

Rosana Satorre Cuerda (Ed.)

El profesorado, eje fundamental de la transformación de la docencia universitaria

Rosana Satorre Cuerda (Ed.)

El profesorado, eje fundamental de la transformación de la docencia universitaria

Octaedro 
Editorial

UA

UNIVERSITAT D'ALACANT
UNIVERSIDAD DE ALICANTE
Vicerectorat de Transformació Digital
Vicerrectorado de Transformación Digital
Institut de Ciències de l'Educació
Instituto de Ciencias de la Educación

El profesorado, eje fundamental de la transformación de la docencia universitaria

EDICIÓN:

Rosana Satorre Cuerda

Revisión y maquetación: ICE de la Universidad de Alicante

Primera edición: octubre de 2022

© De la edición: Rosana Satorre Cuerda

© Del texto: Las autoras y autores

© De esta edición:

Ediciones OCTAEDRO, S.L.

C/ Bailén, 5 – 08010 Barcelona

Tel.: 93 246 40 02 – Fax: 93 231 18 68

www.octaedro.com – octaedro@octaedro.com

ISBN: 978-8-19506-52-8

Producción: Ediciones Octaedro

La revisión de los trabajos se ha realizado de forma rigurosa, siguiendo el protocolo de revisión por pares.

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.

NOTA EDITORIAL: Las opiniones y contenidos de los textos publicados en esta obra son de responsabilidad exclusiva de los autores.

23. Metodología basada en jueces en línea para la enseñanza de programación

Hernández, Guillermo; de la Prieta, Fernando; Rodríguez, Sara; Chamoso, Pablo; Pinto-Santos, Francisco; Hernández Simón, Juan Andrés y González Arrieta, Angélica

Universidad de Salamanca

RESUMEN

En este trabajo se presenta la experiencia docente de la implantación de un juez en línea para el aprendizaje de programación en el contexto de una asignatura de formación básica del Grado en Ingeniería Informática. En este proceso se ha buscado analizar las principales dificultades, ventajas e inconvenientes de la forma de trabajo orientada a la evaluación automática de tareas de programación mediante pares de entrada/salida que describan el compartimiento esperado del algoritmo, como es más habitual en estos sistemas. La herramienta desplegada se ha basado en el software libre DMOJ, adaptando de forma específica la colección de problemas de la asignatura, y se hizo accesible al alumnado a través del navegador, de forma totalmente opcional y no evaluable. Se recogieron resultados de participación, encontrándose una correlación positiva entre la nota de la asignatura y los resultados en la plataforma, así como una puntuación media en la asignatura superior en los participantes. A través de una encuesta realizada al final del curso para conocer la percepción del alumnado sobre el impacto de la herramienta en su aprendizaje se aprecia también una valoración global positiva sobre la misma y un impacto también positivo en el estudiantado que hizo uso de esta.

PALABRAS CLAVE: programación, docencia, aprendizaje, juez en línea, gamificación.

1. INTRODUCCIÓN

La introducción a la programación constituye una materia de formación básica en numerosas titulaciones de ciencias e ingenierías, siendo, además, central a algunos programas como es el Grado en Ingeniería Informática. El fuerte contenido práctico y la exigencia del desarrollo de una forma de pensar algorítmica y formal suponen un importante obstáculo en la adquisición de las competencias vinculadas a la misma, que puede manifestarse de forma heterogénea por los diferentes perfiles de ingreso en la etapa universitaria.

Una herramienta que es particularmente popular en el ámbito de la programación competitiva –competiciones ubicadas en el ámbito de los deportes mentales en las que se resuelven problemas de programación– son los jueces en línea, herramientas capaces de evaluar de forma automática el desempeño de códigos frente a pares de entrada/salida que describen su comportamiento (Halim et al. 2020). La metodología propia de estos conecta de forma directa con el desarrollo orientado a pruebas que se presenta habitualmente en el mundo de la empresa (Beck, 2003), siendo una evidencia de esto la aparición de plataformas con esta filosofía, en las que, además la detección de talentos, se persigue directamente facilitar la contratación de personal (Ínce, 2021).

Este tipo de metodología, pese a su concepción orientada a los programadores altamente especializados del ámbito competitivo, puede encajar también adecuadamente en las primeras etapas del aprendizaje de programación, donde también puede motivar el aprendizaje con los elementos de

gamificación que incorpora como son la resolución de desafíos, la obtención de puntuaciones, la componente social de comparar los resultados en la plataforma, etc. En la literatura se han reportado ya algunos esfuerzos en este sentido, principalmente orientados a la evaluación automática de tareas (Edwards & Perez-Quinones 2008; Polito & Temperini, 2018). Especialmente reseñable es el trabajo de Enström et al. (2011), de cuya experiencia surgió la plataforma Kattis (<https://open.kattis.com/>), actualmente orientada a la programación competitiva y abierta al público general. Otro trabajo especialmente importante, este orientado a la programación competitiva en su fase inicial es el de Revilla et al. (2008), que dio lugar al juez en línea de la UVa, actualmente continuado por el proyecto Online Judge (<https://onlinejudge.org/>). Otros jueces en línea abiertos al público en esta línea competitiva son DMOJ (<https://dmoj.ca/>) –de código abierto, existe alguna implementación basada en este (González et al., 2021), también se ha empleado en este trabajo, como se presenta más adelante–, Codewars (<https://www.codewars.com/>) y CodinGame (<https://codingame.com/>).

En este trabajo se ha realizado la puesta en marcha de un sistema de juez en línea y la adaptación para el mismo de ejercicios correspondientes a un curso universitario de un semestre de introducción a la programación. Se ha analizado particularmente la adaptación del alumnado a la resolución de problemas en el mismo frente al paradigma habitual de resolución en un entorno de desarrollo integrado tradicional con la ejecución de pruebas de carácter manual, detectando así las principales dificultades que puede suponer este cambio de paradigma y su resolución. Los resultados obtenidos a partir del desempeño de los alumnos participantes en la plataforma y sus calificaciones finales muestran una interrelación entre ambas. Adicionalmente, se ha recogido la percepción del alumnado sobre el papel de la plataforma en su aprendizaje, de lo que también se desprende que el uso de esta es globalmente percibido como positivo por aquellos que optaron por emplearla.

2. MÉTODO

En este apartado se exponen las características en que se realizó la investigación, con especial atención al contexto universitario en que se desarrolló el estudio, las herramientas utilizadas y la metodología de obtención de información utilizadas posteriormente en el análisis de resultados.

2.1. Descripción del contexto y de los participantes

La asignatura en que se ha desarrollado este estudio es Programación I del Grado en Ingeniería Informática durante el curso 2021/22. Es una asignatura semestral de formación básica cuyo temario incluye la iniciación a los fundamentos de programación, incluyendo una parte práctica en lenguaje C. El temario incluye varios aspectos centrales de los lenguajes de programación estructurados como son los tipos de datos y variables, las estructuras condicionales y las iterativas; así como otros específicos del lenguaje como el manejo de punteros. La asignatura puede suponer el primer contacto con la programación para muchos alumnos, al ser la primera del programa del Grado que la incluye. El alumnado estaba compuesto de 174 matriculados.

2.2. Instrumentos

La herramienta utilizada para construir la plataforma es el software DMOJ, disponible bajo licencia AGPL (licencia pública general de Affero). El despliegue se ha realizado a partir de las imágenes de Docker, incluyendo tecnologías adicionales como Mathoid para la incorporación de fórmulas matemáticas o Celery para el sistema de colas para la evaluación de problemas. El despliegue se realizó en servidores propios con sistema operativo Debian 10. Se desplegaron dos instancias de evaluación

independientes, que dieron holgadamente soporte a las necesidades encontradas durante la realización de la asignatura, si bien podrían extenderse fácilmente en caso de detectar necesidades mayores.

2.3. Procedimiento

Al comienzo del curso se desplegó una versión cerrada de la plataforma para su uso por parte del personal docente de la asignatura. En esta etapa se empezó a trabajar en la adaptación de la colección de problemas de la asignatura, orientados a la ejecución interactiva por consola, al formato del juez. La definición de estos problemas comportaba:

- Redacción de un enunciado con formato utilizando el lenguaje *markdown*, posiblemente incluyendo recursos como imágenes, expresiones matemáticas en lenguaje LaTeX, vídeos, enlaces externos, etc.
- Definición de pares entrada/salida, algunos orientados a servir de ejemplo en el propio enunciado, otros a la evaluación de la ejecución del programa. Destaca aquí la necesidad de documentar y definir los casos límite (*corner cases*) importante para asegurar la validez de un programa candidato a solución.
- Definición de información estructurada describiendo el problema (tema al que se asocia el problema, palabras clave, valoración de la dificultad, límites a los recursos computacionales dedicados a la evaluación, etc.).

Un ejemplo particularmente simple de estos problemas correspondiente al cálculo del volumen de una esfera, visto desde la interfaz disponible a los alumnos, se muestra en la Figura 1. En todos los problemas se ha seguido este formato, describiendo el enunciado del problema, los formatos de entrada y salida, y varios ejemplos de entradas con sus correspondientes salidas.

Volumen de una esfera (2-12)

Realizar un programa que calcula el volumen de una esfera cuyo radio (r) se introducirá por teclado.

Se recuerda que el volumen de una esfera viene dado por

$$V = \frac{4}{3}\pi r^3.$$

Se puede tomar como aproximación

$$\pi \approx 3.14159.$$

Entrada

El radio de la esfera.

Salida

El volumen de la esfera (en unidades consistentes con las de entrada).

Ejemplo de entrada 1

4

Ejemplo de salida 1

268.082

Submit solution

All submissions
Best submissions

Manage tickets
Edit problem
Edit test data
Manage submissions
Clone problem

✓ Points: 5
⌚ Time limit: 2.0s
📦 Memory limit: 64M

✓ Author: admin

➤ Problem type

▼ Allowed languages
C

🏴‍☠️ Judge: dorado

Figura 1. Ejemplo de problema en JUEZ.

Con una colección de problemas inicial se abrió la plataforma al uso de los alumnos de la asignatura, realizando una pequeña presentación en clase de unos 10 minutos de duración en que se ejemplifica cómo el *feedback* del juez puede servir para autodiagnosticar los defectos en el código. Un ejemplo de esto se muestra en la Figura 2, donde a la izquierda se puede ver un intento de solución

erróneo para el problema del volumen de una esfera (Figura 1). En el lado derecho se puede observar la salida que el programa proporciona para el caso del enunciado. En este caso, el error se debe a que en lenguaje C la operación $4/3$ tiene como resultado 1 , debido a que designa el cociente de la división entera.

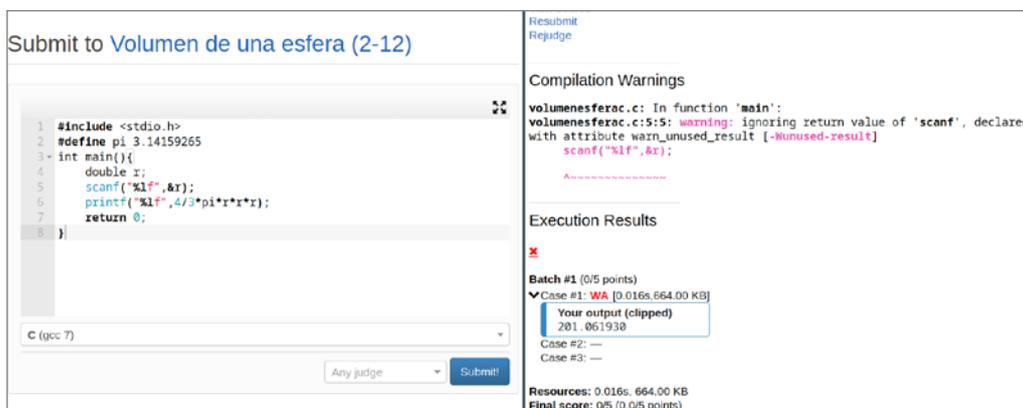


Figura 2. Ejemplo de intento incorrecto de solución a un problema.

El uso de la plataforma se presentó como opcional al alumnado y no evaluable, si bien se incluirían en la misma las actividades realizadas durante las prácticas de la asignatura de forma prácticamente exhaustiva (con las debidas adaptaciones), así como otros ejercicios procedentes de pruebas de evaluación anteriores o realizadas durante el propio curso y algunos problemas especialmente complicados de desafío, orientados a los alumnos más avanzados. La incorporación de problemas nuevos a la plataforma se realizó temporalmente de manera aproximadamente uniforme, de forma paralela al desarrollo de la docencia.

Los datos obtenidos se han recogido de las estadísticas de la propia plataforma, incluyendo número de envíos por ejercicio, resultado de la evaluación (éxito, resultado incorrecto, error de compilación, límite de tiempo de ejecución, etc.), fecha y hora, etc.; así como de una encuesta de evaluación realizada al término de la actividad lectiva y antes de la realización de las pruebas de evaluación final de la asignatura, realizada de forma opcional por el alumnado que se prestó a ello. Esta encuesta se diseñó con preguntas de elección múltiple o simple y valoraciones en escala Likert (1932) de cinco puntos (valoración del grado de adhesión a una afirmación, donde 1 es “totalmente en desacuerdo”; 2, “de acuerdo”; 3 “ni de acuerdo ni en desacuerdo”; 4 “de acuerdo” y 5, “totalmente de acuerdo”). Se proporcionó también una opción “no sabe/no contesta” para facilitar que los estudiantes con poco uso de la plataforma puedan contestarla también, aportando información en los apartados en que se sientan con confianza para opinar.

Por cuestión de extensión, en este trabajo nos centraremos en presentar y analizar los datos relevantes a la interrelación entre la plataforma y el desempeño académico, como se desarrollará en los siguientes apartados.

3. RESULTADOS

Las distribuciones de notas finales en la convocatoria ordinaria del alumnado que ha utilizado el JUEZ y de los que no se muestra en la Figura 3. Se ha utilizado la técnica de *kernel density estimation* para suavizar las distribuciones, facilitando la comprensión de sus diferencias. Las medias se

muestran con las líneas verticales discontinuas. Debe notarse que la distribución del alumnado que ha empleado al juez y la del que ha resuelto más del 75% de sus problemas contiene muestras en común.

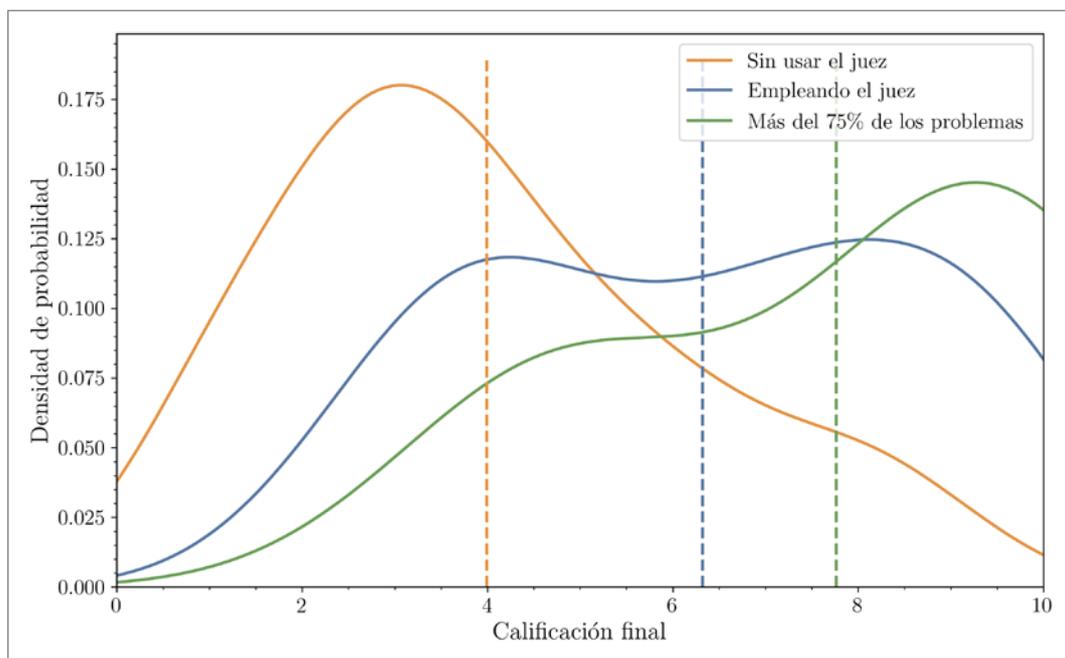


Figura 3. Calificaciones en la asignatura para los diferentes grupos del alumnado.

Un gráfico de dispersión que compara la cantidad de puntos obtenidos en el JUEZ (ponderando los problemas de mayor dificultad y los que menos alumnos resuelven con el algoritmo de DMOJ (2017)) se muestra en la Figura 4, en la que también se incluye un ajuste lineal de tendencia. En la siguiente sección se analizará la significación estadística del mismo.

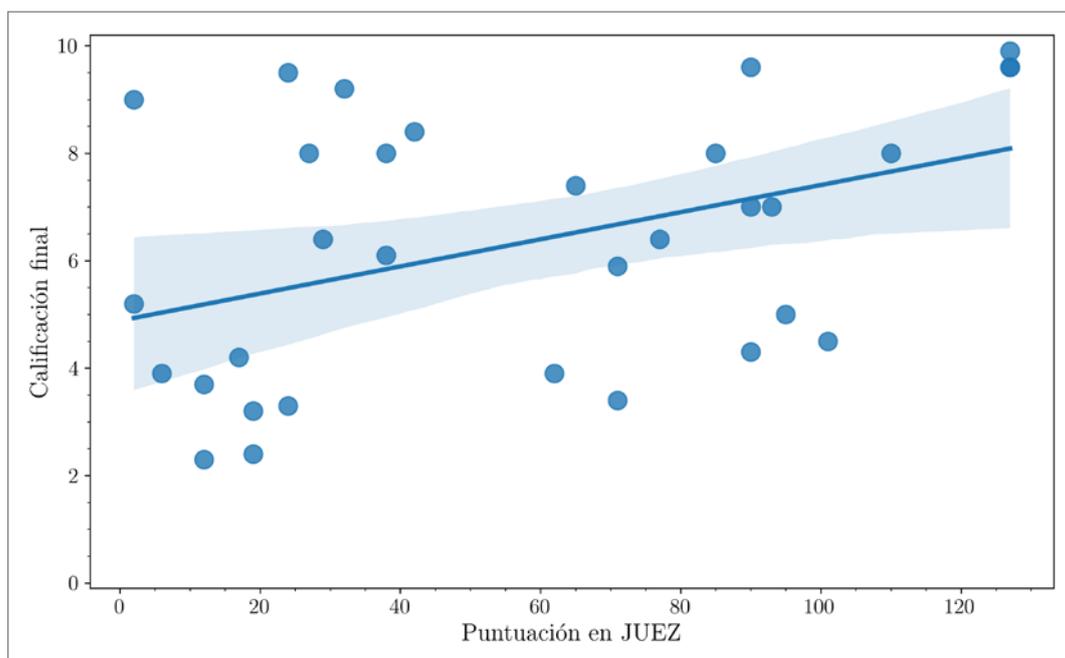


Figura 4. Interrelación entre la puntuación en JUEZ y la nota en convocatoria ordinaria.

Finalmente, respecto a la encuesta realizada al alumnado, se obtuvo un total de 19 respuestas. Aunque este número es significativamente inferior al número de alumnos (144), es comparable al de participantes en la prueba del juez (33 que hayan resuelto al menos un problema). Se reproducen a continuación los resultados relevantes para el alcance del presente trabajo, como se introdujo en el apartado anterior.

Los estudiantes que declararon haber empleado de forma limitada el JUEZ adujeron las razones que se muestran en la Figura 5. Las opciones que aparecen recortadas son “No contribuye a la calificación”, “Quería ceñirme a otros materiales para preparar la asignatura”. Las últimas dos se corresponden a una respuesta libre dada por dos estudiantes, que reproducimos de forma literal: “A pesar de ser entendible, el hecho de que tenga una entrada y una salida exactas a la hora de compilar, hace que no tenga ganas de utilizar mucho la página porque no se ciñe a la forma de programar que tengo” y “Hay veces que no puedes saber cual es el fallo en el resultado que buscas obtener”.

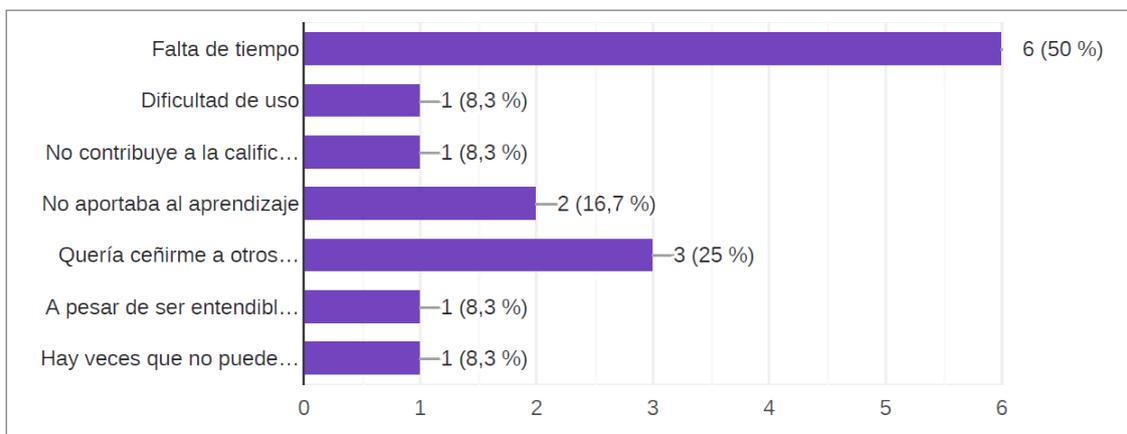


Figura 5. Motivos que han dificultado la adopción de JUEZ para el alumnado.

La Figura 6, la Figura 7, la Figura 8 y la Figura 9, muestran valoraciones en escala Likert (ver apartado anterior) a las preguntas que sus correspondientes títulos recogen.

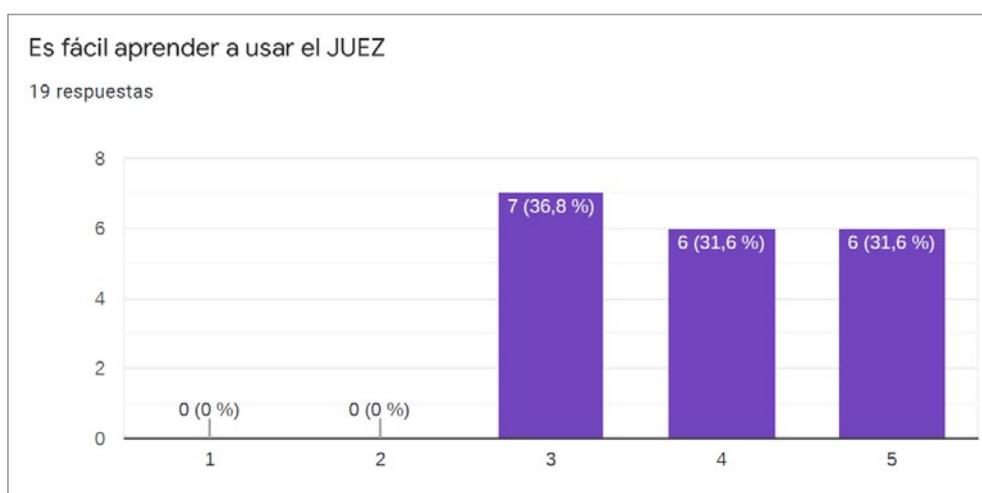


Figura 6. Valoración del alumnado sobre la facilidad de aprendizaje de JUEZ.

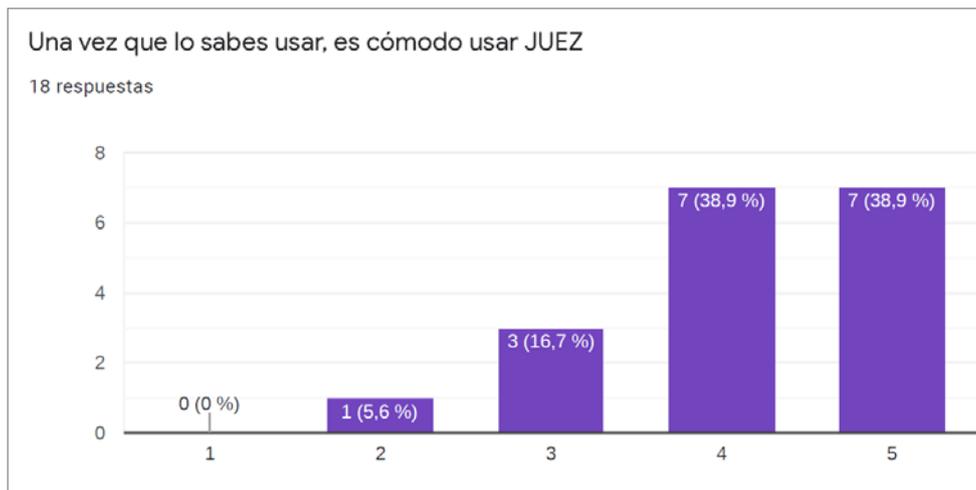


Figura 7. Valoración del alumnado sobre la comodidad de uso de JUEZ tras su aprendizaje inicial.

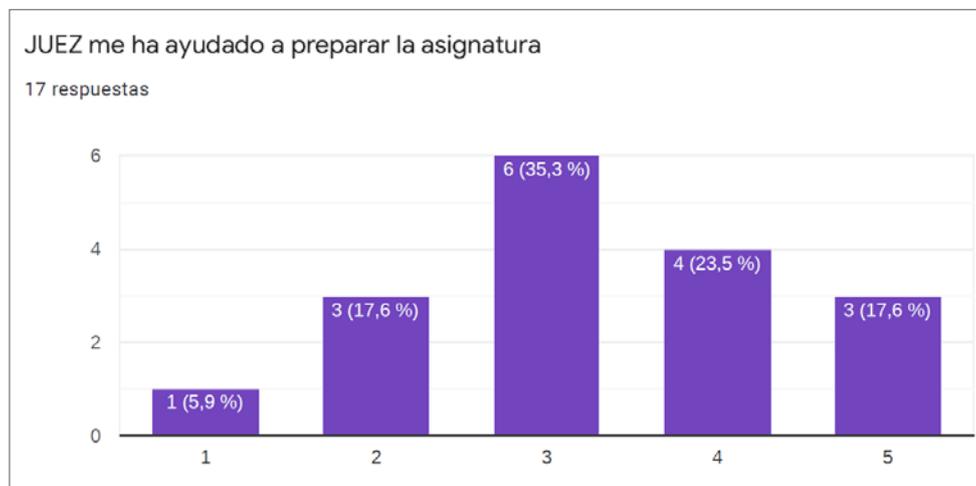


Figura 8. Valoración del alumnado sobre el aporte de JUEZ a la preparación de la asignatura.

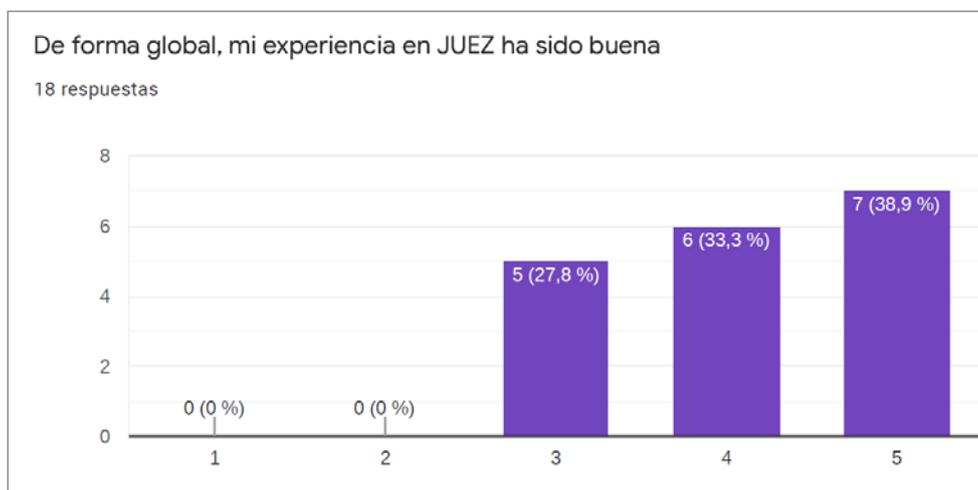


Figura 9. Valoración de alumnado sobre la experiencia con JUEZ.

Como se discutirá más adelante, para hacer una lectura correcta de la Figura 8 es necesario apreciar que tanto estudiantes que han utilizado el JUEZ como otros que no han respondido a la encuesta. La Figura 10 representa de forma simultánea su nivel de participación (reflejada por el encuestado) junto con su valoración del impacto en la preparación de la asignatura.

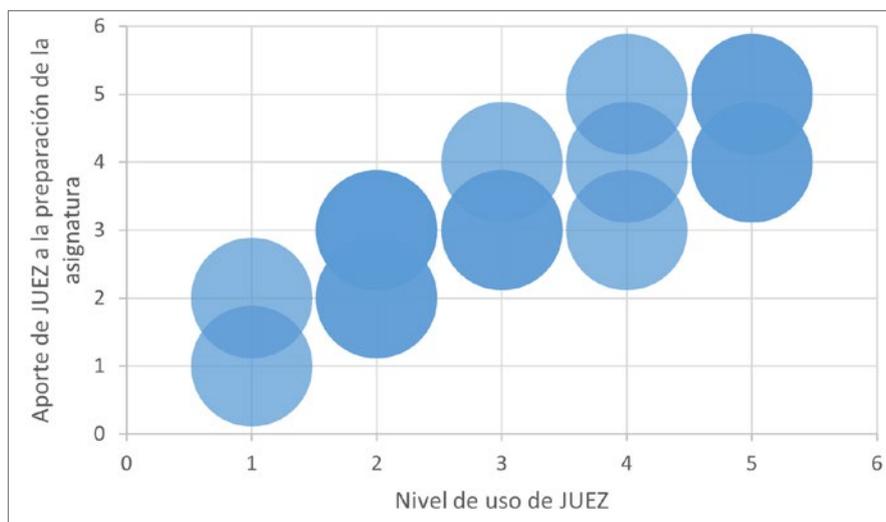


Figura 10. Relación entre el nivel de uso de JUEZ y la preparación de la asignatura.

4. DISCUSIÓN Y CONCLUSIONES

La correlación positiva entre la puntuación en el JUEZ y la calificación de la asignatura se evidencia en la Figura 4. La tendencia se puede explicar con la regresión lineal representada, con un coeficiente de correlación de 0.42 y un p -valor de 0.0086 según la prueba de Wald que tiene por hipótesis alternativa que la pendiente sea positiva. Esto supone que se puede asegurar que existe una interrelación positiva entre la puntuación que el alumnado consigue en la plataforma y sus resultados académicos. Aunque esta correlación no demuestra por sí sola que se mejoren las calificaciones del alumnado gracias al juez (cuestión a la que nos acercaremos con el análisis de los datos de la encuesta), sí nos sirve para validarlo como una herramienta para la detección de casos de talento. La revisión individual de los casos que aparecen en la porción derecha de la Figura 4 (aún antes de disponer de la evaluación de la asignatura que se corresponde con el eje vertical) revelaron varios de estos perfiles, que resolvieron en algunos casos problemas de desafío de la plataforma que, aún adecuándose a los contenidos de la asignatura, tenían una complejidad mayor que la que se espera llegar a poder resolver superando esta.

Una visión más profunda de la diferencia entre los alumnos que participaron en la plataforma y los que es la que ofrece la Figura 3, donde a simple vista se observa que la nota media (líneas discontinuas verticales) de los alumnos que participaron es superior a la de los que no. Este efecto es todavía mayor en caso de considerar aquellos alumnos que resolvieron al menos el 75% de los problemas, lo que se corresponde aproximadamente con la cantidad de trabajo recomendada en las prácticas de la asignatura. Recurriendo a un contraste estadístico podemos verificar que la diferencia entra las calificaciones es significativa: aplicando la prueba t de Welch (1947) –esto es, suponiendo que las muestras son independientes, pero sin hipótesis sobre las distribuciones individuales y sin suponer heterocedasticidad– con hipótesis alternativa de que la media de la porción del alumnado que ha utilizado el JUEZ es superior a la de la que no lo ha hecho, demostrándose así esta con un p -valor de.

Pasamos a continuación a analizar las impresiones del alumnado sobre la plataforma. El principal obstáculo que identifica para la adopción de esta, como muestra la Figura 5, es la “falta de tiempo”, así como la muy relacionada “quería ceñirme a otros materiales para preparar la asignatura”. Este problema de limitación de recursos temporales que dedicar a la asignatura se ha confirmado en varias entrevistas individuales con los estudiantes durante las tutorías, en las que han manifestado que, desde su punto de vista, existe una carga de trabajo continuada generada por las actividades de las diversas asignaturas que requiere su atención. Paradójicamente, solo un estudiante reflejó que el hecho de que no fuera parte directa de la calificación de la asignatura era causa de su falta de participación, lo que choca con la visión de los estudiantes de estar agobiados por otras actividades evaluables. Una posible interpretación de esto es que no se ven como *mercenarios* que trabajan para conseguir una calificación, sino que dedican recursos para aprender una asignatura o para preparar la prueba de evaluación de la misma. Finalmente, merecen especial mención las dos respuestas no preestablecidas que dos estudiantes aportaron, ambas relacionadas con la dificultad que supone trabajar con pares entrada y salida, que constituyó la duda más consultada al equipo encargado de la plataforma y que elaboró una guía específica con ejemplos para facilitar la adaptación al paradigma de trabajo.

Sobre esta última cuestión también aportan más información las opiniones recogidas en la Figura 6 y la Figura 7, donde el alumnado reconoce la existencia de una carga de trabajo adicional para la adopción de la herramienta, que identificamos con el trabajo en base a pares entrada/salida ya presentado.

El papel que el alumnado da al JUEZ sobre la preparación de la asignatura se veía en la Figura 8, que podría dar la impresión de una valoración mixta. Sin embargo, es importante observar que la encuesta la han contestado tanto alumnos que han usado la plataforma como otros que no, y es natural que para estos últimos la herramienta no haya sido de utilidad. En previsión de esto, la encuesta recogía también una valoración en escala Likert del nivel de uso que han hecho en la plataforma. La interrelación de estas variables está representada en la Figura 10, en la que se puede observar que aquellos alumnos que han utilizado lo suficiente la plataforma (tres o más puntos, según su percepción) creen que la experiencia ha sido útil para preparar la asignatura.

Finalmente, la Figura 9 recoge la valoración global del alumnado sobre la experiencia. El 72% refiere una experiencia positiva, mientras que ningún estudiante la refiere negativa. No obstante, esta cifra debe interpretarse con prevención, pues el hecho de que los participantes hayan elegido libremente tomar parte de la experiencia del juez puede hacer que están sobrerrepresentados los más participativos.

Como conclusión, está claro que la experiencia realizada ha sido positiva. Entre las principales virtudes de la herramienta podemos destacar el impacto que produce sobre los estudiantes (apoyados en la percepción que estos reflejan en la encuesta), su valor orientador sobre el nivel del alumnado en la asignatura (por la correlación encontrada y por la diferencia de medias estadísticamente significativa), su utilidad para corregir dudas (como ha mostrado la interacción entre docentes y alumnos durante la experiencia) y la capacidad de detectar talentos entre el alumnado. La principal dificultad que supone es que el alumnado debe aprender a trabajar con pares de entrada/salida, si bien creemos que esta habilidad resulta de interés en sí misma por su interrelación con la técnica del desarrollo orientado a pruebas, por lo que es una competencia que es beneficioso desarrollar. Otra dificultad que se ha de considerar es la necesidad de minimizar la carga de actividades, según la opinión del alumnado. Una forma de afrontar esto puede ser acotar una cantidad reducida de actividades obligatorias, manteniendo el resto opcionales. La disponibilidad del diagnóstico del juez en cualquier momento

del curso puede facilitar también el reparto de tiempo del estudiantado, minimizando así el efecto de la habitual coincidencia temporal de actividades debido al calendario académico y la disponibilidad horaria del profesorado. En cuanto los puntos negativos, creemos que el hecho de que el alumnado tenga la certeza de que una implementación supera la evaluación del juez en línea puede inducirlos a recopilar y compartir respuestas. Este fenómeno no es exclusivo de esta metodología y ya se ha observado anteriormente con otro tipo de entregas de ejercicios, sin embargo; creemos que en caso de que la metodología se implemente de forma que sea parte de la evaluación de la asignatura es conveniente prever contramedidas para detectar este tipo de escenario.

Como potencial trabajo futuro destacamos la extensión del soporte del juez a otras asignaturas, revisando aspectos de diseño de este a partir de los datos recogidos en esta experiencia, así como el estudio del potencial fenómeno de plagio en la plataforma. Para reducir las retenciones del alumnado a su adopción se planteará la demostración de algún ejercicio por parte del equipo docente, ilustrando dificultades como lo estricto del formato de entrada y salida.

5. REFERENCIAS

- Beck, K. (2003). *Test-driven development: by example*. Addison-Wesley Professional.
- DMOJ (2017, febrero, 21) *Point System Rework*. DMOJ. Recuperado el 5 de julio de 2022 de <https://dmoj.ca/post/103-point-system-rework>.
- Edwards, S. H., & Perez-Quinones, M. A. (2008, June). Web-CAT: automatically grading programming assignments. *Proceedings of the 13th annual conference on Innovation and technology in computer science education* (pp. 328-328).
- Enström, E., Kreitz, G., Niemelä, F., Söderman, P., & Kann, V. (2011). Five years with Kattis—using an automated assessment system in teaching. *Frontiers in education conference (FIE)* (p. T3J-1). IEEE.
- Gómez-Martín, P. P., & Gómez-Martín, M. A. (2017). ¡Acepta el reto!: juez online para docencia en español. *Actas de las Jornadas sobre Enseñanza Universitaria de la Informática*, 2, 289-296.
- Halim, S., Halim, F., & Effendy, S. (2018). *Competitive Programming 4: The New Lower Bound of Programming Contests in the 2020s*. Lulu.
- Hernández González, F., Rodríguez Morales, J. D., & Ripoll Méndez, D. A. (2021). The Cuban Olympiad in Informatics: A New Stage from the DMOJ Online Judge.
- Ínce, E. Y. (2021). Students' Perceptions on Learning Programming with CodinGame. *International Journal of Technology in Teaching & Learning*, 17(1).
- Likert, R. (1932). A technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 1-55.
- Polito, G., y Temperini, M. (2018). A gamified approach to automated assessment of programming assignments. *Challenges and Solutions in Smart Learning* (pp. 3-12). Springer.
- Revilla, M. A., Manzoor, S., & Liu, R. (2008). Competitive learning in informatics: The UVa online judge experience. *Olympiads in Informatics*, 2(10), 131-148.
- Welch, B. L. (1947). The generalization of 'STUDENT'S' problem when several different population variances are involved. *Biometrika*, 34(1-2), 28-35. <https://doi.org/10.1093/biomet/34.1-2.28>