

Análisis con IA del modo competición de Wooclap a través de las grabaciones de las clases

Óscar Cánovas Reverte, Pilar González Férez, Félix J. García Clemente, Federico Pardo García
Departamento de Ingeniería y Tecnología de Computadores
Universidad de Murcia
30100 Murcia

ocanovas@um.es, pilargf@um.es, fgarcia@um.es, federico.pardog@um.es

Resumen—Este artículo presenta los hallazgos derivados de la implementación de diferentes modos de utilización de un sistema interactivo de respuesta en una asignatura del Grado en Ingeniería Informática. La investigación se centra en el empleo de la plataforma Wooclap, que permite plantear preguntas a los estudiantes para que respondan en tiempo real mediante sus dispositivos móviles. La experiencia diseñada aborda dos modalidades de uso de la plataforma: una sin competición y otra con competición basada en puntos. La metodología empleada involucra dos grupos distintos de estudiantes que respondieron a las mismas preguntas, dirigidos por el mismo docente en pruebas separadas, utilizando diferentes modos de la plataforma. Este enfoque cuasi-experimental se basa en la recopilación de datos sobre el rendimiento de los estudiantes (respuestas correctas a las preguntas) y el nivel general de interacción en el aula durante las pruebas. Además, se utiliza un sistema de análisis de grabaciones de audio respaldado por inteligencia artificial para caracterizar la participación de los diversos actores involucrados. Para ampliar el análisis, se examinan las transcripciones de las intervenciones orales del docente, investigando la tipología de dichas intervenciones y evaluando si la cantidad y distribución de estos tipos varían entre el modo de competición y el modo sin competición. Los resultados obtenidos revelan patrones distintivos en el nivel de interacción de los estudiantes durante las pruebas, dependiendo de la modalidad utilizada, así como en el tipo de intervenciones que realiza el docente. A pesar de estas variaciones, los resultados indican que la modalidad de uso no ejerce influencia significativa en el rendimiento obtenido por los estudiantes.

Index Terms—Sistema interactivo de respuesta, Wooclap, competición, investigación cuasi-experimental, análisis de grabaciones de audio, inteligencia artificial.

I. INTRODUCCIÓN

LOS sistemas interactivos de respuesta (también denominados de respuesta de audiencia o, simplemente, cuestionarios interactivos) son herramientas de uso habitual en las aulas en todos los niveles educativos, incluyendo la educación superior [1]. Además, estas herramientas pueden emplearse de formas muy diversas en función del enfoque pedagógico seguido [2].

Entre las distintas modalidades de uso, algunas plataformas proporcionan una experiencia de usuario muy cercana al concurso. Se pone especial énfasis en aspectos de gamificación como las puntuaciones, tablas de clasificaciones o una música y estética acordes con la idea de un concurso. La percepción subjetiva de los estudiantes difiere en función de si la experiencia es más competitiva o menos [3], lo que quizá pueda conllevar a un rendimiento distinto por parte de los estudiantes.

Este artículo presenta una experiencia docente basada en el uso de Wooclap en una asignatura del Grado en Ingeniería Informática. En él se analiza la influencia que puede tener usar o no el modo competición en el rendimiento de los estudiantes durante las pruebas realizadas en el aula. Además se indaga acerca del nivel de interacción que genera cada modo durante la dinámica y si tiene alguna influencia sobre el tipo de intervenciones orales que realiza el docente durante el desarrollo de las pruebas. Teniendo todo ello en consideración, estas son las tres preguntas de investigación planteadas:

- Pregunta 1: ¿Existen diferencias significativas en las puntuaciones obtenidas al utilizar el modo competición en Wooclap frente a no usarlo?
- Pregunta 2: ¿Existen diferencias en los patrones de interacción de los estudiantes en el aula dependiendo del modo de uso de Wooclap?
- Pregunta 3: ¿Existen diferencias en los tipos de intervenciones que realiza el docente en función del uso o no del modo competición?

En relación con la dinámica generada en el aula, muchos resultados de investigaciones educativas han demostrado que la interacción exitosa de los estudiantes en un aula, es decir, compartir ideas con el profesor o participar en discusiones, se correlaciona con un aprendizaje de alta calidad [4]. Las interacciones entre profesores y alumnos puede medirse a través de métodos de observación estandarizados que proporcionen a los profesores datos sobre características relevantes de las interacciones en el aula [5]. En nuestro caso analizamos los audios de las grabaciones realizadas en el aula durante el desarrollo de las pruebas que son objeto de estudio, utilizando para ello un sistema de análisis basado en técnicas de aprendizaje computacional.

La principal contribución de este trabajo es analizar si el rendimiento de los estudiantes al usar un sistema interactivo está influenciado por el uso o no del modo competición de la herramienta. Para ello usamos técnicas de inteligencia artificial para el análisis del audio de las grabaciones. Analizamos tanto las características del discurso no verbales como las transcripciones de las intervenciones del docente, y cómo pueden variar en base a usar o no el modo competición.

Los resultados obtenidos en esta experiencia docente con estudiantes de Ingeniería en Informática parecen indicar que el modo competición no produce puntuaciones significativamente

distintas respecto al modo sin competición, pero sí genera mayor interacción en el aula durante el desarrollo de las pruebas y tiene cierta influencia en la forma en la que el docente desarrolla las pruebas usando Wooclap.

Este es un artículo que extiende la investigación publicada en las actas del congreso JENUI 2023 [6] y que fue elegido para su envío a IEEE-RITA al ser reconocido como uno de los mejores de aquel congreso. La ampliación respecto al artículo de partida consiste en la inclusión de una nueva pregunta de investigación (Pregunta 3) acerca de la tipología de las intervenciones orales del docente. El uso de la tecnología de transcripción y el etiquetado de las frases nos ha permitido visualizar y analizar para este trabajo si existen diferencias en función de si se usa o no el modo competición.

II. HERRAMIENTAS UTILIZADAS

En esta sección presentamos los detalles de las dos tecnologías implicadas en la experiencia docente. Primero, describimos brevemente Wooclap, la herramienta que ha sido objeto de estudio. A continuación, exponemos las principales características del sistema de análisis de audio empleado para medir el nivel de interacción en el aula durante el desarrollo de las pruebas.

II-A. Wooclap

Wooclap pertenece a la categoría de los denominados sistemas interactivos de respuesta. Los sistemas de respuesta en el aula no son una tecnología reciente, dado que hay experimentos documentados en la década de 1960 [7], pero es cierto que hoy en día son más fáciles de usar, más completos y sólo se necesita un teléfono inteligente convencional para responder, por lo que se han popularizado mucho en la última década.

En un uso típico de esta tecnología, a los estudiantes se les presentan preguntas de verdadero/falso o de opción múltiple que tienen que responder utilizando la aplicación o el hardware asociado. A continuación, el docente muestra las respuestas y proporciona retroalimentación, que puede implicar una discusión colectiva posterior. Varios estudios, por ejemplo [8], [3], resumen algunos usos comunes de este tipo de tecnologías documentados en la literatura, que no deben reducirse únicamente a plantear cuestiones de repaso.

Existen muchas plataformas, sobradamente conocidas, como Kahoot!, Socrative, Wooclap, Quizizz, Quizlet o Poll Everywhere, por nombrar algunas. Wooclap es una herramienta muy similar a Socrative, tanto en lo funcional como en el aspecto visual, alejada de la estética de concurso de Kahoot!. Sin embargo, Wooclap incluye gran parte de la versatilidad de Kahoot!, como es la gran diversidad de tipos de preguntas a plantear y la posibilidad de insertar las preguntas entre las diapositivas que forman parte de una presentación. También dispone de un modo competitivo, aunque con menos elementos de gamificación que Kahoot!. Este modo competitivo se diferencia del modo sin competición en los siguientes aspectos:

- Los estudiantes deben identificarse siempre, bien con un seudónimo o con el nombre real.

- Los estudiantes obtienen puntuaciones por las respuestas correctas, en función del orden en el que llega su respuesta respecto a la del resto de estudiantes, siendo mayor cuanto antes se envía la respuesta.
- El docente puede mostrar en cualquier momento la clasificación en pantalla, pero solo si pulsa el botón correspondiente, ya que por defecto no se muestra la clasificación.

Podemos encontrar en la literatura algunos trabajos de investigación sobre Wooclap que analizan la implicación del alumnado en función del uso de la funcionalidad de mensajes [9], pero este trabajo es el primero en analizar las consecuencias del modo competición en el rendimiento y en la interacción.

II-B. Análisis de audio

Con el fin de medir el nivel de interacción de los estudiantes durante el desarrollo de las pruebas, empleamos técnicas de análisis de audio basadas en el aprendizaje automático. Este sistema de análisis no verbal se ha utilizado anteriormente [10] para analizar los audios de diferentes actividades en el aula, como las clases magistrales, la resolución de problemas, el aula invertida o el uso de cuestionarios interactivos. Se trata, por tanto, de una herramienta idónea para estudiar las diferencias entre el uso del modo competición y la ausencia de la misma con Wooclap.

El procesamiento del audio realizado por dicho sistema conlleva varias etapas, como queda representado visualmente en la Figura 1. En primer lugar, se utiliza un método de umbralización para separar el silencio de los segmentos de voz y, utilizando pequeños segmentos de audio, se calculan los coeficientes cepstrales de frecuencia de Mel (MFCC), el tono y la energía media de cada segmento de voz. Estas características se utilizan ampliamente en la literatura científica [11] como entrada para la diarización del hablante.

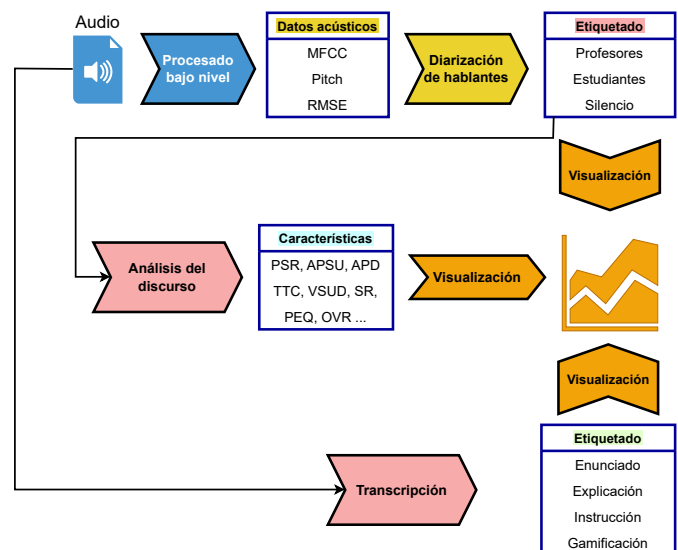


Figura 1: Fases del procesamiento del audio.

La diarización es la segunda etapa y consiste en determinar “quién habló cuándo”, es decir, etiquetar segmentos de la

grabación con el correspondiente hablante. En este artículo no distinguimos qué estudiante concreto está hablando, y todos los estudiantes que intervienen se etiquetan simplemente como Estudiantes. En concreto, distinguimos entre Docente, Estudiantes y Silencio. Este proceso de diarización se basa en técnicas de aprendizaje automático y es un proceso semi-supervisado, ya que hay que entrenar el sistema con algunos fragmentos de audio que se etiquetan como Docente o Estudiantes. Este método proporciona una precisión de clasificación en torno al 95 % con otros audios de prueba. En consecuencia, obtenemos una secuencia de etiquetas ordenada en el tiempo que se utiliza en la siguiente etapa, pero que también puede visualizarse mediante líneas de tiempo. En esta etapa no ponemos en el foco en qué se está diciendo en cada momento, sino simplemente en quién interviene, como hacen otros trabajos previos de análisis no verbal del clima en el aula [12].

El tercer paso es la obtención de las características del discurso no verbal a partir de la secuencia de etiquetas. De todas las características que el sistema proporciona, basadas en los trabajos previos [13], [14], se han empleado las que nos resultan más útiles para este trabajo. Específicamente, empleamos dos características de discurso por rol (profesor, estudiantes):

- Ratio de participación del hablante (PSR). Mide la relación de participación de cada rol durante la grabación.
- Duración promedio de la intervención del hablante (AP-SUD). Medida en segundos, es la duración promedio de las intervenciones de cada tipo de participante.

Además, también se emplean, entre otras, las siguientes características globales de discurso:

- Duración promedio de la pausa (APD). Mide cuán largo es el intervalo de una pausa entre intervenciones.
- Conteo de turnos (TTC). Representa cuántos cambios de turno ocurrieron en el diálogo entre estudiantes y docente.

El cuarto paso consiste en la transcripción del audio a texto, obteniendo así “qué se habló cuándo”. Este proceso se realiza exclusivamente en los segmentos de audio pertenecientes al docente, haciendo uso de la diarización extraída en el segundo paso y obteniendo una versión del audio recortada únicamente con sus intervenciones. La transcripción del audio se lleva a cabo mediante el sistema de reconocimiento Whisper [15], que permite obtener no solo la transcripción textual del habla, sino también marcas de tiempo de inicio y fin de cada intervención. Estas transcripciones generadas por Whisper tienen que ser ajustadas manualmente por un operador humano para ser fieles a lo hablado en el audio, si bien son un buen punto de partida dado que se logra una tasa de error WER (Word Error Rate) [16] del 25 %.

Una vez se obtiene la transcripción, se procede a separarla por oraciones. Dado que no estamos interesados en cada oración individual sino en la función de dichas intervenciones orales, debemos llevar a cabo un proceso de categorización del discurso del docente similar al realizado en trabajos previos para disciplinas como el aprendizaje de las Matemáticas [17] o la enseñanza del proceso de argumentación [18]. En nuestro

caso, se etiqueta cada oración en alguna de las categorías definidas, siendo estas: “Enunciado”, “Explicación”, “Instrucción” y “Gamificación”. Este esquema de etiquetado, explicado en la Sección III-C, permite separar las intervenciones del docente para su análisis, facilitando la comparación en diferentes contextos. Actualmente, para este trabajo, este proceso de etiquetado de las transcripciones se realiza de forma manual por operadores humanos. No obstante, los siguientes pasos irán dirigidos a la automatización de dicho proceso de etiquetado mediante técnicas de análisis automático y modelado del discurso basado en inteligencia artificial, como ya se está explorando en trabajos recientes [19], [20].

Finalmente, el sistema nos permite la visualización de los datos, utilizando diferentes técnicas de representación para mostrar las características calculadas, el etiquetado de las transcripciones y también los resultados del proceso de diarización. Como vemos más adelante, estos gráficos proporcionan información valiosa para analizar las actividades del aula, como los niveles de interacción.

III. CONTEXTO DE LA EXPERIENCIA

III-A. Materia y participantes

Este trabajo presenta la experiencia docente desarrollada durante el curso académico 2022/23 en la asignatura Fundamentos de Computadores, de primer curso del Grado en Ingeniería Informática de la Facultad de Informática de la Universidad de Murcia. Se trata de una asignatura de 6 créditos ECTS, 3 de teoría y 3 de prácticas. En el grupo que se ha realizado el estudio hay un total de 68 estudiantes matriculados, de los cuales han asistido a clase regularmente alrededor de 55. Para las clases prácticas los alumnos se dividen en tres grupos, cada grupo está compuesto por, aproximadamente, 22 alumnos. El docente de todos los grupos es siempre la misma persona.

En esta asignatura la teoría y las prácticas están estrechamente relacionadas, de forma que las clases de prácticas se usan para afianzar conceptos de teoría a través de ejercicios teórico-prácticos. Estos ejercicios algunas veces siguen el modelo tradicional y se desarrollan en pizarra, mientras que otras veces son ejercicios que se apoyan en el ordenador, usando alguna herramienta informática. Del total de clases prácticas, más del 75 % de las mismas usan herramientas informáticas, mientras que en menos del 25 % de las mismas no se usa el ordenador y se realizan ejercicios en pizarra siguiendo el modelo tradicional.

La metodología usada en esta asignatura abarca un gran abanico de opciones. Así, se hace uso de clases magistrales para introducir los conceptos de la misma. En las clases de prácticas, que se llevan a cabo en laboratorio con grupos reducidos, se resuelven problemas en ordenador/pizarra para afianzar dichos conceptos, en este caso se fomenta en todo momento la participación del alumno, tanto al resolver el docente el ejercicio, como proponiendo que los estudiantes resuelvan individualmente parte de los ejercicios (pero con apoyo del profesor cuando es necesario). Además, en dos de las sesiones teóricas del curso se hace uso de la metodología de aula invertida. Por último, al final de cada tema, se utiliza la herramienta Wooclap con cuestiones específicas del tema

correspondiente, lo que permite resolver dudas y comprobar cómo los alumnos han afianzado los conceptos del mismo. El único cambio sustancial respecto a años anteriores es que dos de las pruebas de Wooclap se han movido de las sesiones de teoría a las sesiones de laboratorio, para poder realizar nuestra investigación.

III-B. Diseño de actividades

Usamos la herramienta Wooclap como sistema interactivo de respuesta para hacer nuestro estudio. Las dos pruebas realizadas con Wooclap en las sesiones de prácticas se configuran de la misma manera: están compuestas por 10 preguntas, con un tiempo máximo de respuesta de 60 segundos por pregunta.

La primera prueba está compuesta de nueve preguntas de “Buscar en una imagen” y una pregunta de respuesta múltiple (con tres posibles respuestas y sólo una de ellas válida). En las preguntas de “Buscar en una imagen” se presenta la imagen de una placa base completa, se le pregunta por un componente de la misma, y el alumno tiene que marcar qué zona de la imagen se corresponde con dicho componente.

En la segunda prueba las diez preguntas giran alrededor de los distintos lenguajes del computador. Son de respuesta múltiple, teniendo cada pregunta tres posibles respuestas, de las cuales sólo una es válida.

El estudio se ha llevado a cabo durante las sesiones de prácticas y sólo dos de los tres grupos de prácticas han participado en la investigación aquí presentada. En uno de los grupos prácticas se han tenido que hacer adaptaciones pedagógicas específicas debido al alumnado presente, por lo que se sigue una metodología algo diferente y no se incluye en esta investigación.

Por tanto, en la investigación participan los grupos 2.1 y 2.2, que en adelante llamaremos G. 2.1 y G. 2.2, respectivamente. En ambos grupos, y en ambas sesiones, se realizan las mismas pruebas, salvo en el aspecto de la competición. En concreto, para el G. 2.1 la primera prueba se configura sin competición, mientras que para el G. 2.2 se configura con competición autenticada. Para la segunda prueba, esta configuración se invierte, de forma que el G. 2.1 realiza la segunda prueba con competición autenticada y el G. 2.2 la hace sin competición. El horario de los grupos de prácticas es consecutivo dentro de la misma mañana.

III-C. Procedimiento de obtención de datos

Con el fin de responder a las preguntas de investigación, se han recopilado varios elementos de información a partir de diversas fuentes de datos. En primer lugar, se dispone de los resultados obtenidos a través de Wooclap. Se trata de las puntuaciones numéricas de los estudiantes para cada prueba realizada.

En segundo lugar, se dispone de los audios de clase grabados al realizar la prueba de Wooclap. Respecto a la grabación de los audios, todos los estudiantes fueron informados acerca del procedimiento (aprobado por la institución en el marco de un proyecto de innovación docente), la finalidad del estudio y el compromiso de no desvelar por ningún medio la identidad

Audio hablado	Transcripción de Whisper
¿Dónde está el Puente Norte? Puente Norte. Puente Norte	¿Dónde está el Puente Norte?
A ver, se os ha salido un poco, ¿no?	A ver, ¿se os ha salido un poco no?
El ISA de una máquina CPU RISC...	Elisa de una máquina CPU RISK...

Cuadro I: Ejemplos de errores en la transcripción de Whisper.

de los participantes intervinientes. Para efectuar las grabaciones se situó una grabadora digital de mano (TASCAM DR-07X) a una distancia mínima de 1,5 metros del profesor y de los alumnos de la primera fila. Posteriormente se procesaron los audios para seleccionar únicamente los fragmentos que abarcan desde el instante en el que se lanzaba el enlace para conectar hasta el momento en el que se da por concluida la prueba. El resultado son cuatro grabaciones con duraciones de entre 13 y 15 minutos.

Posteriormente dichas grabaciones se procesan por el sistema de análisis de audios [21] para generar la diarización de las mismas, la extracción de características y la generación y etiquetado de las transcripciones, específicamente de las intervenciones del docente.

La generación de las transcripciones mediante el sistema Whisper, sirven como punto de partida, pero el sistema cometía ciertos errores que se corrigieron a mano. La mayoría de estos errores se centraban en la repetición de frases o palabras por parte del docente, muy comunes en el lenguaje hablado, o en la dificultad para la transcripción de palabras técnicas relacionadas con la materia impartida. En menor medida, existían ciertos errores con la puntuación de las frases, mediante el uso de comas, puntos, exclamaciones o interrogaciones en lugares que no correspondían. Un pequeño ejemplo de cada tipo de error cometido por Whisper se puede observar en el Cuadro I.

El sistema de etiquetado desarrollado consiste en 4 categorías: “Enunciado”, “Explicación”, “Instrucción” y “Gamificación”. El Cuadro II explica brevemente estas etiquetas y muestra un ejemplo de cada una de ellas. Este sistema de etiquetado fue consensuado por 3 investigadores para el uso en el contexto específico de este trabajo:

- **Enunciado:** Esta etiqueta representa la lectura de los enunciados de las cuestiones a los estudiantes.
- **Explicación:** Agrupa todas las explicaciones que realiza el docente a los alumnos después de una pregunta, con el fin de razonar el porqué de la respuesta o para guiar a los alumnos que puedan haber errado en su razonamiento.
- **Instrucción:** Recoge las intervenciones para el buen funcionamiento de la actividad. Esto incluye intervenciones relacionadas con el tiempo restante para responder una pregunta, cómo interactuar con las máquinas durante la actividad o indicaciones para volver a captar la atención de los alumnos antes de pasar a la siguiente pregunta.
- **Gamificación:** Contiene las intervenciones enfocadas en la participación de los estudiantes. Un ejemplo de esto puede ser la comparación del ranking de puntos obtenido. También engloba oraciones en las que el profesor fomenta la competitividad entre los alumnos.

Etiqueta	Descripción	Ejemplo
Enunciado	Lectura de los enunciados de las cuestiones a los alumnos.	“Señala en esta imagen dónde estaría la CPU.” “¿Dónde estaría el Puente Norte?” “¿Cuál de estas instrucciones es la más adecuada para acceder a un array?”
Explicación	Incluye las explicaciones aportadas por el docente acerca de las respuestas correctas a las preguntas, normalmente una vez estas han finalizado.	“Estos son dispositivos para RAM, pero ahí no tenemos la RAM” “North Bridge es la parte rápida. South Bridge para la parte lenta” “Las instrucciones del ensamblador nunca llevan dos puntos”
Instrucción	Engloba las intervenciones relacionadas con el correcto desarrollo de la actividad, como indicaciones sobre cómo responder o tiempo restante en la pregunta.	“Tenéis que pinchar sobre la imagen”. “Quedan 30 segundos” “¿Quién falta por responder?”
Gamificación	Identifica intervenciones en las que se fomenta la competición entre los alumnos o el análisis de las puntuaciones obtenidas.	“Bueno, ¿qué? ¿Alguien se moja?” “Vamos a ver cómo van las puntuaciones” “En primera posición va Julio César, segundo Colón y tercero Turing”

Cuadro II: Descripción de las etiquetas de transcripción.

Para este trabajo, se decide no utilizar la etiqueta “Gamificación” en el caso de los Wooclaps no competitivos, pues no consideramos que la gamificación fuese algo apreciable si no se consideran las puntuaciones de los alumnos. Las intervenciones que hacen referencia a porcentaje de acierto o fallo en este caso se engloban como “Instrucción”.

IV. RESULTADOS

En esta sección se analizan los datos recogidos para responder a las tres preguntas de investigación formuladas en la sección introductoria.

Para analizar el rendimiento obtenido en ambas pruebas se han analizado las distribuciones de las puntuaciones obtenidas. Los Cuadros III y IV muestran el rendimiento alcanzado en las pruebas 1 y 2, respectivamente, junto con el número de estudiantes que participaron. En concreto, en dichas tablas usamos la siguiente notación: N para el número de participantes, Media para la nota media, D.E. para la desviación estándar y la p de Shapiro-Wilk para comprobar el criterio de normalidad en las puntuaciones. “Comp.” indica que la prueba se configuró con competición, mientras que “Sin Comp.” indica que la prueba se configuró sin competición. Para completar nuestro análisis descriptivo la Figura 2 presenta las distribuciones de rendimiento obtenidas por ambos grupos en ambas pruebas.

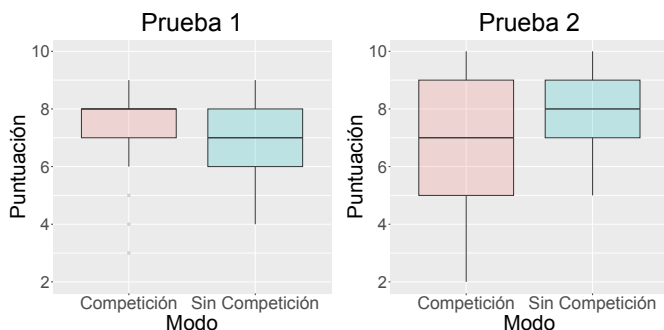


Figura 2: Distribución de puntuaciones para las pruebas 1 y 2 en base al modo competición en Wooclap.

IV-A. Prueba 1

A partir de los datos del Cuadro III podemos comprobar que, dado un número similar de participantes, los resultados

	Grupo	Valor
N	Comp. (G. 2.2)	18
	Sin Comp. (G. 2.1)	20
Media	Comp. (G. 2.2)	7.28
	Sin Comp. (G. 2.1)	6.75
D.E.	Comp. (G. 2.2)	1.74
	Sin Comp. (G. 2.1)	1.33
p de Saphiro-Wilk	Comp. (G. 2.2)	0.004
	Sin Comp. (G. 2.1)	0.34

Cuadro III: Descriptivas del rendimiento de la Prueba 1.

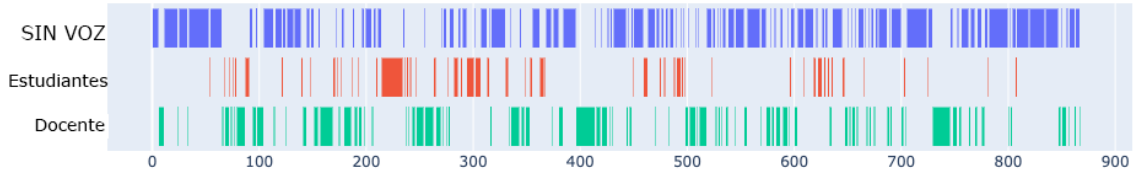
	Grupo	Valor
N	Comp. (G. 2.1)	19
	Sin Comp. (G. 2.2)	17
Media	Comp. (G. 2.1)	6.84
	Sin Comp. (G. 2.2)	8.06
D.E.	Comp. (G. 2.1)	2.22
	Sin Comp. (G. 2.2)	1.43
p de Saphiro-Wilk	Comp. (G. 2.1)	0.168
	Sin Comp. (G. 2.2)	0.271

Cuadro IV: Descriptivas del rendimiento de la Prueba 2.

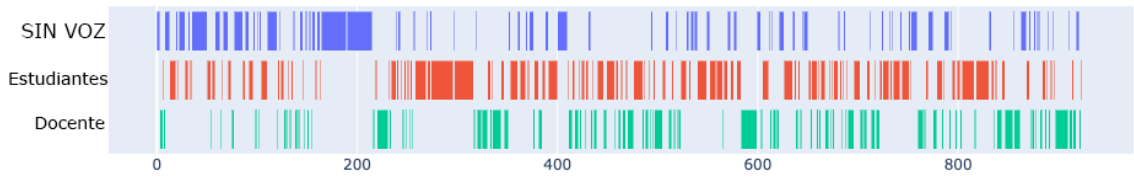
del grupo con competición (G. 2.2) fueron superiores en media a los del grupo sin competición (G. 2.1). No obstante, la Figura 2 parece indicar que las diferencias no son estadísticamente significativas. Para corroborarlo, dado que no se cumple el criterio de normalidad en las puntuaciones del grupo 2.2, se ha empleado el test no paramétrico de Kruskal-Wallis, con $\chi^2 = 2,32$ y $p = 0,127$, lo que confirma que las diferencias no son significativas ($p > 0,05$).

Sin embargo, si analizamos las grabaciones de audio de la Prueba 1 podemos observar diferencias en el patrón de interacción. Las Figuras 3a y 3b muestran líneas temporales en las que se resaltan los instantes en los cuales intervienen los distintos tipos de participantes. Un análisis comparativo de ambas indica un mayor porcentaje de intervención de los estudiantes en el modo con competición. De hecho, el grado de participación casi se triplica, como se aprecia en las características de PSR (ratio de participación de hablante) y TTC (conteo de turnos) que se ilustran en las Figuras 3c y 3d.

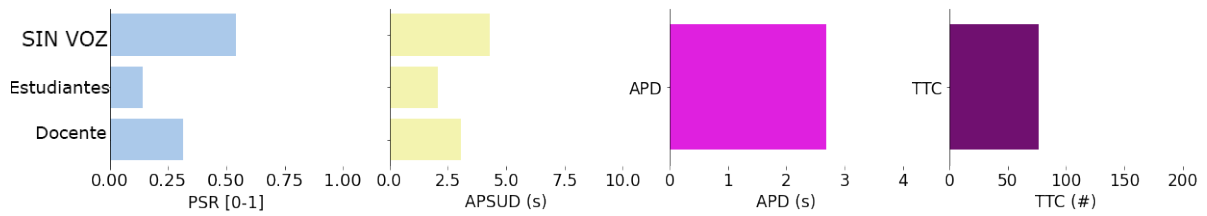
Una vez que hemos comprobado que se registra una mayor cantidad de audio proveniente de los estudiantes en el modo competición, vamos a analizar la naturaleza de las intervenciones del docente para cada tipo de prueba. Las Figuras 3e y 3f, en las que se analizan los instantes en los que se



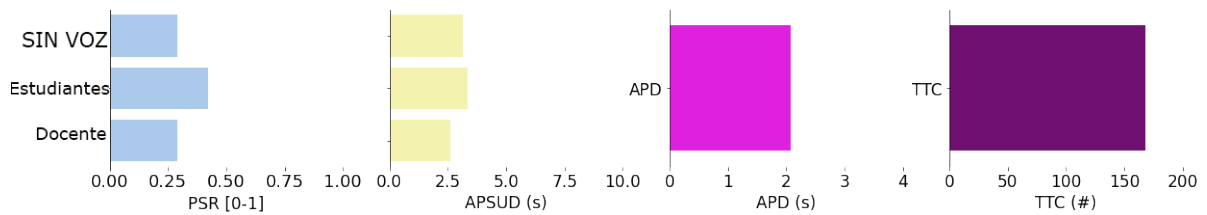
(a) Línea temporal de intervenciones. Sin competición (G. 2.1).



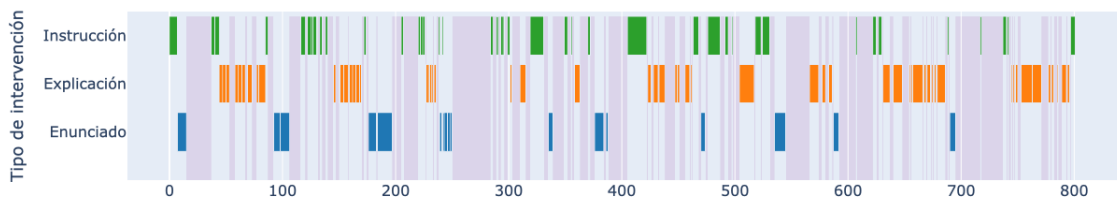
(b) Línea temporal de intervenciones. Con competición (G. 2.2).



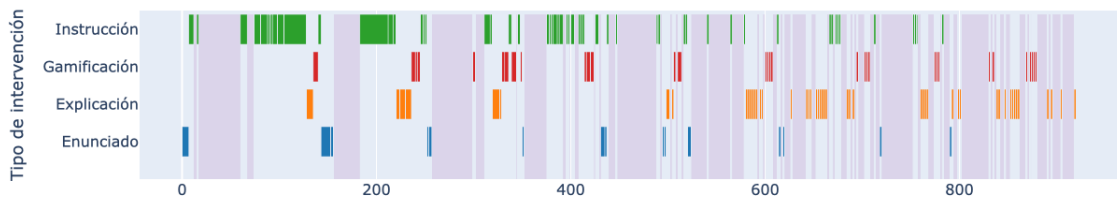
(c) Descriptores de los audios. Sin competición (G. 2.1).



(d) Descriptores de los audios. Con competición (G. 2.2).



(e) Línea temporal del tipo de intervenciones. Sin competición (G. 2.1).



(f) Línea temporal del tipo de intervenciones. Con competición (G. 2.2).

Figura 3: Resultados del procesamiento de los audios de la Prueba 1.

producen intervenciones de cada uno de los 4 tipos explicados en la Sección II-B, muestran patrones bien diferenciados. Por un lado, como cabría esperar, durante las pruebas con competición se realizan comentarios que entran en la categoría de gamificación mientras que no sucede así en la prueba sin competición. Además, parece observarse una secuencia clara de intervenciones para cada pregunta, que comienza con frases relacionadas con el enunciado de la pregunta, un silencio posterior (mostrado mediante un sombreado en el gráfico de color más oscuro), una breve explicación de las respuestas correctas y, en el modo competición, algunas intervenciones de la categoría “Gamificación”. En ambos casos también hay intervenciones distribuidas de forma bastante uniforme de tipo instructivo, para guiar a los estudiantes en el desarrollo de la prueba y uso de la herramienta. Ambos gráficos también parecen indicar que el porcentaje de tiempo dedicado por el docente a cada categoría difiere en función de si hay competición o no. Esta comparativa específica se llevará a cabo de forma más detallada en la Sección IV-E.

IV-B. Prueba 2

Para la segunda prueba, volvemos a disponer de un número similar de participantes en cada grupo. Observamos que la nota media del grupo sin competición (G. 2.2) es superior a la del grupo con competición. En la Figura 2 apreciamos una diferencia en la distribución de puntuaciones, por lo que tenemos que corroborar si la diferencia es estadísticamente significativa. En esta ocasión podemos emplear un test paramétrico dado que se cumple la condición de normalidad en las distribuciones. El resultado con ANOVA de Welch es $F = 3,90$ y $p = 0,057$, por lo que de nuevo se rechaza la hipótesis de que las puntuaciones difieran estadísticamente ($p > 0,05$).

Por tanto, se aprecia que en las pruebas realizadas, el grupo 2.2 siempre saca una nota media más alta, independientemente del modo. Esto queda también reflejado en la Figura 2, ya que para el G. 2.2 la distribución de puntuaciones ocupa siempre un menor rango teniendo el valor mínimo más alto.

Al analizar las grabaciones de audio de la Prueba 2 encontramos diferencias aún más acusadas en los patrones de interacción. Las Figuras 4a y 4b evidencian que el grado de interacción en el modo sin competición es prácticamente inexistente, muy inferior al del grupo 2.1. El silencio es mayoritario durante el 70% del tiempo, como indica la Figura 4d, con una duración de las pausas del docente también notablemente superior. Sin embargo, podemos apreciar que el grupo en modo competición tiene un patrón bastante regular de intervenciones, que en este caso se relaciona con los murmullos y las conversaciones que se producen entre los estudiantes mientras que esperan el resultado de las preguntas o cuando interrogan al docente.

Respecto a la tipología de las intervenciones del docente, las Figuras 4e y 4f evidencian diferencias aún más acusadas entre sí de lo que sucedió en la Prueba 1. Podemos comprobar que la mayoría de las intervenciones del modo sin competición son explicaciones, aunque hay también muchos más intervalos de silencio en esta modalidad. En el caso con competición, también se aprecia una diferencia en el orden del tipo de intervenciones en las primeras preguntas respecto a la Prueba 1. En

este caso se están introduciendo comentarios de gamificación antes de proporcionar explicaciones durante las 6 preguntas iniciales, lo cual evidencia que no se ha seguido el mismo patrón de desarrollo de la competición en ambas pruebas.

IV-C. Respuesta a la pregunta 1

La primera pregunta planteaba si existen diferencias significativas en las puntuaciones obtenidas al utilizar el modo competición en Wooclap frente a no usarlo. Según nuestro análisis la respuesta es no. Los datos indican que estadísticamente no hay diferencias significativas. En este caso, el grupo G. 2.2 obtiene siempre mejores puntuaciones, sin importar si la prueba fue configurada en modo competición o no. Quizá sea importante apuntar que también se analizaron los resultados del examen final de la asignatura (igual para todos los grupos) y que incluye conceptos teórico-prácticos. El objetivo era comprobar si efectivamente los alumnos del grupo G. 2.2 también obtenían mejores calificaciones en media, sin embargo, los resultados fueron similares en ambos grupos. En concreto, calificación obtenida en el examen final por los alumnos del G. 2.1 fue de ($\mu = 5,75, \sigma = 1,51$), mientras que los alumnos del G. 2.2 obtuvieron ($\mu = 5,22, \sigma = 1,95$), sin diferencias estadísticamente significativas entre ambos grupos.

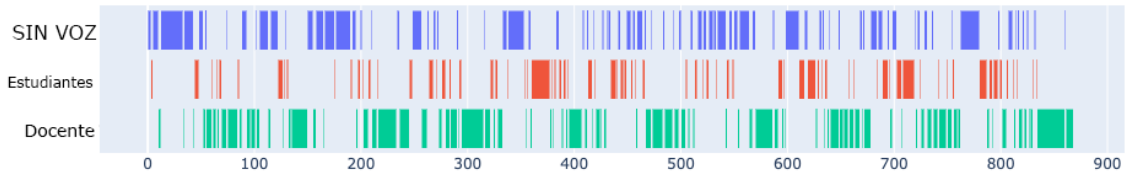
Por tanto, basándonos en los resultados presentados, podemos concluir que usar el modo de competición no influye en el resultado obtenido por los alumnos en nuestra experiencia.

IV-D. Respuesta a la pregunta 2

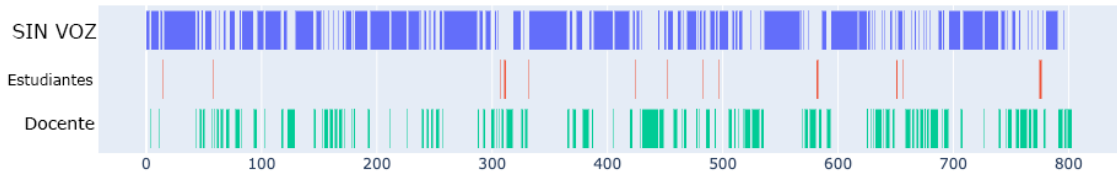
Respecto a la segunda pregunta, acerca de la diferencia en los patrones de interacción de los estudiantes, los datos confirman que sí hay diferencias importantes en las líneas temporales y en las características extraídas de los audios. Las pruebas basadas en competición muestran un nivel de interacción mayor, de forma más uniforme a lo largo del desarrollo de las mismas. La participación de los estudiantes se triplica, al menos, respecto al modo sin competición. También el docente interviene de forma más continua, con pausas más cortas y por tanto de forma más dinámica. Por ejemplo, cuando hay competición el docente tiene que poner los elementos de la gamificación en relieve, como se vio también en las líneas temporales de las Figuras 3f y 4e, lo que incrementa su número de intervenciones. Además, hay más alternancia en las intervenciones entre docente y estudiantes, como indican los valores más elevados de TTC (conteo de turnos).

IV-E. Respuesta a la pregunta 3

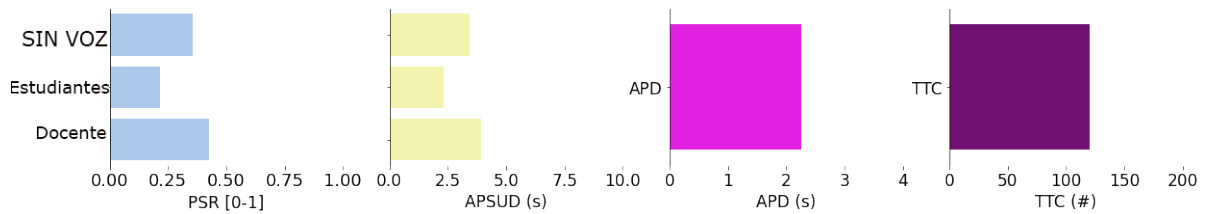
La tercera pregunta planteaba si existen diferencias en los tipos de intervenciones que realiza el docente en función del uso o no del modo competición. Al analizar las líneas de tiempo que mostraban los instantes en los que se realizan los distintos tipos de intervenciones, nos hemos centrado principalmente en la secuencia de los tipos de intervenciones, no en la proporción de las mismas. Para comparar de forma precisa si existen diferencias importantes en el tiempo destinado a cada una de ellas en función del modo, vamos a hacer una valoración empírica de los porcentajes de tiempo resultantes en cada prueba, tal y como muestra la Figura 5.



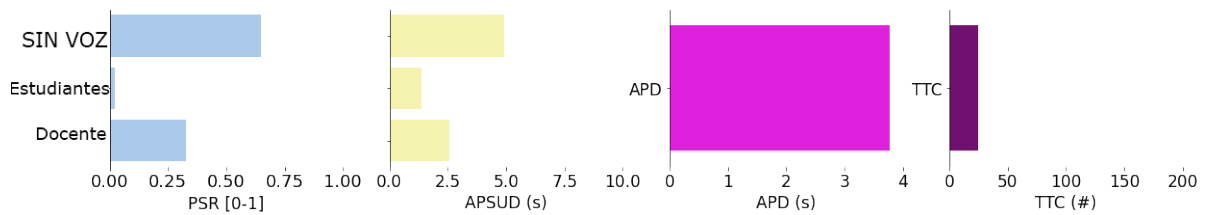
(a) Línea temporal de intervenciones. Con competición (G. 2.1).



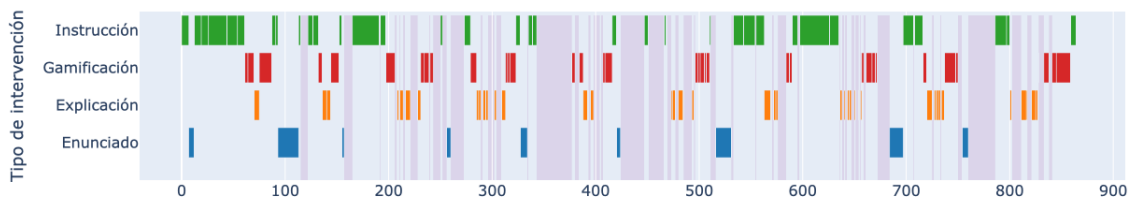
(b) Línea temporal de intervenciones. Sin competición (G. 2.2).



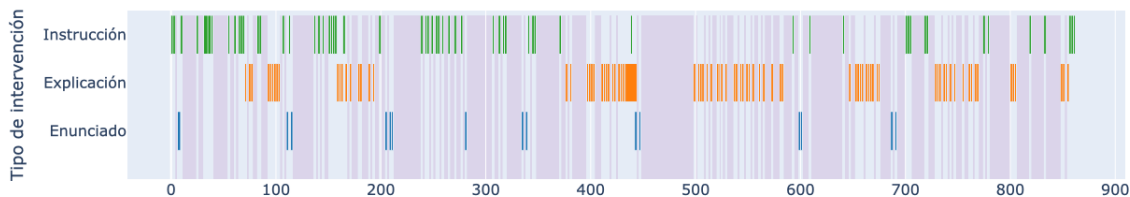
(c) Descriptores de los audios. Con competición (G. 2.1).



(d) Descriptores de los audios. Sin competición (G. 2.2).

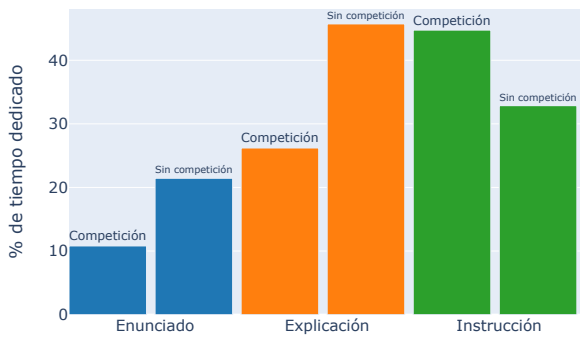


(e) Línea temporal del tipo de intervenciones. Con competición (G. 2.1).



(f) Línea temporal del tipo de intervenciones. Sin competición (G. 2.2).

Figura 4: Resultados del procesamiento de los audios de la Prueba 2.



(a) Tipo de intervenciones para la Prueba 1.



(b) Tipo de intervenciones para la Prueba 2.

Figura 5: Comparación de la proporción de tiempo dedicada para cada tipo de intervención en las Pruebas 1 y 2 en función de si se usa o no el modo competición.

Como se puede apreciar, si comparamos aquellas categorías que son comunes a las dos pruebas en sus dos modalidades, es decir, si excluimos la gamificación, apreciamos que hay diferencias en el porcentaje de cada tipo de intervención a lo largo de la clase. Mientras que en la categoría de “Enunciado” no encontramos una tendencia clara entre las modalidades, en las intervenciones de tipo “Instrucción” se produce un ligero aumento del tiempo en el modo competición y sí encontramos diferencias sustanciales en el porcentaje de tiempo destinado a proporcionar explicaciones acerca de las respuestas correctas de las preguntas. En ambas pruebas el porcentaje se duplica, al menos, en el caso de los cuestionarios Wooclap que se realizan sin competición. Por tanto, para el caso concreto de estas pruebas aquí analizadas, sí podemos afirmar que hay diferencias importantes en el tipo de intervenciones que realiza el docente durante el desarrollo de la clase, especialmente dentro de la categoría “Explicación”.

V. DISCUSIÓN Y CONCLUSIONES

Son varios los estudios que reflejan que el uso de sistemas interactivos de respuesta en clase beneficia el clima generado, favoreciendo la participación activa del alumnado [22] y contribuyendo, en cierta medida, a mejorar los resultados del proceso de aprendizaje.

En este trabajo hemos indagado acerca de si el uso del modo competición o no en estas herramientas influye en el rendimiento de los estudiantes y en el comportamiento

del docente a la hora de conducir las pruebas. Para ello hemos basado nuestro análisis en las evidencias recogidas a través del rendimiento obtenido en unas pruebas con Wooclap configuradas con y sin competición, en los patrones de interacción de los alumnos durante dichas pruebas y en los tipos de intervenciones que realiza el profesorado. Hasta ahora, la mayoría de los trabajos previos se habían limitado a la percepción subjetiva de los estudiantes. En [23], por ejemplo, se contrasta que el uso reiterado de Kahoot! durante varios meses consecutivos no produce un descenso paulatino en la motivación e implicación del alumnado.

Los resultados presentados reflejan aspectos interesantes. En estudios previos se indicaba que los estudiantes percibían que el uso de sistemas de respuesta no gamificados podía ser más apropiado para evaluar conceptos complejos, mientras que los gamificados serían más adecuados para responder a preguntas rápidas sobre conceptos más sencillos [3]. En nuestro caso, el rendimiento obtenido por el estudiante no depende de si la prueba fue configurada en modo competición o no, dado que no hay diferencias estadísticamente significativas. Sin embargo, esta configuración sí influye en el grado de interacción del alumnado en clase, ya que las pruebas basadas en competición muestran un nivel de interacción mayor que las realizadas sin competición. Además, se observa que las intervenciones del docente para proporcionar explicaciones acerca de las respuestas correctas de las preguntas son significativamente mayores al usar el modo sin competición. Este aspecto puede encontrar su justificación tanto en el hecho de que el docente dedica parte del tiempo a intervenciones sobre aspectos de la gamificación y, por tanto, reduce su dedicación a dar explicaciones, como también a que el docente desea centrar la actividad en la competición y no alargar la actividad con explicaciones.

Nuestra experiencia docente a lo largo de las pruebas realizadas indica que el clima generado en clase con las pruebas con competición es más lúdico y ameno, el alumnado está más integrado en las mismas, y parece que muestra más interés aunque sólo sea por el hecho de rivalizar entre compañeros. Esto ya fue resaltado en [24], donde Oigara y Keengwe estudian la percepción del alumnado y muestran que los estudiantes se muestran satisfechos con el uso de estas herramientas, en particular para mejorar su participación e implicación en las clases. Con el modo competición se genera más interacción en todos los sentidos, es decir docente-estudiante y estudiante-estudiante. Desde el punto de vista del docente, como se ha visto en el análisis de las transcripciones, el modo competición introduce nuevos elementos en sus intervenciones, dado que tiene que hacer referencia explícita a las puntuaciones, clasificaciones y los nombres o los seudónimos de los estudiantes, lo que le sitúa en otro rol distinto al habitual por momentos. Por otro lado, los estudiantes también participan más, comentan los resultados y la clasificación. Da la sensación subjetiva de que los estudiantes están más motivados y atentos.

Destacar que este ambiente también puede llevar a un mayor nivel de distracción de los alumnos y un mayor nivel de ruido en el aula entre cada pregunta de la prueba, lo que puede conllevar un mayor esfuerzo por parte del docente para continuar

con la prueba. De hecho, los resultados obtenidos del análisis de las intervenciones muestra que el docente debe dedicar un tiempo significativo a aspectos de la gamificación e incluso aumentar las intervenciones dedicadas al buen funcionamiento de la actividad (es decir, de tipo “Instrucción”) en comparación con el modo sin competición. Este nivel de activación está en la línea de lo apuntado en la revisión sistemática en [25] respecto a la idoneidad de este tipo de herramientas para gestionar la ansiedad de los estudiantes. En el caso de esa revisión, específica para Kahoot!, se comprobaba que el setenta por ciento de los artículos que abordan la ansiedad estudiantil informan de una reducción de dicha ansiedad al jugar a Kahoot! canalizada a través de la participación de los estudiantes

Cuando la prueba se lleva a cabo sin competición, el clima es menos participativo. En este modo los estudiantes parecen prestar más atención a las explicaciones del docente tras cada pregunta formulada, pero es una apreciación subjetiva. Sí parece estar contrastado que Wooclap, se use o no en modo competición, aumenta la participación y la motivación del alumnado de forma generalizada, tanto cuando se usa de forma presencial [26] como en docencia online [27].

La conclusión final es que debe ser el docente el que elija el modo con competición o sin competición según el clima que quiera crear en clase, dado que no parece afectar a las puntuaciones obtenidas. Así, si el docente quiere tener un clima más interactivo, con alumnos más participativos, aunque ello pueda conllevar una posible pérdida de atención a la realimentación que el docente realiza tras las preguntas, debe configurar la prueba en modo competición. Por el contrario, si el docente quiere un ambiente más relajado, en el que los alumnos puedan centrar más la atención en sus explicaciones, es una buena alternativa configurar la prueba sin competición. De hecho, un estudio previo realizado con Kahoot! en el que se examinaba el impacto de usar puntuaciones frente a no usarlas indicaba que los estudiantes perciben el cuestionario igualmente útil y divertido incluso sin puntos [28].

Como vías de trabajo futuro, por un lado, queremos extender el estudio al uso de distintas plataformas de respuesta interactiva, dado que hay diferencias de diseño importantes en el sistema de gamificación que cada una incorpora. Esto nos permitirá analizar si el rendimiento y la interacción también están condicionados por el tipo de plataforma. Por otro lado, estamos desarrollando un proceso automatizado de etiquetado de las transcripciones para reducir el tiempo de intervención de un operador humano.

REFERENCIAS

- [1] J. I. Castillo-Manzano, M. Castro-Nuño, L. López-Valpuesta, M. T. Sanz-Díaz, and R. Yñiguez, “Measuring the effect of ars on academic performance: A global meta-analysis,” *Computers & Education*, vol. 96, pp. 109–121, 2016.
- [2] Ó. Cánovas, “Sistemas interactivos de respuesta: hay vida más allá de los concursos,” *Actas de las XXVII JENUI*, pp. 243–250, 2021.
- [3] W. Wang, S. Ran, L. Huang, and V. Swigart, “Student perceptions of classic and game-based online student response systems,” *Nurse educator*, vol. 44, no. 4, pp. 6–9, 2019.
- [4] T. D. Nguyen, M. Cannata, and J. Miller, “Understanding student behavioral engagement: Importance of student interaction with peers and teachers,” *The Journal of Educational Research*, vol. 111, no. 2, pp. 163–174, 2018.
- [5] J. Archer, S. Cantrell, S. L. Holtzman, J. N. Joe, C. M. Tocci, and J. Wood, *Better feedback for better teaching: A practical guide to improving classroom observations*. John Wiley & Sons, 2016.
- [6] O. Cánovas and P. González Férrez, “Rivalizar o no: análisis del modo competición de wooclap basado en rendimiento y procesamiento de audio,” *Actas de las XXIX Jornadas sobre la Enseñanza Universitaria de la Informática (JENUI)*, pp. 65–72, 2023.
- [7] E. Dill, “Do clickers improve library instruction? Lock in your answers now,” *The Journal of Academic Librarianship*, vol. 34, no. 6, pp. 527–529, 2008.
- [8] R. H. Kay and A. LeSage, “Examining the benefits and challenges of using audience response systems: A review of the literature,” *Computers & Education*, vol. 53, no. 3, pp. 819–827, 2009.
- [9] J. Hutain and N. Michinov, “Improving student engagement during in-person classes by using functionalities of a digital learning environment,” *Computers & Education*, vol. 183, p. 104496, 2022.
- [10] O. Canovas, F. J. García-Clemente, and F. Pardo, “Ai-driven teacher analytics: Informative insights on classroom activities,” in *2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. IEEE, 2023, pp. 1–8.
- [11] T. J. Park, N. Kanda, D. Dimitriadis, K. J. Han, S. Watanabe, and S. Narayanan, “A review of speaker diarization: Recent advances with deep learning,” *Computer Speech & Language*, vol. 72, p. 101317, 2022.
- [12] A. James, Y. H. V. Chua, T. Maszczyk, A. M. Núñez, R. Bull, K. Lee, and J. Dauwels, “Automated classification of classroom climate by audio analysis,” in *9th International Workshop on Spoken Dialogue System Technology*. Springer, 2019, pp. 41–49.
- [13] I. Bhattacharya, M. Foley, N. Zhang, T. Zhang, C. Ku, C. Mine, H. Ji, C. Riedl, B. F. Welles, and R. J. Radke, “A multimodal-sensor-enabled room for unobtrusive group meeting analysis,” in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018, pp. 347–355.
- [14] C. Lai, J. Carletta, and S. Renals, “Modelling participant affect in meetings with turn-taking features,” in *Proc. Workshop of Affective Social Speech Signals*, 2013.
- [15] O. I. (2023) Introducing whisper. Fecha de acceso: 18/12/2023. [Online]. Available: <https://openai.com/research/whisper>
- [16] A. Ali and S. Renals, “Word error rate estimation for speech recognition: e-wer,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2018, pp. 20–24.
- [17] A. Suresh, T. Sumner, J. Jacobs, B. Foland, and W. Ward, “Automating analysis and feedback to improve mathematics teachers’ classroom discourse,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 9721–9728.
- [18] T. Nazaretsky, J. N. Mikeska, and B. Beigman Klebanov, “Empowering teacher learning with ai: Automated evaluation of teacher attention to student ideas during argumentation-focused discussion,” in *LAK23: 13th International Learning Analytics and Knowledge Conference*, 2023, pp. 122–132.
- [19] E. Jensen, S. L. Pugh, and S. K. D’Mello, “A deep transfer learning approach to modeling teacher discourse in the classroom,” in *LAK21: 11th international learning analytics and knowledge conference*, 2021, pp. 302–312.
- [20] D. Schlotterbeck, P. Uribe, A. Jiménez, R. Araya, J. Van der Molen Moris, and D. Caballero, “Tarta: teacher activity recognizer from transcriptions and audio,” in *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part I 22*. Springer, 2021, pp. 369–380.
- [21] Ó. Cánovas and F. J. García, “Analysis of classroom interaction using speaker diarization and discourse features from audio recordings,” in *Learning in the Age of Digital and Green Transition*, 2023, pp. 67–74.
- [22] D. Filer, “Everyone’s answering: Using technology to increase classroom participation,” *Nursing education perspectives*, vol. 31, no. 4, pp. 247–250, 2010.
- [23] A. I. Wang, “The wear out effect of a game-based student response system,” *Computers & Education*, vol. 82, pp. 217–227, 2015.
- [24] J. Oigara and J. Keengwe, “Students’ perceptions of clickers as an instructional tool to promote active learning,” *Education and Information Technologies*, vol. 18, no. 1, pp. 15–28, 2013.
- [25] A. I. Wang and R. Tahir, “The effect of using kahoot! for learning – a literature review,” *Computers Education*, vol. 149, p. 103818, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360131520300208>
- [26] B. Catalina and M. d. C. García-Galera, “Innovación y herramientas hi-tech en la docencia del periodismo. El caso de Wooclap,” *Doxa comunicación*, 2022.

- [27] M. Pericacho, J. A. Rosado, J. P. de Villanueva, L. Arbea *et al.*, “Experiencias de docencia virtual en facultades de medicina españolas durante la pandemia covid-19 (i): Anatomía, fisiología, fisiopatología, oncología,” *Revista Española de Educación Médica*, vol. 1, no. 1, pp. 32–39, 2020.
- [28] A. I. Wang and A. Lieberoth, “The effect of points and audio on concentration, engagement, enjoyment, learning, motivation, and classroom dynamics using kahoot,” in *European conference on games based learning*, vol. 20. Academic conferences international limited, 2016, pp. 738–746.

Óscar Cánovas se licenció en Ingeniería en Informática por la Universidad de Murcia en 1998 y se doctoró en Informática en 2003. Actualmente es Profesor Titular de Universidad en la Universidad de Murcia. Ha participado en varios proyectos de investigación y publicado artículos en ámbitos diversos como la seguridad de la información, la autenticación y autorización de usuarios, las redes definidas por software, el posicionamiento en interiores, la investigación en educación y la aplicación de la tecnología educativa.

Pilar González Férez es Profesora Titular de Universidad en el área de Arquitectura y Tecnología de Computadores de la Facultad de Informática de la Universidad de Murcia. Obtuvo su doctorado en Informática en el año 2012 en dicha Universidad. Su investigación se centra en ámbitos diversos como los sistemas operativos, los sistemas distribuidos, el diseño de bases de datos clave-valor y la investigación en educación. Ha participado en varios proyectos de investigación y publicado sus resultados de investigación tanto en revistas como en congresos.

Félix J. García Clemente es Doctor en Informática y Catedrático de Universidad en el área de Arquitectura y Tecnología de Computadores de la Facultad de Informática de la Universidad de Murcia. Su actividad investigadora se centra en la ciberseguridad, la gestión dinámica de servicios, la industria 4.0 y la tecnología educativa. Como resultado de esta actividad es autor de más de 120 publicaciones, incluyendo revistas y conferencias, y es miembro activo de diversos proyectos de I+D nacionales e internacionales.

Federico Pardo licenciado en Ingeniería en Informática por la Universidad de Murcia en 2020 y Máster en Big Data en 2021. Actualmente estudiante de Doctorado en la Universidad de Murcia. Ha participado en proyectos de clasificación de sonidos mediante Inteligencia Artificial y reconocimiento de imágenes.