

Propiedades psicométricas de una encuesta de evaluación para el área docente de psiquiatría.

Psychometric properties of a teaching evaluation questionnaire for an Academic Psychiatric Area.

Pedro Saz¹; Antonio Campayo¹; Elena Lobo¹; Eduardo Aguilar²; Sonia De-La-Fuente²; Antonio Lobo¹

1. Departamento de Psiquiatría y de Medicina Preventiva. Universidad de Zaragoza, Hospital Clínico e Instituto Aragonés de Ciencias de la Salud. 2. Estudiantes de 5º Curso de Medicina.

Introducción: La aplicación de los estándares internacionales de calidad en los estudios de Medicina es necesario en el contexto del actual proceso de Convergencia al Espacio Europeo de Educación Superior (CEEES) en Ciencias de la Salud; de ello se deriva la importancia de la mejora de la calidad docente. El empleo de cuestionarios es una de las herramientas más empleadas para evaluar la calidad de la docencia. El objetivo principal de esta publicación es presentar las propiedades psicométricas de fiabilidad y validez de una encuesta desarrollada para evaluar las asignaturas del Área de Psiquiatría de la Facultad de Medicina de Zaragoza.
Material y Métodos: Durante el curso académico 2005-2006, en la asignatura "optativa" Psicósomática y Psiquiatría de Enlace Especial, se distribuyó a los alumnos asistentes a las dos últimas clases previas al examen, una encuesta confidencial diseñada por los autores. Se evaluaron las propiedades psicométricas

de fiabilidad test-retest, consistencia interna, validez de contenido y aproximación a la validez de constructo. **Resultados:** La encuesta ofreció un coeficiente de fiabilidad alfa de Cronbach de 0,865 (Meritorio-Excelente) y un índice medio de acuerdo KappaW test-retest=0,655 (I.C. al 99% 0,562-0,748). Así mismo se documenta una adecuada factibilidad e indicios de validez.

Conclusiones: Los datos obtenidos documentan la factibilidad y fiabilidad de la nueva encuesta y sugieren que su validez de constructo es adecuada. Consecuentemente, puede resultar un instrumento útil en la toma de decisiones sobre modificaciones en la metodología docente del Área de Psiquiatría; además, generalizable a otras disciplinas de las ciencias biomédicas.

Palabras clave: Educación Médica; Fiabilidad y Validez; Psicometría; Docencia.

Correspondencia:

P. Saz, A. Lobo
Hospital Clínico Universitario, Servicio de Psiquiatría, planta 3ª
Calle San Juan Bosco, 15
50009 Zaragoza
Teléfono: 976 55 97 95 Fax: 976 76 57 88
e-mail: psaz@unizar.es
alobo@unizar.es

Introduction: In the context of the process to establish the European Higher Education Area (EHEA), the implementation of international quality standards in Medical Education is essential; and the improvement of teaching effectiveness is therefore of great importance. Students' rating questionnaires are

the teaching evaluation tool in most widespread use. The main objective of this study is to present psychometric properties, such as reliability and validity, of a questionnaire developed for the assessment of subjects in the Psychiatry Department of the Faculty of Medicine in Zaragoza.

Methods: Students attending the last two Psychosomatics and Consultation-Liaison Psychiatry classes in 2005-2006, were invited to fill out a confidential questionnaire designed by the authors. Test-retest reliability, internal consistency and content validity were tested, and an approximation to construct validity was also approached.

Results: Cronbach's α coefficient was 0.865 (good to excellent according to the standard) and test-retest kappaW was 0.655 (C.I. 99% .562-.748). Feasibility was appropriate and data about correlation of measures of satisfaction and the outcome of teaching support the construct validity of the questionnaire.

Conclusions: The feasibility and reliability of the new scale for medical students evaluation is documented, and preliminary data support its construct validity. It is suggested that the new tool may be useful in deciding about teaching methods in the Area of Psychiatry. Furthermore, its usefulness might be extended to other Biomedical Areas.

Key words: Education, Medical; Reliability and Validity; Psychometrics; Teaching

INTRODUCCIÓN

La aplicación de los estándares internacionales de calidad en la formación de pre-grado en Medicina es un camino necesario en el contexto del actual proceso de Convergencia al Espacio Europeo de Educación Superior (CEEES) en Ciencias de la Salud. Este proceso es ya patente en muchas Universidades españolas, aunque en distintas fases de desarrollo¹. En el mismo sentido, desde el ámbito estatal se pretende promover la coordinación y dinamización de las políticas de gestión de la calidad en las Universidades mediante la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA). Resulta claro, en este contexto, que la mejora de la calidad docente no sólo es importante sino que, en algunos aspectos, es imperativa².

Además, en el citado proceso CEEES queda clara-

mente establecido, especialmente en las últimas dos conferencias bianuales de Ministros de Educación: Berlín 2003 y Bergen 2005, que "la calidad de la enseñanza superior es el ideal central en torno al que gira la creación del Espacio Europeo" y, entre otros fundamentos, defienden: 1) una adaptación de las enseñanzas a modelos en que el alumno sea el centro del proceso de aprendizaje³; y 2) que se evalúe dicho proceso como un sistema de "retro-alimentación"⁴.

Una de las herramientas básicas y ampliamente utilizadas para intentar esa "retroalimentación", aunque no siempre con el fin de mantener un proceso dinámico de mejora en la calidad docente⁵, es el empleo de cuestionarios completados por los estudiantes. Sin embargo, la fiabilidad y validez de este método no se ha demostrado sistemáticamente^{6,7}, a pesar de que algunos investigadores han tratado de dar luz a una adecuada elaboración y uso de dichos cuestionarios. Por ejemplo, Metcalfe y Matharu⁸ han documentado que un criterio para asegurar la fiabilidad y validez de la información obtenida con este tipo de cuestionarios es la tasa de respuesta, de manera que una tasa baja de respuesta puede indicar *baja* validez ya que las preguntas, por ejemplo, pueden no reflejar lo que realmente concierne a los estudiantes⁸.

Además, algunos autores secundan la necesidad de evaluar las propiedades psicométricas de este tipo de instrumentos porque, a pesar de su extendida utilización, advierten de la persistencia de malentendidos sobre las evaluaciones de la docencia por parte de los estudiantes⁹.

A este respecto, revisando la bibliografía nacional e internacional, encontramos diversos trabajos que tratan de evaluar la fiabilidad y validez de encuestas de evaluación^{6,10-12}, pero en su mayoría centrados en aspectos parciales de un proceso de aprendizaje¹³, la intervención eventual de tutorizaciones^{14,15} o modelos de actividad experimentales¹⁶. Además, una de las propiedades menos atendidas es la de la fiabilidad test-retest, a pesar de que resulta fundamental por las características intrínsecas de variabilidad en la población universitaria⁷.

En el contexto de la implementación del nuevo plan de estudios en la Facultad de Medicina de Zaragoza, el objetivo principal de esta publicación es presentar los aspectos psicométricos de fiabilidad y validez de una encuesta desarrollada para la evaluación global de las asignaturas del área docente de Psiquiatría: Psicología Médica, Psiquiatría General y la asignatura optativa de Psicopatología y Psiquiatría de Enlace.

MATERIAL Y MÉTODOS

Descripción y desarrollo de la escala:

La encuesta consta de un total de 24 ítems: Los primeros 21 ítems son puntuados en una escala de 5 grados y los 3 restantes son cualitativos, siendo el último de ellos reservado para las observaciones y comentarios libres por parte del alumno. La encuesta está subdividida en apartados referentes a (ver Anexo 1*): a) la asignatura; b) el profesorado c) los criterios de evaluación y d) una valoración global.

El desarrollo de la encuesta se sustenta en la amplia experiencia previa de los autores en el ejercicio de modelos de evaluación para la asignatura troncal de Psiquiatría de la Facultad de Medicina de Zaragoza. Así mismo, para incrementar la probabilidad de validez de la encuesta, los ítems fueron reevaluados en distintos cursos académicos y modificados de acuerdo con la bibliografía¹⁰ y sometidos a la impresión de varios estudiantes, con el objeto de asegurar la adaptabilidad y factibilidad.

Muestra y procedimiento de evaluación del instrumento:

El proceso de validación de la encuesta se realizó durante el curso académico 2005-2006, año en que se empezó a impartir la nueva asignatura: Psicopatología y Psiquiatría de Enlace Especial.

La encuesta se distribuyó entre los alumnos asistentes durante las últimas dos clases, previo al examen de evaluación, siguiendo un procedimiento de test-retest: La encuesta se administró, por tanto, en dos ocasiones (test 1 y test 2). Ambas fueron cumplimentadas por los alumnos asistentes de forma anónima e identificadas por cada uno de ellos por un código (número de identificación académica o similar), para hacer posible el emparejamiento posterior.

Para asegurar la confidencialidad de las encuestas, todo el proceso fue supervisado por cuatro alumnos representantes elegidos entre ellos mismos, que aseguraron el emparejamiento de las encuestas y posterior eliminación de los códigos de identificación.

Finalmente, las encuestas emparejadas fueron guardadas en un sobre sellado y el análisis de datos se realizó posteriormente a la publicación de las calificaciones y presentación de actas.

Tratamiento estadístico:

Fiabilidad: Se calcularon los siguientes parámetros.

a) La consistencia interna fue calculada mediante el coeficiente alfa de Cronbach¹⁷.

b) La fiabilidad test-retest se calculó, para los 21 primeros ítems, mediante los coeficientes de acuerdo global (OA), el Kappa (corregido por la probabilidad de acuerdo debido al azar) y el coeficiente KappaW (ponderado por proximidad a la diagonal en la matriz de acuerdos; dado que estos ítems corresponden a variables ordinales)^{18,19}.

Validez: La validez de contenido se propició inicialmente mediante la adecuación de los ítems para hacerlos comprensibles a los alumnos, además de modificarlos de acuerdo con las encuestas validadas publicadas, como ha sido sugerido por Copeland¹⁰.

Con respecto a la validez de constructo, teniendo en cuenta las dificultades ya comentadas en la introducción, se pretendió una aproximación a la misma de acuerdo con el concepto de "concurrency" de Regier y Burke²⁰: con lo que consideramos que determinados ítems debían estar relacionados con las calificaciones obtenidas en la asignatura.

RESULTADOS

El total de alumnos matriculados fue de 43 (todos ellos de 4º curso de medicina, ya que era el primer año en que se establecía el nuevo plan de estudios en el segundo ciclo de la licenciatura). De ellos, el 75% eran mujeres. La asistencia a clase fue regular durante el periodo en que se impartió la asignatura (en torno al 80%).

El test 1 fue cumplimentado por N=39 alumnos. Estas encuestas fueron las utilizadas para realizar el análisis descriptivo de la distribución de respuestas en los ítems ordinales y en los ítems cualitativos. El test 2 fue completado por N=34 alumnos (discreta diferencia atribuible a la mayor proximidad de los exámenes finales de la convocatoria de Junio de 2006).

El número de encuestas emparejadas fue de N=31 (alumnos que completaron el estudio realizando la encuesta en las dos ocasiones). Sobre esta muestra, se realizó el análisis de fiabilidad test-retest cuyos resultados se muestran en la tabla 1. Con un acuerdo KappaW global=0,655 con un intervalo de confianza al 99% de 0,562-0,748 que, según la interpretación de Landis y Koch²¹, resulta un acuerdo moderado a bueno. Así mismo, el coeficiente de fiabilidad alfa de Cronbach ofreció un resultado de 0,865 (Meritorio-Excelente)¹⁷ lo que nos habla de la consistencia interna del estudio.

* A disposición de los interesados contactando con los autores

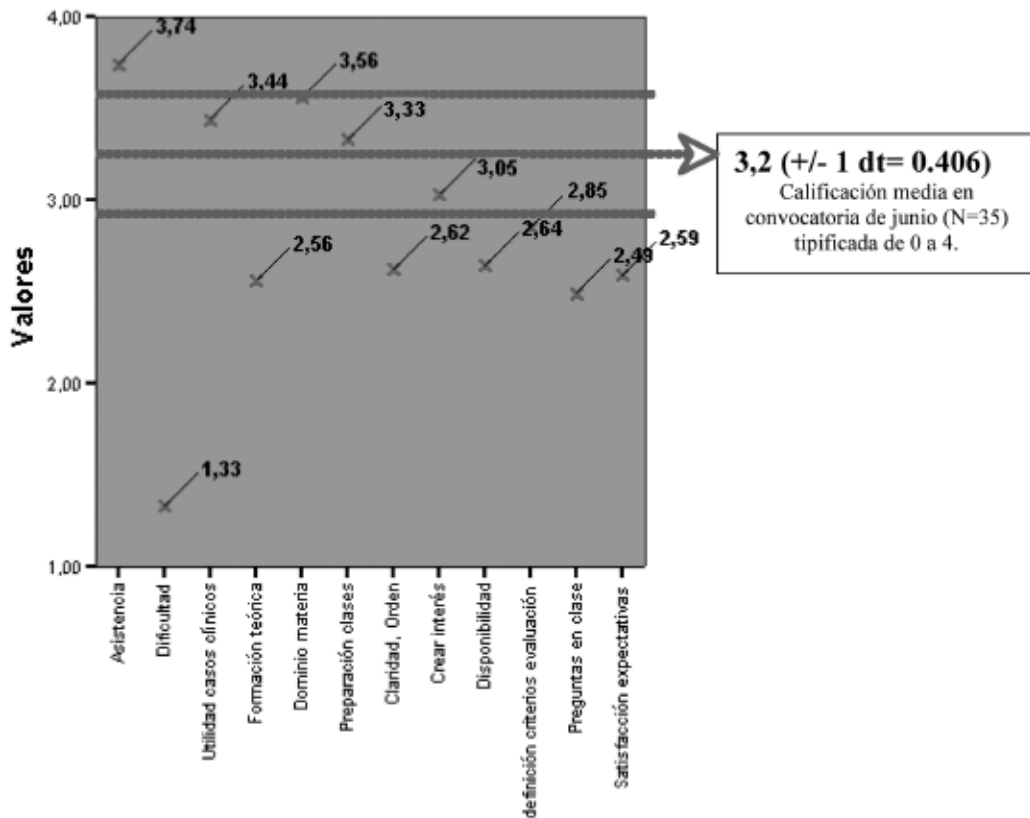
Tabla 1. Fiabilidad Test-Retest de la nueva encuesta de evaluación docente

	Acuerdo Global (OA)	Kappa	kappaW	SIGNIFICACIÓ
Item 1	0,903	0,363	0,542	P<0.001
Item 2	0,806	0,595	0,649	P<0.001
Item 3	0,689	0,472	0,541	P<0.001
Item 4	0,766	0,513	0,568	P<0.001
Item 5	0,838	0,704	0,726	P<0.001
Item 6	0,733	0,614	0,699	P<0.001
Item 7	0,548	0,295	0,385	P<0.001
Item 8	0,806	0,410	0,425	P<0.001
Item 9	0,806	0,692	0,770	P<0.001
Item 10	0,709	0,572	0,650	P<0.001
Item 11	0,741	0,589	0,671	P<0.001
Item 12	0,741	0,623	0,718	P<0.001
Item 13	1,000	1,000	1,000	P<0.001
Item 14	0,866	0,730	0,737	P<0.001
Item 15	0,741	0,622	0,722	P<0.001
Item 16	0,612	0,366	0,435	P<0.001
Item 17	0,700	0,472	0,528	P<0.001
Item 18	0,806	0,693	0,752	P<0.001
Item 19	0,806	0,640	0,660	P<0.001
Item 20	0,709	0,540	0,688	P<0.001
Item 21	0,900	0,854	0,901	P<0.001
Media Global	0,772	0,588	0,655	P<0.001

En las dos ocasiones en que se aplicó la encuesta, el grado de cumplimentación de los ítems fue cercano al 100% y no sobrepasó los 15 minutos de duración.

Finalmente, respecto a la validez de la encuesta, se ofrecen resultados aproximativos a la misma, basándonos en ciertas premisas referentes a la validez de constructo: Dados los altos resultados de fiabilidad observados para la encuesta y suponiendo que ésta mida adecuadamente la impresión global de los alumnos sobre la asignatura; los ítems se relacionarán con los resultados en las calificaciones de los alumnos. Puesto que existe la limitación, común a estos procedimientos, de que se debe asegurar el anonimato en la realización de los cuestionarios, en la figura 1 se representa la correlación entre los ítems más significativos y la calificación global de los alumnos en la convocatoria de Junio 2006 tipificada por escala de puntuación de la encuesta (0 a 4 puntos). Como se puede observar, existe correlación directa con todos los ítems expuestos excepto el de "dificultad", que es inversa (lo cual apoya la validez, ya que la mayoría de alumnos evaluaron la asignatura como "fácil").

Figura 1. Relación de los principales ítems (Asignatura y Profesorado) con respecto a la calificación media tipificada.



DISCUSIÓN

Los datos mostrados anteriormente confirman que la encuesta de evaluación global presenta una alta consistencia interna y adecuada fiabilidad test-retest. Así mismo, la validez de contenido es adecuada y la correlación de los resultados con las calificaciones globales del curso apoya la validez por “concurrency” de la misma.

Todo ello, junto con la factibilidad observada aseguran que la encuesta permita en nuestro departamento un adecuado “feed-back” entre el alumnado y favorece un modelo dinámico de docencia⁴.

Como se ha comentado en la introducción, otros autores han documentado resultados similares. Entre ellos, cabe destacar el trabajo de Copeland y Hewson¹⁰ que obtiene resultados comparables en cuanto a validez y consistencia interna para una encuesta de un departamento, pero estandarizada dentro del ámbito clínico y careciendo de un contraste de fiabilidad test-retest.

Con respecto a los estudios que sí han aportado un test-retest en el ámbito de las habilidades en formación médica preclínica, Holloway y cols.²² obtuvieron resultados moderados; y también observaron una fiabilidad similar a la documentada en este artículo Torsheim y cols.²³ para una breve autoescala que evalúa la percepción de apoyo por parte de los profesores y los compañeros de clase.

Los autores creen que éste es un aspecto a destacar, ya que la variabilidad demostrada en la opinión de los alumnos se debe tener en cuenta para futuras metodologías de evaluación. Las razones de dicha variabilidad están por clarificar, aunque ha sido documentada en relación con las dificultades para medir validez⁷. No obstante, se podría hipotetizar que el hecho de que existan amplias diferencias en las metodologías docentes utilizadas en las asignaturas de la licenciatura influye, incluso en cuestión de horas, en la relatividad de las opiniones de los alumnos.

Como limitaciones al estudio presentado destaca: a) la persistencia de algunos ítems con baja fiabilidad que, eventualmente, deberían ser revisados; b) la ausencia de un adecuado “patrón oro” para contrastar la validez de constructo, aunque se ha diseñado un seguimiento de la evolución formativa de los alumnos de la asignatura optativa en la asignatura de Psiquiatría General, durante el presente año, y potencialmente se podría seguir en la futura formación especializada; y c) la posibilidad de generalización de la encuesta a otras áreas de conocimiento médico, ya que ha sido estandarizada en una signatura optativa,

aunque ha sido aplicada en otras dos asignaturas troncales, a cargo del Departamento de Psiquiatría.

A pesar de las limitaciones expuestas, los autores creen que esta metodología presenta un valor divulgativo considerable y puede contribuir a la comentada CEES en Ciencias de la Salud y la adaptación a estándares europeos.

En concreto, el cuestionario permite: a) la evaluación individual de cada curso; b) la medición sistemática entre distintos cursos c) comparación en el tiempo de la evolución de distintos cursos d) detección de deficiencias para facilitar la puesta en marcha de intervenciones y control de cambios con una nueva administración.

Se puede resumir, como conclusión, que se dispone de un instrumento potencialmente válido, que nos aporta una retroalimentación fiable con los alumnos, que resulta útil en la toma de decisiones sobre modificaciones en la metodología docente en el contexto de la adquisición de los estándares europeos y que podría ser, eventualmente, generalizable a otras disciplinas.

BIBLIOGRAFÍA

1. Alfonso-Roca MT. El aprendizaje de la Medicina en la universidad de Castilla la Mancha: Una experiencia educativa basada en el aprendizaje de las competencias profesionales. *Educ Méd* 2005; 8(2):09-12.
2. Baron Maldonado M. La razón de ser de la Acreditación de las Facultades de Medicina. *Educ Méd* 2003; 6(3):12-3.
3. Keeling R. The Bologna Process and the Lisbon Research Agenda: the European Commission's expanding role in higher education discourse. *European Journal of Education* 2006; 41(2):203-23.
4. Richardson JTE. Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education* 2005; 30(4):387-415.
5. Bandaranayake RC. Utilization of feedback from student evaluation of teachers. *Med Educ* 1978; 12(4):314-20.
6. Peluso MA, Tavares H, D'Elia G. Assessment of medical courses in Brazil using student-completed questionnaires. Is it reliable? *Rev Hosp Clin Fac Med Sao Paulo* 2000; 55(2):55-60.
7. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med* 2005; 20(12):1159-64.
8. Metcalfe DH, Matharu M. Students' perception of good and bad teaching: report of a critical incident study. *Med Educ* 1995; 29(3):193-7.
9. Leamon MH, Fields L. Measuring teaching effectiveness in a pre-clinical multi-instructor course: a case study in the development and application of a brief instructor rating scale. *Teach Learn Med* 2005; 17(2):119-29.

10. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 2000; 75(2):161-6.
11. Solomon DJ, Speer AJ, Rosebraugh CJ, DiPette DJ. The reliability of medical student ratings of clinical teaching. *Eval Health Prof* 1997; 20(3):343-52.
12. Irby D, Rakestraw P. Evaluating clinical teaching in medicine. *J Med Educ* 1981; 56(3):181-6.
13. Dolmans DH, De Grave W, Wolfhagen IH, van der Vleuten CP. Problem-based learning: future challenges for educational practice and research. *Med Educ* 2005; 39(7):732-41.
14. Dolmans DH, Ginns P. A short questionnaire to evaluate the effectiveness of tutors in PBL: validity and reliability. *Med Teach* 2005; 27(6):534-8.
15. Berlanga E, Monteagudo M., Pineda V., et al. Experiencia en la evaluación de tutores de especialistas en formación. *Educ Méd* 2004; 7(2):65-9.
16. Santibáñez M, Tamayo G, Meana JJ. Percepción de los estudiantes sobre un modelo experimental de actividad práctica en farmacología : un análisis controlado, aleatorizado y bajo condiciones de enmascaramiento. *Educ Méd* 2000; 3(2):69-76.
17. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16:297-334.
18. Larraga L, Saz P, Dewey ME, Marcos G, Lobo A. Validation of the Spanish version of the EURO-D scale: an instrument for detecting depression in older people. *Int J Geriatr Psychiatry* 2006.
19. Lobo A, Campos R, Perez-Echeverria MJ, Izuzquiza J, García Campayo J, Saz P, et al. A new interview for the multiaxial assessment of psychiatric morbidity in medical settings. *Psychol Med* 1993; 23(2):505-10.
20. Regier DA, Burke JD. Epidemiology--Quantitative and experimental methods in psychiatry. In: Kaplan HI, Sadock BJ (eds.) *Comprehensive textbook of psychiatry*/VI. 6th edition. Baltimore, MD: Williams & Wilkins, 1995:347-52.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1):159-74.
22. Holloway R, Nesbit K, Bordley D, Noyes K. Teaching and evaluating first and second year medical students' practice of evidence-based medicine. *Med Educ* 2004; 38(8):868-78.
23. Torsheim T, Wold B, Samdal O. The teacher and classmate support scale: factor structure, test-retest reliability and validity in samples of 13- and 15-year-old adolescents. *School Psychology International* 2000; 21(2):195-212.

agenda

ASME Annual Meeting

Lugar: Keele, Reino Unido
Fecha: 11-13 julio 2007
Contacto: www.asme.org.uk
e.mail: jenniferb@asme.org.uk

IAMSE Meeting

Lugar: Cleveland, USA
Fecha: 21-24 julio 2007
Contacto: www.iamse.org

18th Wonca World Conference

Lugar: Singapur
Fecha: 27 Julio 2007
Contacto: <http://wonca2007.com/index.html>

AMEE, 2007

Lugar: Trondheim, Noruega
Fecha: 27-29 Agosto 2007
Contacto: www.amee.org

XVIII Congreso de la SEDEM

Lugar: Santa Cruz de Tenerife
Fecha: 24-26 octubre 2007
Contacto: www.sedem.org

Discovery Courses in Medical Education 'Assessment'

Lugar: Dundee, UK
Fecha: 30 julio-3 agosto 2007
Contacto: www.dundee.ac.uk/meded/frames/courses.html#
e.mail: c.m.e.courses@dundee.ac.uk

The Network: Towards Unity for Health Conference,

Lugar: Kampala, Uganda
Fecha: 15-20 Septiembre 2007
Contacto: www.the-networktuffh.org/conference

XIIIth Regional Conference of Wonca, Europe, Paris, FRANCE

Lugar: Paris, Francia
Fecha: 17-20 Octubre.
Contacto: www.woncaeurope2007.org
e.mail: contactwonca2007@cngc.fr

AAMC Meeting

Lugar: Washington DC, USA
Fecha: 2-7 Noviembre 2007
Contacto: www.aamc.org