

## **DALALearn: Una plataforma para la enseñanza y la investigación en programación**

Federico Alliani<sup>1</sup>, Lucas Raposeiras<sup>1</sup>, Gonzalo Pablo Fernández<sup>1,2</sup> y Christian Cossio-Mercado<sup>1,3</sup>

<sup>1</sup> Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Departamento de Computación. Buenos Aires, Argentina.

{fedealliani, raposeiras.lucas}@gmail.com,  
{gpfernandez, ccossio}@dc.uba.ar

<sup>2</sup> Universidad Nacional de Quilmes. Departamento de Ciencia y Tecnología. Bernal, Buenos Aires, Argentina.

<sup>3</sup> CONICET-Universidad de Buenos Aires. Instituto de Ciencias de la Computación (ICC). Buenos Aires, Argentina.

**Resumen** En los últimos años han surgido herramientas digitales que facilitan el aprendizaje interactivo y adaptativo, como es el caso de Duolingo, lo cual muestra que es posible promover la ejercitación y el seguimiento autónomo del progreso individual. Sin embargo, en la enseñanza de la programación aún se carece de herramientas de este tipo. Por otra parte, resulta muy complejo llevar adelante experimentos educativos rigurosos con un gran número de participantes, una dificultad que también se manifiesta en la investigación en educación en programación. Este tipo de estudios requiere evaluaciones en contextos reales, en los que las condiciones experimentales no interfieran con el desarrollo habitual de un curso. Además, los investigadores no disponen de plataformas digitales que permitan realizar un registro detallado del comportamiento estudiantil a lo largo de un curso completo. En este contexto surge DALALearn, una plataforma concebida como complemento al trabajo en el aula, que facilita la enseñanza y la práctica de la programación a través de una aplicación móvil, y que además ofrece a docentes e investigadores un entorno para el diseño de experimentos educativos, con métricas detalladas de uso y rendimiento. La propuesta busca complementar cursos existentes mediante lecciones breves, ejercitación y evaluaciones, en una herramienta moderna orientada a la recolección de datos empíricos. La plataforma fue evaluada a nivel general para poner a prueba su usabilidad y también fue utilizada para presentar contenidos de dos versiones de un curso introductorio de programación. De este modo, fue posible emplear DALALearn para evaluar una hipótesis relacionada con el uso, o no, de actividades de metacognición y su impacto en el aprendizaje.

**Palabras claves:** Plataforma educativa; Enseñanza de programación; Investigación educativa; Experimentación educativa

Received May 2026; Accepted June 2026; Published July 2026



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

## ***DALALearn*: A platform for teaching and research in programming**

**Abstract.** In recent years, digital tools that support interactive and adaptive learning, such as Duolingo, have emerged showing that it is possible to promote practice and the autonomous monitoring of individual progress. However, similar tools are still largely absent in the teaching of programming. At the same time, conducting rigorous educational experiments with large numbers of participants is highly challenging, a difficulty that also affects research in computer science education. Such studies require evaluation in real educational settings, where experimental conditions do not interfere with the regular development of a course. Moreover, researchers lack digital platforms that enable detailed tracking of student behavior throughout a course. In this context, DALALearn was developed as a platform that accompanies classroom instruction and facilitates the teaching and practice of programming through a mobile app, while also providing teachers and researchers with an environment for designing educational experiments supported by fine-grained usage and performance metrics for educational research. The platform is intended to complement existing courses by offering short lessons, practice, and assessments within a modern tool centered on empirical data collection.

The platform was evaluated at a general level to assess its usability and was also used to deliver content from two versions of an introductory programming course. This allowed DALALearn to be employed to test a hypothesis regarding the use of metacognitive activities and their impact on learning outcomes.

**Keywords:** Educational platform; Programming education; Educational research; Educational experimentation

## 1 Introducción

La educación en programación se ha convertido en una competencia fundamental en los últimos años, siendo considerada como una habilidad esencial para el desarrollo profesional y personal (Nambiar, 2020). Sin embargo, presenta desafíos únicos que requieren enfoques pedagógicos especializados y herramientas tecnológicas que faciliten tanto la práctica como la investigación educativa.

En el contexto educativo actual, existe una creciente necesidad de herramientas digitales interactivas para ejercitar y practicar contenidos complejos, como es el caso de la programación. Plataformas como Duolingo<sup>1</sup> han demostrado el potencial de las herramientas educativas interactivas para mejorar la motivación, favorecer la autonomía del estudiante y facilitar el seguimiento del progreso (Criollo-C et al., 2021; Shortt et al., 2021). En el contexto específico de la programación, plataformas como Mimo<sup>2</sup> han mostrado efectividad al combinar técnicas de visualización, resolución de problemas y aprendizaje basado en juegos (Amro y Romli, 2019). Asimismo, aplicaciones gamificadas como Moravec<sup>3</sup>, desarrollada en Argentina, han demostrado el potencial de las herramientas móviles para recopilar datos masivos y motivar la práctica extensiva mediante diseño gamificado (Zimmerman et al., 2016).

Un gran desafío en la investigación educativa es realizar experimentos con muchos participantes sin interferir en las clases. Los estudios deben llevarse a cabo en entornos reales, lo cual es difícil, especialmente al comparar enfoques didácticos. Las plataformas actuales no ofrecen las herramientas necesarias para controlar variables o monitorear el comportamiento estudiantil. En particular, en la enseñanza de la programación aún faltan plataformas que combinen esta interactividad con la posibilidad de realizar investigaciones educativas rigurosas y a gran escala.

De acuerdo con lo anterior, planteamos el desarrollo de DALALearn, una plataforma diseñada especialmente para beneficiar a estudiantes y docentes de programación e investigadores en educación.

### 1.1 Estado del Arte en Plataformas Educativas para Programación

Para contextualizar la propuesta de DALALearn, es necesario analizar las plataformas existentes según los tres objetivos planteados: facilitar la ejercitación, complementar la enseñanza y permitir la investigación educativa.

- **Aplicaciones Móviles de Práctica.** Diversas revisiones sistemáticas han identificado una relación positiva entre el uso de aplicaciones móviles y la mejora en habilidades de programación, destacando el potencial de estrategias de gamificación para aumentar el *engagement* y la motivación estudiantil (Alfarsi et al., 2024). Aplicaciones como Mimo y Codecademy ejemplifican esta tendencia, ofreciendo ejercicios con *feedback* inmediato y rutas de aprendizaje personalizadas. Sin embargo, estas plataformas están orientadas exclusivamente al estudiante individual, ya que no ofrecen herramientas para que los docentes gestionen grupos ni permiten

<sup>1</sup> Duolingo - Sitio web: <https://www.duolingo.com/>

<sup>2</sup> Mimo - Sitio web: <https://mimo.org/>

<sup>3</sup> Moravec - El Gato y la Caja: <https://elgatoylajaja.com/notas/moravec>

la exportación de datos para análisis e investigación educativa. En 2023, la discontinuación de Grasshopper<sup>4</sup>, una aplicación desarrollada por Google para enseñar fundamentos de programación en JavaScript, ilustró la fragilidad inherente de depender de soluciones comerciales para iniciativas educativas a largo plazo.

- **Sistemas de Gestión del Aprendizaje.** Moodle es el sistema de gestión de aprendizaje (en adelante, LMS, del inglés Learning Management System) de código abierto más utilizado, con capacidades de seguimiento del progreso y analíticas de aprendizaje. Sin embargo, se encontró que Moodle se utiliza principalmente para exámenes calificados con preguntas de opción múltiple, dejando de lado su potencial para actividades de aprendizaje más diversas (Walpita Gamage et al., 2022). En el mismo trabajo se indica que apenas el 5 % de los estudios aplicaron teorías pedagógicas en el diseño de sus cursos, y la dificultad de configuración inicial sigue siendo un obstáculo importante para docentes sin conocimientos técnicos.
- **Plataformas de Desafíos de Programación.** LeetCode<sup>5</sup> y HackerRank<sup>6</sup> ofrecen miles de problemas de programación con corrección automática y *rankings*. No obstante, están orientadas principalmente a la preparación para entrevistas laborales en la industria del software, con contenidos descontextualizados del currículum académico, y no ofrecen herramientas para que los docentes gestionen grupos de estudiantes, ni dan acceso a datos para investigación educativa externa.
- **Herramientas de Evaluación Automática.** En una revisión sistemática (Messer et al., 2024) de 121 herramientas de corrección automática, se encontró que el 66 % de las herramientas sólo evalúa si el código es correcto, con *feedback* limitado a indicar si las pruebas pasaron o fallaron, y sin orientación sobre cómo corregir los errores. Además, el 84 % carece de conjunto de datos disponibles de manera pública, lo que dificulta la reproducibilidad y el avance de la investigación educativa. Estas limitaciones reflejan un enfoque centrado en la corrección automática más que en el aprendizaje, y evidencian la falta de herramientas que combinen retroalimentación pedagógica con capacidades de recolección de datos para investigación.

En resumen, el análisis realizado revela una fragmentación clara del ecosistema. Por un lado, las aplicaciones móviles priorizan la experiencia del estudiante, pero ignoran las necesidades docentes e investigativas, mientras que los LMS como Moodle ofrecen gestión académica, pero carecen de especificidad para programación. Además, las plataformas de desafíos, como es el caso de HackerRank, se orientan a la industria sin considerar el contexto educativo, mientras que las herramientas de corrección automática no constituyen plataformas integrales.

Así, ninguna alternativa existente integra simultáneamente una experiencia de aprendizaje móvil interactiva, herramientas de gestión para docentes y capacidades de experimentación controlada para investigadores, lo que motiva el desarrollo de DALALearn.

---

<sup>4</sup> Noticia de la baja del servicio por parte de Google (en inglés): [chromeunboxed.com/google-shutting-down-grasshopper-coding-app](https://chromeunboxed.com/google-shutting-down-grasshopper-coding-app)

<sup>5</sup> LeetCode - Sitio web: <https://leetcode.com/>

<sup>6</sup> HackerRank - Sitio web: <https://hackerrank.com>

## 1.2 Problema de Investigación

La falta de herramientas específicas para la investigación educativa en programación limita la capacidad de los educadores e investigadores en las siguientes tareas:

- **Realizar experimentos controlados:** En entornos educativos tradicionales, múltiples variables simultáneas (características del estudiante, estilo del docente, condiciones del aula, factores temporales) interfieren con la capacidad de aislar el efecto de metodologías pedagógicas específicas. Esta falta de control experimental impide establecer relaciones causales claras y realizar estudios rigurosos que permitan validar científicamente la efectividad de diferentes enfoques didácticos.
- **Recopilar datos empíricos:** Si bien existen plataformas, como Moodle, que registran distintos tipos de eventos, se trata de sistemas de propósito general, cuya complejidad dificulta su uso para experimentación educativa. Se requieren herramientas más simples y específicas para complementar cursos y realizar experimentación haciendo un seguimiento detallado del comportamiento estudiantil.
- **Realizar investigaciones a escala:** La complejidad logística dificulta coordinar experimentos con múltiples participantes en diferentes contextos educativos.
- **Evaluar efectividad pedagógica:** La falta de métricas detalladas sobre el proceso de aprendizaje dificulta la evaluación de diferentes enfoques didácticos.

## 1.3 Objetivo y propuesta del trabajo

El objetivo principal de este trabajo es **desarrollar y evaluar** una plataforma educativa que facilite la ejercitación y práctica de programación. Así, se busca que complemente las acciones de enseñanza y permita la realización de investigaciones educativas rigurosas, proporcionando herramientas para la recolección de datos empíricos y la experimentación pedagógica.

Así, presentamos DALALearn, una plataforma educativa diseñada para servir simultáneamente a tres grupos de usuarios: estudiantes que buscan practicar programación de forma autónoma con *feedback* inmediato, docentes que necesitan herramientas para complementar sus cursos y hacer seguimiento del progreso estudiantil, e investigadores que requieren entornos controlados para realizar experimentos educativos y validar hipótesis pedagógicas.

DALALearn combina la interactividad de aplicaciones educativas modernas con capacidades avanzadas de recolección de datos, permitiendo tanto la ejercitación efectiva como la investigación educativa rigurosa en una misma plataforma.

En la sección 2 se desarrolla el diseño general de la plataforma, además de incluir detalles sobre el comportamiento de la app y la versión web de la misma. Posteriormente, en la sección 3 se presentan el diseño de la prueba de usabilidad y los resultados de la misma, mientras que en la sección 4 se detalla la prueba integral que se hizo para evaluar una hipótesis de investigación en educación por medio de DALALearn. Por último, en la sección 5 se resumen las principales conclusiones del trabajo, y en la sección 6 se detallan los próximos trabajos que pueden realizarse.

## 2 La plataforma DALALearn

En esta sección se describe la arquitectura general y el funcionamiento de DALALearn, explicando cómo están estructurados los componentes, cómo se relacionan entre sí, y cómo operan tanto desde la perspectiva del estudiante (aplicación móvil) como desde la de los educadores e investigadores (panel web).

### 2.1 Componentes de DALALearn

La unidad principal de organización en DALALearn es un **curso**. Cada curso está compuesto por múltiples módulos y representa un conjunto completo de ejercicios y contenidos sobre un tema específico de programación. Los cursos pueden encontrarse en dos estados: **publicado** (disponible para los estudiantes) o **borrador** (visible solo para educadores e investigadores). Cada curso incluye una descripción que permite a los usuarios entender su contenido y objetivos antes de inscribirse. En la figura 1 se muestra un esquema general con la estructura de un curso.



Figura 1: Esquema general de un curso

Un **módulo** dentro de un curso es una unidad temática compuesta por una lista ordenada de actividades. Los módulos pueden tener cualquier cantidad de actividades y de cualquier tipo, y con un orden definido por el educador o investigador. Los módulos se relacionan entre sí mediante un sistema de dependencias que forma un grafo dirigido, donde la completitud de un módulo puede desbloquear el acceso a otros módulos dependientes. Siempre debe existir al menos un módulo inicial que no tenga dependencias, permitiendo que los usuarios puedan comenzar el curso.

Una **actividad** es la unidad más pequeña de contenido educativo. Las actividades dentro de un módulo mantienen un orden fijo definido por el creador del curso, que es el mismo para todos los usuarios. Los usuarios navegan secuencialmente por las actividades de un módulo mediante botones de avanzar y retroceder, pero deben completar la actividad actual antes de poder avanzar.

Existen tres tipos de actividades:

- Las **lecciones** presentan contenido teórico en formato *markdown*<sup>7</sup> extendido. Además de la sintaxis estándar de markdown (texto formateado, imágenes y enlaces externos), DALALearn agrega soporte para videos de YouTube integrados directamente en el contenido de las lecciones.
- Los **ejercicios** son actividades interactivas donde los usuarios deben resolver problemas o completar tareas. Los enunciados de los ejercicios soportan formato *markdown* para incluir texto formateado, imágenes y otros elementos. DALALearn soporta actualmente dos tipos de ejercicios: *fill the gaps*, en los que los usuarios deben completar fragmentos de código arrastrando bloques de código a los espacios correspondientes), y *multiple choice*, en los que los usuarios responden seleccionando una o múltiples opciones de una lista.
- Las **evaluaciones** consisten en una lista ordenada de ejercicios que permiten evaluar el aprendizaje del usuario. Cada evaluación permite definir un sistema de puntuación, un conjunto de ejercicios obligatorios y un criterio de aprobación.

## 2.2 Experiencia de uso de los estudiantes

Al ingresar a la aplicación móvil, los usuarios deben registrarse o iniciar sesión utilizando correo electrónico y contraseña (ver figura 2a). Una vez autenticados, acceden a un listado de cursos disponibles (i.e., cursos en estado *publicado*). Para cada curso, el sistema muestra si el usuario está inscripto, o no, y si ya lo completó, como se aprecia en la figura 2b.

Al seleccionar un curso, el usuario puede ver su descripción, su lista de módulos y un botón para inscribirse al mismo (si aún no está inscripto). Si el usuario está inscripto, los módulos aparecen bloqueados o desbloqueados según su progreso. Como se muestra en la figura 2c, los módulos bloqueados muestran claramente qué módulos deben completarse primero para desbloquearlos.

Cuando un usuario ingresa a un módulo accede a la primera actividad sin completar (si es la primera vez que accede, será la primera actividad del módulo). En caso de haber completado todas las actividades del módulo, puede ingresar y ver todo desde el comienzo. El usuario navega secuencialmente por las actividades del módulo usando botones de avanzar y retroceder, pero debe completar la actividad actual antes de poder avanzar. El orden de las actividades es fijo. Finalmente, cuando un usuario completa un módulo, el sistema verifica las dependencias y desbloquea automáticamente los módulos que dependían de su completitud.

<sup>7</sup> Markdown es un lenguaje de marcado ligero creado por J. Gruber: <https://daringfireball.net/projects/markdown/>

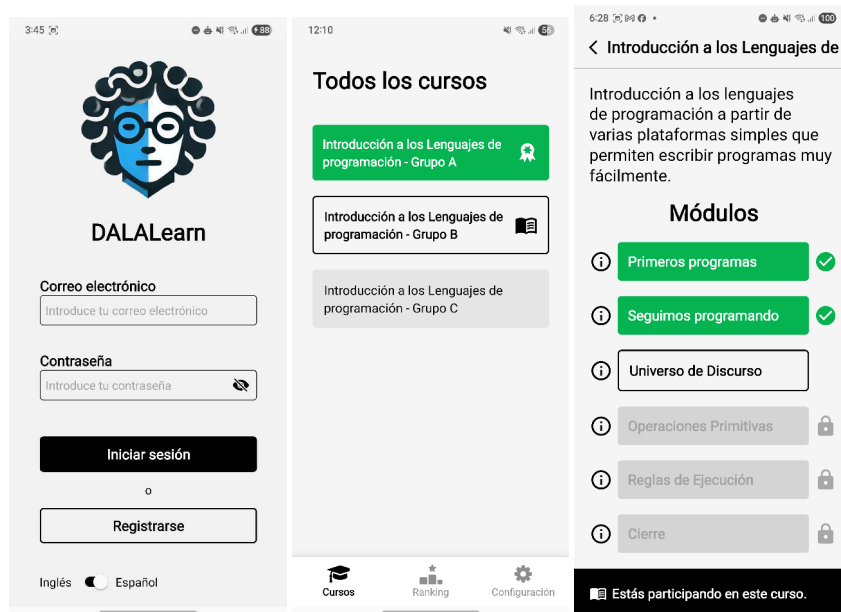


Figura 2: App DALALearn: (a) pantalla de login, (b) listado de cursos, (c) progreso en un curso.

### 2.3 Experiencia de uso de docentes e investigadores

El panel web proporciona herramientas completas para la gestión de cursos, módulos y actividades, así como para el análisis de datos de usuarios. En la página principal se pueden crear nuevos cursos (ya sea importando un archivo JSON o usando la interfaz) o editar los existentes (figura 3).

Dentro de la página de un curso se puede modificar toda su información (la descripción, el estado y su conjunto de módulos), así como ver y exportar datos y estadísticas del mismo (cantidad de usuarios registrados, estadísticas de uso y eventos generados por los usuarios al interactuar con los módulos y las actividades). En la página de un módulo se puede modificar toda su información (la descripción, el estado y su secuencia de actividades), así como ver y exportar estadísticas de uso del mismo.

En cuanto a las actividades, desde la página de una lección se puede modificar su contenido a la vez que se visualiza el resultado renderizado (figura 4). Al ver el detalle de un ejercicio de tipo *fill the gaps* se pueden modificar el enunciado, el texto que hay que completar y las opciones que se ofrecen para cada hueco, así como el puntaje asociado a cada una (figura 5). Por otro lado, para un ejercicio de tipo *multiple choice* se pueden modificar el enunciado y las opciones que se ofrecen, así como el puntaje asociado a cada una y si se permite o no seleccionar más de una opción (figura 6). Finalmente, para la evaluación se pueden modificar el listado de ejercicios, el puntaje relativo de cada ejercicio y el criterio de aprobación, definido a partir del puntaje míni-

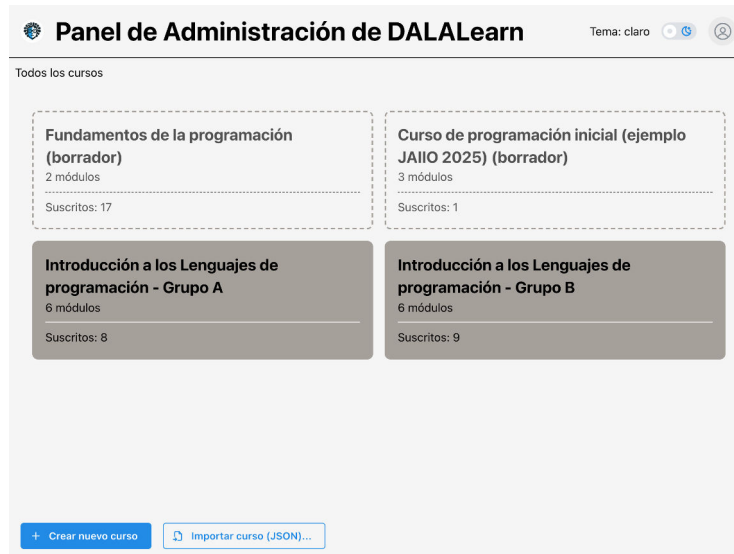


Figura 3: Gestión de cursos en el panel web

mo a obtener y los ejercicios que deben estar bien, necesariamente, para aprobar (figura 7).

#### 2.4 Características de la Plataforma DALALearn

**Recolección de datos** Una característica fundamental de DALALearn es su capacidad de seguimiento del progreso de los estudiantes. Esta funcionalidad permite registrar información detallada sobre el comportamiento y avance de los usuarios en la plataforma, proporcionando datos valiosos que pueden utilizarse para diferentes propósitos según el perfil del usuario. Para docentes, estos datos facilitan el seguimiento del estado del curso y el progreso individual de cada estudiante, permitiendo identificar quién necesita apoyo adicional y cómo avanzan los participantes a lo largo del contenido. Para investigadores, esta información permite realizar experimentos controlados y análisis empírico de metodologías pedagógicas, facilitando la validación de hipótesis educativas en entornos reales de aprendizaje. Los datos recolectados por la plataforma se descargan en formato CSV desde el panel de administración para su posterior análisis estadístico y visualización en herramientas especializadas.

DALALearn implementa un sistema de métricas de progreso educativo que opera en tres niveles jerárquicos: curso, módulo y actividad. A nivel de curso, la plataforma registra información completa sobre cada usuario incluyendo si completó el curso en su totalidad, la cantidad de módulos terminados, el puntaje acumulado a lo largo de todo el curso, la fecha de inscripción inicial y la fecha de finalización del curso. Estas métricas permiten tanto a docentes como a investigadores evaluar la efectividad general del curso, identificar patrones de abandono y analizar los tiempos de completación según diferentes variables demográficas o pedagógicas.

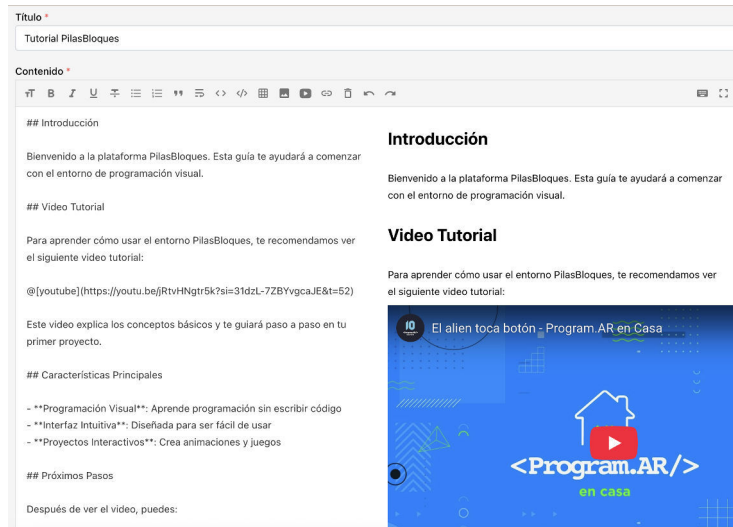


Figura 4: Edición de una actividad de tipo lección en el panel web

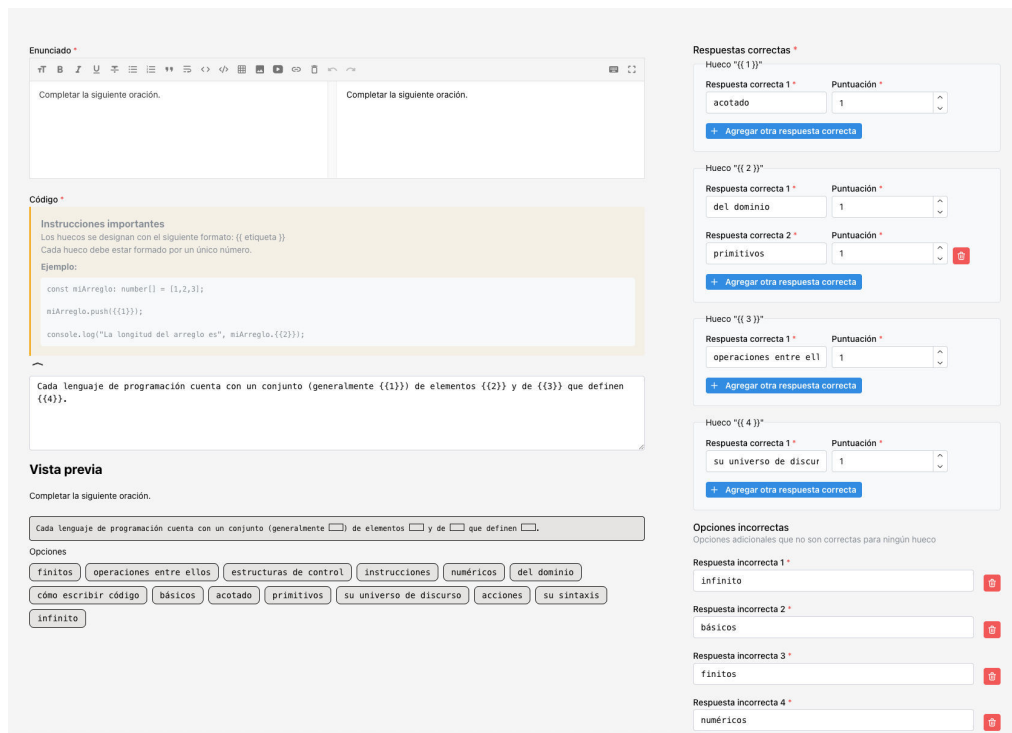


Figura 5: Edición de una actividad de ejercitación *fill the gaps*

**Título \***  
Evaluá tu propio aprendizaje (1)

**Enunciado \***

Si tuvieras que calificar del 0 al 10 tu aprendizaje sobre los conceptos que aquí trabajamos (\_universo de discurso\_, \_elementos primitivos\_, \_elementos de dominio\_), ¿qué puntaje te darías?

Si tuvieras que calificar del 0 al 10 tu aprendizaje sobre los conceptos que aquí trabajamos (*universo de discurso, elementos primitivos, elementos de dominio*), ¿qué puntaje te darías?

**Opciones**

**Importante**  
Las opciones se mostrarán al usuario en orden aleatorio.

Opción 1

Contenido \* Puntuación \*

0 1

Esta opción es correcta

Opción 2

Contenido \* Puntuación \*

Figura 6: Edición de una actividad de ejercitación *multiple choice*

**Título \***  
Evaluación

Porcentaje mínimo para aprobar (0-100) \*

60%

**Ejercicios \***

**Ejercicio 1** (MULTIPLE CHOICE)

Puntuación máxima

1

Requerido para aprobar la evaluación

¿Qué pasa en Code Slime si se intenta mover al slime hacia una posición sobre la que se encuentra un obstáculo?

Seleccione la opción que considere correcta

- El slime choca contra el obstáculo y continúa la ejecución de la siguiente instrucción.
- El slime atraviesa el obstáculo y continúa la ejecución de la siguiente instrucción.
- La instrucción se ignora y finaliza la ejecución.
- El slime choca contra el obstáculo y finaliza la ejecución.
- La instrucción se ignora y continúa la ejecución de la siguiente instrucción.
- El slime se queda en lugar.

[Editar ejercicio](#) [Duplicar ejercicio](#) [Eliminar ejercicio](#)

Figura 7: Edición de una actividad de tipo evaluación

A nivel de módulo, el sistema recolecta datos específicos sobre el progreso de cada usuario dentro de cada unidad temática del curso. Esto incluye el estado de finalización del módulo, la cantidad de actividades completadas dentro del módulo, el puntaje acumulado en todas las actividades del módulo, y las fechas de inicio y finalización del trabajo en ese módulo específico. Esta granularidad permite identificar qué módulos presentan mayor dificultad, cuáles generan más *engagement*, y cómo varía el ritmo de aprendizaje entre diferentes temas o conceptos de programación, información útil tanto para el seguimiento docente como para análisis de investigación.

A nivel de actividad individual, DALALearn registra el desempeño detallado de cada usuario en cada ejercicio, lección o evaluación. Los datos incluyen si el usuario completó la actividad, el puntaje obtenido en esa actividad específica, y las fechas de inicio y finalización de la actividad. Esta información permite identificar patrones en tipos específicos de ejercicios, evaluar la efectividad de diferentes formatos de actividades, y detectar conceptos que requieren refuerzo adicional, facilitando tanto la toma de decisiones pedagógicas como el análisis de investigación educativa.

Complementando las métricas de progreso, DALALearn implementa un sistema completo de registro de eventos de interacción del usuario que captura el comportamiento navegacional y educativo en tiempo real. Los eventos de navegación incluyen cuando un usuario se inscribe en un curso, accede a un curso específico, navega a un módulo determinado, o ingresa a una actividad particular. Estos eventos proporcionan información valiosa sobre patrones de uso, frecuencia de acceso, y comportamientos de navegación que pueden correlacionarse con resultados de aprendizaje.

Los eventos de interacción educativa capturan información detallada sobre las acciones específicas de aprendizaje, registrando cuando un usuario completa una actividad e incluyendo la respuesta enviada en caso de ejercicios y evaluaciones. La combinación de eventos de navegación e interacción educativa permite reconstruir la secuencia completa de acciones de cada usuario, identificar estrategias de resolución de problemas, detectar patrones de error recurrentes, y analizar la evolución del desempeño a lo largo del tiempo, complementando las métricas de progreso con información sobre el proceso de aprendizaje.

**Feedback inmediato** El sistema de *feedback* inmediato de DALALearn es una característica fundamental que proporciona retroalimentación en tiempo real a los estudiantes durante la práctica de programación. Esta funcionalidad se aplica específicamente a ejercicios y evaluaciones, ofreciendo información instantánea sobre el desempeño del usuario y facilitando el proceso de aprendizaje a través de la corrección inmediata de errores. El sistema está diseñado para simular evaluaciones reales, mostrando criterios de aprobación claros, notas mínimas requeridas y el aporte de cada ejercicio a la calificación final, contribuyendo al aprendizaje al proporcionar retroalimentación constructiva que permite a los estudiantes identificar y corregir sus errores de manera autónoma.

En los ejercicios el sistema proporciona *feedback* detallado que muestra al usuario qué elementos completó correctamente y cuáles requieren corrección, junto con la nota resultante de la actividad en formato de estrellas. Para obtener la nota máxima, el estudiante debe completar la actividad de manera totalmente correcta. En los ejercicios de tipo *fill the gaps*, el sistema indica para cada hueco si la respuesta proporcionada es

correcta o incorrecta, permitiendo al usuario identificar específicamente qué elementos necesitan ser corregidos. De manera similar, en los ejercicios de selección múltiple, el sistema muestra para cada opción seleccionada si la elección es correcta o incorrecta, facilitando la comprensión de los errores cometidos. En la figura 8 se muestran las pantallas de *feedback* resultantes de la resolución de un ejercicio de tipo *multiple choice* (a la izquierda resuelto de forma correcta y a la derecha resuelto de forma incorrecta).



Figura 8: *Feedback* de la app tras resolver un ejercicio. A la izquierda resuelto de forma correcta y a la derecha resuelto de forma incorrecta

Una característica pedagógica del sistema de *feedback* en ejercicios es que no revela las respuestas correctas cuando el usuario comete errores, permitiendo que los estudiantes vuelvan a intentar la actividad. Esta decisión se fundamenta en el aprendizaje iterativo, donde el conocimiento se construye mediante ciclos de retroalimentación que permiten identificar y corregir progresivamente las dificultades (Shute, 2008). Sin embargo, esta aproximación presenta riesgos: ciclos repetidos de intentos fallidos sin orientación suficiente pueden generar frustración o reforzar concepciones erróneas. Por esta razón, el sistema complementa la retroalimentación automática indicando específicamente qué elementos requieren corrección, buscando balancear la autonomía del estudiante con el soporte necesario para un aprendizaje efectivo.

En las evaluaciones, el sistema de *feedback* opera de manera diferente para simular un entorno de evaluación más formal. En lugar de mostrar la calificación en formato de estrellas, el sistema indica claramente si el usuario aprobó o reprobó la evaluación, junto con la nota obtenida en formato numérico (e.g., 10/30). Esta presentación más for-

mal refleja la naturaleza evaluativa de estas actividades y proporciona una experiencia similar a las evaluaciones académicas tradicionales.

El *feedback* en evaluaciones incluye información detallada sobre el desempeño en cada ejercicio individual. Para cada ejercicio de la evaluación, el sistema indica si el usuario cometió algún error, sin especificar cuál fue el error específico, manteniendo la integridad de la evaluación. Si el ejercicio se completó correctamente, el sistema muestra cuánta nota aportó ese ejercicio específico a la calificación final de la evaluación.

Los usuarios deben aprobar la evaluación obligatoriamente para poder avanzar en el curso. Si no aprobaron la evaluación, la intención es que los usuarios vuelvan a practicar los ejercicios, lecciones y módulos anteriores hasta poder realizar bien la evaluación. El sistema muestra en todo momento cuál es la nota mínima que se requiere para aprobar y cuál es la nota máxima que puede aportar cada ejercicio, proporcionando transparencia total sobre los criterios de evaluación.

**Gamificación** La gamificación se define como el uso de elementos de diseño de juegos en contextos no lúdicos (Deterding et al., 2011). Esta técnica incorpora componentes como sistemas de puntuación, tablas de clasificación, insignias, niveles y desafíos en actividades que originalmente no poseen características de juego, con el objetivo de aumentar la motivación y el compromiso de los participantes.

En el ámbito educativo, la investigación empírica ha demostrado que la gamificación puede tener efectos positivos significativos. Un meta-análisis concluyó que tiene un impacto positivo en el rendimiento cognitivo, motivacional y conductual de los estudiantes, aunque la efectividad varía según el diseño específico de los elementos gamificados y el contexto de implementación (Sailer y Homner, 2020). En plataformas de educación en línea, la gamificación adquiere particular relevancia al abordar el desafío del abandono estudiantil y la falta de compromiso, promoviendo el aprendizaje autorregulado y la adquisición de competencias. Además, la gamificación puede contribuir a la motivación estudiantil al proporcionar incentivos tangibles para el progreso y la participación activa en las actividades de aprendizaje (Narasareddygarí et al., 2018).

El sistema de gamificación de DALALearn está diseñado para fomentar la participación y motivación del usuario en la plataforma educativa. Esta funcionalidad se implementa a través de un sistema de puntos y *rankings* que busca mantener a los estudiantes dentro de la plataforma y promover la práctica continua de la programación. El sistema está diseñado para reconocer tanto el esfuerzo de participación como la calidad del desempeño, creando un balance entre motivación por completar actividades y motivación por realizarlas correctamente. Los puntos se otorgan de manera automática al finalizar cada actividad, proporcionando retroalimentación inmediata sobre la actividad completada.

El sistema de puntos opera con diferentes valores según el tipo de actividad completada. Las lecciones otorgan 1 punto al ser completadas, reconociendo el esfuerzo de revisar el contenido teórico. Los ejercicios pueden otorgar hasta 10 puntos, donde la cantidad de puntos ganados es proporcional a la calificación obtenida en el ejercicio, incentivando a los estudiantes a esforzarse por obtener la máxima puntuación. Las evaluaciones pueden otorgar hasta 30 puntos, también de manera proporcional a la calificación obtenida, reflejando su mayor complejidad y valor educativo. Adicional-

mente, completar un módulo otorga 50 puntos, mientras que completar un curso otorga 100 puntos, proporcionando incentivos significativos para la finalización de unidades de aprendizaje más amplias.

Los usuarios pueden consultar su puntaje total acumulado y acceder a un sistema de *ranking* que permite comparar su desempeño con otros usuarios de la plataforma. Actualmente, el *ranking* está disponible únicamente de manera global, mostrando la posición entre todos los usuarios de la plataforma. Sin embargo, se espera ampliar el sistema para hacer tablas limitadas a entornos más específicos, como por ejemplo, un *ranking* por cada curso. Esta funcionalidad de comparación social puede contribuir a la motivación estudiantil al proporcionar un contexto de competencia amigable y reconocimiento del progreso individual.

### **3 Prueba de Usabilidad de DALALearn**

El objetivo principal de esta prueba fue evaluar la usabilidad de la aplicación DALALearn desde la perspectiva de los estudiantes, con la idea de garantizar una excelente experiencia de usuario para la práctica y ejercitación de conceptos de programación. El estudio se centró en observar la interacción de los estudiantes con la aplicación a lo largo de su recorrido educativo, identificando tanto las fortalezas como las posibles fricciones en la experiencia de uso.

A partir de una base de datos de personas interesadas para colaborar, se seleccionaron participantes con conocimientos previos intermedios o avanzados en programación, con el propósito de que puedan concentrarse en el análisis de la interfaz y la navegación sin que la falta de familiaridad con los contenidos condicione la evaluación. Esta decisión metodológica permitió mantener el foco en los aspectos propios del diseño de la aplicación.

Para generar un contexto de uso realista y suficientemente amplio que abarcara todos los componentes de la plataforma, se desarrolló un curso básico de programación estructurado en dos módulos. Este curso incluyó un total de 31 actividades, diseñadas específicamente para cubrir la diversidad de recursos disponibles en la aplicación y evaluar la experiencia en diferentes escenarios de interacción.

#### **3.1 Metodología de Evaluación**

Los usuarios debieron completar el conjunto de ejercicios utilizando exclusivamente la aplicación, lo que permitió observar la experiencia de práctica de principio a fin. Esta dinámica ofreció la posibilidad de detectar problemas técnicos o dificultades de navegación que pudieran surgir en un uso cotidiano, al mismo tiempo que brindó información sobre la percepción subjetiva de la interfaz y el diseño visual. Es importante señalar que el foco de la prueba no estuvo en la calidad educativa del contenido, sino en la forma en que los usuarios navegan, acceden y se relacionan con los distintos recursos de la aplicación. El interés central fue analizar hasta qué punto la aplicación resulta intuitiva, clara y satisfactoria como entorno de práctica de programación.

Los problemas detectados durante esta prueba de usabilidad sirvieron como insumo para implementar mejoras en la interfaz, la navegación y la experiencia general del

usuario, con el objetivo de garantizar que los estudiantes puedan ejercitar conceptos de programación sin fricciones en el futuro uso de DALALearn, y que los futuros experimentos puedan realizarse sobre una aplicación con usabilidad funcional y experiencia de usuario optimizada.

Concluida la prueba de usabilidad en la aplicación, se le envió a los participantes una encuesta post-curso, con el fin de recolectar información sobre sus percepciones y experiencias con el uso de la aplicación. Este instrumento permitió profundizar tanto en los aspectos cuantitativos como cualitativos de la usabilidad y satisfacción. Se recibieron un total de nueve (9) respuestas de usuarios que completaron el curso, que constituyeron la base del análisis presentado a continuación.

### 3.2 Resultados y análisis

Los resultados de la prueba de usabilidad indicaron que la aplicación DALALearn tiene una usabilidad robusta en aspectos fundamentales como el acceso, inscripción y navegación básica. Los usuarios con conocimientos previos en programación encontraron la plataforma intuitiva y con una estructura clara que facilita el avance en el curso. La figura 9 muestra los puntajes promedios con los que los usuarios calificaron cada una de las categorías.

Así, se identificaron seis áreas relevantes de mejora:

1. **Diseño visual:** Requiere más elementos gráficos y opciones para mejor personalización visual (colores y temas).
2. **Navegación:** Podría optimizarse para que la retomada de actividades sea más fluida y evitar redundancias en la valoración de contenidos.
3. **Ejercicios de drag and drop:** Los usuarios reportaron dificultades específicas para colocar elementos en la posición deseada, especialmente en ejercicios largos donde las opciones no cabían completamente en pantalla, requiriendo mejoras en la precisión y funcionalidad de arrastrar y soltar.
4. **Experiencia móvil:** Los problemas principales se relacionaron con contenido externo no optimizado para móviles, que llevó a usuarios a cambiar a computadora, sugiriendo la necesidad de integración nativa de recursos externos.
5. **Seguimiento del progreso:** Mejorar la percepción del progreso con indicadores claros y retroalimentaciones específicas puede aumentar la motivación y el seguimiento del avance.
6. **Integración de recursos:** La integración de recursos externos debería ser más nativa para evitar discontinuidades en la experiencia.

Esta prueba de usabilidad permitió tener una sólida base para continuar iterando en el diseño de la aplicación, apuntando a mejorar la satisfacción y efectividad en la experiencia del usuario. Pese a que la cantidad de personas que respondieron la encuesta no fue grande ( $n=9$ ), en tanto los participantes tenían conocimientos medios o avanzados de programación, los hallazgos obtenidos permitieron realizar los ajustes necesarios antes de utilizar la plataforma en un contexto de investigación educativa.

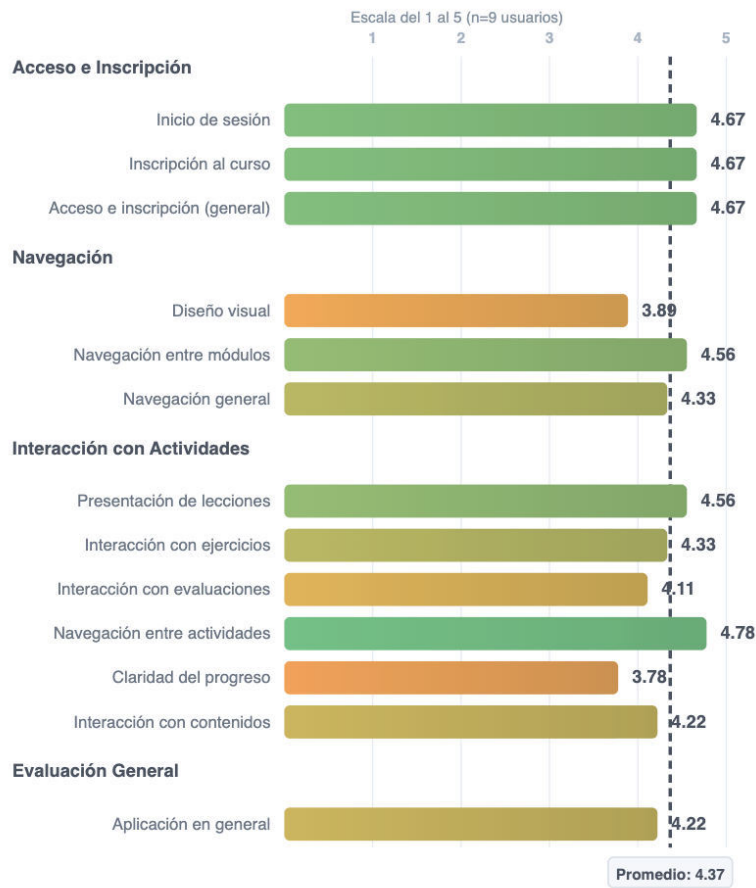


Figura 9: Evaluación de Usabilidad - Puntuaciones Promedio

## 4 Prueba de uso de DALALearn para un experimento de Investigación Educativa

Uno de los objetivos fundamentales de DALALearn es servir como herramienta para la realización de investigaciones educativas rigurosas en el ámbito de la enseñanza de la programación. Desde su concepción, la plataforma fue diseñada no solo para facilitar el aprendizaje de los estudiantes, sino también para proporcionar a investigadores y docentes un entorno controlado donde puedan diseñar, ejecutar y recopilar datos de experimentos educativos. Esta capacidad resulta particularmente valiosa en un campo donde existe una necesidad constante de evidencia empírica que respalde las decisiones didáctico-pedagógicas.

La presente prueba de uso tuvo como propósito principal demostrar que DALALearn permite a investigadores llevar a cabo estudios educativos controlados, facilitando la validación de hipótesis pedagógicas en entornos reales de aprendizaje. A diferencia de otros estudios que requieren la implementación de herramientas específicas para cada investigación, DALALearn ofrece una infraestructura ya preparada que simplifica significativamente el proceso de investigación educativa.

Para demostrar estas capacidades de manera concreta, se diseñó un experimento sobre la plataforma basado en la metodología de A/B testing, una técnica ampliamente utilizada en investigación experimental y educativa que consiste en comparar dos versiones de un mismo tratamiento para evaluar el efecto de una variable específica (Chen et al., 2016). En este caso, se compararon dos versiones de un curso de introducción a la programación con diferencias controladas.

Es importante destacar que este tipo de diseño experimental es solo uno de los múltiples enfoques que DALALearn permite implementar. La flexibilidad de la plataforma permite adaptarse a diferentes necesidades de investigación educativa, desde estudios exploratorios hasta experimentos controlados con múltiples condiciones experimentales.

### 4.1 Diseño del Experimento

La metacognición es, en términos simples, la capacidad de “pensar sobre cómo pensamos” (Flavell, 1979). Cuando un estudiante se detiene a preguntarse “¿realmente entendí esto?” o “¿qué estrategia me funcionó mejor para resolver este problema?”, está ejercitando habilidades metacognitivas.

En el contexto de la enseñanza de la programación, la metacognición adquiere particular relevancia. Aprender a programar requiere no solo comprender conceptos técnicos, sino también desarrollar habilidades de resolución de problemas, depuración de errores y evaluación de la propia solución (Cossio-Mercado et al., 2024).

Sin embargo, los estudiantes principiantes frecuentemente carecen de estas habilidades metacognitivas. No saben qué no saben, no reflexionan sobre sus estrategias de resolución de problemas y tienden a avanzar sin evaluar si realmente comprendieron los conceptos. Esta carencia puede explicar parte de las dificultades que enfrentan los estudiantes en cursos introductorios de programación.

Una pregunta natural que surge es: **¿Puede una intervención explícita que promueva la reflexión metacognitiva mejorar el aprendizaje de programación?** Para

explorar esta pregunta, diseñamos un experimento que compara dos grupos de estudiantes: uno que recibe únicamente el contenido del curso, y otro que además realiza actividades periódicas de autoevaluación y reflexión sobre su propio aprendizaje.

*Hipótesis de investigación:* La inclusión de actividades de metacognición y autoevaluación del aprendizaje en un curso de introducción a la programación mejora el desempeño de los estudiantes en la evaluación final y promueve una mayor reflexión sobre el proceso de aprendizaje.

Para investigar el efecto de las actividades metacognitivas en el aprendizaje de programación, se implementó un diseño experimental de dos grupos con asignación equitativa. Este diseño clásico de grupo control vs. grupo experimental permite comparar directamente los resultados de ambos grupos y atribuir las diferencias observadas a la variable manipulada (presencia o ausencia de actividades metacognitivas).

Es importante destacar que ambos grupos recibieron exactamente el mismo contenido teórico y los mismos ejercicios de práctica. La única diferencia entre las condiciones experimentales fue la inclusión de las actividades de autoevaluación y reflexión sobre el propio aprendizaje en el Grupo B. De esta manera, cualquier diferencia observada en los resultados puede atribuirse con mayor confianza al efecto de estas actividades metacognitivas.

El Grupo B incluyó actividades adicionales diseñadas para promover la reflexión sobre el propio proceso de aprendizaje. Estas actividades fueron cuidadosamente distribuidas a lo largo del curso, intercalándose con el contenido regular en momentos estratégicos para maximizar su efectividad sin sobrecargar a los participantes.

Al finalizar el curso, los estudiantes del Grupo B fueron invitados a completar una encuesta opcional de reflexión general sobre todo lo aprendido durante el curso. Esta encuesta incluyó preguntas abiertas que permitieron a los participantes expresar libremente sus percepciones sobre el proceso de aprendizaje experimentado.

## 4.2 Resultados

**Puntaje por ejercicio** En la figura 10 se muestran los promedios de los puntajes obtenidos por cada grupo en cada ejercicio del curso y la comparación de promedios entre grupos (la última columna, que muestra la resta entre el promedio del grupo B y el del grupo A). Los primeros 6 ejercicios corresponden a los ejercicios de tipo *multiple choice* intercalados entre las lecciones del curso. El puntaje en cada uno de estos es 1 si la respuesta fue correcta y 0 si no. Los siguientes 8 ejercicios (E1 a E8) corresponden a cada uno de los ejercicios de la evaluación final. Uno de estos ejercicios es de tipo *fill the gaps*, con 4 huecos para completar. En este ejercicio el puntaje se calcula sumando un punto por cada hueco completado correctamente y el ejercicio tenía 4 huecos así que el puntaje va de 0 a 4. El resto son todos *multiple choice*, pero en algunos hay que elegir todas las opciones correctas. En estos ejercicios el puntaje se calcula sumando uno por cada respuesta correcta y restando 0.3 por cada respuesta incorrecta así que el puntaje va de 0 a la cantidad de respuestas correctas. Para estandarizar, todos los puntajes de estos ejercicios se dividieron por el puntaje máximo de cada uno, dando como resultado un número entre 0 y 1.

Los colores se usan para identificar rápidamente qué tan cerca de 0 (en rojo) o de 1 (en verde) está cada valor. La última fila corresponde al promedio de las calificaciones

obtenidas en la evaluación final (un número entre 0 y 10). En la última columna se resaltan las diferencias entre grupos, con verde aquellas actividades en las que el grupo B, en promedio, obtuvo mayor puntaje y con naranja aquellas en las que, en promedio, obtuvo menor puntaje.

#	Grupo A	Grupo B	B - A
1	0.86	0.86	0.00
2	0.86	0.57	-0.29
3	0.57	0.57	0.00
4	0.43	0.71	0.29
5	0.71	0.71	0.00
6	0.71	0.71	0.00
E1	0.29	0.14	-0.14
E2	0.14	0.14	0.00
E3	0.71	0.57	-0.14
E4	0.61	0.57	-0.04
E5	0.67	0.85	0.18
E6	0.59	0.46	-0.13
E7	0.83	0.93	0.10
E8	0.64	0.66	0.02
<b>E</b>	<b>6.21</b>	<b>5.86</b>	<b>-0.36</b>

Figura 10: Puntajes promedio en cada ejercicio del curso por grupo y comparación entre grupos.

**Tiempo necesario para completar cada actividad** El tiempo que le llevó a cada sujeto completar una actividad en particular se calculó como la resta entre el tiempo de finalización de la misma (obtenido a partir del evento generado por completarla) y el tiempo de acceso a ella (obtenido a partir del evento generado por solicitarla).

En la tabla de la figura 11 se muestran los promedios de los tiempos calculados para cada grupo y la comparación de promedios entre grupos (la última columna, que muestra la resta entre el promedio del grupo B y el del grupo A) para cada actividad. Las actividades que sólo figuran en el grupo B son las de metacognición.

Se resaltan con distintos colores los tiempos que parecen ser mayores a lo que tendría sentido que alguien demore en realizar una actividad. La actividad 7 en particular demoró en promedio más de una hora en el grupo B porque un sujeto de dicho grupo registró una duración de más de 24 horas (seguramente porque abrió la actividad, se distrajo, y la retomó al día siguiente). También se resaltan las diferencias en la columna de comparación, con verde aquellas actividades en las que el grupo B, en promedio, demoró menos y con naranja aquellas en las que, en promedio, demoró más.

#	Tipo de actividad	Grupo A	Grupo B	B - A
1	Lección	0:00:19	0:00:20	0:00:01
2	Lección	-	0:00:20	-
3	Lección	0:03:13	0:06:41	0:03:28
4	Multiple Choice	0:00:30	0:00:23	-0:00:07
5	Multiple Choice	0:00:19	0:00:24	0:00:05
6	Lección	0:00:46	0:03:25	0:02:39
7	Lección	0:04:02	1:17:56	1:13:54
8	Multiple Choice	0:00:35	0:00:22	-0:00:13
9	Multiple Choice	0:00:22	0:00:20	-0:00:02
10	Lección	0:00:51	0:00:52	0:00:01
11	Lección	0:02:35	0:00:48	-0:01:47
12	Multiple Choice	0:00:13	0:01:38	0:01:25
13	Multiple Choice	0:00:44	0:00:45	0:00:01
14	Lección	0:00:51	0:00:45	-0:00:06
15	Lección	-	0:00:10	-
16	Multiple Choice	-	0:00:17	-
17	Multiple Choice	-	0:00:13	-
18	Multiple Choice	-	0:00:18	-
19	Multiple Choice	-	0:00:11	-
20	Multiple Choice	-	0:00:16	-
21	Multiple Choice	-	0:00:10	-
22	Multiple Choice	-	0:00:11	-
23	Lección	0:01:19	0:01:30	0:00:11
24	Lección	0:00:10	0:00:08	-0:00:02
25	Lección	0:00:21	0:00:51	0:00:30
26	Multiple Choice	-	0:00:13	-
27	Multiple Choice	-	0:00:23	-
28	Lección	0:00:40	0:00:47	0:00:07
29	Lección	0:00:05	0:00:10	0:00:04
30	Lección	0:00:11	0:00:21	0:00:10
31	Multiple Choice	-	0:00:10	-
32	Multiple Choice	-	0:00:08	-
33	Lección	0:00:28	0:00:28	-0:00:01
34	Lección	0:00:35	0:00:35	0:00:01
35	Lección	0:00:36	0:04:18	0:03:42
36	Multiple Choice	-	0:00:07	-
37	Multiple Choice	-	0:00:07	-
38	Lección	0:01:14	0:00:57	-0:00:17
39	Lección	-	0:02:02	-
40	Evaluación	0:07:01	0:07:29	0:00:28
<b>Total:</b>		<b>0:28:00</b>	<b>1:57:29</b>	<b>1:29:29</b>

Figura 11: Tiempos promedio de realización de cada actividad por grupo y comparación entre grupos.

**Respuestas reportadas en las actividades de metacognición** En las tablas de la figura 12 se muestran cada una de las respuestas dadas por cada sujeto (identificados como B1 a B7) a cada una de las actividades de metacognición del curso B.

El primer conjunto de preguntas se presenta luego de haber realizado dos desafíos de PilasBloques y de haber respondido preguntas relacionadas a ellos (del estilo “cuántas instrucciones se necesitan para completar el desafío”). Los otros 3 conjuntos de preguntas se presentan luego de haber tenido la respectiva lección sobre cada tema (dominio, operaciones primitivas y reglas semánticas), donde se explica el concepto en cuestión y luego se lo identifica en los lenguajes trabajados (CodeSlime y PilasBloques).

La última columna muestra el promedio de respuestas de cada pregunta. En todos los casos los colores se usan para identificar rápidamente qué tan cerca del puntaje mínimo (en rojo) o del máximo (en verde) está cada valor. Las preguntas del primer conjunto se evaluaron entre 1 y 5 mientras que las del resto se evaluaron entre 1 y 10.

<b>Después de PilasBloques (evaluar de 1 a 5):</b>	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>	<b>Promedio</b>
Consideré varias alternativas al problema antes de probar.	5	5	1	3	4	5	5	4.00
Pensé una posible solución antes de comenzar a manipular los bloques.	3	5	5	5	5	1	5	4.14
Antes de ejecutar el programa sabía si iba a funcionar o no.	4	5	4	4	3	1	4	3.57
Concentré conscientemente mi atención en la información importante.	3	5	4	3	5	5	5	4.29
Me hice preguntas sobre el desafío antes de empezar a ejecutarlo.	4	5	5	2	1	5	5	3.86
Pensé en varias formas de resolver el desafío y elegí la mejor.	4	4	2	2	1	5	3	3.00
Fui buena/o juzgando cuán bien iba a resolver el desafío.	3	3	4	3	1	1	4	2.71

<b>Después de la lección de Dominio (evaluar de 0 a 10):</b>	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>	<b>Promedio</b>
Aprendizaje sobre los conceptos	3	9	7	4	4	6	10	6.14
Podría identificar esos conceptos en un lenguaje diferente.	2	4	8	3	7	8	8	5.71

<b>Después de la lección de Operaciones Primitivas (evaluar de 0 a 10):</b>	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>	<b>Promedio</b>
Aprendizaje sobre los conceptos	5	7	8	4	9	10	9	7.43
Podría identificar esos conceptos en un lenguaje diferente.	2	4	8	3	7	8	7	5.57

<b>Después de la lección de Reglas Semánticas (evaluar de 0 a 10):</b>	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>	<b>Promedio</b>
Aprendizaje sobre los conceptos	2	8	8	6	7	7	9	6.71
Podría identificar esos conceptos en un lenguaje diferente.	2	6	8	3	6	8	8	5.86

Figura 12: Respuestas a cada actividad de metacognición del curso B reportada por cada sujeto.

La prueba de uso realizada demostró de manera práctica que DALALearn es una herramienta viable y funcional para la realización de investigaciones educativas en el ámbito de la enseñanza de la programación. A lo largo de todo el proceso experimental, la plataforma cumplió satisfactoriamente con los requisitos básicos que un investigador necesita para llevar a cabo un estudio controlado.

DALALearn ha demostrado ser una plataforma con potencial real para facilitar la investigación educativa en programación. Si bien existen áreas de mejora identificadas, la infraestructura básica está en funcionamiento y permite a investigadores diseñar, ejecutar y analizar experimentos educativos de manera efectiva.

## 5 Conclusiones

Este trabajo abordó el problema de enseñar programación y realizar investigaciones científicas en educación en programación con una misma plataforma. Las herramientas existentes, aunque efectivas para la práctica, no ofrecen la infraestructura necesaria para que investigadores y docentes puedan diseñar experimentos controlados y recopilar datos empíricos de manera sistemática. Para resolver esta necesidad insatisfecha se propuso la plataforma DALALearn.

La contribución principal de este trabajo es presentar y evaluar DALALearn como una plataforma integrada que articula enseñanza de la programación, recolección sistemática de datos y experimentación educativa controlada en contextos reales.

En cuanto al desarrollo de la plataforma móvil, se diseñó e implementó una aplicación funcional para dispositivos iOS y Android que permite a los estudiantes practicar conceptos de programación de forma estructurada. La plataforma incorpora tres tipos de actividades (lecciones, ejercicios y evaluaciones), dos modalidades de ejercicios interactivos (*fill the gaps* y *multiple choice*), un sistema de dependencias entre módulos que guía el progreso del estudiante, y elementos de gamificación que incentivan la práctica continua. El diseño modular de la aplicación permite la incorporación de nuevos tipos de ejercicios sin grandes complicaciones.

Para el desarrollo de herramientas de administración y recolección de datos, se implementó un panel web completo que permite a docentes e investigadores crear y gestionar cursos, definir relaciones de dependencia entre módulos mediante una interfaz visual, y configurar diferentes condiciones experimentales. El sistema de recolección de datos opera en tres niveles de granularidad (curso, módulo y actividad), registrando automáticamente las interacciones de los usuarios, tiempos de actividad, respuestas a ejercicios y resultados de evaluaciones. La funcionalidad de exportación en formato CSV permite el análisis de datos con herramientas estadísticas estándar.

La evaluación de usabilidad se realizó con 9 participantes que completaron un curso de prueba y respondieron una encuesta detallada. Los resultados fueron positivos: el acceso e inscripción obtuvo un promedio de 4.67/5, la navegación general 4.22/5, y la interacción con contenidos 4.0/5. La prueba permitió identificar áreas de mejora específicas, particularmente en los ejercicios de drag and drop y en los indicadores de progreso. Estas mejoras fueron implementadas antes de la siguiente fase de evaluación.

La prueba de uso para investigación educativa validó la capacidad de DALALearn como herramienta de investigación. Se diseñó y ejecutó un experimento A/B con 14 participantes (7 en cada grupo) para evaluar el efecto de actividades metacognitivas en el aprendizaje de programación. La plataforma demostró su capacidad para crear condiciones experimentales diferenciadas, recolectar datos de manera automática y sistemática, estandarizar la evaluación entre grupos, y exportar información para análisis externo. Los investigadores pudieron completar todo el ciclo de investigación utilizando las herramientas provistas por DALALearn.

DALALearn combina la interactividad de una aplicación educativa con herramientas de recolección de datos para investigación. Esto permite realizar experimentos sin que los participantes perciban diferencias respecto a una app educativa común. La plataforma queda disponible para su uso, y las mejoras identificadas durante las evaluaciones se detallan en la sección de trabajo futuro.

## 6 Trabajo Futuro

Con base en los resultados de la prueba de usabilidad y el uso para investigación educativa, así como en los comentarios de los usuarios, docentes e investigadores, se identificaron varias líneas de trabajo futuro para mejorar DALALearn.

En términos de experiencia de usuario, se identificó la necesidad de implementar un modo oscuro (*dark mode*) en la aplicación. Además, se requiere implementar un sistema más completo de registro de eventos que capture cambios de contexto de la aplicación (apertura, cierre, minimización, regreso a primer plano) e interacciones de interfaz (tocar botones, navegar entre pantallas, desplazamiento). Esto permitiría calcular tiempos de uso efectivo y realizar análisis más detallados del comportamiento del usuario.

Actualmente, los investigadores exportan los datos de actividad en formato CSV y los procesan con herramientas externas. Para facilitar este proceso, se espera desarrollar un analizador de datos integrado dentro del panel web que permita visualizar y analizar datos de experimentos en forma directa. Esta herramienta incluiría dashboards con gráficos y estadísticas descriptivas, herramientas de segmentación para comparar grupos experimentales, y opciones de exportación de datos ya procesados con métricas calculadas (tiempos de completitud, promedios, etc.).

La experiencia en la creación de cursos reveló la necesidad de nuevos tipos de actividades. Así, se podría implementar una escala numérica configurable (tipo Likert) donde solo se defina el valor mínimo y el máximo, y la cantidad de valores intermedios, de forma de evitar la creación manual de múltiples opciones. Este tipo de actividad sería útil para formularios de evaluación y encuestas dentro de los cursos.

En las actividades de tipo ejercicio, se propone mejorar el *feedback* que recibe el usuario. Una vez evaluada la respuesta, además de indicar si es correcta o incorrecta, se mostraría una explicación de por qué es así. Para esto, el panel web permitiría configurar un texto de *feedback* positivo y uno negativo por cada respuesta completada.

Se espera la incorporación de un entorno de ejecución de código integrado dentro de la aplicación, similar a herramientas como PilasBloques o Gobstones. Este tipo de actividad permitiría a los usuarios ordenar bloques de código y ejecutarlos para ver el resultado, sin necesidad de salir de DALALearn. Así, se mejoraría la experiencia de aprendizaje práctico sin salir de la plataforma.

El sistema de gamificación actual ofrece oportunidades de expansión. Se propone agregar nuevas mecánicas como *rankings* por curso y por actividad, un sistema de rachas que otorgue puntos bonus al completar varias actividades correctas de manera consecutiva, y mensajes de motivación más frecuentes durante el uso de la aplicación.

## Referencias

- Alfarsi, G., Tawafak, R., Alyoussef, I., & Al-Rahmi, W. (2024). Systematic Literature Review for Smart Mobile Learning in Programming Courses. *International Journal of Interactive Mobile Technologies (IJIM)*, 18, 129-142. <https://doi.org/10.3991/ijim.v18i12.46299>
- Amro, J., & Romli, R. (2019). Investigation on the Learning Programming Techniques via Mobile Learning Application, 1-7. <https://doi.org/10.1109/ICRAIE47735.2019.9037764>

- Chen, Z., Chudzicki, C., Palumbo, D., Alexandron, G., Choi, Y.-J., Zhou, Q., & Pritchard, D. (2016). Researching for better instructional methods using AB experiments in MOOCs: results and challenges. *Research and Practice in Technology Enhanced Learning*, 11. <https://doi.org/10.1186/s41039-016-0034-4>
- Cossio-Mercado, C., Fernández, G. P., & Pérez, G. (2024). ¿Cómo llevar lo que sabemos de Metacognición a las clases de programación? *Jornadas Argentinas de Didáctica de las Ciencias de la Computación 2024*.
- Criollo-C, S., Guerrero-Arias, A., Jaramillo-Alcázar, Á., & Luján-Mora, S. (2021). Mobile Learning Technologies for Education: Benefits and Pending Issues. *Applied Sciences*, 11(9). <https://doi.org/10.3390/app11094111>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining Gamification. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, 11, 9-15. <https://doi.org/10.1145/2181037.2181040>
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906-911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Messer, M., Brown, N. C. C., Kölling, M., & Shi, M. (2024). Automated Grading and Feedback Tools for Programming Education: A Systematic Review. En *ACM Trans. Comput. Educ.* (Vol. 24). Association for Computing Machinery. <https://doi.org/10.1145/3636515>
- Nambiar, R. (2020). Coding as an Essential Skill in the Twenty-First Century, 237-243. [https://doi.org/10.1007/978-981-15-7018-6\\_29](https://doi.org/10.1007/978-981-15-7018-6_29)
- Narasareddygar, M. R., Walia, G., & Radermacher, A. (2018). Gamification in Computer Science Education: a Systematic Literature Re-view Gamification in Computer Science Education: A Systematic Literature Review. <https://doi.org/10.18260/1-2--30554>
- Sailer, M., & Homner, L. (2020). The Gamification of Learning: a Meta-analysis. *Educational Psychology Review*, 32, 77-112. <https://doi.org/10.1007/s10648-019-09498-w>
- Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2021). Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, 36. <https://doi.org/10.1080/09588221.2021.1933540>
- Shute, V. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78, 153-189. <https://doi.org/10.3102/0034654307313795>
- Walpita Gamage, S., Ayres, J., & Behrend, M. (2022). A systematic review on trends in using Moodle for teaching and learning. *International Journal of STEM Education*, 9. <https://doi.org/10.1186/s40594-021-00323-x>
- Zimmerman, F., Shalom, D., González, P., Garrido, J., Heduan, F., Dehaene, S., Sigman, M., & Rieznik, A. (2016). Arithmetic on Your Phone: A Large Scale Investigation of Simple Additions and Multiplications. *PLOS ONE*, 11, e0168431. <https://doi.org/10.1371/journal.pone.0168431>