

Utilidad de la evaluación formativa en cursos clínicos integrativos en estudiantes de pregrado de medicina

JAIME LABARCA¹, CATALINA FIGUEROA²,
BÁRBARA HUIDOBRO², ANA CECILIA WRIGHT³,
ARNOLDO RIQUELME^{3,4}, RODRIGO MORENO³

Perception of medical students about formative assessments during clinical courses

Background: Formative evaluation is a range of formal assessment employed by professors during the teaching process in order to modify curriculum activities, to improve student attainment. For students, it is helpful to evaluate their learning process. Although recommended, it is seldom used. **Aim:** To evaluate the perception and performance of medical students subjected to formative assessments during an integrative clinical course. **Material and Methods:** Fourth year medical students that participated in a multiple choice formative assessment, similar to the final exam during 2007 and 2008, responded a survey about the usefulness and quality of such assessment. Student achievement was expressed as the percentage of correct answers of the tests. **Results:** The formative assessment was answered by 99% of students. In 2007 97% of students considered the experience as excellent or very good and 92% evaluated it as useful or very useful. During 2008 the figures were 89% and 79%. The students outlined that this assessment oriented their study, allowed them to discover their weaknesses and have a perception of the degree of difficulty of the final exam. Over 90% of students that took the formative evaluation, improved their academic achievement. **Conclusions:** Formative assessments are well evaluated by medical students and improve their academic achievement.

(Rev Med Chile 2014; 142: 1193-1199)

Key words: Education, medical; Educational measurement; Students, medical.

Las evaluaciones formativas son métodos diagnósticos estructurados aplicados durante el desarrollo de un curso que forman parte del plan de evaluación de una actividad curricular. Pueden ser de utilidad tanto para profesores como para los alumnos. Pueden servir para evaluar el currículo de un determinado programa. Se conciben en estrecha relación con los objetivos y actividades de aprendizaje programados, permitiendo retroalimentar distintas etapas del proceso al identificar fortalezas y debilidades¹⁻³. También sirven a los alumnos para evaluar sus competencias

y conocimientos, constituyendo una excelente herramienta para facilitar el aprendizaje de estudiantes de pre-grado de ciencias de la salud^{4,5}. Los estudiantes que se someten a evaluaciones formativas a mitad del período académico tienen mejores tasas de éxito en las evaluaciones sumativas que quienes no lo hacen⁶. Más aun, la sola participación en estas evaluaciones es un mejor predictor del resultado final que el rendimiento en ellas mismas⁶, lo cual apoya el rol clave que ocupa la información recuperada por los alumnos en las evaluaciones formativas. Sin embargo, la evalua-

¹Departamento de Enfermedades Infecciosas del Adulto.

²Escuela de Medicina

³Centro de Educación Médica.

⁴Departamento de Gastroenterología.

Pontificia Universidad Católica de Chile, Santiago, Chile.

Este trabajo fue presentado en el 2009 Association for Medical Education in Europe (AMEE) Annual Conference, Agosto-Septiembre 2009. Málaga, España.

Agradecimientos: Artículo parcialmente financiado por el Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) proyecto N° 1120652 (A.R.).

Recibido el 29 de marzo de 2014, aceptado el 2 de septiembre de 2014.

Correspondencia a:

Dr. Jaime Labarca

Departamento de Enfermedades

Infecciosas del Adulto

Escuela de Medicina

Pontificia Universidad Católica de Chile

Lira 63, Santiago, Chile.

Casilla 114-D.

Teléfono: 56-2-26863508

Fax: 56-2-26339820

jlabarca@med.puc.cl

ción formativa es una práctica no suficientemente integrada en escuelas de Medicina. Es así como en acreditaciones, este aspecto es frecuentemente considerado poco desarrollado y es recomendado en los planes de mejoría⁷.

Los cursos clínicos integrados abarcan múltiples disciplinas con una importante cantidad de contenidos. Es por ello que para los alumnos las evaluaciones sumativas finales son difíciles de preparar. Así, evaluaciones formativas al medio del período o cercanas al final de éste, permiten a los alumnos un mejor diagnóstico de sus conocimientos, jerarquizar mejor el estudio, llevándolos a concretar mejor los objetivos finales del curso^{8,9}.

El currículo de pregrado de la Escuela de Medicina de la Pontificia Universidad Católica de Chile dura siete años, con un número variable de entre 100 y 120 alumnos por curso^{10,11}. Específicamente, cuarto año tiene un curso integrado que incluye una parte práctica y una teórica. El curso teórico consta de 14 capítulos: Medicina Familiar, Traumatología, Neurología, Respiratorio, Hematología, Oncología, Gastroenterología, Nutrición, Diabetes, Geriátrica, Cardiología, Reumatología, Nefrología y Endocrinología, con una evaluación sumativa al final de cada uno. Asimismo, todo el curso teórico es evaluado con una evaluación sumativa final de aplicación clínica con temas fundamentales de cada especialidad. Se ha definido un total de 77 temas a evaluar, y al finalizar el año académico se efectúa una prueba de 59 preguntas de elección múltiple.

Dentro del diseño curricular del curso existen instancias estándar (*core curriculum*) e instancias electivas. Desde el año 2007, como parte de las instancias electivas, se ha implementado una evaluación formativa voluntaria en el curso de Integrado de cuarto año de clínicas médicas, un mes antes de la evaluación sumativa final. El objetivo de este trabajo fue evaluar la percepción y desempeño de los estudiantes de un sistema de evaluación formativa en un curso integrado clínico en pregrado de medicina.

Métodos

Con el objetivo de ayudar a los alumnos en el proceso de integración final de los contenidos del Curso Integrado, el año 2007 se efectuó por primera vez la evaluación formativa con preguntas

de aplicación clínica, incluyendo una pregunta por cada tema relevante definido para este curso y con grado de complejidad similar a la evaluación sumativa (mismo número y tipo de preguntas, mismo número de opciones por preguntas y mismos temas). Se realizó 4 semanas antes de la evaluación sumativa con asistencia voluntaria. Se extendió una invitación vía correo electrónico y mediante una reunión del profesor jefe con los alumnos. Además se realizó en 2 horarios diferentes, para facilitar la asistencia de los alumnos.

Aquellos alumnos que asistieron voluntariamente consintieron participar de esta actividad. El currículo del curso, incluyendo esta actividad de evaluación formativa, fue aprobado por el Comité de Currículo de la Escuela. Por lo anterior, no se solicitó consentimiento informado para el análisis de los resultados, como tampoco fue sometido a revisión por comité de ética, dado que esta instancia formativa se planificó como un instrumento de docencia y con el consentimiento de todos los jefes de programa (comité interno del curso).

Inmediatamente después de rendir la evaluación sumativa se aplicó una encuesta a los alumnos con preguntas abiertas acerca de la experiencia de la evaluación formativa, incluyendo una evaluación de la calidad del instrumento, de la utilidad de la experiencia y sugerencias para mejorar.

La evaluación formativa incluyó entre 3 y 6 preguntas por cada capítulo teórico, con un total de 59 preguntas de elección múltiple. Posteriormente, en un plazo de 72 h después del examen, se entregó un informe por alumno, con rendimiento por capítulo y rendimiento global, evaluando con una nota del 1,0 a 7,0. El rendimiento se definió por el método holístico, donde el 60% corresponde a nota 4,0^{9,12-13}, método utilizado según los estándares habituales de nuestra escuela. Además se envió un correo electrónico a todo el curso con comentarios finales acerca de la experiencia y se ofreció a aquellos alumnos que tuvieron dificultades, la posibilidad de discutir con el jefe del curso las causas de su mal rendimiento y establecer estrategias remediales en el período restante del curso.

Los resultados de la evaluación formativa y sumativa tanto del año 2007 como del año 2008 se expresaron en porcentaje (%) de respuestas correctas y se comparó el desempeño de las cohortes en relación a sus promedios mediante test de McNemar y test t de Student.

El mismo examen de evaluación formativa se

aplicó nuevamente, en iguales condiciones, el año 2008. La evaluación sumativa final se modificó en cuanto al contenido de las preguntas, pero conservando el mismo estilo, número de preguntas y grado de dificultad con respecto al año anterior.

Resultados

En la evaluación formativa participaron voluntariamente un total de 107 alumnos de cuarto año el año 2007 y 112 alumnos el año 2008, correspondientes al 99 y 98,2% del total de alumnos de cada año respectivamente.

En cuanto a la evaluación del instrumento, en

2007 el 97% de los alumnos calificaron la prueba como muy buena o excelente, porcentaje que disminuyó a 88,5% en el año 2008. En cuanto a la utilidad, 92% apreció la evaluación formativa como útil o muy útil en 2007; en 2008 el porcentaje de alumnos con la misma opinión fue de 79%. Los años 2007 y 2008, 31% y 27% respectivamente afirmaron que les sirvió para orientar el estudio, 22% y 27% para darse cuenta de sus debilidades, 18% y 17% para conocer el estilo de la prueba y el grado de dificultad, 22% y 3% para tranquilizar y disminuir la ansiedad, 0% y 20% para crear falsas expectativas y 7% y 9% postularon otras opciones como integrar conocimientos, ensayar el tiempo, entretenimiento, motivar el estudio (Figuras 1 y 2). En

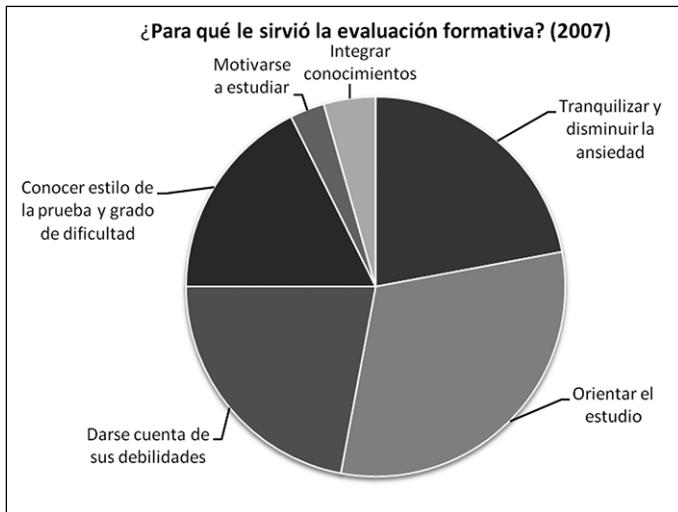


Figura 1. Opinión de los alumnos al preguntarles en qué les ayudó la evaluación formativa el año 2007.

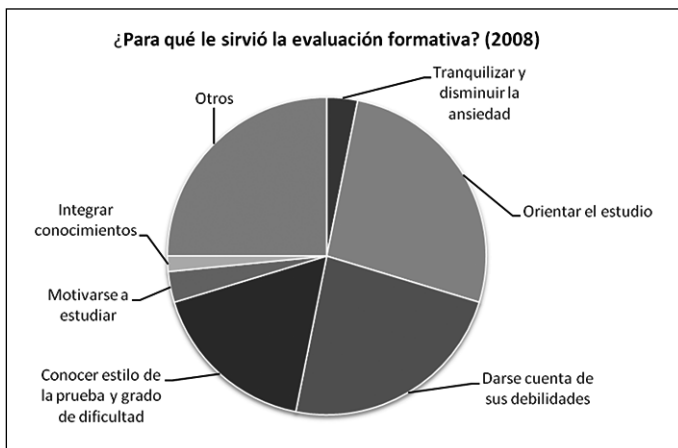


Figura 2. Opinión de los alumnos al preguntarles en qué les ayudó la evaluación formativa el año 2008.

cuanto a las sugerencias para el próximo año, 68% el 2007 y 78% el año 2008 pidió mayor similitud en el grado de dificultad entre la prueba formativa y la prueba final sumativa. Otras sugerencias incluyeron repetir la actividad los próximos años y hacer más de una prueba formativa durante el año, entre otras.

No se observaron diferencias significativas en el rendimiento entre las pruebas formativas de ambos años, como tampoco de ambas pruebas sumativas finales, tanto al comparar el rendimiento global como por capítulos. El promedio de logro en ambas evaluaciones de los años 2007 y 2008 se muestra en la Tabla 1. La diferencia de promedio de logro entre la prueba formativa y la sumativa

en ambos años es estadísticamente significativa ($p < 0,05$). El 93% y 95% de los alumnos aumentó su nota de la evaluación sumativa con respecto a la evaluación formativa en los años 2007 y 2008 respectivamente.

La correlación del porcentaje de logro de cada alumno, entre la evaluación formativa y sumativa de los años 2007 y 2008, se muestra en las Figuras 3 y 4 respectivamente (coeficiente de correlación R 0,54 y 0,5). Destaca que el 50% de alumnos de mejor rendimiento en la evaluación formativa del año 2007 tuvo una mejoría de 7,7% en la evaluación sumativa, mientras que el grupo de alumnos del 50% de más bajo rendimiento mejoró en 15,2% el mismo año. En tanto, el año 2008 se observó

Tabla 1. Participación y rendimiento en evaluaciones formativa y sumativa (2007 y 2008)

	2007 (n = 108)		2008 (n = 112)	
Participación (n,%)	107	(99%)	112	(98,2%)
Promedio de logro en evaluación formativa (% , rango)	74	(52,5-89,8)	72,4	(47,5-94,9)
Promedio de logro en evaluación sumativa (% , rango)	85,4	(64,0-98,0)	87	(71,4-100)
Diferencia promedio entre evaluación formativa y sumativa (%)	15,7	($p < 0,05$)	19	($p < 0,05$)
Alumnos que aumentaron su nota (%)	93		95	

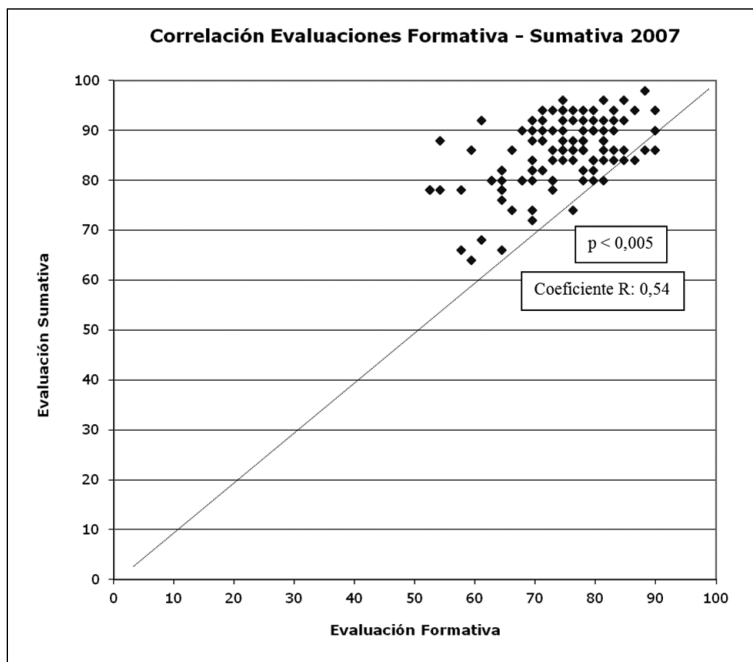


Figura 3. Correlación del porcentaje de rendimiento de cada alumno en las evaluaciones formativa y sumativa del año 2007.

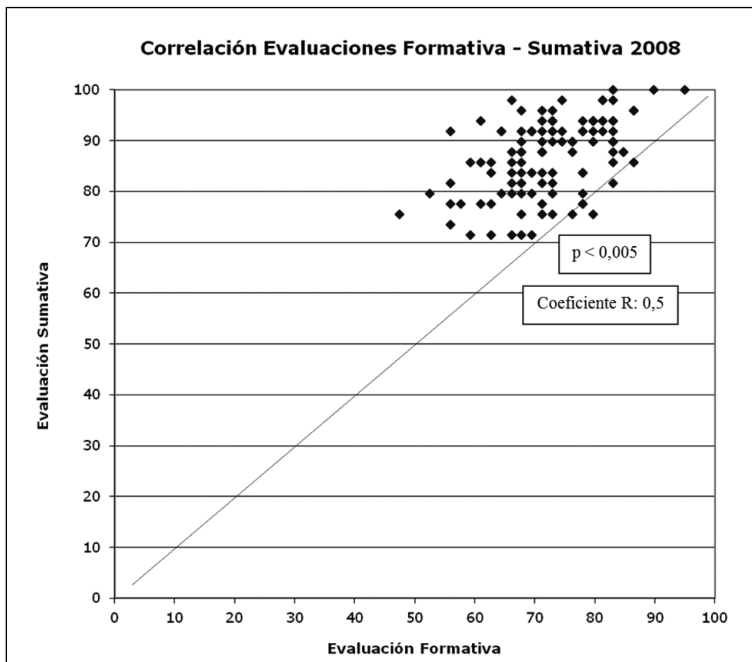


Figura 4. Correlación del porcentaje de rendimiento de cada alumno en las evaluaciones formativa y sumativa del año 2008.

una mejoría de rendimiento de 10,8% en el grupo correspondiente al 50% de mejor rendimiento, porcentaje que asciende a 18,3% en el 50% de peor rendimiento, resultados que son estadísticamente significativos ($p < 0,005$).

Discusión

Si bien las evaluaciones formativas han demostrado ser de utilidad para los profesores en la confección y mejoría de programas de un curso, son también relevantes para ayudar a fortalecer el aprendizaje de los estudiantes. En este caso, la evaluación formativa se implementó como una actividad de beneficio para los alumnos, que les permitiera evaluar la adquisición de conocimientos fundamentales y familiarizarse con la evaluación sumativa final. En ese sentido lo primero que destaca es la asistencia masiva a una actividad de tipo voluntaria, demostrando el interés de los alumnos por una actividad de un tema generativo y de alto interés para aprendizaje.

Al ser un curso integrado, se implementó hacia el final de curso pues busca revisar los contenidos fundamentales adquiridos durante todo el año, para lo cual es necesario que los alumnos hayan

pasado por la mayor parte de los capítulos. Al mismo tiempo se consideró realizarlo con suficiente anticipación a la evaluación sumativa final (1 mes) para dar tiempo a los alumnos a trabajar en sus debilidades y orientar la preparación de su examen final. Esto se logró gracias a una retroalimentación oportuna, con un informe detallado del porcentaje de logro de cada área y del rendimiento global, como también la posibilidad de revisar las preguntas y establecer estrategias para mejorar debilidades en conjunto con los docentes del curso.

En cuanto a la confección de la prueba, se cumple con la validez de contenido, dado que todos los capítulos teóricos tuvieron una adecuada representación en la evaluación formativa. El objetivo fue evaluar la adquisición de conocimientos generales de cada capítulo y no los aspectos puntuales de las distintas patologías, por lo cual se seleccionaron e hicieron preguntas de los temas relevantes de cada capítulo, tanto en la evaluación formativa como sumativa, lo cual le da a esta intervención una buena validez de muestreo (*sampling validity*). Para evaluar de manera más amplia y específica los conocimientos de cada capítulo, el curso Integrado cuenta con evaluaciones teóricas al finalizar cada uno de ellos.

Analizando el resultado de la encuesta aplicada a los alumnos, tanto en 2007 como en 2008 hubo un alto grado de satisfacción en la aplicación y la utilidad de esta herramienta metodológica¹⁴. En 2008, 20% de los alumnos sugiere a la evaluación formativa como creadora de falsas expectativas, hecho que ningún alumno del año 2007 menciona. Probablemente esto se debe a la percepción del grado de dificultad, a que los grupos humanos evaluados en ambos años son diferentes y debido a que los docentes o compañeros del curso anterior pueden haber creado muchas expectativas sobre la base de la experiencia previa. Sin embargo, ambos grupos, en su gran mayoría, valoran esta actividad como útil o muy útil, además de mencionar que repetirían la experiencia. Esto es importante debido a que al mantenerse la percepción de utilidad mejora la eficiencia (una misma prueba formativa protegida por los mismos alumnos para ser usada cada año)¹⁵. A la fecha se ha seguido realizando esta actividad, con una asistencia voluntaria sobre 95%, siendo bien evaluada y valorada por los alumnos.

En cuanto al rendimiento de los alumnos, en ambos años hubo una mejoría de más de un 10% al comparar la evaluación formativa con la sumativa, lo que se explicaría por una mayor preparación y estudio personal antes de la evaluación final. Por lo anterior es difícil valorar en cuanto contribuye al aprendizaje. Sin embargo, la importancia de la evaluación formativa radica en ayudar a dirigir y orientar el estudio para la evaluación final, pues mediante una retroalimentación (*feedback*) formativa efectiva los alumnos descubren qué mejorar^{8,13,16,17}. El hecho de que los alumnos con rendimiento más bajo sean los que cuantitativamente mejoren más su rendimiento y que esta diferencia sea estadísticamente significativa sugiere que precisamente en ellos sería más útil la aplicación de la evaluación formativa. Esto contribuye como evidencia para desarrollar estrategias individualizadas y efectivas para la enseñanza de los alumnos de aprendizaje más lento¹⁸. Pero pese a que los alumnos de mayor rendimiento mejoran levemente (pequeño aumento porcentual), probablemente lo valoran tanto como sus compañeros con mayor margen de ganancia¹⁹. Así esta evaluación formativa tendría una buena validez predictiva, pues constituye un buen predictor del desempeño individual y grupal en la evaluación sumativa.

Basándonos en las 4 etapas del modelo de Kirkpatrick: en la primera *–Reacción–* se debe evaluar en qué grado les gustó el proceso de aprendizaje a los alumnos. En este caso la percepción de los estudiantes es muy buena de acuerdo a la encuesta de satisfacción. En la segunda *–Aprendizaje–* se evalúa qué fue lo que aprendieron (conocimientos y/o habilidades). En este caso se evidencia un incremento significativo de los aprendizajes terminales definidos según objetivos de curso. En la tercera *–Comportamiento–* se debe evaluar qué cambios en el rendimiento del alumno resultaron del proceso de aprendizaje. Por último, en la cuarta etapa *–Resultados–*, se evalúan los efectos medibles del proceso de aprendizaje en términos de costo, eficiencia, calidad, cambio o modificación de sistemas. Este último nivel sería el más difícil de demostrar, pero es un hecho que se generó un cambio perdurable en el tiempo ya que esta instancia formativa es parte del programa del curso integrado de cuarto año hasta la fecha^{20,21}. En una próxima instancia sería interesante completar la evaluación de impacto de esta intervención en el nivel 3 y el 4 a más largo plazo.

En cuanto a las líneas futuras a desarrollar, sería importante aumentar la dificultad de la evaluación formativa, considerando que una de las mayores críticas que hacen los alumnos es la discrepancia en el grado de dificultad entre la evaluación formativa y sumativa. Además se evaluará realizar esta experiencia dos veces por año en el curso de cuarto año, al finalizar el primer semestre para detectar alumnos de bajo rendimiento y al finalizar el segundo semestre para estimular la preparación del examen final⁴.

En suma, ésta es una experiencia que apoya el uso de las evaluaciones formativas en pregrado, especialmente en cursos complejos integradores en que se abarcan múltiples áreas de la medicina. Su implementación se alinea con el concepto de evaluación programática, que sugiere un desplazamiento desde la evaluación del aprendizaje hacia la evaluación para el aprendizaje, abriendo nuevas perspectivas al otorgarle un papel activo en el proceso del aprendizaje a las instancias de evaluación formativa y sumativa²². Además, muchos de los alumnos lo reconocen como un esfuerzo adicional de los docentes por mejorar su aprendizaje, lo que confirma esta herramienta como útil y valorada en el proceso de enseñanza de los estudiantes de medicina.

Referencias

- Al-Wardy NM. Assessment methods in undergraduate medical education. *Sultan Oaboos Univ Med J* 2010; 10 (2): 203-9.
- Ellwart T, Konradt U. Formative versus reflective measurement: an illustration using work-family balance. *J Psychol* 2011; 145 (5): 391-417.
- Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: A systematic review of reviews. *Int J Med Inform* 2012; 81 (1): 1-11.
- Pittenger AL, Lounsbery JL. Student-generated questions to assess learning in an online orientation to pharmacy course. *Am J Pharm Educ* 2011; 75 (5): 94.
- Diker A, Cunningham-Sabo L, Bachman K, Stacey JE, Walters LM, Wells L. Effective Training Design: Use of Theory and Formative Assessment. *Health Promot Pract* 2012; 13 (4): 496-505.
- Carrillo-de-la-Peña MT, Bailles E, Caseras X, Martínez A, Ortet G, Pérez J. Formative assesment and academia achievement in pre-graduate students of health sciences. *Adv Health Sci Educ Theory Pract* 2009; 14 (1): 61-7.
- Sánchez I, Riquelme A, Moreno R, García P, Salas SP. International accreditation process at a Latin American medical school: a 10-year experience. *Med Teach* 2010; 32 (3): 271.
- Carless D. 'Conceptualizing pre-emptive formative assessment', *Assessment in Education: Principles, Policy & Practice* 2007; 14 (2): 171-84.
- George S, Hague MS, Oyeboode F. Standard setting: comparison of two methods. *BMC Med Educ* 2006; 14 (6): 46.
- Bitran M, Wright AC, Zúñiga D, Mena B, Velasco N, Moreno R. [Improvement of medical student's academic performances in times of curricular reform]. *Rev Med Chile* 2002; 130 (4): 437-45.
- Rosso P, Velasco N, Moreno R. [Undergraduate curriculum reform at the Pontifical Catholic University Medical School: aims, methodology and advance status]. *Rev Med Chile* 1997; 125 (7): 796-807.
- Jalili M, Hejri SM, Norcini JJ. Comparison of two methods od standard setting: the performance of the three-level Angoff method. *Med Educ* 2011; 45 (12): 1199-208.
- Hess BJ, Weng W, Lynn LA, Holmboe ES, Lipner RS. Setting a fair performance standard for physicians' quality of patient care. *J Gen Intern Med* 2011; 26 (5): 467-73.
- Ross S, Poth CN, Donoff M, Humphries P, Steiner I, Schipper S, Janke F, Nichols D. Competency-based achievement system: using formative feedback to teach and assess family medicine residents' skills. *Can Fam Physician* 2011; 57 (9): e323-30.
- Haffling AC, Beckman A, Edgren G. Structured feedback to undergraduate medical students: 3 years' experience of an assessment tool. *Med Teach* 2011; 33 (7): e349-57.
- Nicola D, Macfarlane-Dickb D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud High Educ* 2006; 31 (2): 199-218.
- Bienstock JL, Katz NT, Cox SM, et al. Association of Professors of Gynecology and Obstetrics Undergraduate Medical Education Committee. To the point: medical education reviews-providing feedback. *Am J Obstet Gynecol* 2007; 196 (6): 508-13.
- Krishnakumar P, Geeta MG, Ramakrishnan P. Effectiveness of Individualized Education Program for Slow Learners. *Indian J Pediatr* 2006; 73 (2): 135-7.
- Schuwirth LW, Van der Leuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach* 2011; 33 (6): 478-85.
- Kirkpatrick DL. *Evaluating Teaching Program*. San Francisco: Berrett-Koehler Publishers; 1994.
- Kirkpatrick D. Great Ideas Revisited: Revisiting Kirkpatrick's four-level model. *Training Dev* 1996; 50 (1): 54-7.
- Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach* 2011; 33 (6): 478-85.