

# ¿Realmente aprenden los estudiantes generando preguntas de test?

Arturo Jaime<sup>1</sup>, César Domínguez<sup>1</sup>, José Miguel Blanco<sup>2</sup>,  
Rosa Arruabarrena<sup>2</sup> y Patricia Jaime<sup>3</sup>

<sup>1</sup>Dto. de Matemáticas y Computación, Universidad de La Rioja, Logroño.

<sup>2</sup>Dto. de Lenguajes y Sistemas Informáticos, Universidad del País Vasco/Euskal Herriko

Unibertsitatea, Donostia-San Sebastián. <sup>3</sup>Psicología GeldiAdi, Donostia-San Sebastián

arturo.jaime@unirioja.es, josemiguel.blanco@ehu.eus,  
cesar.dominguez@unirioja.es, rosa.arruabarrena@ehu.eus,  
patricia@psicologiageldiadi.es

## Resumen

La generación de preguntas por parte de estudiantes es un método de aprendizaje cuya eficacia ha sido avalada por diferentes estudios. Sin embargo, los experimentos se suelen reducir a unas pocas horas. Creemos que todavía queda por aclarar su utilidad real y sus limitaciones, sobre todo si se desea extender su uso a un curso completo. El método admite múltiples variantes tanto en el tipo de pregunta como en el ciclo de generación que puede ser individual, colaborativo, con aportación o no del instructor, etc. En este trabajo presentamos un ciclo concreto de generación y tres formas de aplicarlo. Los resultados muestran mejoras en los aciertos de los test de examen y diferencias significativas en las calificaciones de los estudiantes que adelantan por partes el test del examen final. Los estudiantes tienen una opinión neutra del método y una opinión muy positiva tanto de adelantar el test como de disponer de formularios con preguntas para poder estudiar.

## Abstract

Several studies have supported the effectiveness of student-generated questions as a learning method. However, experiments are usually reduced to a few hours. We believe that its real usefulness and limitations need to be clarified, especially if it is to be extended to an entire course. The method admits multiple variations both in the type of question and in the generation cycle, which can be individual, collaborative, with or without instructor contribution, etc. In this paper we present a specific generation cycle and three ways of applying it. The findings show significant improvements in the results of the exam tests and significant differences in the scores of

students who anticipate the final exam test in parts. Students have a neutral opinion of the method and a very positive opinion of having the test ahead of time and also of having forms with questions to study.

## Palabras clave

Preguntas generadas por estudiantes, revisión por pares, aprendizaje colaborativo y entre iguales.

## 1. Introducción

Según algunos autores la estrategia educativa de las *preguntas generadas por estudiantes* (PGE) puede ser adecuada para desarrollar el pensamiento cognitivo y la comprensión y promover la autorregulación [1, 3]. En pocas palabras, el método PGE consiste en poner a los estudiantes a crear preguntas sobre los materiales de estudio tratando de distinguir lo importante de cuestiones secundarias [11, 14].

Aunque la mayoría de estudios sobre PGE se centra en la creación de preguntas de test [11, 14], se pueden considerar otros tipos de pregunta, como las de respuesta abierta [1] o el diseño de juegos sencillos con pistas [3]. La experiencia que se analizará en este trabajo se circunscribe a preguntas de test de cuatro opciones y una de ellas correcta.

Normalmente se busca que la calidad de las preguntas generadas permita que eventualmente formen parte del examen de la asignatura [7, 11, 12]. Para lograrlo, el estudiante debe entender bien el concepto por el que pregunta y que la redacción cumpla ciertas condiciones de calidad [8, 10], que suelen incluir cuatro aspectos. El primero es *preguntar por algo concreto* y suficientemente tratado en el material de estudio, donde se debería poder comprobar si cada respuesta es o no correcta.

El segundo aspecto es *evitar introducir pistas*. Lo más laborioso al crear preguntas de test es crear distractores (respuestas falsas) plausibles para quienes desconocen el material. Las respuestas no deben ser evidentes, hacer referencia a las demás respuestas, ser excluyentes, o que la verdadera tenga mayor longitud. Además, han de concordar gramaticalmente con el enunciado. En tercer lugar, la pregunta debe ser *fácil de entender* (clara), sin dobles negaciones, ni introducir trampas en la lectura. Hay que evitar imprecisiones (como “posiblemente”...) y extremos (como “todo”, “nunca”...). Por último, las preguntas deben ser *breves*. Para ello conviene usar un lenguaje directo, simple y evitar las repeticiones o introducir cuestiones irrelevantes.

Los estudiantes no suelen tener formación previa en creación de preguntas (como nos pasa a muchos profesores). Así que se sorprenden de la dificultad que conlleva hacer buenas preguntas [1, 12, 14]. Por tanto, es necesario dedicar tiempo y esfuerzo a formarles y a guiarles [3]. A pesar de ello, no es raro encontrar preguntas que incumplan requisitos de calidad o con interpretaciones erróneas de los materiales. Las preguntas con errores de concepto pueden afectar negativamente al aprendizaje [10]. Por ello, se plantea que los profesores las verifiquen e incluso el uso de sistemas de verificación automática [7, 10]. La verificación puede ser un trabajo arduo para el profesor, pero la tarea se puede limitar a una selección de preguntas hecha por los estudiantes. Esta actividad permite al profesor detectar errores de concepto y ambigüedades en los materiales de estudio. También ocurre que, al seguir este método, los profesores mejoran la calidad de sus preguntas, aumentando así la dificultad de sus exámenes, sobre todo al eliminar pistas de los distractores.

Con respecto a la calidad de las preguntas creadas por los estudiantes, encontramos en la literatura resultados contradictorios con criterios de calidad diversos. Tomemos por ejemplo dos estudios sobre la calidad de preguntas de opción múltiple. En el primero el 86% de las preguntas enviadas por estudiantes de primero de química eran de *alta calidad* [6] mientras que en el segundo se reducía al 14,5% [10], en este caso con estudiantes de primero de varias especialidades (química, bioquímica, estadística,...). Nuestra experiencia se acerca más al segundo caso, ya que recibimos un 17% de preguntas sin fallos de las preguntas enviadas para revisión, pero los estudiantes crean muchas más [4].

La mayoría de experiencias con PGE utiliza alguna herramienta informática que automatiza la gestión y publicación de las preguntas creadas. Quizá la más conocida sea *PeerWise* [6]. También se destaca el interés de usar herramientas que guíen a los

estudiantes en el proceso de PGE [2, 13, 14]. Pero hay que tener en cuenta que las herramientas suelen fijar un ciclo concreto de PGE, que admite múltiples alternativas de implementación. En nuestro caso usamos una herramienta propia [8].

En este trabajo se analiza nuestra experiencia aplicando PGE durante varios cursos académicos. Los objetivos son dilucidar hasta qué punto el método PGE es útil para el aprendizaje y qué opinión les merece este método a los estudiantes.

Tras una revisión bibliográfica de estudios que abordan estos objetivos (Sección 2) presentamos cómo se ha implementado el ciclo PGE en nuestra experiencia (Sección 3). Después se presentan los materiales y métodos usados (Sección 4), y se discuten los resultados obtenidos (Sección 5) para finalizar con algunas conclusiones (Sección 6).

## 2. Trabajos relacionados

En muchas experiencias de aplicación de PGE se observa una mejora clara del aprendizaje o de los resultados académicos [11, 3]. Por ejemplo, una revisión bibliográfica del uso de PGE en medicina [9] encontró dos trabajos donde los estudiantes que creaban preguntas de opción múltiple obtenían mejores resultados en sus exámenes que los que no las hacían. Además, a mayor número de preguntas creadas mejores resultados se obtenían, coincidiendo en esto con dos estudios en primero de química y cuarto de informática [6, 4]. Un estudio cuantitativo y cualitativo hizo un análisis pre-post con preguntas abiertas y de opción múltiple durante 10 semanas [13] con estudiantes de primaria (9-11 años). Observaron mejoras en ambos grupos respecto a su grupo de control (sin preguntas). La mayor significación se daba en el grupo con preguntas de opción múltiple. En otro experimento con estudiantes de 11 y 12 años, estos incluían preguntas en un juego sencillo que iban diseñando con el objetivo de aprender conceptos de informática [3]. Los resultados mostraron que los estudiantes que diseñaron los juegos aprendieron significativamente más que quienes sólo los usaron para aprender. Encontramos muchos otros estudios que observan mejoras en el aprendizaje usando preguntas de múltiples opciones con estudiantes de biología y otras titulaciones incluyendo ingeniería informática [5, 11, 12].

A pesar de lo apabullante de estos resultados, en una experiencia con estudiantes de ciencias de 22 años de media con preguntas cerradas (entendemos que de opciones múltiples) y abiertas, se pudo constatar la *ausencia de mejoras*, con la excepción de la mejora significativa en las respuestas a las preguntas de mayor dificultad [1]. En otro caso, con

estudiantes de licenciatura y máster de matemáticas, se observó que el PGE *no mejoraba los resultados académicos*, aunque sí aumentó el uso de los materiales de estudio [2]. Estos hallazgos siembran dudas sobre el método, sobre el tipo de asignatura al que se adapta y sobre cómo conviene aplicarlo.

Varios estudios recogen *opiniones de los estudiantes* sobre el método a través de encuestas o entrevistas. Una de las preguntas más relevantes es *qué utilidad tiene el PGE para el aprendizaje*. La revisión de PGE aplicado a medicina [9] encuentra tres trabajos con opiniones positivas al respecto, pero también encuentra un trabajo donde los estudiantes manifiestan que PGE no les ayudó ni les aportó valor para el aprendizaje. Otro estudio de PGE con estudiantes de física de varios grados de ingeniería creando preguntas de opción múltiple, recoge una opinión positiva mayoritaria (93%) sobre la utilidad del método [7]. Por otro lado, sólo el 31% de los estudiantes de ciencias que generaban preguntas cerradas y abiertas tenían una opinión positiva de la utilidad del PGE mientras que el 74% tenía una opinión positiva de la utilidad para el estudio del banco de preguntas conseguido [1].

Una cosa es la utilidad percibida para el aprendizaje y otra la *satisfacción creando, revisando y contestando preguntas*. Esta pregunta es más difícil de encontrar en los trabajos revisados. Hay algunas experiencias donde se pregunta explícitamente por ello. El 81% de los estudiantes de física que hicieron preguntas de opción múltiple afirmaron que estaban satisfechos con la experiencia [7]. En el trabajo sobre la creación de juegos con preguntas hablan de mejoras en la motivación y la confianza en lo aprendido en lugar de hablar sobre satisfacción [3].

Otra cuestión importante es la *dificultad que conlleva la creación de preguntas* a los estudiantes. En varios estudios que tratan este tema [2, 12] destacan que a los estudiantes les parece una actividad difícil y en uno de ellos, donde puntuaron la dificultad entre 1 y 5, se obtuvo una media de 4,5 [5].

Pocos trabajos analizan el comportamiento de los estudiantes en las actividades del PGE. En un estudio reciente detectamos desviaciones del trabajo de los estudiantes respecto al proceso por fases previsto y una tendencia inesperada hacia *comportamientos no colaborativos*. Además, los estudiantes más activos en las actividades de PGE (colaborando con terceros) conseguían crear más preguntas de calidad y mejores calificaciones en sus exámenes de test [4].

### 3. Ciclo de PGE utilizado

En esta sección se van a explicar las actividades que forman parte de nuestro ciclo de creación de

preguntas. Este ciclo ha sido diseñado teniendo en cuenta tres cuestiones. La primera es que haya una participación *individual* y otra en *equipo*. La segunda es que los estudiantes propongan *nuevas preguntas* y además *colaboren* con sus iguales aportándoles ideas. La tercera es que los *profesores* participen en la publicación de un conjunto reducido de preguntas.

En cada ciclo se trabajan uno o dos temas y se realizan las siguientes fases:

1. *Lectura del material*: tras exponer el material en clase, el profesor publica unos apuntes en formato textual (no son diapositivas) y pide a los estudiantes que contesten un test con algunas preguntas de los temas del ciclo.
2. *Generar pregunta*: se dividen los temas en partes y se asignan varios equipos de estudiantes a cada una. Se pide a cada estudiante que cree una pregunta de la parte del tema asignada. En el resto del ciclo, la actividad se restringe a ese trozo de tema. Para esta fase, y para las tres siguientes, se establece un plazo de realización.
3. *Seleccionar preguntas*: en esta actividad y la siguiente se trabaja con preguntas hechas por otras personas. Se pide a cada estudiante que vote cuáles son, a su parecer, las tres mejores preguntas de su parte. Cada estudiante puede dar un voto a su propia pregunta, a preguntas de miembros de su equipo o de otros equipos. Para que la votación surta efecto se debe aportar la respuesta correcta a la pregunta votada.
4. *Mejorar preguntas*: cada estudiante puede identificar y proponer mejoras a enunciados o a respuestas de preguntas de otras personas. También debe aportar la respuesta correcta.
5. *Reescribir preguntas*: esta fase se realiza en equipo. Se seleccionan las dos o tres mejores preguntas del equipo y se elaboran versiones definitivas, teniendo en cuenta las propuestas de mejora recibidas y posibles nuevas ideas.
6. *Revisar preguntas*: el profesor revisa las preguntas enviadas por los equipos. Dependiendo de la calidad de cada pregunta la puede aceptar, modificar o rechazar.
7. *Reutilizar preguntas*: El profesor se compromete a poner algunas preguntas en el test del examen.

Las preguntas creadas y sus datos asociados (votos obtenidos, propuestas de mejora...) son accesibles por los estudiantes. Cada curso realizamos seis de estos ciclos, donde los dos primeros sólo tocan alguno de los pasos, como selección o mejora, utilizando como base preguntas de cursos anteriores. En estos ciclos iniciales lo más importante es que los estudiantes aprendan a hacer buenas preguntas, para lo que creamos y publicamos hace tiempo un vídeo formativo de 10 minutos ([youtu.be/mU2OqtvwsJI](https://youtu.be/mU2OqtvwsJI)).

## 4. Materiales y método

Utilizaremos *datos extraídos de los exámenes finales* de los diez cursos comprendidos entre el 14-15 y el 23-24 en una asignatura de diseño de bases de datos de la Universidad de La Rioja. Hemos calculado el porcentaje de *aciertos en preguntas de test* y la nota obtenida en los *ejercicios de desarrollo* (entre ocho y nueve ejercicios relacionados con el temario). Los cursos donde los estudiantes generaron preguntas también se contabilizó, de forma separada, el número de preguntas de test acertadas en el examen que habían sido publicadas durante el curso. La evaluación de la asignatura consta del examen final (70%), de un proyecto de diseño de una base de datos (15%) y de las actividades de PGE (15%). El test tiene un peso de un 20% del examen final. En los cursos sin PGE el peso del proyecto era del 30%.

En los seis primeros cursos se presentaron al examen final 274 estudiantes. La primera prueba era un conjunto de preguntas de test de cuatro opciones. Durante el curso no había actividades diseñadas para crear o estudiar preguntas de test. Tampoco se publicaban preguntas de test de ejemplo. Sin embargo, si se publicaban algunos modelos de examen con ejercicios de desarrollo resueltos, que se mantuvieron con la introducción de PGE.

Los cuatro últimos cursos, se introdujo PGE como *actividad obligatoria*. Se hicieron tres variaciones de PGE a las que llamaremos de la siguiente forma:

1. *Sólo PGE*. El curso 20-21 se presentaron al examen 60 estudiantes que hicieron PGE en la mayor parte de los temas. Las preguntas creadas se publicaron en formato textual y sin marcar las respuestas correctas.
2. *PGE con adelanto*. El curso 21-22 se presentaron 55 estudiantes. Se extendió el método anterior a todos los temas. Se ofreció a los estudiantes la posibilidad de anticipar la realización del test del examen (y no el resto de preguntas) como forma de impulsar el estudio temprano de los conceptos de la asignatura y motivar el PGE. El test se dividió en tres partes que se hicieron la octava, duodécima y decimoquinta semanas. Los estudiantes que no se presentaron a alguna de estas partes las realizaron en el examen final. El conjunto de las partes del test tiene idéntico peso en el examen (20%) tanto si se adelanta alguna parte como si no.
3. *PGE con adelanto y formularios*. En los cursos 22-23 y 23-24, se presentaron 105 estudiantes. Además de la posibilidad de adelantar el test del examen final, las preguntas creadas se publicaron en formularios interactivos que mostraban

los errores cometidos y las respuestas correctas. El objetivo era facilitar e impulsar el estudio de conceptos teórico-prácticos.

Todos los cursos estudiados (con y sin PGE) *comparten en el test del examen final el mismo subconjunto de 28 preguntas* (las llamaremos *preguntas desconocidas*). Esto nos va a permitir comparar el aprendizaje logrado por unos y otros.

<i>Nº de preguntas PGE</i>	<i>20-21</i>	<i>21-22</i>	<i>22-23</i>	<i>23-24</i>
Creadas por estudiantes	169	334	264	263
Verificadas estud.+profe.	43+28	85+52	114+52	111+60
Conocidas examen	10	17	21	20
Desconocidas comun+otras	28+2	28+8	28+6	28+14
Número de estudiantes	61	62	61	61

Cuadro 1: Preguntas en los cursos con PGE.

Como parte del método seguido se incluyen en el examen final algunas preguntas creadas por los estudiantes durante el curso (las llamaremos *preguntas conocidas*). El porcentaje de aciertos en estas preguntas nos permite identificar qué uso se les da para el estudio. El cuadro 1 muestra el número de preguntas creadas en cada curso, cuántas verificó el profesor, cuántas estaban elaboradas por los profesores y de cuántas preguntas conocidas y desconocidas constaba el test del examen. Tanto los test adelantados como los del examen final tuvieron la misma configuración (número de preguntas conocidas y desconocidas).

Además de los exámenes, disponemos de los resultados de una encuesta de opinión sobre PGE, pasada a los estudiantes a partir del curso 21-22, que se presenta en la siguiente sección. La encuesta se realiza en las últimas sesiones presenciales del curso y antes del examen final (donde completarán su test los estudiantes que no adelantaron esa prueba).

## 5. Resultados y discusión

### 5.1. Preguntas y ejercicios de examen

En el cuadro 2 se comparan los resultados del test y los ejercicios de desarrollo entre los seis cursos sin PGE y los cuatro cursos con diferentes variantes de PGE. Se puede observar una *mejora significativa* en la respuesta al conjunto de preguntas común (a las que hemos llamado preguntas desconocidas). Esto viene acompañado de una *bajada significativa* en los ejercicios de desarrollo. De estos resultados podemos deducir que las actividades de PGE mejoran los test y, por lo tanto, este aprendizaje. Como contrapartida, los estudiantes parecen relajar la preparación de ejercicios de desarrollo. Al final de esta sección analizaremos con más detalle este último resultado.

	<i>Preguntas test desconocidas</i>	<i>Ejercicios desarrollo</i>
<i>Sin PGE</i>	44,35(13,03)	50,38(21,74)
<i>Con PGE</i>	50,29(14,91)	41,58(21,67)
<i>U de Mann-Whitney</i>	$Z=-4,475^{***}$	$Z=-4,386^{***}$

\*\*\*  $p < 0,001$

Cuadro 2: Medias (desviaciones) de los porcentajes de acierto/nota sobre 100 en preguntas de test y ejercicios de desarrollo, introduciendo, o no, PGE.

<i>Variante PGE</i>	<i>Preg. test conocidas</i>	<i>Preg. test desconocidas</i>	<i>Ejercicios desarrollo</i>
<i>Sólo PGE (P)</i>	52,33(22,05)	46,94(12,23)	42,33(24,60)
<i>Con adelanto (A)</i>	65,45(14,81)	53,94(16,38)	47,33(20,57)
<i>Con adel.+form. (F)</i>	70,87(21,06)	50,51(15,21)	38,51(20,25)
<i>H de Kruskal-Wallis</i>	$H=28,257^{***}$	$H=6,907^*$ $p=0,032$	$H=6,790^*$ $p=0,034$
<i>Tras Bonferroni</i>	$P < A, F$	$P < A$	$F < A$

\*\*\*  $p < 0,001$ , \* $p < 0,05$

Cuadro 3: Medias (desviaciones) de los porcentajes de acierto/nota sobre 100 en preguntas de test y desarrollo en las tres variantes de PGE.

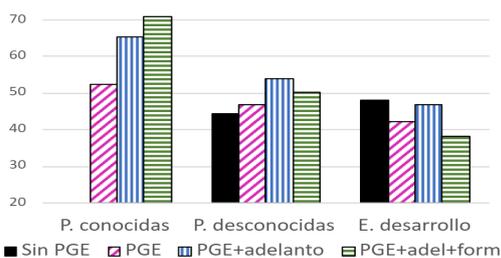


Figura 1: Porcentaje de aciertos/nota sobre 100 de preguntas de test y desarrollo sin PGE y con cada una de las tres variantes de PGE.

El cuadro 3 compara las tres variantes del método PGE explicadas en la sección anterior. Se observan diferencias significativas entre todas ellas para todos los tipos de pregunta. Tras la corrección de Bonferroni permanecen las mejoras en las *preguntas conocidas* cuando se da la posibilidad de adelantar el test (variantes A y F). En el grupo de *preguntas desconocidas* se mantienen las mejoras significativas en la versión con adelanto y sin formulario frente a no ofrecer adelanto. Las conocidas tienen medias superiores a las desconocidas, síntoma de que se memorizan. En los *ejercicios de desarrollo* se observa una caída significativa en los últimos cursos, donde se ofrecía formulario y adelanto del test. La figura 1 muestra gráficamente los datos incluidos en los cuadros 2 y 3. Considerando estos resultados los formularios estarían produciendo mejores resultados en preguntas conocidas con antelación, resultados

intermedios en el test de preguntas desconocidas pero peores resultados en los ejercicios de desarrollo. Nuevamente observamos el empeoramiento inesperado de los ejercicios de desarrollo.

<i>PGE SÓLO ADELANTO</i>	<i>Preg. test conocidas</i>	<i>Preg. test desconocidas</i>	<i>Ejercicios desarrollo</i>
<i>Adelantan (87%)</i>	66,18(14,59)	55,21(15,72)	47,33(20,64)
<i>Posponen (13%)</i>	60,50(16,54)	45,24(19,42)	42,98(21,21)
<i>U de Mann-Whitney</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>

#### *PGE ADELANTO Y FORMULARIOS*

<i>PGE ADELANTO Y FORMULARIOS</i>	<i>Preg. test conocidas</i>	<i>Preg. test desconocidas</i>	<i>Ejercicios desarrollo</i>
<i>Adelantan (70%)</i>	77,31(19,84)	53,75(14,91)	43,49(19,67)
<i>Posponen (30%)</i>	56,52(15,86)	42,50(12,83)	25,53(17,47)
<i>U de Mann-Whitney</i>	$U=663,50^{***}$	$U=466,00^{***}$	$U=612,50^{***}$

\*\*\*  $p < 0,001$

Cuadro 4: Medias (desviaciones) de los porcentajes de acierto/nota sobre 100 en preguntas de test y desarrollo con PGE con y sin adelanto del test.

La posibilidad de adelantar el test del examen final durante el curso, nos permite distinguir dos estilos de estudiante. Por una parte están los que adelantan el test, llevando la asignatura más *al día*. Por otra parte, están quienes deciden posponer esta prueba hasta el examen final. El cuadro 4 compara los resultados obtenidos por ambos estilos de estudiante en los tres tipos de pregunta y en las dos variantes de PGE que admiten adelantar el test. No observamos diferencias significativas entre los dos estilos de estudiante en la variante de PGE sólo con adelanto. Esto parece deberse a que pospuso el test un número insignificante de estudiantes (siete). Sólo se observa un tamaño del efecto medio (0,56) en las preguntas desconocidas. Sin embargo, en la variante de PGE que publicó las preguntas en formularios, los resultados logrados en todos los tipos de pregunta por los que adelantan el test son significativamente mejores. También se hizo el estudio separado del primer curso que se introdujeron formularios y no se encontraron diferencias significativas por posponer el test un número muy pequeño de estudiantes (cinco), aunque sí había un tamaño del efecto alto en preguntas conocidas (1,09) y ejercicios de desarrollo (0,88) y medio en las desconocidas (0,56).

En el cuadro 5 se comparan los resultados obtenidos en las dos variantes de PGE con posibilidad de adelanto, pero separando lo que ocurre en ambos estilos de estudiante. Para los estudiantes que *adelantan* el test no hay diferencias significativas en el conjunto común de preguntas desconocidas ni en los ejercicios de desarrollo. Sin embargo, sí se aprecian los efectos de los formularios interactivos en una *mejora significativa en las preguntas conocidas*. En los estudiantes que *posponen* el test no hay

diferencias significativas en las preguntas de test. Esto nos lleva a pensar que utilizan menos los formularios interactivos para estudiar. Ello podría deberse a que disponen de menos tiempo para preparar el examen. También obtienen resultados sustancialmente peores en los ejercicios de desarrollo si se publican formularios. Este es un resultado inesperado que tratamos al final de esta sección.

	<i>Preg. test conocidas</i>	<i>Preg. test desconocidas</i>	<i>Ejercicios desarrollo</i>
<b>ADELANTAN</b>			
<i>PGE sólo adelanto</i>	66,18(14,59)	55,21(15,72)	47,33(20,64)
<i>PGE adel. y formu.</i>	77,31(19,84)	53,75(14,91)	43,49(19,67)
<i>U de Mann-Whitney</i>	<i>U=983,000***</i>	<i>n.s.</i>	<i>n.s.</i>
<b>POSPONEN</b>			
<i>PGE sólo adelanto</i>	60,50(16,54)	45,24(19,42)	42,98(21,21)
<i>PGE adel. y formu.</i>	56,52(15,86)	42,50(12,83)	25,53(17,47)
<i>U de Mann-Whitney</i>	<i>n.s.</i>	<i>n.s.</i>	<i>U=52,000* p=0,022</i>

\*\*\* $p < 0,001$ , \* $p < 0,05$

Cuadro 5: Medias (desviaciones) de porcentajes de acierto/nota sobre 100 en preguntas de test y desarrollo con PGE con opción de adelanto separando los estudiantes que adelantan o no el test.

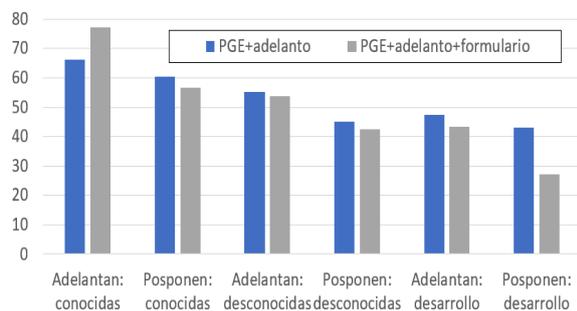


Figura 2: Porcentaje de acierto/nota sobre 100 distinguiendo quienes adelantan el test o no y si las preguntas se publican en formularios interactivos (gris) o de forma textual sin soluciones (azul).

En la figura 2 podemos observar de manera gráfica los resultados de los cuadros 3 y 4, con unos resultados similares en todas las parejas de columnas salvo en la primera y la última. El primer par de columnas muestra cómo los estudiantes que adelantan el test mejoran en las preguntas conocidas gracias al formulario interactivo y en la última se observa una bajada significativa en los ejercicios de desarrollo por parte de los estudiantes que posponen su test cuando disponen de formularios interactivos. También se observan diferencias significativas entre las parejas de columnas que corresponden a quienes adelantan y posponen el mismo concepto (las comparaciones entre columnas grises son significativas).

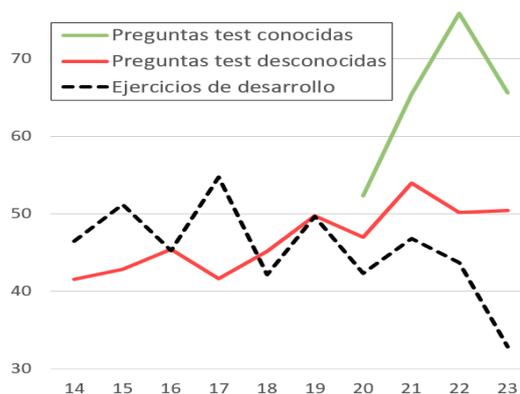


Figura 3: Notas medias durante los cursos en preguntas de test (conocidas y desconocidas) y ejercicios de desarrollo.

No encontramos explicaciones convincentes que relacionen el peor rendimiento en ejercicios de desarrollo y la publicación o no de formularios interactivos para estudiar el test. Entendemos que la importante caída en este concepto en estudiantes que posponen su test debe atribuirse a causas exógenas (coincidencia de exámenes, uso abusivo de IA en proyectos, alta inasistencia...). La Figura 3 apoya esta hipótesis ya que la severa caída en los ejercicios de desarrollo (línea discontinua) se produce el último curso. Sin embargo, no hay una caída semejante en el penúltimo curso, donde se aplicó PGE de forma idéntica. También se observa una caída en las preguntas conocidas (las publicadas en formularios), obteniéndose resultados similares a cuando se publicaban en texto sin marcar las correctas. Esto último es indicativo de que se han utilizado menos para el estudio. También hay que tener en cuenta que el porcentaje de estudiantes que pospusieron su test creció considerablemente el último curso, pasando de un 13% el curso 21-22 y un 9% el 22-23 al 52% el 23-24. Esto ha impactado de forma muy negativa en el resultado global.

Los resultados globales observados en el cuadro 2 parecen sembrar dudas sobre si PGE influye en la mejora de los test a costa de perjudicar, de forma significativa, los resultados en ejercicios de desarrollo. Volviendo a la figura 3, se observa que, en efecto, los tres primeros cursos con PGE se produce un descenso moderado en la línea discontinua (ejercicios de desarrollo) respecto a la etapa sin PGE, pero el enorme descenso localizado en el último curso ha influido drásticamente en esta estadística.

Sin embargo, este declive no afecta al test. Por un lado, los estudiantes que adelantan su test mejoran significativamente los resultados del mismo (preguntas desconocidas) en relación a cuando no

había PGE, esto es desde 44,35 sin PGE (cuadro 2) a 55,21 y 53,75 en las opciones con adelanto (cuadros 4 y 5). Por otro lado, quienes lo *posponen* obtienen resultados similares a los estudiantes sin PGE en sus dos versiones (sin y con formularios), con tamaños del efecto  $-0,05/bajo$  y  $0,14/bajo$ . Esta situación tiene otra lectura si tomamos a los ejercicios de desarrollo como punto de contraste de la calidad de los estudiantes. En tal caso, aunque los estudiantes del último curso logran peores resultados en este apartado, PGE les ha ayudado a mantener los estándares de cursos previos en las preguntas de test.

## 5.2. Opinión de los estudiantes

El cuadro 6 muestra un conjunto de preguntas contenidas en una encuesta pasada a los estudiantes (PGE con sólo adelanto y con adelanto y formulario) los últimos días del semestre. Todas las preguntas se valoran entre 1 y 5. En el momento de completarse la encuesta, los estudiantes que así lo hubieran decidido, han podido realizar las tres partes adelantadas del test del examen final. Las preguntas se presentan en el cuadro 6 organizadas en tres grupos. El primero trata sobre el *estudio realizado* por el estudiante. Observamos que las preguntas que más se utilizan para estudiar son las *creadas por estudiantes y verificadas* por el profesor. A estas le siguen las elaboradas por el *profesor* y, bastante por detrás, las elaboradas por estudiantes pero *sin verificar* por el profesor. La opinión sobre el esfuerzo que conlleva la creación de preguntas es medio-alto.

<i>Pregunta</i>	<i>Media (d.t.)</i>
<i>Estudio realizado con preguntas</i>	
1. Verificadas por el profesor	4,05(1,26)
2. Creadas por el profesor	3,74(1,19)
3. Esfuerzo hecho creando preguntas	3,41(0,99)
4. Sin verificar	2,15(1,04)
<i>Útil para aprender</i>	
5. Tener las preguntas en formularios	4,27(0,87)
6. Crear preguntas y ayudar a otros	2,95(1,17)
7. Crear preguntas en más asignaturas	2,57(1,24)
<i>Satisfacción</i>	
8. Poder adelantar el test del examen	4,66(0,74)
9. Evaluación con ejerc. de desarrollo	4,27(0,79)
10. Evaluación mediante test	3,82(1,08)
11. Creando preguntas de test	2,57(1,18)
12. Ayudando a otros con sus preguntas	2,31(1,08)

Cuadro 6: Opiniones sobre estudio/esfuerzo realizado, utilidad para el aprendizaje y satisfacción.

Respecto a la *utilidad para el aprendizaje*, destaca positivamente el disponer de las preguntas en *formularios* interactivos. Esta pregunta sólo la pudieron responder los estudiantes de los cursos donde se puso en marcha esta opción. La opinión

sobre la utilidad para el aprendizaje del PGE es neutra y la idea de extenderlo a otras asignaturas es moderadamente negativa.

El último grupo de preguntas trata sobre la *satisfacción*. Destaca de forma muy positiva la posibilidad de *adelantar el test* respecto al examen final. Los dos tipos de pregunta del examen final (test y ejercicios de desarrollo) tienen opiniones bastante positivas pero con preferencia por los *ejercicios de desarrollo*. Por último, la creación de preguntas de test obtiene opiniones moderadamente negativas, siendo más negativa la opinión sobre la *parte colaborativa* del PGE.

## 6. Conclusiones

La mayoría de estudios sobre PGE reducen esta actividad a unas pocas horas. En este trabajo hemos extendido el método a un curso completo, analizando su utilidad para el aprendizaje y la opinión que les merece a los estudiantes. El método seguido exige un esfuerzo de revisión a los profesores que implica, en muchos casos, hacer aportaciones a preguntas, sobre todo sustituyendo distractores evidentes. Con este método se consigue un conjunto amplio de preguntas de test de buena calidad que son útiles para el estudio.

Los resultados obtenidos muestran con claridad que PGE mejora sustancialmente los aciertos en los test, sobre todo si se permite adelantarlos respecto al examen final. Como contrapartida, se observa cierto retroceso en los resultados de ejercicios de desarrollo. Entendemos que ello se debe a que los estudiantes reparten su tiempo de estudio entre conceptos teórico-prácticos y ejercicios prácticos permitiendo un cierto equilibrio entre ambas cuestiones. Nuestra conclusión, por tanto, es que el método es válido, sobre todo haciendo test adelantados y publicando las preguntas con una herramienta interactiva.

Los estudiantes que optan por *adelantar* el test respecto al examen final obtienen mejores resultados y son los que mejor aprovechan la disponibilidad de formularios interactivos con preguntas de test. Los estudiantes que *posponen* el test al examen final consiguen peores resultados tanto en el test como en los ejercicios de desarrollo. El último curso se ha producido un aumento de este segundo tipo de estudiante acompañado de un más que notable descenso en sus calificaciones en ejercicios de desarrollo. Ello ha afectado a los resultados globales del curso. Sin embargo, los resultados en los test se han mantenido por encima de los que había sin PGE.

Respecto a las opiniones de los estudiantes sobre el método, estos prefieren estudiar con preguntas de compañeros verificadas por el profesor antes que con las del profesor y dicen usar mucho menos las de

compañeros sin verificar. También destacan la utilidad de los formularios interactivos con las preguntas generadas. Tienen una opinión neutra del método y moderadamente negativa de extenderlo a otras asignaturas. Por último, les ha gustado mucho poder adelantar el test del examen final durante el curso y les gusta más que se les evalúe con ejercicios de desarrollo en el examen final que con preguntas de test, aunque en ambos casos la opinión es positiva.

En el futuro queremos comprobar si se mantienen los resultados de los test usando preguntas hechas en cursos previos en lugar de crearlas. También nos gustaría estudiar si hay alguna relación entre la actividad en PGE y los resultados en preguntas de desarrollo. Por último, nos preocupa el posible cambio de tendencia de los estudiantes hacia la postergación del estudio, observado el último curso, por lo que analizaremos su evolución futura.

## Agradecimientos

Este trabajo ha sido subvencionado parcialmente por el Ministerio de Ciencia e Innovación MCIN/AEI/10.13039/501100011033 y la Unión Europea “NextGenerationEU”/PRTR, proyecto PDC2021-121128-I00 (ReCREA) y por el proyecto AFIANZA 2022/05 concedido por la Comunidad Autónoma de La Rioja.

## Referencias

- [1] Ester Aflalo. Students generating questions as a way of learning. *Active Learning in Higher Education*, 22(1), 63-75, 2021.
- [2] Sima Caspari-Sadeghi, Brigitte Forster-Heinlein, Jutta Maegdefrau y Lena Bachl. Student-generated Questions: Developing Mathematical Competence through Online-Assessment. *International Journal for the Scholarship of Teaching and Learning*, 15(1), 2021.
- [3] Yu-Ping Cheng, Chin-Feng Lai, Yun-Ting Chen, Wei-Sheng Wang, Yueh-Min Huang a y Ting-Ting Wu. Enhancing student’s computational thinking skills with student-generated questions strategy in a game-based learning platform. *Computers & Education*, 200, 2023.
- [4] César Domínguez, Arturo Jaime, Beatriz Pérez, Ángel Luis Rubio y María A. Zapata. Using Process Mining to Analyze Tasks Involvement and Collaboration in a Student Generated Questions Activity. Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning. Canadá. 2023.
- [5] Dale Hancock, Nicole Hare, Paul Denny y Gareth Denyer. Improving Large Class Performance and Engagement Through Student-Generated Question Banks. *Biochemistry and Molecular Biology Education*, 46(4), 2018.
- [6] Kyle Galloway y Simon Burns. Doing it for themselves: students creating a high quality peer-learning environment. *Chemistry Education Research and Practice*, 16, 2015.
- [7] Julia Gil y Luis Manuel Tobajas. An Educational Method Based on Student-Generated Questions. *International Journal of Educational Methodology*, 9(2), 2023.
- [8] Arturo Jaime, César Domínguez, José Miguel Blanco y Ana Sánchez. Preguntas generadas por estudiantes: una solución online para supervisar la aplicación sistemática del método. *La innovación como motor para la transformación de la enseñanza universitaria*. Universidad de La Rioja, 2022.
- [9] Jacob Lahti, Miriam Salamon, Jordan Farhat y Thomas Varkey. Multiple choice question writing and medical students: a systematic literature review. *MedEdPublish*, 13(34), 2023.
- [10] Steven Moore, Huy Nguyen, Tianying Chen y John Stamper. Assessing the Quality of Multiple-Choice Questions Using GPT-4 and Rule-Based Methods. *18th European Conference on Technology Enhanced Learning*, 2023.
- [11] Marcos Sanchez-Elez, Inmaculada Pardines, Pablo García, Guadalupe Miñana, Sara Román, Margarita Sánchez y José Luis. Risco. Enhancing Students’ Learning Process Through Self-Generated Tests. *Journal of Science Education and Technology*, 23, 2014.
- [12] Max Teplitski, Tracy Irani, Cory J. Krediet, Mariachiara Di Cesare y Massimiliano Marvasi. Student-Generated Pre-Exam Questions is an Effective Tool for Participatory Learning: A Case Study from Ecology of Waterborne Pathogens Course. *Journal of Food Science Education*, 17(3), 2018.
- [13] Longkai Wu, Yaohuizhuo Liu, Meng-Leong How y Sujin He. Investigating Student-Generated Questioning in a Technology-Enabled Elementary Science Classroom: A Case Study. *Education Sciences*, 13(158), 2023.
- [14] Fu-Yun Yu y Chun-Ping Wu. The effects of an online student-constructed test strategy on knowledge construction. *Computers & Education*, 94, 89–101, 2016.