

Evaluación para el aprendizaje: experiencia en un curso teórico de pregrado en medicina

GONZALO PÉREZ¹, EDUARDO KATTAN¹, LUZ COLLINS^{2,b}, ANA CECILIA WRIGHT^{2,b}, TOMÁS RYBERTT¹, AGUSTÍN GONZÁLEZ¹, MARISOL SIRHAN^{2,3}, NANCY SOLÍS^{3,a}, MARGARITA PIZARRO^{3,a}, MARCO ARRESE³, ALBERTO SARFATIS¹, NICOLE LUSTIG¹, JUAN PABLO ARAB³, JAIME LABARCA¹, ARNOLDO RIQUELME^{2,3}

Assessment for learning: experience in an undergraduate medical theoretical course

Background: Assessment for learning is a paradigm that is taking shape in the field of medical education. This approach aims to embed the assessment process within the educational and learning process. **Aim:** To evaluate the impact of curricular changes, from a focus of assessment of learning to one of assessment for learning, in the perception of undergraduate students of medicine and their final grades obtained in a theoretical course (TCG). **Material and Methods:** In the year 2011 lectures were reduced and intermediate assessments followed by a feedback session were introduced. The activities of each program course, surveys about student perceptions of the course and the final grades of students (assessments with multiple choice questions) were compared between the periods prior and after curricular changes (2005-2010 and 2011-2013). **Results:** As a consequence of curricular changes, time for lectures was reduced by 19.5%, time for summative assessments was increased by 8.5%, and feedback activity, occupying 7.3% of the course time was added. There were significant improvements in student is perceptions in all areas assessed by surveys, emphasizing feedback and assessments. The overall grade assigned to the course dictated after implementing the changes increased from 6.18 to 6.59 ($p < 0.001$, 1-7 scale). The grades of students also improved from an average of 5.78 to 6.43 ($p < 0.001$, 1-7 scale). **Conclusions:** Assessment for learning achieved the desired educational impact without increasing the assigned curricular time. Programmatic assessment is favorably perceived by students.

(Rev Med Chile 2015; 143: 329-336)

Key words: Competence, clinical; Education, medical; Measurement, educational; Students, medical.

La Escuela de Medicina de la Pontificia Universidad Católica de Chile (EMPUC) ha realizado cambios curriculares desde los años 90, adoptando las tendencias internacionales en Educación Médica¹⁻⁵. El foco de la innovación curricular se basa en la aplicación de enseñanza en pequeños grupos y ambientes simulados con

entrega de retroalimentación (*feedback*) estructurado a los estudiantes. En términos de evaluación se ha avanzado en la mejoría de la calidad de las preguntas utilizadas para medición de conocimientos mediante capacitación docente y apoyo de parte del Centro de Educación Médica (CEM). La innovación en evaluación se ha centrado en la

¹Escuela de Medicina.

²Centro de Educación Médica.

³Departamento de Gastroenterología, Escuela de Medicina, Pontificia Universidad Católica de Chile.

^aBioquímico.

^bPedagogo.

Recibido el 30 de abril de 2014, aceptado el 12 de enero de 2015.

Correspondencia a

Arnoldo Riquelme, MMedEd
Departamento de

Gastroenterología, Centro de Educación Médica Pontificia Universidad Católica de Chile. Marcoleta 367, Casilla 114-D. Santiago, Chile.

Teléfono: 56-2-3543820

Fax: 56-2-6397780

a.riquelme.perez@gmail.com

incorporación de instrumentos de evaluación de destrezas clínicas (ECOEs) y de actitudes y profesionalismo, como por ejemplo, el Portafolio introducido el año 2007⁶.

La evaluación es un área de la educación médica de particular relevancia y que no se basa en una sola teoría que la sostenga, sino que toma elementos de la educación general, psicología cognitiva, teorías de toma de decisiones y teorías psicométricas^{7,8}.

En evaluación ha surgido con fuerza un nuevo paradigma, el paso desde la evaluación *del* aprendizaje a la evaluación *para* el aprendizaje^{9,10,11}. Este marco teórico se basa en algunos elementos distintivos destacados en la Tabla 1.

En el diseño conceptual de la evaluación *para* el aprendizaje, el proceso de evaluación está intrínsecamente ligado al proceso educacional, incorporando elementos e información de varias fuentes para identificar las fortalezas y debilidades de cada estudiante, con el propósito de proveer al alumno de *feedback* efectivo e individualizado para optimizar su aprendizaje¹¹. Esta aproximación está alineada con los principios FAIR, que se han planteado para lograr mayor aprendizaje efectivo: entrega de *Feedback*, comprometer al estudiante en aprendizaje *Activo*, *Individualizar* el aprendizaje a las necesidades del estudiante y hacer que el aprendizaje sea *Relevante*¹².

La EMPUC ha incorporado aspectos de esta visión de la evaluación en los últimos años. Durante el IV año del pregrado los alumnos realizan el Curso de Integrado de Clínicas (CIC), el cual contiene rotaciones clínicas ambulatorias y hospitalarias evaluadas con pautas aplicadas por los docentes

a cargo de grupos de 4 a 6 alumnos y ECOEs. El componente teórico se subdivide por módulos, cada uno de éstos aborda distintas especialidades y sub-especialidades. Específicamente, el módulo correspondiente al Curso Teórico de Gastroenterología (CTG) fue sometido a profundos cambios en el programa del curso aplicando los conceptos relacionados con el enfoque de evaluación *para* el aprendizaje a partir del año 2011. Este módulo ya había presentado mejoras en evaluación, con la creación de un manual para el estudiante y la formulación adecuada de preguntas de selección múltiple (PSM) utilizadas en sus pruebas, a partir del período 2002-2006¹³.

El objetivo del presente artículo fue evaluar el impacto del cambio curricular del CTG de la EMPUC, desde un enfoque de evaluación *del* aprendizaje a uno de evaluación *para* el aprendizaje. Los resultados de esta experiencia de innovación curricular incluyeron la percepción de los propios estudiantes y el impacto académico reflejado en las calificaciones finales.

Métodos

Se utilizaron los registros del CEM relacionados con el CTG desde los años 2005 hasta 2013, incluyendo los programas del curso, encuesta de evaluación del curso en base a la percepción de los alumnos y calificaciones finales de los alumnos.

Programa del curso

El programa del CTG incluye objetivos, descripción general, y calendario de actividades con

Tabla 1. Ideas centrales evaluación para el aprendizaje

- El propósito central de un currículo es que los alumnos estudien bien y aprendan lo que más puedan
- En la evaluación del aprendizaje, se busca responder a si un alumno A es mejor que el alumno B, mientras que en la evaluación para el aprendizaje la pregunta central es si el alumno A es mejor que antes, y en paralelo si el alumno B lo es también
- La evaluación cambia de un enfoque en que se asume uniformidad de los estudiantes a uno más acorde a las características de cada uno de ellos
- Se enfatiza que los instrumentos de evaluación sean los óptimos para cada competencia específica que se quiera medir
- Para lograr resultados de aprendizaje, es necesario que el proceso de evaluación se base en un enfoque programático que contenga diferentes elementos, tanto cualitativos como cuantitativos, en donde el *feedback* ocupe un lugar relevante, logrando una caracterización más completa del estudiante
- El juicio y criterio de los docentes son indispensables en la evaluación

la distribución en tiempo de clases lectivas, seminarios y evaluaciones. El CTG se realiza durante las tardes de 3 semanas académicas y tiene carácter obligatorio, siendo parte de los contenidos mínimos necesarios para aprobar el IV año.

Encuesta de evaluación del curso en base a la percepción de los alumnos

Esta encuesta se aplica al finalizar el curso, posterior a la evaluación sumadora final. La realización de esta encuesta es común a todos los cursos teóricos de la EMPUC, siendo responsabilidad del CEM su tabulación y registro de manera anónima.

La encuesta cuenta con 3 dominios pre-determinados: Dominio 1. Tiempo dedicado (h/semana). Dominio 2. Logro de Objetivos (%). El Dominio 3 consta de 7 ítems: Métodos, Profesores, Fuentes de Información, *Feedback*, Notas, Organización, Infraestructura y Materiales (escala Likert de 1 a 7). Al final de la encuesta el alumno asigna una Nota Global (que el alumno pone al ramo en su conjunto, con escala 1 a 7) y espacio en blanco para comentarios libres.

Calificaciones finales de los alumnos

Tienen como universo todos los alumnos cursando el CTG. Se basan en una escala de 1 al 7, con un porcentaje de exigencia de 60% (método holístico) para la nota mínima aprobatoria, común para todos los módulos teóricos del curso integrado de cuarto año (equivalente a nota 4,0)¹⁴.

El cambio curricular en el CTG se realizó a contar del año 2011. Las intervenciones realizadas al programa, incluyendo aspectos psicométricos de los instrumentos de evaluación^{13,15,16}, se describen en la Tabla 2.

El modelo de evaluación programática, contempla ciclos de aprendizaje que incluyen actividades de aprendizaje o entrenamiento, evaluación formativa y de soporte (*feedback*). Cada ciclo culmina con una evaluación intermedia y se repite las veces que se estime conveniente según los objetivos del curso, para terminar con una evaluación final¹⁷. En la Figura 1 se ilustra una adaptación del modelo implementado en el CTG.

Para estimar si las intervenciones al programa resultaron en una mejoría en la percepción de parte de los alumnos al curso y/o una mejora en las calificaciones finales, se dividieron los registros de los cursos en dos grupos: el grupo pre-intervención curricular, del año 2005 al 2010, y el grupo post-intervención curricular, del año 2011 al 2013.

Se realizó una comparación entre los programas de ambos grupos. Se agruparon las actividades descritas en el calendario de los programas en: clases lectivas, integración (clases/seminarios), evaluación sumadora, y *feedback*. Se cuantificó el tiempo destinado a cada actividad en horas y se calculó su porcentaje respecto al total de tiempo del curso.

Se compararon las evaluaciones realizadas por los alumnos entre ambos grupos, promediando las

Tabla 2. Intervenciones realizadas al Programa del CTG

<ul style="list-style-type: none"> Las 2 evaluaciones sumadoras luego de las clases de "Digestivo alto/bajo" y de "Hepatología", que cada una de ellas, correspondían a 50% de la nota final, pasaron a corresponder a evaluaciones intermedias (10% de la nota final cada una) con intención formativa. Estas evaluaciones consistieron en pruebas de conocimiento en base a PSM, con viñetas basadas en casos clínicos, sin mayores modificaciones de 2006 a la fecha, ya que del 2002 al 2006 se logró una mejoría sistemática de las pruebas con 100% de cobertura de temas (validez de contenido), ausencia de violaciones de construcción de PSM y confiabilidad de 0,78 (Alfa de Cronbach)
<ul style="list-style-type: none"> Luego de cada evaluación intermedia se realizó una sesión de <i>feedback</i> respecto a lo contestado por los alumnos en las pruebas 1 y 2, en sesión con todo el curso, dirigido por el docente encargado del CTG
<ul style="list-style-type: none"> Se agregó una tercera evaluación sumadora de integración con una ponderación de 80%
<ul style="list-style-type: none"> Se contaba desde el año 2011 con autoinstructivos disponibles en el escritorio virtual del curso, consistentes en presentaciones sobre los temas tratados en clase. Estos contienen preguntas con formato EUNACOM y MIR, utilizadas de manera formativa para acercarse al nivel de dificultad y tipo de preguntas que encontrarían los alumnos en las evaluaciones, promoviendo aprendizaje independiente
<ul style="list-style-type: none"> Se programaron sesiones de integración (clases y seminarios) en la tercera semana del curso, posterior a las evaluaciones intermedias, a modo de repaso general para la evaluación sumadora final

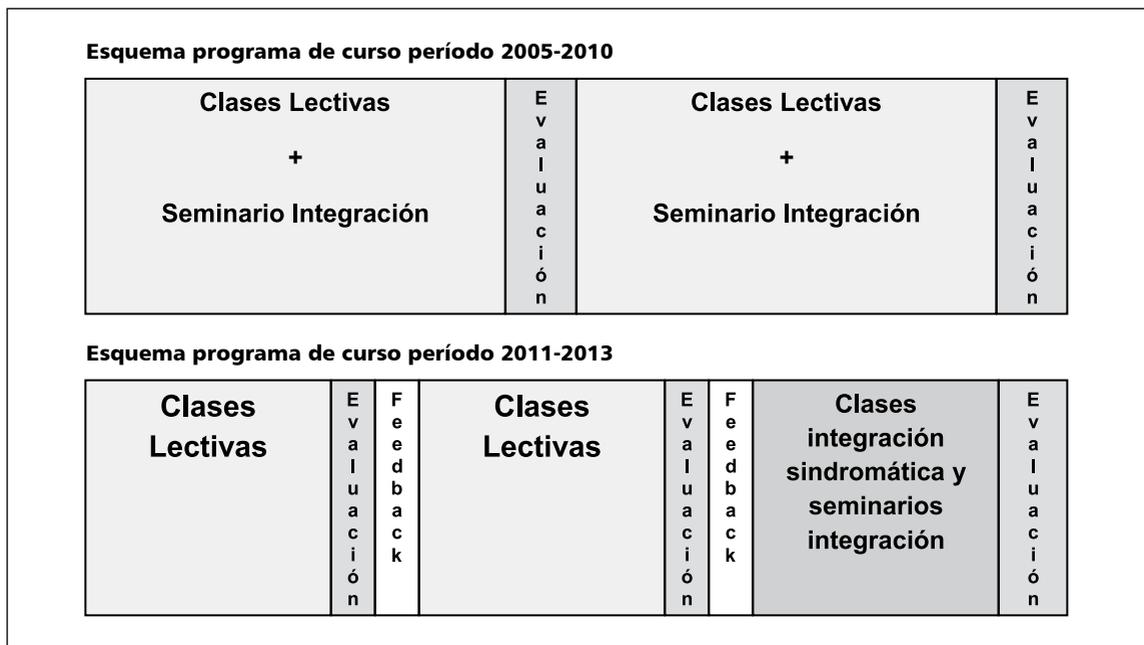


Figura 1. Esquemas de distribución de actividades en ambos programas. Período pre intervención curricular: 2005-2010. Período post intervención curricular: 2011-2013. *Feedback*: retroalimentación.

notas de las distintas áreas predeterminadas. Los aspectos cualitativos de la encuesta se basaron en la revisión de los comentarios libres.

Para comparar las calificaciones en ambos grupos, se promediaron las notas de cada alumno en cada año estipulado, obteniendo una nota promedio por año. Luego se promediaron las notas de los años 2005 a 2010, (pre-intervención curricular), y se compararon con el promedio de notas entre los años 2011 y 2013 (post-intervención).

Para descartar posibles efectos confundentes, se compararon las calificaciones obtenidas por los alumnos en otros cursos teóricos de Medicina Interna del CIC, que no tuvieron cambios curriculares (manteniendo el sistema de evaluación al final del módulo). Las cohortes analizadas fueron las correspondientes a los años 2005-2010 y 2011-2013.

Métodos estadísticos

Para comparar la distribución del tiempo en ambos programas se realizó una comparación gráfica basada en los porcentajes de tiempo asignado a cada actividad. Se utilizó Test Z para dos proporciones, con un intervalo de confianza de 95%.

Para comparar las evaluaciones de los alumnos se utilizó test de t de student para muestras independientes. Para ver las diferencias entre las calificaciones en ambos grupos se utilizó test de t de student. Se consideró una diferencia estadísticamente significativa un valor $p < 0,05$.

Resultados

Programa del Curso

En ambos programas, pre y post intervención curricular, la duración total del curso fue similar, y sin diferencias significativas (32,41 y 34,08 horas cronológicas respectivamente, cuatro tardes semanales durante 3 semanas calendario), por lo que la implementación de la evaluación *para* el aprendizaje no implicó un aumento del tiempo curricular. Ambos programas presentaron clases lectivas ($n = 26$ y $n = 24$), seminarios y clases de integración ($n = 3$ y $n = 5$) y evaluaciones sumadoras ($n = 2$ y $n = 3$). El programa post-intervención presentó además 2 sesiones de *feedback* posteriores a las pruebas 1 y 2. La distribución de los seminarios y clases de integración fue distinta, siendo en

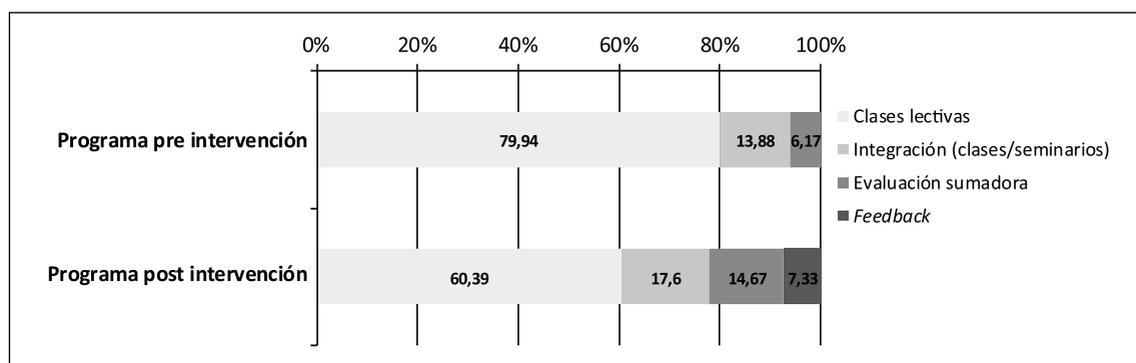


Figura 2. Expresión gráfica de la distribución porcentual de las actividades en ambos programas. Pre intervención: años 2005 a 2010. Post intervención 2011 a 2013.

Tabla 3. Calificaciones por áreas de la encuesta de percepción de los alumnos al finalizar el curso. Pre intervención: años 2005 a 2010. Post Intervención años 2011 a 2013

Parámetro	Pre intervención		Post intervención		Valor P
	n	Media ± DE	n	Media ± DE	
Dominio 1					
Tiempo	522	10,23 ± 5,68	279	11,81 ± 6,52	0,001
Dominio 2					
Objetivos	571	88,1 ± 9,05	298	91,5 ± 10,39	0,001
Dominio 3					
1. Métodos	591	6,27 ± 0,75	326	6,43 ± 0,81	0,004
Profesores	590	6,52 ± 0,67	325	6,63 ± 0,65	0,013
Fuentes de información	591	6,36 ± 0,90	325	6,66 ± 0,84	< 0,001
Feedback	350	5,57 ± 1,33	278	6,83 ± 0,45	< 0,001
Notas	583	5,71 ± 1,30	326	6,70 ± 0,66	< 0,001
Organización	590	6,49 ± 0,75	326	6,67 ± 0,68	< 0,001
Infraestructura y materiales	587	6,58 ± 0,63	326	6,76 ± 0,61	< 0,001
Nota global	581	6,18 ± 0,76	316	6,59 ± 0,63	< 0,001

el programa post intervención en la última semana del curso, previo a la evaluación sumadora final.

En ambos programas la mayoría del tiempo fue destinado a clases lectivas (79,94% y 60,39% del tiempo total, respectivamente). Luego ocupan lugar las actividades de integración (13,88% y 17,6%). En tercer lugar el tiempo se destina a evaluación (6,17 y 14,67%). En el programa post intervención se agregó la instancia de *feedback*, que ocupó 7,33% del total del tiempo.

Al comparar ambos programas, respecto al total del tiempo del curso en horas, en el programa post intervención se redujo 19,55% ($p < 0,001$) el tiempo destinado a clases lectivas, aumentó 3,72% el tiempo de integración ($p < 0,001$) y aumentó 8,5% el tiempo de evaluaciones sumadoras ($p < 0,001$). Esto se ve graficado en la Figura 2.

Encuesta de evaluación del Curso

En cuanto a las áreas evaluadas, hubo diferencias estadísticamente significativas en todas ellas, al comparar los períodos pre y post-intervención (2005-2010).

En el ítem "Tiempo", los alumnos en el programa post-intervención reportaron un mayor tiempo de dedicación al curso, a pesar de tener la misma duración en semanas. Las diferencias más importantes se observaron en las áreas donde los alumnos asignaban una nota en escala de 1 a 7, especialmente en las áreas de *Feedback*, Notas y Nota Global. En *Feedback* hubo un aumento de 1,26 puntos en promedio, en Notas aumentó la calificación en 0,99 puntos y en Nota Global hubo un aumento de 0,41 puntos. Las evaluaciones por dominios se presentan en la Tabla 3.

Tabla 4. Resumen de calificaciones finales de los alumnos en ambos períodos observados. Pre intervención: años 2005 a 2010. Post intervención: años 2011 a 2013

Parámetro	Pre intervención	Post intervención	Valor p
Número de alumnos	664	348	
Media (desviación estándar)	5,78 (0,57)	6,43 (0,38)	< 0,001
Mínimo-Máximo	3,8-7	4,4-7	

En cuanto a los comentarios en texto libre, en el período pre-intervención destacan comentarios como “considerar más tiempo para las evaluaciones”, “faltó revisión de las pruebas” “deberían dar feedback después de la primera prueba, comentar los casos que son buenos”, “me habría gustado que se trabajara más en seminarios”. En el período post-intervención se rescatan comentarios como: “el método en enseñanza con las 3 pruebas encuentro que es muy efectivo”, “fue buena la realización del feedback”, “muy buen método hacer 2 pruebas antes, la materia se asimila mucho mejor” “encuentro que el capítulo se encuentra muy bien estructurado, de esta manera se favorece el aprendizaje”.

Calificaciones finales de los alumnos

Durante los años 2005 a 2010, los alumnos del CTG fueron calificados 664 alumnos, a los que se les aplicaron 2 evaluaciones sumadoras con igual ponderación (50%), con un promedio de notas final de 5,78. Entre los años 2011 a 2013, fueron evaluados 348 alumnos, bajo el nuevo sistema de evaluación para el aprendizaje del CTG, aplicándose tres evaluaciones con ponderación diferenciada (10-10-80%) con un promedio final de 6,43. Esto implica un aumento de 0,65 puntos en escala de 1 a 7, siendo un cambio estadísticamente significativo ($p < 0,001$) (Tabla 4).

Al comparar las calificaciones de los otros cursos teóricos de Medicina Interna del CIC entre los años 2005-2010 vs 2011-2013, se aprecia que no existieron cambios estadísticamente significativos ($6,0 \pm 0,70$ vs $6,01 \pm 0,65$; $p = 0,56$).

Discusión

Los resultados del presente estudio son promisorios en cuanto al impacto educacional de la implementación del nuevo paradigma de evaluación para el aprendizaje. Los cambios curriculares introducidos en el CTG fueron bien recibidos por los alumnos en cuanto a su percepción y lo que es

más relevante, tuvieron un impacto directo en sus calificaciones finales.

El nuevo programa curricular destinó más tiempo a evaluación sin un incremento significativo en la duración del curso, acompañado además de una actividad nueva, el *feedback*, realizado en sesión general con todos los alumnos posterior a las evaluaciones intermedias, en el que se discutía con el docente jefe del curso las preguntas de las pruebas y se daba espacio para aclarar errores de concepto. El impacto educacional de esta actividad se basa en la entrega inmediata de *feedback*, permitiendo al alumno reconocer sus fortalezas y debilidades, para así poder dirigir su estudio independiente en base a sus necesidades individuales, distinguiendo los contenidos de mayor relevancia en el desarrollo de competencias vinculadas al manejo de pacientes con situaciones clínicas específicas. Además, se disminuyó significativamente el tiempo destinado a clases lectivas, aumentando el espacio destinado para clases y seminarios de integración en grupo pequeño, donde se discuten casos clínicos con apoyo audiovisual y los alumnos solucionan las dudas que tienen de las clases previas o de la bibliografía¹⁸.

En cuanto a la percepción de los alumnos, hubo mejoría en todos los ítems evaluados en el período post intervención, destacando las áreas de Notas, *Feedback* y Nota global. Es llamativo notar que los alumnos reportaron una mayor dedicación al curso en tiempo, siendo el período asignado al CTG igual en ambos programas, reflejando un mayor tiempo dedicado al aprendizaje independiente basado en autoinstructivos, en función de sus necesidades individuales identificadas en el proceso de revisión de las pruebas 1 y 2 (*feedback* inmediato, estructurado, sistemático)¹⁹. Las evaluaciones intermedias con posterior corrección fueron bien recibidas, ya que al tener una pequeña ponderación en la nota final, los alumnos se veían incentivados a estudiar. En el momento de hacer la tercera evaluación, de carácter integrativo, ya habían estudiado los con-

tenidos al menos una vez previamente. Además los alumnos lograron reforzar sus conocimientos durante la tercera semana del curso, dedicada para actividades de integración, siguiendo un modelo curricular iterativo²⁰. El ítem *Feedback* aumentó en más de 1 punto su calificación, lo que habla de la gran valoración que le otorgaron los alumnos, lo que incidió directamente en una mejor calificación en el ítem Nota global, que evalúa el curso de manera general²¹.

El impacto de la intervención curricular no fue sólo evidente en una mejor percepción del curso por parte de los alumnos, sino también en un cambio significativo en las calificaciones finales, aumentando las notas promedio del curso en 0,65 puntos (incremento de 9,28% absoluto), que no fue explicado por un cambio en las calificaciones de los otros cursos teóricos de Medicina Interna del CIC. Lo central de la intervención curricular fue introducir *feedback* y aumentar el tiempo destinado a evaluación (de 2 a 3 evaluaciones) reforzando la importancia de las evaluaciones en el desempeño de los alumnos²².

El CTG incluye la evaluación de los 2 primeros niveles de la pirámide de Miller (*Saber-Saber Cómo*)²³ explorando niveles cognitivos superiores en relación a resolución de escenarios clínicos¹³. El impacto educacional, dentro de las cuatro etapas del modelo de Kirkpatrick (*Reacción-Aprendizaje-Comportamiento-Resultados*)²⁴, correspondería a las etapas de *Reacción*, ya que se evaluó en qué grado les gustó el proceso de aprendizaje a los alumnos; *Aprendizaje*, pues se evidenciaron cambios en el rendimiento del alumno producto del proceso de aprendizaje, y *Comportamiento*, ya que los alumnos reportaron haber dedicado más tiempo al CTG post-intervención. Sería interesante evaluar la validez predictiva de los resultados de este cambio curricular en un seguimiento de estas cohortes a nivel nacional (ej: EUNACOM)²⁵ y finalmente el impacto en el cuidado del paciente con enfermedades gastrointestinales y hepatológicas. Sin embargo, reconocemos las limitaciones de este estudio y no contamos con acceso a dicha información. Otra limitación del presente estudio, es que el análisis se basa en la implementación de cambios en la estructura curricular, entrega de *feedback* y cambios en el sistema de evaluación en diferentes cohortes. Al comparar el rendimiento de las cohortes 2005-2010 con las cohortes de 2011-2013, observamos que el rendimiento sólo

presenta un cambio significativo en el CTG y no en los otros módulos teóricos del CIC, reforzando el impacto educacional del cambio a un sistema de evaluación *para* el aprendizaje. El diseño ideal para demostrar el impacto de esta modalidad sería a través de un grupo control y otro experimental distribuidos de manera aleatorizada. Sin embargo, este tipo de estudios son difíciles de implementar en intervenciones educacionales.

El *feedback* ha demostrado ser una herramienta útil en distintos escenarios de la educación médica. En el ámbito quirúrgico y en la realización de procedimientos se ha visto su beneficio en estudios clínicos aleatorizados^{26,27}. En la simulación como experiencia de aprendizaje es considerado un tópico relevante para lograr resultados efectivos²⁸. Hay experiencias de su utilidad en evaluaciones formativas. En un estudio reciente en un curso de ciencias básicas (fisiología) se observó una alta correlación del rendimiento entre la prueba final del curso y preguntas *online* que hacían los alumnos previo a la evaluación, con *feedback* inmediato e individualizado²⁹. Una de las virtudes del presente estudio es la demostración de las fortalezas y beneficios del *feedback* de las evaluaciones sumadoras en el dominio cognitivo (conocimientos).

Como complemento al *feedback* de evaluación, los alumnos podrían recibir información individualizada respecto al cumplimiento de los objetivos, esto permitiría al alumno dirigir su estudio focalizado en los aspectos por mejorar³⁰.

En conclusión, la evaluación para el aprendizaje ha resultado efectiva en lograr impacto en el aprendizaje de los alumnos a raíz de la experiencia en un acotado curso teórico de pregrado de la EMPUC. Es de esperar que a la luz de estos resultados la evaluación programática se expanda no sólo a otros cursos teóricos del IV año, sino también a cursos teóricos y prácticos de toda la EMPUC.

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

Referencias

1. Frank JR, Jabbour H. Report of the Can MEDS Phase IV Working groups. Ottawa, ON: Royal College of Physicians and Surgeons of Canada; 2005.
2. General Medical Council. Tomorrow's doctors: outco-

- mes and standards for undergraduate medical education. London: GMC; 2009. September. Disponible en www.gmc-uk.org/education [Consultado el 19 de enero de 2014].
3. Proyecto MECESUP: Desarrollo y evaluación de un perfil común del egresado de las escuelas de medicina de Chile. Disponible en: https://www.ucursos.cl/medicina/2007/0/INNCURRME/9/material_docente/previsualizar?id_material=138916 [Consultado el 19 de enero de 2014].
 4. Cumming A, Ross M. The Tuning Project for Medicine-learning outcomes for undergraduate medical education in Europe. *Medical Teacher* 2007; 29: 636-41.
 5. Sánchez I, Riquelme A, Moreno R, Mena B, Dagnino J, Grebe G. Revitalising medical education: the School of Medicine at the Pontificia Universidad Católica de Chile. *The Clinical Teacher* 2008; 5: 1-5.
 6. Riquelme A, Méndez B, de la Fuente P, Padilla O, Benaglio C, Sirhan M, et al. Desarrollo y validación de encuestas de percepción del portafolio en estudiantes de medicina de pregrado. *Rev Med Chile* 2011; 139: 45-53.
 7. Hart IR. (1992). Trends in clinical assessment, En: Harden, RM, Hart IR & Mulholland H. (Eds) International Conference proceedings: Approaches to the Assessment of Clinical Competence, Dundee, Centre for Medical Education.
 8. Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment: AMEE Guide N° 57. *Med Teach* 2011; 33 (10): 783-97.
 9. Martínez ME, Lipson JI. 1989. Assessment for learning. *Educ Leader* 47: 73-5.
 10. Shepard L. 2009. The role of assessment in a learning culture. *Educ Res* 29: 4-14
 11. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach* 2011; 33 (6): 478-85.
 12. Harden RM, Laidlaw JM. Be FAIR to students: four principles that lead to more effective learning. *Med Teach* 2013; 35 (1): 27-31.
 13. Kattan E, Pérez G, Le Roy C, Sirhan M, González A, Rybertt T, et al. Proceso de mejoría de pruebas de conocimiento con preguntas de selección múltiple en un curso teórico de Pregrado de Medicina. RECS 2014; Artículo aceptado para publicación. Volumen 12 (2).
 14. Angoff W. Scales, norms and equivalent scores. In: Thorndike, editor. *Educational measurements*. Second ed. Washington DC: American Council of Education; 1971. p. 508-600.
 15. Ebel RL. Estimation of the reliability of ratings. *Psychometrika* 1951; 16: 407-24.
 16. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334.
 17. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach* 2012; 34 (3): 205-14.
 18. Harden RM, Sowden S, Dunn WR. Educational strategies in curriculum development: the SPICES model. *Med Educ* 1984; 18 (4): 284-97.
 19. Embo MP, Driessen EW, Valcke M, Van der Vleuten CP. Assessment and feedback to facilitate self-directed learning in clinical practice of Midwifery students. *Med Teach* 2010; 32 (7): e263-9. doi: 10.3109/0142159X.2010.490281.
 20. Harden RM, Davies MH. AMEE Guide N° 5. The core curriculum with options or special modules. *Medical Teacher* 1995; 20: 317-22.
 21. Harrison CJ, Könings KD, Molyneux A, Schuwirth LW, Wass V, van der Vleuten CP. Web-based feedback after summative assessment: how do students engage? *Med Educ* 2013; 47 (7): 734-44. doi: 10.1111/medu.12209.
 22. Cilliers FJ, Schuwirth LW, Herman N, Adendorff HJ, van der Vleuten CP. A model of the pre-assessment learning effects of summative assessment in medical education. *Adv Health Sci Educ Theory Pract* 2012; 17 (1): 39-53.
 23. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65 (9 Suppl): S63-7.
 24. Kirkpatrick DL. *Evaluating teaching program*. San Francisco: Berret-Koehler Publishers; 1994.
 25. Bastías G, Villarroel L, Zúñiga D, Marshall G, Velasco N, Mena B. Academic performance of medical students: a predictable result? *Rev Med Chile* 2000; 128 (6): 671-8.
 26. Herrmann-Werner A, Nikendei C, Keifenheim K, Bosse HM, Lund F, Wagner R, et al. "Best Practice" Skills Lab Training vs a "see one, do one" Approach in Undergraduate Medical Education: An RCT on Students' Long-Term Ability to Perform Procedural Clinical Skills. *PLoS One* 2013; 8 (9): e76354.
 27. Farquharson AL, Cresswell AC, Beard JD, Chan P. Randomized trial of the effect of video feedback on the acquisition of surgical skills. *Br J Surg* 2013; 100 (11): 1448-53.
 28. Motola I, Devine LA, Chung HS, Sullivan JE, Issenberg SB. Simulation in healthcare education: a best evidence practical guide. AMEE Guide N° 82. *Med Teach* 2013; 35 (10): e1511-30.
 29. Marden NY, Ulman LG, Wilson FS, Velan GM. Online feedback assessments in physiology: effects on students' learning experiences and outcomes. *Adv Physiol Educ* 2013; 37 (2): 192-200.
 30. Burr SA, Brodier E, Wilkinson S. Delivery and use of individualised feedback in large class medical teaching. *BMC Med Educ* 2013; 13: 63.