5%.

RESULTADOS

Ingresaron al estudio 358 niños durante 1997, que constituyeron dos grupos: uno de 206 casos, que fueron evaluados por médicos APE en su primer año del Programa de Especialización en Pediatría y otro de 152 niños que fueron examinados por internos de la Carrera de Medicina. Cuatro evaluadores experimentados calificaron la situación de 15 niños en la misma forma que los educandos, obteniendo índices kappa de concordancia para presencia o ausencia de la manifestación clínica entre 0,60 y 0,75 con un valor del conjunto, para todos los elementos clínicos, de 0,65 e IC 95% de 0,41 a 0,88.

La distribución por edad género y estado nutricional de los enfermos en cada grupo se muestra en la tabla 1. Es posible observar que los dos grupos estudiados resultaron comparables en estas variables.

Tabla 1

Distribución por edad, género y estado nutricional de 358 niños examinados por internos de medicina o médicos alumnos de primer año del Programa de Especialización en Pediatría (APE)

Edad (m)	Grupo examinado por					
	Internos		Médicos APE		Total	
	n	%	n	%	n	%
< 6	45	29,6	80	38,8	125	34,9
6-11	51	33,6	60	29,1	111	31,0
12 y >	56	36,8	66	32,1	122	34,1
Totales	152	100,0	206	100,0	358	100,0

Ji² = 3,28 , 2 g.l. p = 0,19

Género

Masculino	86	63,2	103	50,0	199	55,6
Femenino	66	36,8	103	50,0	159	44,4
Totales	152	100,0	206	100,0	358	100,0

Ji² Yates = 1,27 , 1 g.l. p = 0,26

*Estado nutricional**

O	6	3,9	8	3,9	14	3,9
SP	24	15,8	45	21,9	69	19,3
N	98	64,5	109	52,9	207	57,8
DI	15	9,9	32	15,5	47	13,1
DII-DIII	9	5,9	12	5,8	21	5,9
Totales	152	100,0	206	100,0	358	100,0

Ji² = 5,83 , 4 g.l. p = 0,21

* O,SP, N, DI, DII DIII = Respectivamente: obesidad, sobrepeso, normal, desnutrición grado I-II-III.

En la tabla 2 destaca el hecho que para las 5 manifestaciones clínicas investigadas la evaluación por internos mostró niveles de concordancia muy bajos, fluctuando entre -0,037 y 0,171. Solo este último índice kappa resultó diferente de cero y correspondió a la valoración de dificultad respiratoria.

En el conjunto de médicos APE se encontraron índices kappa entre 0,133 y 0,680, correspondiendo este último a la calificación de compromiso de conciencia. En este grupo de médicos, la totalidad de los índices kappa resultó significativamente diferente de cero. Su nivel, sin embargo, estuvo bajo

0,20 para 3 de los 5 elementos clínicos estudiados y en los restantes, dados por calificación de gravedad y compromiso de conciencia alcanzó, respectivamente, 0,396 y 0,680.

Al comparar ambos grupos solo se encontró un significativo mayor grado de concordancia en médicos respecto a internos en el estudio de compromiso de conciencia y calificación de gravedad inmediata.

La situación detectada, de baja concordancia general entre examinadores, en la calificación de las manifestaciones clínicas, podría atribuirse en parte al número de cate-

Tabla 2

Concordancia clínica en internos de Medicina y médicos alumnos de primer año del Programa de Especialización en Pediatría (APE)

Manifestación clínica (4 o 5 categorías)*	Grupo	Índice Kappa	IC 95%	I vs. M
Tiraje	I	0,069	-0,067 a 0,206	ns
	M	0,133	0,038 a 0,228	
Crepitaciones	I	0,018	-0,098 a 0,135	ns
	M	0,148	0,056 a 0,239	
Dificultad respiratoria	I	0,171	0,044 a 0,298	ns
	M	0,160	0,051 a 0,269	
Compromiso de conciencia	I	0,154	-0,030 a 0,338	p < 0,05
	M	0,680	0,457 a 0,903	
Gravedad inmediata	I	-0,037	-0,202 a 0,128	p < 0,05
	M	0,396	0,289 a 0,503	

* Categorías: (-), (+), (++) , (+++) para las cuatro primeras manifestaciones clínicas
En la última: (-), obnubilación, somnolencia, sopor, coma

I: Internos. Examinaron 152 niños

M: Médicos APE. Examinaron 206 niños

ns: no significativo

gorías. Por ello, se muestra en la tabla 3 el resultado de establecer solo dos categorías en cada elemento clínico: presente o ausente. Para el grupo de internos esencialmente no hay cambio en el grado de concordancia excepto un ascenso discreto en la valoración de dificultad respiratoria (kappa de 0,171 a 0,231) y de compromiso de conciencia (kappa de 0,154 a 0,280). Los índices que resultaron significativamente diferentes de cero fueron los de crepitaciones, dificultad respiratoria y compromiso de conciencia. El de crepitaciones indica una concordancia negativa, inferior a la que podría obtenerse por azar. Ello indicaría franco desacuerdo entre examinadores. La situación con los médicos APE sería mejor si consideramos el ascenso general de los índices, excepto en compromiso de conciencia donde mantiene el valor de 0,680. Al comparar los dos grupos en estas nuevas condiciones, se encuentra que hay diferencia significativa en la valoración de todos los elementos clínicos, excepto en dificultad respiratoria.

COMENTARIO

El procedimiento de evaluación propuesto sería un complemento interesante de los

más tradicionales y permitiría monitorear la marcha y progreso de un mismo conjunto en los aspectos de destrezas clínicas. Mayor información se lograría al disponer del estándar o patrón de comparación para establecer el nivel de aproximación a este⁷, aunque no es el fin primario de este tipo de análisis.

El estudio de concordancia genera usualmente índices en el rango de 0 a 1, los que se suelen interpretar como (3): 0,00-0,20 pobre, 0,21-0,40 regular, 0,41-0,60 buena, 0,61-0,80 substancial y 0,81-1,00 casi perfecta o perfecta.

En el estudio presentado resultó sorprendente encontrar grados tan escasos de concordancia en el nivel de internos frente a manifestaciones clínicas que son tan común y repetidamente evaluadas. Se podría pensar que tuvo un efecto negativo el desconocimiento, por los examinadores, de todos los antecedentes del paciente, pero ello no debiera influir notoriamente sobre las conclusiones del examen físico. También se podría argumentar que, al no tener una responsabilidad directa sobre el paciente, el examen físico podría diferir del efectuado en otras condiciones. Sin embargo, no se comunicó el objetivo perseguido con la maniobra soli-

Tabla 3

Concordancia clínica de internos de Medicina y médicos alumnos de primer año del Programa de Especialización en Pediatría (APE)

Manifestación clínica (4 o 5 categorías)*	Grupo	Índice Kappa	IC 95%	I vs. M
Tiraje	I	-0,040	-0,199 a 0,119	p < 0,05
	M	0,311	0,181 a 0,441	
Crepitaciones	I	-0,163	-0,322 a -0,004	p < 0,05
	M	0,279	0,138 a 0,420	
Dificultad respiratoria	I	0,231	0,029 a 0,434	ns
	M	0,191	0,051 a 0,331	
Compromiso de conciencia	I	0,280	0,020 a 0,540	p < 0,05
	M	0,680	0,453 a 0,902	
Gravedad inmediata	I	0,018	-0,182 a 0,218	p < 0,05
	M	0,485	0,366 a 0,605	

* Categorías: Presente, ausente

I: Internos. Examinaron 152 niños

M: Médicos APE. Examinaron 206 niños

ns: no significativo

citada a cada participante y bien pudiera haber correspondido a una evaluación docente, lo que debería estimular a un buen desempeño. Como sea, es interesante que el tiempo que demoró la valoración por internos fue aproximadamente el doble del empleado por médicos APE. Un problema de definición debe ser contemplado, ya que se pudo constatar en varias oportunidades que los examinadores no compartían las mismas ideas sobre algunas manifestaciones como tiraje o niveles de conciencia.

Un grado bajo de concordancia no tiene entre sus opciones de interpretación una favorable ya que, independientemente del hecho de no tener establecida la respuesta ideal en cada caso (con lo que se elimina la variable "docente examinador"), los participantes no están de acuerdo frente a la clínica del paciente. Si el índice kappa señala buena concordancia en el grupo, la interpretación es favorable, en el sentido de opiniones acordes ante el problema clínico común. Si estas son acertadas, no se puede asegurar porque no se comparó con un patrón, pero hay que recordar que los docentes examinadores como estándar ideal no tienen siempre buena concordancia entre ellos. Por tal razón es que se estima que los

estudios de confiabilidad de los métodos de evaluación clínica deben contemplar este aspecto, cosa con frecuencia olvidada^{1, 2}. Sin embargo, el nivel de acuerdo global entre los evaluadores expertos que analizamos fue en el grado de "substantial" con intervalo de confianza de "buena a "casi perfecta", según la nomenclatura previamente detallada.

Con un procedimiento de evaluación individual el resultado del grupo de educandos, expresado en calificación promedio, con una determinada dispersión no proporciona una idea de la homogeneidad del conjunto. Es concebible así dos conjuntos con iguales calificaciones de resumen y con dos índices de concordancia dispares.

Los resultados en internos, que pese a mostrar muy insuficiente concordancia no parecen generar consecuencias serias aparentes, enfatizarían: a) la importancia que tiene la supervisión de actividades en el grupo para corregir lo necesario y, desde luego, el entrenamiento en destrezas clínicas y b) el hecho que el examen del paciente grave hospitalizado es generalmente repetido varias veces en el día, dando opción a un cambio en la calificación del enfermo y en las decisiones que se derivan.

Del presente estudio surge la necesidad de una revisión crítica de los objetivos y metodología utilizada en la docencia de pre y postgrado conducente a la adquisición de destrezas clínicas, puesto que es insatisfactorio el escaso nivel de concordancia alcanzado, especialmente para varias categorías de las manifestaciones clínicas investigadas, tanto entre internos como entre médicos APE. Los primeros vienen de un curso de Pediatría que contempla como actividad práctica el énfasis en lo semiológico y los segundos han contado con tiempo adicional de práctica supervisada en sala, lo que debiera asegurar la adquisición de estas y otras destrezas clínicas.

REFERENCIAS

1. Harden RM, Stevenson M, Downie WW, et al: Assessment of clinical competence using objective structured examination. *BMJ* 1975; 1: 447-52.
2. Harden RM, Gleeson FA: Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979; 13: 41-54.
3. Armitage P, Berry G: Estadística para la investigación biomédica. 3ª edición, Harcourt Brace S.A., Barcelona, España 1997: 419-24.
4. República de Chile, Ministerio de Salud, Comisión Coordinadora PNAC. Gráficas de evaluación antropométrica para niños(as) menores de 6 años. 1993.
5. Blas N, Ortega C, Frankena K, Noordhuizen J: Win Episcopo 1.0. University Wageningen. 1995.
6. Cohen J: A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37-46.
7. Selby C, Osman L, Davis M, Lee M: Setup and run an objective structured clinical exam. *BMJ* 1995; 310: 1187-90.