



# EVALUAR EL PENSAMIENTO CRÍTICO EN EDUCACIÓN PARA LA CIUDADANÍA. PROPUESTA PARA CONTEXTOS MASIFICADOS

Recepción: 18/09/2017 | Revisión: 19/10/2017 | Aceptación:24/10/2017

**Daniel David MARTÍNEZ ROMERA**

Universidad de Málaga  
ddmartinez@uma.es

**Resumen:** Se propone una metodología de evaluación que aprovecha la singularidad de la Educación para la Ciudadanía. Para ello, el alumnado ha sido instado a emitir juicio crítico sobre sus pares, creando un escenario en el que la objetividad es difícil. Se plantea una metodología basada en la evaluación de 360° y el aprendizaje colaborativo. Las herramientas utilizadas incluyen las TIC, el análisis estadístico exploratorio y la minería de datos. El caso se ha llevado a cabo sobre 47 estudiantes de Magisterio que cursan dicha materia en dos lenguas. La propuesta ha permitido obtener resultados interesantes, tanto en comportamientos individuales como colectivos; y muestra un gran potencial como herramienta para asistir en la reflexión docente sobre por qué existen determinados patrones de comportamiento y cómo de sólida es la capacidad de juicio crítico de sus estudiantes.

**Palabras clave:** didáctica de las Ciencias Sociales; innovación; patrones de comportamiento.

## CRITICAL THINKING ASSESSMENT IN EDUCATION FOR CITIZENSHIP. PROPOSAL FOR OVERCROWDED CONTEXTS

**Abstract:** An evaluation methodology is proposed that takes advantage of the singularity of Education for Citizenship. To that end, the students have been urged to issue critical judgment on their peers, creating a scenario in which objectivity is difficult to maintain. We propose a methodology based on 360° evaluation and collaborative learning. The tools used include ICT, exploratory statistical analysis and data mining. The case has been developed over 47 students of Teacher Training studies of this subject taught in two languages. The proposal has yielded interesting results, both in individual and collective behavior; and shows great potential as a tool to assist in teacher reflection on why there are certain patterns of behavior and how solid is the critical judgment capacity of their students.

**Keywords:** didactics of Social Sciences; innovation; behavior patterns.

## AVALUAR EL PENSAMIENTO CRÍTICO EN EDUCACIÓN PER A LA CIUDADANÍA. PROPOSTA PER A CONTEXTOS MASSIFICATS

**Resum:** Es proposa una metodologia d'avaluació que aprofita la singularitat de l'Educació per a la Ciutadania. Amb aquest fi, s'insta l'alumnat a emetre un judici crític sobre els seus parells, creant un escenari en el que l'objectivitat és difícil. Es planteja una metodologia basada en l'avaluació de 360° i l'aprenentatge col·laboratiu. Les eines utilitzades inclouen les TIC, l'anàlisi estadística exploratòria i la mineria de dades. El cas s'ha dut a terme en 47 estudiants de Magisteri que estudien aquesta matèria en dues llengües. La proposta ha permès obtenir resultats interessants, tant en comportaments individuals com col·lectius, i mostra un gran potencial com a eina per ajudar en la reflexió docent sobre per què existeixen determinats patrons de comportament i com de sòlida és la capacitat de judici crític dels seus estudiants.

**Paraules clau:** didàctica de les Ciències Socials; innovació; patrons de comportament.

## Introducción

La sociedad contemporánea ha refinado a lo largo de los años una serie de elementos básicos sobre lo que considera un buen proceso educativo: correcta formación docente, con buena motivación e implicación, extensibles al alumnado, recursos y entornos adecuados para el desarrollo del acto educativo... Estos son algunos de los aspectos más habituales. Pero lo cierto es que ninguna concreción curricular se limita a definir las piezas del puzzle educativo, sino que también tiene buen cuidado de establecer mecanismos de control, de todo tipo, para garantizar el desarrollo formativo en los términos que ellas mismas establecen (Pérez Gómez, 2014; Tiana Ferrer, 2014).

Tal vez la expresión más emblemática en la actualidad de esta preocupación la represente el Espacio Europeo de Educación Superior (López, Pérez-García, y Rodríguez, 2015; Molero López-Barajas, 2007), que subraya la necesidad de mayor grado de innovación, planteamientos más cercanos al desempeño futuro en los procesos de aprendizaje y, en consecuencia, un mayor énfasis en las tareas y dinámicas individuales y de pequeño grupo. Exigencias a las que no es ajena España (Perales Montolío, Jornet Meliá, y González-Such, 2014), tanto por la necesidad de convergencia educativa explicitada en dicho espacio de educación como de mejora constante en la calidad de la enseñanza que se ofrece.

Así, cuanto más diversa y profunda es una materia, como ocurre en Ciencias Sociales, dos cuestiones acaban surgiendo de forma recurrente sobre su evaluación: en primer lugar, cómo obtener datos sobre la progresión del alumnado, respecto a la adquisición objetiva de contenidos así como a su integración subjetiva; y en segundo lugar, cómo desarrollar estrategias eficientes de análisis sobre los mismos, ya que incluso un grupo reducido de estudiantes puede generar, a través de las diversas tareas y dinámicas de clase, una cantidad significativa de información. Semejante dificultad explica la recurrencia de la prueba final como instrumento docente de mayor peso relativo.

Por ello, la evaluación es uno de los aspectos más trabajados en la investigación e innovación educativa, ya sea desde planteamientos formales (Pérez Pueyo, Hortigüela Alcalá y Gutiérrez García, 2016; Perales Montolío, Jornet Meliá y González-Such, 2014; Valliant, 2008) o en la búsqueda de propuestas específicas que le otorguen valor práctico (Arribas Estebanz, Manrique Arribas y Taberner Sánchez, 2016; Moreno Olivos, 2011; Karacapilidis, Ng, y Raisinghani, 2011; Poggi, 2008).

Sin embargo, el reto docente permanece constante más allá de los marcos epistemológicos o matizaciones semánticas que se le puedan introducir (Hamodi, Pastor, y Pastor, 2015): cuanta mayor profundidad analítica sea necesaria alcanzar, mayor inversión en tiempo requerirá (Adesina, Stone, Batmaz y Jones, 2015; Wilcox, 2015; Wojcicki, Izumi, y Chang, 2015; Stevens y Levi, 2013). Algo que para ratios bajas, en el entorno de los 10-15 estudiantes, no es demasiado gravoso, pero que se puede convertir en todo un reto para grupos de mayor tamaño, especialmente si se apuesta por formas de evaluación continua junto con dinámicas en pequeño y gran grupo basadas en el aprendizaje mediante tareas.

Bajo esta realidad, el reto de la evaluación demanda mecanismos procedimentales eficaces que asistan al profesorado en su tarea, algo que invariablemente exige un acercamiento al ámbito de la informática (Kanashiro Medina, 2012). Afortunadamente, los instrumentos digitales útiles a la educación se encuentran cada vez más difundidos en los contextos formativos, especialmente

entre las nuevas generaciones de estudiantes, quienes viven su inclusión de una forma natural aunque en absoluto homogénea (Escofet Roig, López Costa y Álvarez, 2014). Es el profesorado quien se encuentra en clara desventaja comparativa, pues si no le viene dado por especialidad o por ámbito de investigación le supone un reto formativo de primer nivel (Correa Gorospe et al., 2015).

Por todo ello, consideramos pertinente la investigación e innovación sobre nuevas formas de análisis evaluador con base tecnológica en educación superior (Flavin, 2016; Marchis, 2016); en este caso, centrada en contextos masificados y por ende articulada alrededor de tres características clave: eficiencia, en términos de tiempo requerido hasta obtener conclusiones; agilidad, en el manejo de los datos, para hacer posible su descripción, exploración y análisis; y escalabilidad, de modo que la estrategia se aplicable, con un mínimo de adaptaciones, a contextos livianos, densos o saturados de estudiantes, e independientemente de la naturaleza de la materia o de su carácter presencial o virtual. Todo ello, con objeto de identificar aspectos relevantes para el proceso de evaluación que de otro modo pasarían desapercibidos o requerirían un consumo de tiempo extendido.

## 1. Método

La tarea principal del artículo es ofrecer un procedimiento analítico sintetizador y sincretizador cuando el volumen de datos y el universo de referencia son densos, apoyándose en el análisis exploratorio (Baglin, 2014; Beavers et al., 2013) y la minería de datos (Algarni, 2016; Papamitsiou y Economides, 2014). Con ello, se pretende ofrecer una alternativa de trabajo que permita reducir significativamente la carga temporal de análisis, a la par que se fortalece la capacidad para discernir patrones y relaciones ocultos en la matriz de datos. Dicho en términos didácticos, aprovechar mejor el significado de la información registrada por el docente y descubrir relaciones internas que aporten valor añadido a la evaluación como instrumento diagnóstico (Sadler, 2016; Sánchez Giraldo y Escobar Hoyos, 2015; Evans, 2013).

Para ilustrar su pertinencia y aplicabilidad, se ha usado el método en una clase con 47 personas del grado en Magisterio de Educación Primaria, correspondientes a la asignatura bilingüe de Educación para la Ciudadanía y los Derechos Humanos, durante el curso 2016/17, en términos generales orientada hacia el trabajo en pequeño y gran grupo. Por tanto, estamos ante un planteamiento de materia habitual en Magisterio, pero con una ratio claramente alejada de los ideales de cualquier teoría educativa, razones por la que consideramos pertinente la elección de dicho grupo como ejemplo de caso.

La componente didáctica de la asignatura se aborda de dos formas complementarias. La primera se centra en aprovechar una de las principales habilidades que se deben promocionar en la asignatura: el pensamiento crítico y la toma de decisiones en contextos de incertidumbre. Esto es, sin que se cuente con todos los datos (una cosa es una exposición y otra distinta es el documento escrito del que mana), ni con todo el contexto (solo los miembros del grupo conocen las dinámicas personales y de trabajo que finalmente expresaron la intervención). La ventaja que presenta esta asignatura es que dicha habilidad es también fundamental para la formación de docentes en una materia, como es la ciencia social, en la que conviven razón y emoción.

Teniendo esto presente, se han seleccionado dos de sus principales tareas:

1. El diseño de una intervención didáctica completa orientada hacia la educación cívica y los derechos humanos. Aquí, los distintos grupos deben trabajar en la elaboración de una propuesta de intervención educativa sobre una parte de los contenidos (ciudadanía local y global, democracia y modelos democráticos, derechos humanos universales, derechos infantiles, integración y diversidad cultural, diálogo y convivencia, igualdad, minorías, acoso...). Desde una perspectiva docente, esto permite trabajar de forma práctica habilidades propias de la materia, tales como sociabilización, tolerancia, respeto, argumentación de ideas o participación constructiva en debates.

2. La evaluación mediante rúbrica colaborativa de su exposición. Tanto como futuros docentes como ciudadanos, una de las cuestiones sociales menos gratas es la de la toma crítica de decisiones sobre iguales. Analizar el trabajo del resto de estudiantes y emitir un juicio derivado de ello, evaluarlos, es algo insoslayable si pretendemos hacer de la educación ciudadana un conocimiento práctico. Implica asumir una responsabilidad sobre alguien que se conoce. Esto lleva a quien juzga a tomar consciencia de su dualidad intelectual, racional e irracional: nepotismo, afinidad de gustos, intereses académicos comunes, enfrentamientos personales o el más elemental desinterés por los demás llevan a conclusiones ilusorias, injustas en un sentido u otro. Esto es algo no siempre fácil de detectar y mostrar al alumnado.

Para la ejecución de la primera se organizaron en grupos no dirigidos de 4 a 6 personas. Para la segunda, se planteó una evaluación entre pares orientada hacia la metodología de 360° (Bisquerra Alzina et al., 2006; Brutus y Gorriti, 2005) sobre una rúbrica digital. Esta última se construyó en clase a partir de las propuestas discentes y docentes como parte de un ejercicio de debate y búsqueda de consenso en gran grupo. En ella, se definieron los elementos fundamentales: categorías, variables y evidencias de observación asociadas a escala de medición, así como la ponderación de los dos primeros elementos y la normalización de los últimos (Cebrián de la Serna y Bergman, 2014).

En este último caso, pese a ser conscientes de que no es un instrumento exento de crítica, especialmente por situaciones de excesiva recurrencia (Cano, 2015), entendemos que al tratarse de una propuesta singular, a la que se dedicó un tiempo adecuado en todas sus fases, constituye en sí mismo un instrumento de evaluación y formación en competencias docentes (Tójar Hurtado y Velasco Martínez, 2015). La herramienta usada para su diseño y aplicación fue CoRubric (<http://corubric.com/>), aplicación online gratuita para rúbricas federadas, desarrollada por el Grupo de Tecnologías Educativas (GTEA) de la Universidad de Málaga (España). La estructura resultante queda recogida en la Figura 1.

El instrumento así definido fue aplicado de manera individual por todos los miembros de la clase, tanto a personas como a grupos. Al tratarse de una herramienta en línea federada, el alumnado contaba con la posibilidad de hacerlo al final de la sesión mediante cualquier dispositivo móvil con conexión a Internet, en las aulas de informática de la facultad, o en última instancia en casa. De manera ideal, se pretende la sincronización entre exposiciones y evaluaciones, sin embargo las limitaciones horarias y logísticas no siempre lo hacían posible. La influencia que pudiera tener dicho desfase no ha sido considerada aquí, para evitar la deriva excesiva del objeto del presente estudio; sin embargo, sería plausible abordarla mediante el registro de eventos con que cuenta la aplicación. Con todo, se obtuvo una tipología de evaluaciones que podemos resumir del siguiente modo:

- Si la persona se encuentra en fase de exposición:  
Se produce una autoevaluación (autocrítica)  
Y una coevaluación en pequeño grupo
- Si la persona se encuentra en fase de asistente:  
Se produce una evaluación individual entre pares (a cada miembro)  
Y una coevaluación en gran grupo (a la exposición en su conjunto)

1. Citizenship 2016/17 Evaluation rubric on final group task explanation				
1.1. Participation (Within the group that exposes)				
1 No one works as a team, most of the members remain silent	2 Intermediate situation (Poor)	3 Only half of the team is involved and tries to push forward the task	4 Intermediate situation (Good)	5 All members work together and with balanced participation
1.2. Class dynamics (External - openness, interaction and motivation of students)				
1 The explanation is like one-way street, no interaction with students nor motivation attempts	2 Intermediate situation (Poor)	3 The explanation sticks frequently, making hard or barely useful the attempts for interaction and motivation	4 Intermediate situation (Good)	5 The explanation is fluid and naturally bidirectional, implying high levels of interaction and motivation
1.3. Contents (Theoretical foundations, practical application and their concomitance)				
1 Theory and practice walk apart together all the time, nothing is integrated in the experience	2 Intermediate situation (Poor)	3 Frequently, theory and practice relationship is not clear, nor internally objectives, contents, task, evaluation and adaptation elements	4 Intermediate situation (Good)	5 Theory and practice walk together all the time: objectives, contents, tasks, evaluation and adaptation to educational level are perfectly integrated
1.4. Oral expression (Proper use of voice to reinforce the message)				
1 Bad. All the time listening becomes difficult, and makes hard to understand what is being said	2 Good. Most of the time, only occasionally it suppose a difficult to follow the explanation	3 Perfect. Use of the voice makes easier the learning process all the time		
1.5. Body expression (Gestures and movements)				
1 Bad. There is no gestures and movements at all	2 Good. Most of the time, only occasionally it suppose a difficult to follow the explanation	3 Perfect. Gestures and movements emphasise the message making it easier to follow the learning process		
1.6. Time (Adjustment to the given time)				
1 Bad (±20') - Stopped by teacher.	2 Adequate (±15')	3 Perfect (±10')		
1.7. English? (Extra point)				
1 No used at all	2 Most of the time. Used by all the members, but much as a speech than a talk	3 Most people, most of the time uses English as main language in a practical way (not only like in a speech)	4 Used as only language and in practical way: talks, debates, asks/answers contexts	

\* Este ítem es necesario para pasar la rúbrica

Figura 1. Rúbrica digital colaborativa para la evaluación de exposiciones. Fuente: Elaboración propia.

Un tercer nivel sería determinado por la heteroevaluación docente. Sin embargo, dado que supone un procedimiento directo y habitual, cuando no el único aplicado, preferimos centrarnos en el análisis de datos a los que no se suele descender, dada la complejidad de su manipulación, tanto por el número de personas a estudiar como por la multiplicidad de criterios emitidos a cotejar, ya que en última instancia es lo que da sentido a la propuesta. Se obtuvo así una matriz de datos con 2.318 registros sobre un valor esperado de 2.632, lo que supone un 88.8% de cobertura. Siguiendo la estructura de la rúbrica, cada registro posee 7 variables útiles al estudio, de modo que se cuenta con un universo de 16.266 datos individuales.

El primer nivel de análisis viene dado por el instrumento de recogida, en nuestro caso la aplicación permite exportar las evaluaciones en formato de hoja de cálculo, para ello en el presente análisis se ha utilizado Google Sheets (<https://goo.gl/kp6GLM>) y SPSS v23. Con ello se está en condiciones de generar tablas dinámicas de manera inmediata, una herramienta de análisis

estadístico eficiente y rápida (Reyes Dixson y Nuñez Maturel, 2015; Slater, Beal-Hodges, y Reed, 2014; Salafranca Cosialls et al., 2005), llave aquí del análisis exploratorio independiente de las dimensiones de la fuente de datos.

Estas permiten explorar el cruce entre variables, así como el cálculo elemental de datos y su resumen; todo ello bajo una interfaz gráfica de tablero pivotante de triple entrada (filas/columnas/datos). Desde un punto de vista docente o investigador, cuenta con dos fortalezas claras: una exigencia baja de teoría estadística y ausencia de procedimientos secundarios que posterguen la obtención de resultados. Por contra, la capacidad para hacer las preguntas adecuadas y detectar patrones interesantes queda en manos del docente o investigador, algo que no siempre constituye una solución eficiente.

Dos tareas básicas se esperan resolver con esta aproximación: conocer el grado de cumplimiento de la tarea, el comportamiento como evaluadores y evaluados de forma individual y en pequeño grupo, así como describir la relevancia de los diferentes apartados de la rúbrica.

El segundo nivel de análisis, pese a su sencillez, implica necesariamente el dominio de un procedimiento de trabajo para preparar la matriz de cara a su explotación. La minería de datos ha demostrado ser fructífera en educación (Koedinger et al., 2015; Sin y Muthu, 2015), en buena medida gracias a la difusión de aplicaciones destinadas a ello de calidad contrastada, como Weka v3 (Sebastian y Puthiyidam, 2015; Ahmadi y Ahmad, 2013), una opción gratuita y aprehensible a investigadores no especialistas desarrollada por la Universidad de Hamilton (Nueva Zelanda).

Tras una exploración previa, se eligieron los algoritmos que mostraron una mayor precisión y menor probabilidad de error: el algoritmo de aprendizaje de reglas proposicionales RIPPER (Cohen, 1995), y una versión modificada del algoritmo de árboles de decisión C4.5 (Eibe y Witten, 1998). Como tarea previa a dicho análisis, toda la información cuantitativa fue tabulada en diez intervalos regulares, usando como base los valores extremos, ya que este tipo de análisis debe realizarse sobre variables nominalizadas.

Solo una tarea es asignada a esta aproximación: discernir patrones de comportamiento y relaciones en la matriz de datos, determinando en su caso la certeza o fiabilidad de las mismas. Gracias a ello se podrán orientar, apoyándose en la experiencia docente, las posibles explicaciones, o dudas, que puedan surgir al tratar de argumentar las mismas. Esto es, hará posible cruzar la experiencia docente basada en el seguimiento diario de la clase con las relaciones encontradas, con objeto de encontrar explicaciones plausibles o, al menos, espacios que inviten a la reflexión docente.

## 2. Resultados

El análisis exploratorio permitió describir el comportamiento del alumnado tanto desde la perspectiva de quien evalúa como de quien es evaluado; se aplicó el mismo análisis a las 7 categorías que alberga: recuento de evaluaciones, media, mediana y desviación estándar. En la Figura 2 se ilustra el proceso de construcción para la serie de datos de evaluaciones recibidas mediante una captura de pantalla; la dinámica es similar para las evaluaciones emitidas, así como para las distintas categorías de la rúbrica y los grupos de trabajo:



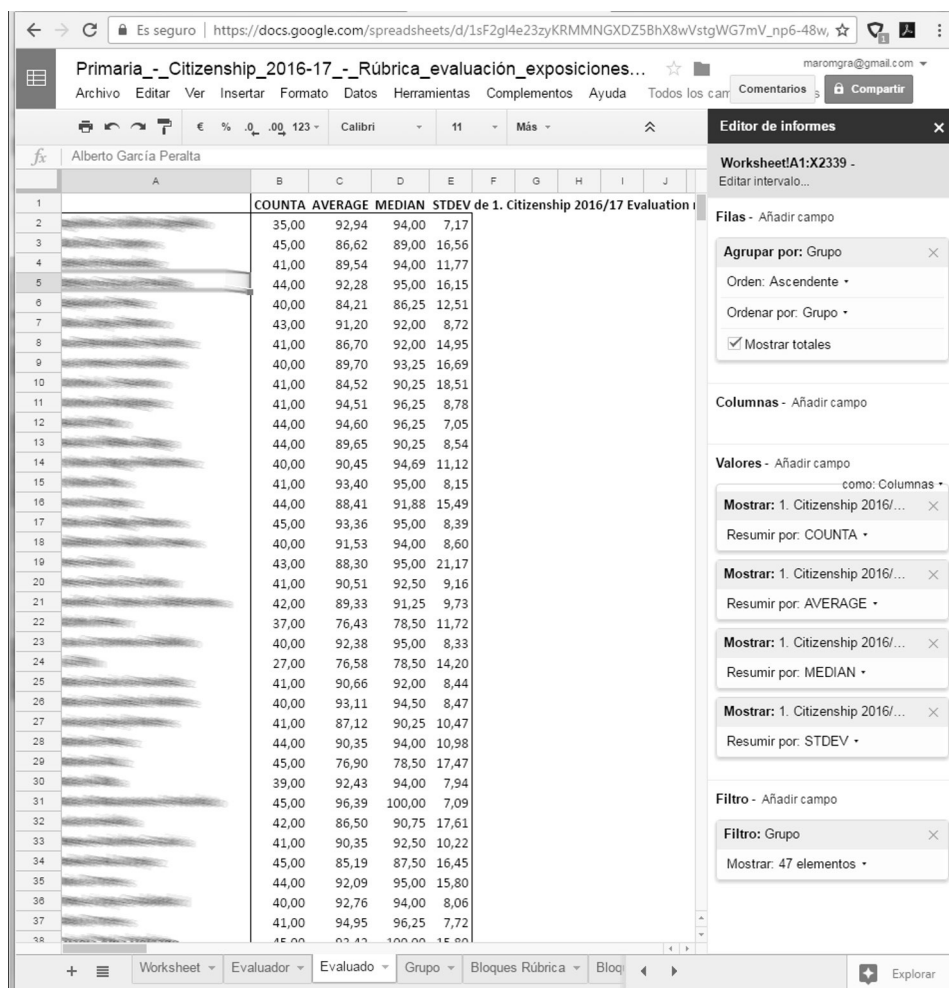


Figura 2. Obtención de datos mediante tablas dinámicas. Captura anonimizada. Fuente: Elaboración propia.

Lo primero que podemos comprobar es el grado de cumplimiento de la tarea por parte de cada persona. En el caso de evaluaciones emitidas, sobre el ideal de 47 (las 9 correspondientes a grupos se estudian aparte) tenemos un histograma con valor mínimo de 6 representado por ID21 (2.13%), 35 personas que alcanzan el umbral fijado (74,5%) y 12 que no llegan a realizar el 75% de la previsión (25.53%), lo que arroja una media de algo más de 39 evaluaciones por persona. La calificación media emitida fue de 89.77 (sobre 100), si bien el comportamiento individual demostró ser bastante heterogéneo, con valores que oscilan entre el 62.76 de ID40 a 97.78 de ID39.

Cabe esperar que el alumnado emita juicios homogéneos por lo que la distribución normal correspondiente no es demasiado excéntrica, haciendo que media y mediana se encuentren relativamente próximas. Sin embargo, la desviación estándar cuenta dos historias distintas: por un lado las personas con baja dispersión en sus valoraciones, para el 91.49% fue inferior a 13 puntos, mientras que para la minoría restante ofrece un rango de heterogeneidad que llega a superar los 35.

Una cuestión razonable aquí sería plantearse qué han visto estas personas a diferencia del resto, algo que se mueve en el terreno de lo subjetivo y a lo que el docente sólo puede añadir su cotejo personal. Afortunadamente, este instrumento sí nos permite averiguar fácilmente dónde se han producido las discrepancias.

Categoría	A	B	C	D	E	G	H	I	J	$\bar{x}$
Participación	95,35	81,98	93,13	85,53	88,13	95,27	85,53	90,85	79,73	88,45
Dinámica de clase	89,53	89,53	94,38	90,13	88,13	91,22	85,53	92,68	86,49	89,78
Contenidos	91,86	91,28	90,63	92,76	90,63	95,27	89,47	92,68	89,19	91,53
Expresión oral	91,63	94,42	94,00	92,63	89,00	96,76	91,58	93,17	83,78	91,93
Expresión corporal	89,77	82,33	87,00	88,42	84,00	92,43	85,26	89,51	81,08	86,64
Gestión tiempo	96,51	97,67	97,50	97,37	98,75	95,95	98,68	96,34	98,65	97,48
Uso del inglés	96,51	68,60	38,75	2,63	32,50	95,27	49,34	59,76	79,73	58,40
Media ponderada	92,37	89,87	92,73	91,30	89,78	94,64	89,46	92,58	86,49	91,04
Evaluaciones	43	43	40	38	40	37	38	41	37	39,67

Tabla 1. Evaluaciones medias recibidas en pequeño grupo, por categorías de rúbrica.  
Fuente: Elaboración propia.

El análisis sintético con tablas dinámicas permite estudiar el comportamiento de los grupos de trabajo. La Tabla 1 refleja cómo han sido valorados por categoría, y la puntuación parcial y final (media ponderada) de cada uno. El mejor valorado es G y el de menor puntuación J. El primero obtuvo las puntuaciones máximas absolutas en Contenidos, Expresión Oral y Expresión Corporal, y el segundo las mínimas absolutas en Participación, Contenidos, Expresión Oral y Corporal.

El uso del inglés como vehículo de comunicación se muestra como la categoría más heterogénea, con un rango que oscila entre 2.63 y 93.88 puntos, lo que se explica por la existencia de grupos en los que no se usó, o se hizo de manera testimonial al principio de la intervención. Obviando esto, Participación, Expresión Oral y Expresión Corporal son las categorías en las que existe una mayor amplitud en el rango de puntuaciones, con 15.62, 12.97 y 11.35 puntos respectivamente.

Todas tienen en común que miden el desempeño individual de los miembros respecto a la clase o al grupo de trabajo, mientras que las categorías más centradas en evaluar el comportamiento medio de los grupos exponentes, como Dinámica de Clase, Contenidos y Gestión del Tiempo, reducen su horquilla a 8.85, 6.08 y 2.80 puntos en cada caso. Por tanto, las discrepancias de criterio del alumnado, en tanto que coevaluadores, son mayores cuanto más pequeña es la unidad de análisis, algo previsible y coherente con las dificultades de formación del profesorado (Cano, 2016; Vázquez Cano, 2016).

El análisis basado en minería de datos con Weka se centró en la determinación de correlaciones entre variables en un proceso secuencial por lotes en 3 pasos para la obtención de asociaciones entre variables: derivadas de la aproximación RIPPER con y sin las desviaciones estándares, y derivadas del algoritmo C4.5 modificado. En nuestro caso, el segundo arrojó una mejor clasificación de instancias: 89.4%, con una Kappa de 0.88, frente a 85.1%, 0.83 y 66.0%, 0.61 respectivamente para los dos análisis RIPPER.

Seguidamente se comentan dos ejemplos de patrones asociados a comportamientos extremos (patrones evidentes) y otro a situaciones intermedias (patrones abstrusos):

1. (Rc\_Recibidas = '(43.2-inf)') and (Rc\_Dev. Stand. = '(-inf-8.462]') => Grupo=A (4.0/0.0). Este grupo, y la mayoría de sus miembros, se caracterizan por ser percibidos de forma nítida, ya que no hay grandes discrepancias entre evaluadores; algo especialmente significativo en tanto que la práctica totalidad de la clase les evaluó.



2. (4e\_Media = '(-inf-64]') => Grupo=J (3.0/0.0). Si se es especialmente exigente, grado mínimo, a la hora de evaluar la expresión oral se pertenece al grupo J. Tres de los cinco miembros muestran este comportamiento, y tampoco se hayan personas ajenas que concurren en el patrón.

3. Rc\_Recibidas = '(43.2-inf)' AND 3c\_Media = '(88.55-90.345]': C (5.0). En este caso estamos ante la detección de un patrón en lo que denominamos zona gris, alejada de los extremos de la escala de evaluación, y tiene interés en la determinación de subgrupos de comportamiento homogéneo, con objeto de detectar posibles ámbitos de dificultad, o de dominio relativo de la materia, que pueden ayudar a orientar mejor el esfuerzo docente.

En general, los patrones detectados permiten caracterizar la práctica totalidad de grupos en relación a su desempeño en la exposición, o la evaluación de la misma. La mayor parte de reglas relacionadas con valores extremos redundan en los grupos G y J, los mejor y peor valorados respectivamente. Sin embargo, invita a la reflexión que el grupo J sea, a su vez, especialmente exigente a la hora de emitir juicio, como se ha ilustrado en la regla 2.

Las situaciones intermedias han sido más difíciles de acotar, en la medida que una característica de estas reglas es tanto la cobertura parcial del grupo al que afectan como la frecuencia de personas ajenas; esto sucede especialmente con los grupo C y H, en cuyas normas de comportamiento concurren con habitualmente miembros de otros grupos: (6c\_Media = '(97.558-98.372]') => Grupo=H (7.0/3.0); 3c\_Media = '(88.55-90.345]' AND 7e\_Mediana = '(70-80]': D (4.0/1.0).

Con todo, el proceso de caracterización de individuos y grupos singulares pone en evidencia la existencia tanto de patrones diferenciadores como de tendencias homogeneizadoras. Explicarlos desde el análisis de datos no es posible, pero sí ha ayudado a detectarlos, así como la necesidad de formularle preguntas adicionales. Es aquí donde, como docentes, podemos argumentar sobre el elevado grado de cohesión de la clase en las tareas presenciales, que claramente parece haber trascendido el aula, en tanto que la concomitancia de notas indica convergencia respecto a las claves principales de la exposición objeto de evaluación: el número de evidencias cumplidas o no cumplidas por estos grupos es semejante.

## Discusión

Las tablas dinámicas introducen la posibilidad de organizar y reorganizar tanto la relación entre variables como los estadísticos resultantes hace que la descripción general del trabajo individual y grupal en clase sea algo prácticamente inmediato. Como se apuntara al inicio, para los docentes se trata de la aproximación estadística menos compleja y más amable, pero que ha demostrado utilidad en sí misma incluso antes de pasar a la siguiente fase de análisis, siendo más determinante en este sentido la capacidad para formular preguntas significativas que el propio utillaje numérico. Desde una perspectiva didáctica, aporta un claro valor añadido como instrumento para la detección de dificultades de aprendizaje por parte del alumnado en la denominada zona gris. Como se ha mostrado, hay estudiantes con criterio convergente al del gran grupo en determinadas categorías, pero no en otras. Además, al crear un contexto activo sobre el proceso de enseñanza en la materia propuesta, cada miembro y grupo desarrolla comportamientos sociales y de trabajo que le son propios, que sin embargo ahora pueden ser detectados de forma indirecta para usarlos como elemento de reflexión: por qué determinados estudiantes, como se ha ilustrado, son más severos

o generosos en la calificación de un miembro de su grupo que el conjunto de la clase. Determinar si se ha exhibido una capacidad suficiente de pensamiento crítico, de juicio equilibrado, es más sencillo y se puede hacer de forma más robusta. Algo especialmente pertinente en el marco de la asignatura. Y abre la posibilidad a nuevas acciones docentes de contraste, corrección o ampliación. Además, debemos tener en cuenta que, aunque descartadas por motivos de concreción en el presente trabajo, existen correspondientes evaluaciones docentes. Lo que permitiría añadir un elemento de reflexión sobre si las discrepancias y las zonas grises detectadas realmente se fundamentan en criterios razonados, condicionados o abiertamente emocionales.

Debemos señalar algunos aspectos a tener en cuenta para una aplicación eficiente de la propuesta. En primer lugar, la escasa diversidad en la naturaleza de los datos, ambas tareas se relacionaban estrechamente como parte de único objetivo, algo que se hizo de forma intencionada con objeto de hacer más sencilla la comprensión del proceso. No obstante, esto ha hecho que la minería ofrezca relaciones no siempre interesantes para la determinación de patrones; es en contextos como la evaluación final de curso, especialmente, donde la presencia de tareas diversas, individuales y grupales, puede aprovechar mejor su utilidad.

La capacidad de conocer determinadas situaciones puede constituirse en una clave para la orientación de la acción docente. Sobre todo si, superando nuestra propuesta, se considera su aplicación recurrente en fases más tempranas de la asignatura, pudiéndose desarrollar en consecuencia estrategias de orientación, recuperación o mejora tales como entrevistas de grupo que permitan la superación dificultades de funcionamiento interno detectadas, o la reorientación de los centros de interés a partir de la percepción real del alumnado. Y, a nuestro entender, se refuerza significativamente el ideal de la evaluación como instrumento diagnóstico y formativo, y no estrictamente calificador.

Desde una perspectiva analítica, se abre la posibilidad de comprobar en qué medida los autómatas de aprendizaje asisten en el proceso evaluador, la utilidad de construir modelos de clasificación para orientar el acto educativo y hacerlo más efectivo, o la pertinencia de incluir nuevas aproximaciones como los tutores inteligentes de tareas (Najar, Mitrovic y McLaren, 2016). En este sentido, la formación de los no especialistas constituye una dificultad que es necesario convertir en oportunidad a través de la curiosidad y la evidencia de su utilidad. Estrategia en la que consideramos necesario seguir profundizando.

## Conclusión

Consideramos que la metodología desarrollada ha demostrado su utilidad en términos de eficiencia temporal. Se han manejado grandes cantidades de datos, se han creado o derivado otros nuevos y se han podido cotejar a diferentes escalas de análisis (individual/grupal/clase), algo que de otro modo llevaría asociados unos costes de dedicación probablemente inasumibles. Pero el verdadero interés radica en la aportación de valor añadido al análisis, es decir, el tiempo liberado ha quedado disponible para estudiar aspectos de la evaluación que habrían sido difíciles de abordar de otro modo.

Surge así, a nuestro criterio, un campo de aplicación especialmente relevante en el ámbito de la Didáctica de las Ciencias Sociales, ya que integra cuestiones relativas a la formación del pen-

samiento crítico o la determinación consensuada de criterios de juicio, y capacita al docente para gestionar y evaluar dichos datos de manera eficaz. En un marco más amplio, con modificaciones menores, podría aplicarse esta metodología a la evaluación de comentarios de texto históricos, objetos artísticos o el análisis de un paisaje.

Finalmente, las pretensiones establecidas al inicio del trabajo, entendemos, han quedado cubiertas. El análisis básico centrado en estadística descriptiva y manipulación asistida de matrices ha permitido no sólo describir las variables individuales más relevantes, sino ensayar diferentes simplificaciones y comprobar de forma interactiva su coherencia. La minería de datos ha demostrado ser un instrumento de interés, tanto por el potencial demostrado para el refinamiento de los análisis como por su capacidad para descubrir relaciones no previstas. Por todo lo expuesto, entendemos pertinente seguir profundizando en estas cuestiones

### Referencias bibliográficas

- Adesina, A. O., Stone, R. G., Batmaz, F., y Jones, I. (2015). A semi-automatic computer-aided assessment approach for marking and providing feedback comments. En M. Helfert, M. T. Restivo, S. Zvacek y J. Uhomoihi (Eds.), *Proceedings of the 7th international conference on computer supported education*, 93-100. Lisbon, Portugal: 7th International Conference on Computer Supported Education. doi:10.5220/0005447000930100
- Ahmadi, F., y Ahmad, S. (2013). Data mining in teacher evaluation system using WEKA. *International Journal of Computer Applications*, 63(10), 14-18. doi:10.5120/10501-5268
- Algarni, A. (2016). Data mining in education. *International Journal of Advanced Computer Science and Applications*, 7(6), 456-461. doi:10.14569/IJACSA.2016.070659
- Arribas Estebanz, J. M., Manrique Arribas, J. C., y Tabernero Sánchez, B. (2016). Instrumentos de evaluación utilizados en la formación inicial del profesorado y su coherencia para el desarrollo de competencias profesionales en los estudiantes: visión del alumnado, egresados y profesorado. *Revista Complutense de Educación*, 27(1), 237. doi:10.5209/rev\_RCED.2016.v27.n1.45724
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment*, 19(5), 1-15. Recuperado de <https://doaj.org/article/0f815b85142b4145931d38f74a0fa676>
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., y Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment*, 18(6), 1-13. Recuperado de <https://doaj.org/article/395bf90f786a47fca6e67e917c2e7e96>
- Bisquerra Alzina, R., Martínez Olmo, F., Obiols Soler, M., y Pérez Escoda, N. (2006). *Evaluación de 360º: una aplicación a la educación emocional*. Recuperado de <http://hdl.handle.net/10201/45230>
- Brutus, S., y Gorriti, M. (2005). La Evaluación Multifuente Feedback 360º. *Revista de Psicología del Trabajo y de las Organizaciones*, 21(3), 235-252. Recuperado de <https://doaj.org/article/b99d22c270f1495bafb77853b8bc2d96>
- Cano, E. (2016). Factores favorecedores y obstaculizadores de la transferencia de la formación del profesorado en educación superior. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 14(2), 133-150. Recuperado de <http://hdl.handle.net/10486/670641>
- Cano, E. (2015). Las rúbricas como instrumento de evaluación de competencias en educación

- superior: ¿uso o abuso? *Profesorado, Revista De Curriculum y Formación Del Profesorado*, 19(2), 265-280. Recuperado de <http://digibug.ugr.es/handle/10481/37376>
- Cebrián de la Serna, Manuel, y Bergman, M. E. (2014). Evaluación formativa con e-rúbrica: aproximación al estado del arte. *REDU: Revista de Docencia Universitaria*, 12(1), 15-22. Recuperado de <https://doaj.org/article/da814e16cb4740ab998950e7b9d9dc92>
- Cohen, W. W. (1995). Fast effective rule induction. *Twelfth International Conference on Machine Learning*, 115-123.
- Correa Gorospe, J. M., Fernández Olaskoaga, L., Gutiérrez-Cabello Barragán, A., Losada Iglesias, D., y Ochoa-Aizpurua Aguirre, B. (2015). Formación del Profesorado, Tecnología Educativa e Identidad Docente Digital. *Revista Latinoamericana de Tecnología Educativa*, 14, 45-56. doi://dx.doi.org/10.17398/1695-288X.14.1.45
- Eibe, F., y Witten, I. H. (1998). Generating accurate rule sets without global optimization. 144-151. *Fifteenth International Conference on Machine Learning*.
- Escofet Roig, A., López Costa, M., y Álvarez, G. (2014). Una mirada crítica sobre los nativos digitales. Análisis de los usos formales de TIC entre estudiantes universitarios. *Revista Q*, 9(17), 1-19. Recuperado de <https://doaj.org/article/28ae56213d344c5089daca92e1336146>
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70-120. doi:10.3102/0034654312474350
- Flavin, M. (2016). Technology-enhanced learning and higher education. *Oxford Review of Economic Policy*, 32(4), 632-645. doi:10.1093/oxrep/grw028
- Hamodi, C., Pastor, V. M. L., y Pastor, A. T. L. (2015). Medios, técnicas e instrumentos de evaluación formativa y compartida del aprendizaje en educación superior. *Perfiles Educativos*, 37(147), 146-161. doi:10.1016/j.pe.2015.10.004
- Kanashiro Medina, A. (2012). Uso de Microsoft Excel para análisis y elaboración de informes de resultados de evaluación de aprendizajes. Recuperado de <http://hdl.handle.net/10757/285367>
- Karacapilidis, N., Ng, E. M. W., y Raisinghani, M. S. (2011). *Dynamic advancements in teaching and learning based technologies*. US: Information Science Reference (Isr).
- Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., y Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333-353. doi:10.1002/wcs.1350
- López, M. C., Pérez-García, P., y Rodríguez, M. J. (2015). Concepciones del profesorado universitario sobre la formación en el marco del espacio europeo de educación superior. *Revista de Investigación Educativa*, 33(1), 179-194. doi://dx.doi.org/10.6018/rie.33.1.189811
- Marchis, G. (2016). New horizons for innovation in European higher education institutions. *EIRP Proceedings*, 11(1), 405-411. Recuperado de <https://doaj.org/article/019d0593f8ae4ed78e82e2bcb532e01d>
- Molero López-Barajas, D. (2007). Rendimiento académico y opinión sobre la docencia del alumnado participante en experiencias piloto de implantación del Espacio Europeo de Educación Superior. *Relieve: Revista Electrónica de Investigación y Evaluación Educativa*, 13(2) doi:IHIE.0000008148.71650.e6
- Moreno Olivos, T. (2011). Consideraciones éticas en la evaluación educativa. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*. Recuperado de <http://hdl.handle.net/10486/661391>
- Najar, A., Mitrovic, A., y McLaren, B. (2016). Learning with intelligent tutors and worked examples: Selecting learning activities adaptively leads to better learning outcomes than a fixed curriculum. *User Modeling and User-Adapted Interaction*, 26(5), 459-491. doi:10.1007/s11257-016-9181-y

- Papamitsiou, Z., y Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, 17(4), 49-64. Recuperado de <http://search.proquest.com/doc-view/1660157007>
- Perales Montolío, M. J., Jornet Meliá, J. M., y González-Such, J. (2014). *Tendencias en las políticas de formación y evaluación del profesorado en la educación superior en España*. Recuperado de <http://hdl.handle.net/10486/661852>
- Pérez Gómez, Ángel I. (2014). Evaluación externa en la LOMCE. Reválidas, exclusión y competitividad. *Revista Interuniversitaria de Formación del Profesorado*, 81(28.3) (2014), 59-71.
- Pérez Pueyo, Á, Hortigüela Alcalá, D., y Gutiérrez García, C. (2016). Reflexión sobre la evaluación en la formación inicial del profesorado en España. En búsqueda de la concordancia entre dos mundos. *IEYA, Revista Infancia Educación y Aprendizaje*, 2(2), 39-75.
- Poggi, M. (2008). Evaluación educativa: sobre sentidos y prácticas. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, doi:10.15366/riec
- Reyes Dixson, Y., y Nuñez Maturel, L. (2015). El análisis estadístico aplicado a la gestión de la enseñanza para la toma de decisiones. *Revista Cubana de Ciencias Informáticas*, 9(3), 113-127. Recuperado de <https://doaj.org/article/a22f895bca174e4fbc623e5ac4dcc61d>
- Sadler, D. R. (2016). Three in-course assessment reforms to improve higher education learning outcomes. *Assessment & Evaluation in Higher Education*, 41(7), 1081-1099. doi:10.1080/02602938.2015.1064858
- Salafranca Cosialls, L., Sierra Olivera, V., Núñez Peña, M. I., Solanas Pérez, A., y Leiva Ureña, D. (2005). *Análisis estadístico mediante aplicaciones informáticas. SPSS, StatGraphics, MiniTab y Excel* (1ª ed.). Barcelona: Universidad de Barcelona.
- Sánchez Giraldo, D. C., y Escobar Hoyos, G. (2015). La evaluación formativa en los escenarios de educación superior. *Revista De Investigaciones UCM*, 15(2), 204-213. doi:10.22383/ri.v15i2.56
- Sebastian, S., y Puthiyidam, J. J. (2015). Evaluating students' performance by artificial neural network using WEKA. *International Journal of Computer Applications*, 119(23), 36-39. doi:10.5120/21380-4370
- Slater, R. D., Beal-Hodges, M., y Reed, A. (2014). Using excels pivot table function for visual data analysis of exam results: A supplemental procedure to classical test theory. *Academy of Educational Leadership Journal*, 18(4), 221. Recuperado de <http://search.proquest.com/doc-view/1645851108>
- Sin, K., y Muthu, L. (2015). Application of big data in education data mining and learning analytics - A literature review. *ICTACT Journal on Soft Computing*, 5(4), 1035-1049. doi:10.21917/ijsc.2015.0145
- Stevens, D. D., y Levi, A. J. (2013). *Introduction to rubrics : An assessment tool to save grading time, convey effective feedback, and promote student learning* (2nd ed.). Virginia: Stylus Publishing.
- Tiana Ferrer, Alejandro (2014). Veinte años de políticas de evaluación general del sistema educativo en España. *Revista de Evaluación de Programas y Políticas Públicas*, 2(2014). DOI: <http://dx.doi.org/10.5944/reppp.2.2014.12029>
- Tójar Hurtado, J. C., y Velasco Martínez, L. (2015). La rúbrica como recurso para la innovación educativa en la evaluación de competencias. En A. Matas, J. J. Leiva, N. M. Moreno, A. H. Martín y E. López (Eds.), *I seminario internacional de innovación docente e investigación educativa*, 20-51. Sevilla: AFOE.
- Valliant, D. (2008). Algunos Marcos Referenciales en la Evaluación del Desempeño Docente. *Revista Iberoamericana de Evaluación Educativa*, 1(2), 7-22. doi:10.15366/riec



- Vazquez Cano, E. (2016). Dificultades del profesorado para planificar, coordinar y evaluar competencias claves. Un análisis desde la inspección de educación. *Revista Complutense De Educación*, 27(3), 1061-1083. doi:10.5209/rev\_RCED.2016.v27.n3.47400
- Wilcox, C. (2015). *The role of automation in undergraduate computer science education*. 90-95. doi:10.1145/2676723.2677226
- Wojcicki, E., Izumi, L., y Chang, A. (2015). *Moonshots in education: Launching blended learning in the classroom*. United States: Pacific Research Institute.