

Análisis de la opinión del alumnado sobre la aplicación del método de la Campana de Gauss en su evaluación.

Francisco Javier Quirós Tomás
Universidad de Sevilla
quiros@us.es

Resumen

Existen múltiples formas de evaluar a los alumnos. Uno de los enfoques es el conocido como Campana de Gauss, que es un método de medida relativa de los resultados. En el presente trabajo, se comenta la opinión de los alumnos de dos asignaturas de la Universidad de Sevilla sobre la conveniencia y utilidad de dicho método para el caso concreto de su calificación en las mismas.

Palabras clave: *Evaluación, fiabilidad, campana de Gauss, curva de la vitalidad.*

Abstract:

There are many ways to evaluate students. One way is the campana de Gauss (as well known as stack ranking o rank and yank), a relative measurement method. This paper analyses the opinion of students of two subjects at Seville University about the use of this method for their qualification.

Key words: *Evaluation, reliability, stack ranking, rank and yank.*

1. INTRODUCCIÓN

Todo proceso formativo debe ir acompañado de la correspondiente evaluación de los conocimientos adquiridos si se quieren conocer los resultados del proceso. Igualmente, esta evaluación es imprescindible si los conocimientos, habilidades, aptitudes y capacidades adquiridas deben ser acreditados ante terceros.

Por ello es necesario un sistema adecuado y fiable para la evaluación de los discentes. Como indican Landy y Farr (1980), la medición debería tener en cuenta índices objetivos,

personales y juicios de valor. Desafortunadamente es difícil tanto obtener incidencias objetivas como emplear los personales, siendo por ello muy común que los evaluadores recurran a juicios de valor.

Estos tienen su origen en la participación de los evaluadores, siguiendo su propio juicio, en las diversas fases del proceso de evaluación, entre ellas en el establecimiento de criterios, la recogida de información suficiente y relevante sobre la enseñanza-aprendizaje en función de los objetivos curriculares preestablecidos, la interpretación valorativa de dicha información, la emisión de un juicio valorativo y la toma de las decisiones correspondientes (Arredondo et al, 2010).

El empleo de juicios de valor lleva a la frecuente comisión de unos errores típicos (entre ellos el efecto halo, el error de proyección, los estereotipos o los errores de limitación de escala) que han sido estudiados por la Teoría de la Perspectiva, entre cuyos investigadores destacan el premio nobel Daniel Kahneman y Amos Tversky y su trabajo en coautoría "Prospect theory: An analysis of decision under risk" (Kahneman y Tversky, 1979). El caso concreto de la subjetividad cuando se recurre a métodos de evaluación basados en rankings como la campana de Gauss ha sido analizada por autores como Wherry y Bartlett (1982).

Otro problema relacionado con la medición del aprendizaje es el derivado de la falta de fiabilidad. Un sistema de evaluación será fiable si las mediciones obtenidas son consistentes, esto es, si proporcionan la misma calificación para cada alumno, siempre y cuando la característica que se está midiendo (los conocimientos adquiridos por el alumno) no haya cambiado (Alcaide et al, 2011).

Existen distintos tipos de fiabilidades en función de los procedimientos seguidos para su análisis, siendo los más relevantes en el caso analizado la fiabilidad entre evaluadores y la de las formas paralelas (Quirós 2015). Estas fiabilidades se relacionan con la existencia de dos o más evaluadores en el caso de la primera y del empleo de dos o más pruebas con igual formato pero diferente contenido para diversos grupos de alumnos en el de la fiabilidad de formas paralelas.

2. OBJETIVOS

Cada método de evaluación conlleva sus propios problemas de subjetividad a la hora de medir el aprendizaje de los alumnos. En este trabajo se diferencia entre métodos absolutos, que miden el rendimiento de cada alumno de manera individual y métodos relativos que lo miden en comparación con el resto de estudiantes, centrándonos dentro de los métodos de este último tipo en la campana de Gauss.

El objetivo del presente trabajo es conocer la opinión de los alumnos de la Universidad de Sevilla tanto sobre su preferencia general entre un método subjetivo de evaluación, que son los tradicionalmente usados, y otro subjetivo, más concretamente la campana de Gauss, así como sobre su posible aplicación en una asignatura concreta. Igualmente se analizarán sus preferencias por un tipo u otro de método en función de sus características.

2. METODOLOGÍA

Con el fin de conocer la opinión de los alumnos se ha procedido a realizar una encuesta. Se recabó la opinión de los alumnos asistentes a tres grupos de dos asignaturas impartidas en dos facultades de la Universidad de Sevilla en el curso 2017-2018. Dos de los grupos correspondían a alumnos de la asignatura Dirección de Recursos Humanos impartida en el segundo semestre del tercer curso del Grado en Administración de Empresas y el otro a un grupo de Gestión de R.R. H.H. y Habilidades Directivas impartido como optativa en el

Grado en Finanzas y Contabilidad en el segundo cuatrimestre de cuarto curso. El número de encuestas obtenidas es de 66, aunque cuatro de ellas están incompletas en un apartado. De ellas, 46 pertenecen a la primera asignatura (21 y 25 por grupo) y 20 a la segunda.

De entre las preguntas de la encuesta, cuatro de ellas eran relativas a las preferencias entre el método de evaluación absoluto y el relativo (campana de Gauss), en tanto que en otras nueve se testaba su acuerdo o desacuerdo con una serie de cuestiones relativas a las ventajas y desventajas de uno u otro método. La respuesta a las preguntas se hizo mediante la manifestación de las preferencias de los alumnos en una escala Likert de 7 puntos o niveles, siendo por tanto la valoración central 4 y las extremas 1 y 7.

Las respuestas a ambos tipos de items serán analizadas en tres apartados distintos. Por una parte se estudiará la preferencia entre métodos absolutos y relativos, en concreto la campana de Gauss, por otra se contrastarán las opiniones sobre los motivos para elegir uno u otro y, por último, se analizará la tipología de los alumnos mediante un análisis cluster.

3. PREFERENCIA DE MÉTODO DE EVALUACIÓN

Con el fin de analizar la opinión de los alumnos sobre la forma de ser evaluados se incluyeron en la encuesta cuatro preguntas relativas a su preferencia entre el empleo de un método relativo o de uno absoluto (en concreto el de la campana de Gauss).

Dos de las preguntas eran relativas al empleo en general como método de evaluación (codificadas como a1 y a7) y otras dos sobre su empleo en la asignatura en la que se les pasó la encuesta (a4 y a6). Con el fin de ver si sus opiniones se veían afectadas por un conocimiento más profundo de ambos tipos de métodos se repitieron las preguntas antes y después de una explicación de las diferencias entre ellos así como de sus ventajas y desventajas (a1 y a4 se contestaron previamente y a6 y a7 con posterioridad).

En los cuatro items los valores inferiores de la escala Likert se correspondían con la preferencia por un método de valoración absoluta y los elevados con la de uno relativo. Como se puede apreciar en la Tabla 1, el conjunto de respuestas muestran una clara predilección por métodos absolutos, con una mediana y moda de 3 en todos los casos. Igualmente la desviación estándar es reducida, variando entre 1 y 1,4 puntos según la pregunta.

		Estadísticos			
		a1	a4	a6	a7
N	Válido	66	66	66	65
	Perdidos	0	0	0	1
Media		2,85	2,77	2,79	2,69
Mediana		3,00	3,00	3,00	3,00
Moda		3	3	3	3
Desviación estándar		1,384	1,287	1,031	1,103
Mínimo		1	1	1	1
Máximo		7	6	6	6
Percentiles	25	2,00	2,00	2,00	2,00
	50	3,00	3,00	3,00	3,00
	75	4,00	3,00	3,00	3,00

Tabla 1: Estadísticos sobre la preferencia medición absoluta-relativa Fuente: elaboración propia

Es de destacar que ningún alumno valoró con un 7 tres de las preguntas. De igual modo, más del 75% de los alumnos eligieron valores entre el 2 y el 4 en la primera de las preguntas analizadas y entre el 2 y el 3 en las otras 3.

Las preferencias previas y posteriores a la explicación no sufrieron cambios apreciables. En relación con su aplicación a la asignatura la puntuación media varió en 0,15 puntos sobre 7 y la relativa a su aplicación en general en 0,01. De igual modo, aquellos alumnos que cambiaron de opinión fueron relativamente pocos y sus cambios no fueron sustanciales. Así, el 91% de ellos no variaron su valoración en más de un punto en el caso de la signatura y el 83% en el caso de su aplicación general. Hubo únicamente un total de tres casos (2,3%) en los que un alumno cambió en más de dos niveles entre la respuesta previa y la posterior en alguno de los pares de preguntas asociados.

Se ha analizado la correlación entre las cuatro respuestas empleando el coeficiente de correlación de Spearman, que es adecuado para variables ordinales medidas con escalas Likert (Martínez et al, 2009). Como se recoge en la Tabla 2, existe una correlación positiva significativa bilateral en el nivel 0,01 en los seis pares posibles, variando el Rho de Spearman entre un mínimo de 0,480 y un máximo de 0,717, mostrándose con ello la consistencia en las respuestas por parte de los encuestados.

Rho de Spearman

		a1	a4	a6	a7
a1	Coeficiente de correlación	1	,717**	,538**	,601**
	Sig. (bilateral)	.	0	0	0
	N	66	66	66	65
a4	Coeficiente de correlación	,717**	1	,702**	,480**
	Sig. (bilateral)	0	.	0	0
	N	66	66	66	65
a6	Coeficiente de correlación	,538**	,702**	1	,607**
	Sig. (bilateral)	0	0	.	0
	N	66	66	66	65
a7	Coeficiente de correlación	,601**	,480**	,607**	1
	Sig. (bilateral)	0	0	0	.
	N	65	65	65	65

** La correlación es significativa en el nivel 0,01 (bilateral).

Tabla 2: Correlaciones de Spearman a1-a4-a6-a7. Fuente: elaboración propia

Como complemento al análisis estadístico se ha procedido a un análisis gráfico de las respuestas a dichas preguntas, recogido en los Gráficos 1 y 2.

El primero de ellos recoge las respuestas a los dos items referidos al método de evaluación preferido antes (a4) y después (a6) de la explicación en clase de ambos sistemas. Unos valores bajos muestran preferencia por un método de medición absoluto y unos elevados preferencia por uno relativo.

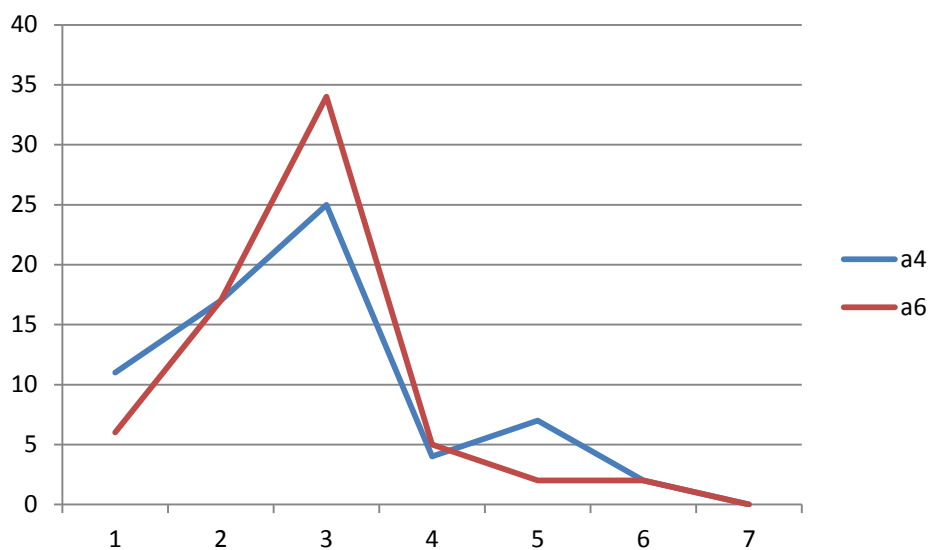


Gráfico 1: Método de evaluación preferido. Fuente: elaboración propia

Como se puede observar, en ambos momentos temporales hay una clara preferencia por los métodos absolutos con una alta concentración de las respuestas en los valores 3 y 2. También puede observarse que las opiniones se concentran en torno al valor 3 una vez conocidos con mayor profundidad ambos modelos, como muestra la disminución de la desviación estandar (de 1,287 a 1,031).

Respecto a su predilección por aplicar uno u otro tipo de método en la asignatura ocurre algo parecido. Las respuestas 3 y 2 son las más frecuentes, siendo mayor su frecuencia y menos la dispersión una vez explicados ambos sistemas (Gráfico 2).

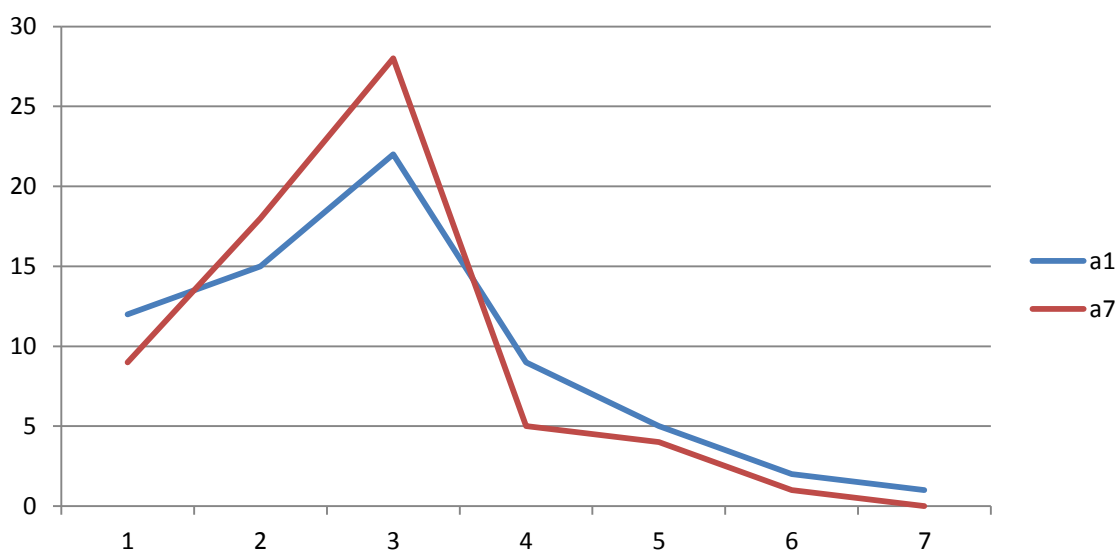


Gráfico 2: Método de evaluación preferido para la asignatura. Fuente: Elaboración propia

4. CARACTERÍSTICAS DE AMBOS METODOS DE EVALUACIÓN

Múltiples autores han analizado las ventajas e inconvenientes de los diversos métodos de evaluación, tanto absolutos como relativos y algunos específicamente los de la campana de

Gauss. Entre ellos se encuentran Giumetti et al. (2015), Osborne y McCann (2004), Stewart et al. (2010), Berger et al. (2013), Blume et al. (2013) o Brown (2011). En este epígrafe se analizan las opiniones de los encuestados relativas a varias de dichas ventajas e inconvenientes, así como sobre las características específicas de cada método.

En el caso de la campana de Gauss algunas de sus características tienen su origen en la base del método, siendo la mayoría de ellas comunes a otros métodos relativos. Entre estas hay tres en las que se ha requerido la opinión a los encuestados. Dos de ellas hacen referencia a la valoración del aprendizaje no en función de las características o el esfuerzo individual de cada alumno sino en función de su comparación con los del grupo. La tercera es que este método parte de que la variable aprendizaje es una función normal, forzando la distribución de las notas obtenidas para que adopten una distribución predeterminada.

Entre las ventajas y desventajas de ambos tipos de métodos sobre cuya opinión se preguntó en la encuesta se encuentran el verse afectados en mayor o menor medida los resultados por la variada dificultad de distintas pruebas de evaluación, los efectos de la heterogeneidad entre los diversos grupos de alumnos, las diferencias en las calificaciones debidas al uso de distintos sistemas de evaluación, el fomento de la competencia o de la cooperación entre el alumnado, evitar diferencias debidas a falta de fiabilidad entre evaluadores, la creencia en la justicia del método o sus efectos sobre las notas del conjunto de los alumnos. La encuesta recoge la opinión sobre todas estas variables, salvo la última, para la cual se ha analizado la correlación entre las preferencias de los alumnos por un método u otro y las notas obtenidas en la asignatura correspondiente en la evaluación continua.

Los items consisten en una proposición sobre la cual los alumnos deben mostrar su acuerdo o desacuerdo. Las respuestas con valores bajos muestran un alto desacuerdo con la proposición, en tanto que los elevados muestran un alto acuerdo.

Preguntados por su preferencia entre un método en el que la evaluación de cada alumno dependiera de sus características (absoluto) o de las del grupo (relativo) los encuestados decantaron claramente por la preferencia de un método de medición absoluta que usa valoraciones individuales. Como se puede observar en el Gráfico 3 la moda fue 6, siendo el segundo valor más frecuente 7.

"Prefiero el método absoluto por ser independiente la valoración de cada alumno"

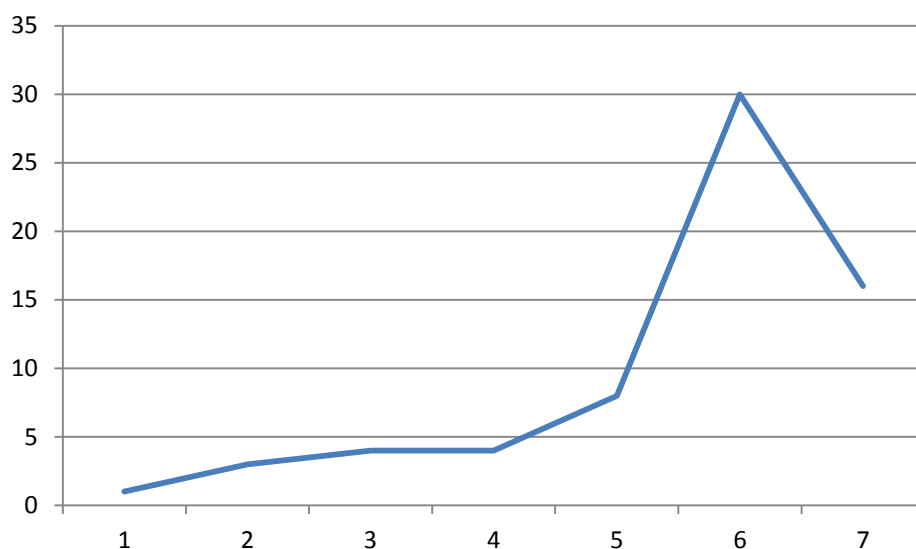


Gráfico3: Prefiero el método absoluto por ser independiente la valoración de cada alumno." Fuente: elaboración propia

El método de la campana de Gauss parte de que la variable aprendizaje es una función normal. Existen múltiples modelos que analizan el aprendizaje recurriendo a innumerables variables. Entre ellas se pueden destacar las del modelo el Modelo 3P de Biggs (1987) o los dos grandes tipos de variables propuestas por Schunk (1997): variables asociadas al medio y al estudiante. Independientemente del modelo preferido, resulta muy complicado dilucidar y ponderar la influencia específica de cada una de las posibles variables en el aprendizaje de un alumno concreto (López et al., 2007).

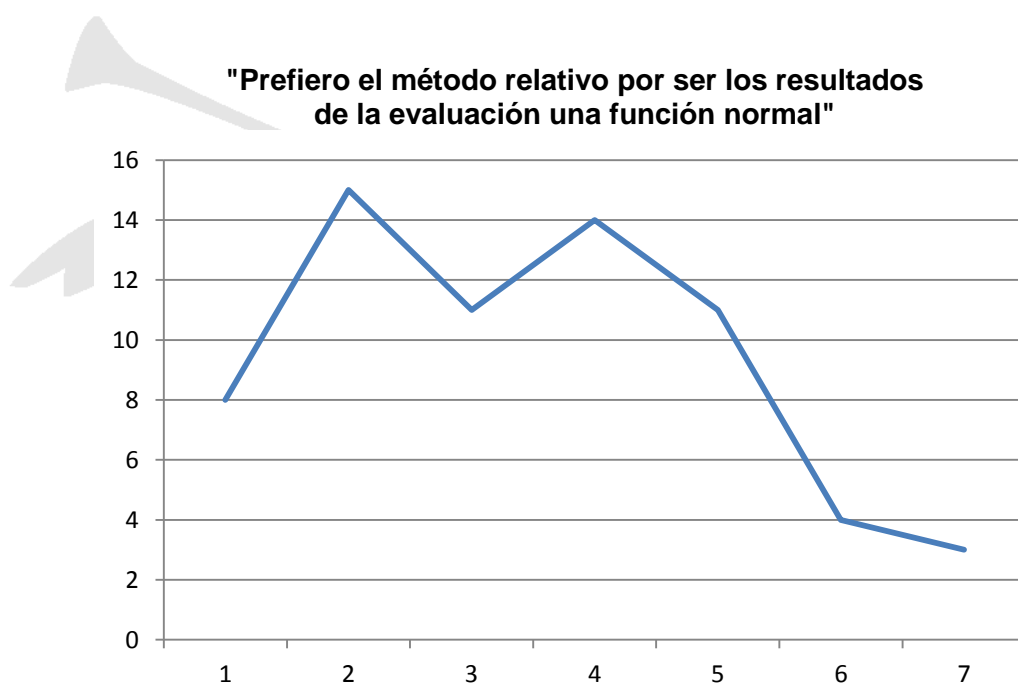


Gráfico 4: *"Prefiero el método relativo por ser los resultados de la evaluación una función normal!"* Fuente: elaboración propia

Por ello, es posible partir de la base de que, aún desconociendo los mecanismos que subyacen a la variable aprendizaje, cada observación será efecto de la suma de una serie de causas independientes y por tanto el aprendizaje adoptará la forma de una función normal. La campana de Gauss fuerza la distribución normal de las calificaciones obtenidas en un proceso de evaluación, adaptándose por tanto al supuesto de normalidad.

Preguntados sobre la preferencia de un método que tenga en cuenta la normalidad de la variable "resultados de la evaluación" (Gráfico 4), la respuesta de los alumnos muestra una

gran dispersión, obteniendo entre 11 y 15 repuestas cada uno de los valores comprendidos entre el 2 y el 5, con una mediana de 3.

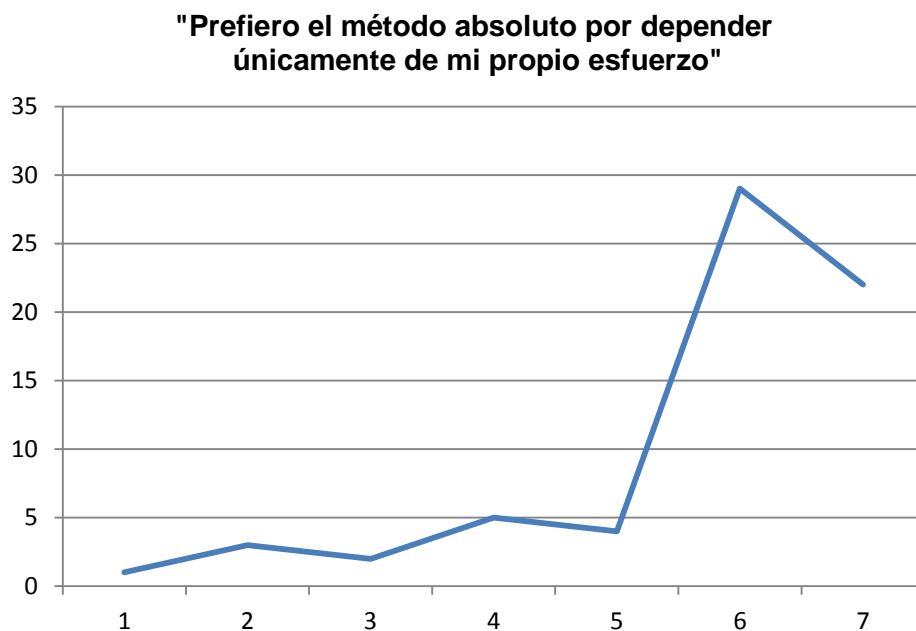


Gráfico 5: "Prefiero el método absoluto por depender únicamente de mi propio esfuerzo". Fuente: elaboración propia

Preguntados sobre si debería primar el esfuerzo individual de alumno o el esfuerzo relativo los encuestados se decantaron claramente por un método de medición absoluto, que tiene entre sus características básicas que los resultados de la evaluación dependen del esfuerzo individual (Gráfico 5). Los valores más frecuentes fueron 6 y 7, ambos con más de 20 respuestas, siendo la mediana 6.

El empleo de diversos sistemas de evaluación, como por ejemplo al procederse sucesivamente primero a una evaluación continua y posteriormente a una final en el caso de los estudios universitarios de grado, puede dar origen a agravios comparativos entre los alumnos que participen en uno y otro sistema por el diferente nivel de dificultad de las pruebas correspondientes. Esto puede evitarse empleando sistemas de evaluación como la campana de Gauss. En el ítem correspondiente los alumnos adoptaron una posición intermedia, siendo la mediana 4 y las respuestas más frecuentes 4 y 5, con 15 casos cada una (Gráfico 6).

Los métodos de evaluación relativos presentan un claro problema de objetividad en el caso de que distintos grupos de alumnos evaluados no sean homogéneos en su composición. En dicho caso las posiciones relativas en uno y otro grupo pueden no ser comparables. Así, por ejemplo, en el caso de los estudios de grado podrían existir diferencias entre grupos de mañana y tarde o en función de la convocatoria a la que se presentan por la diferente tipología de los alumnos.

"Prefiero el método relativo por no verse afectado por la mayor o menor dificultad de las pruebas de evaluación"

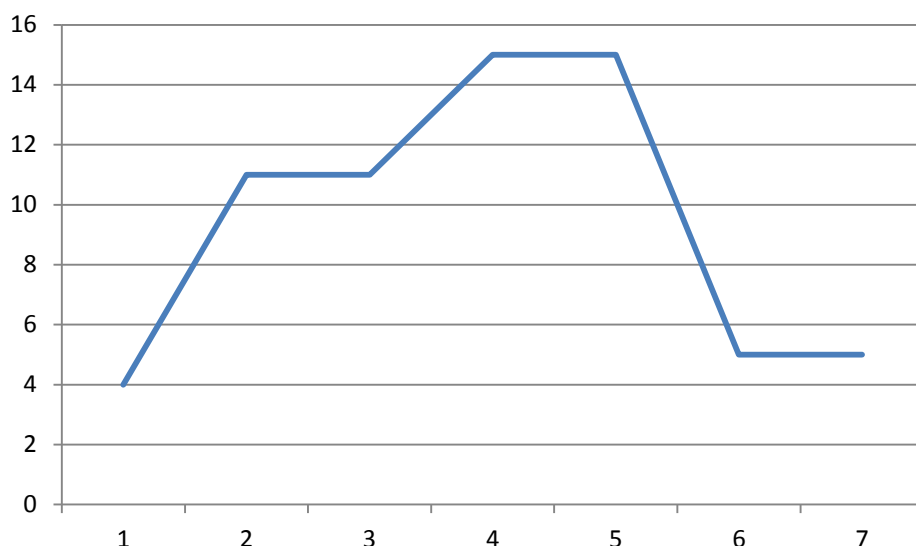


Gráfico 6: *Prefiero el método relativo por no verse afectado por la mayor o menor dificultad de las pruebas de evaluación.* Fuente: elaboración propia

Preguntados por si preferían un método donde la heterogeneidad entre alumnos se tuviera en cuenta (método absoluto) o uno en el que no (método relativo), los alumnos se decantaron claramente por un método de evaluación que no se viese afectado por la falta de homogeneidad intergrupos (Gráfico 7), siendo 6 la moda con más de 20 respuestas y la mediana 5. Esta pregunta se caracteriza por ser la que tiene una menor dispersión en la respuesta al obtener una desviación estandar de 1,3.

"Prefiero el método absoluto por no verse afectado por la posible heterogeneidad entre los grupos de alumnos"

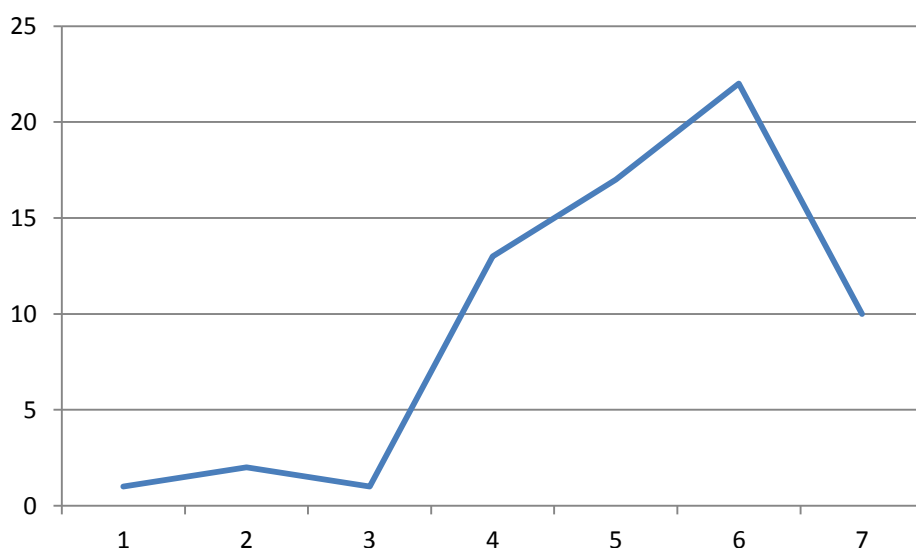


Gráfico 7: *Prefiero el método absoluto por no verse afectado por la posible heterogeneidad entre los grupos de alumnos.* Fuente: elaboración propia

Es frecuente que, aún cuando se utilice un mismo sistema de evaluación, dos pruebas distintas no tengan la misma dificultad para el alumnado. Esto es especialmente relevante en el caso de procesos de evaluación continua cuando se realizan pruebas paralelas para cada

grupo de alumnos, o en el caso de pruebas realizadas en distintas convocatorias. Los resultados pueden variar en mayor o menor medida en función de en cuál de ellas se ha evaluado un alumno, independientemente de que su nivel de conocimientos. Si se quiere que las evaluaciones obtenidas por dos alumnos que han sido sometidos a pruebas paralelas sean realmente comparables se debe exigir una elevada fiabilidad de las formas paralelas, algo que no siempre ocurre.

La fiabilidad entre formas paralelas se analiza mediante la correlación entre resultados obtenidos en dos pruebas similares (paralelas) por un mismo grupo de sujetos, normalmente dejando entre una y otra un corto periodo para evitar cambios, pero suficientemente largo como para evitar la fatiga de los sujetos (Díaz et al, 2003). Para ello se requiere la realización de dos pruebas similares, es decir, usando la misma técnica pero con un contenido distinto, de forma que se mida la misma variable, obteniendo el coeficiente de equivalencia entre las mismas (Alcaide y González, 1997). La campana de Gauss es muy efectiva a la hora de evitar problemas derivados de una baja fiabilidad entre formas paralelas.

La opinión de los alumnos respecto a esta característica se decanta por el empleo de métodos relativos como la campana de Gauss. La mayoría puntúa la variable en valores entre el 4 y el 6 en la correspondiente escala Likert (Gráfico 8), obteniendo estos valores entre 12 y 14 respuestas.

"Prefiero el método relativo por evitar el problema de usar distintos modelos de examen que pueden tener distinto grado de dificultad en una misma evaluación"

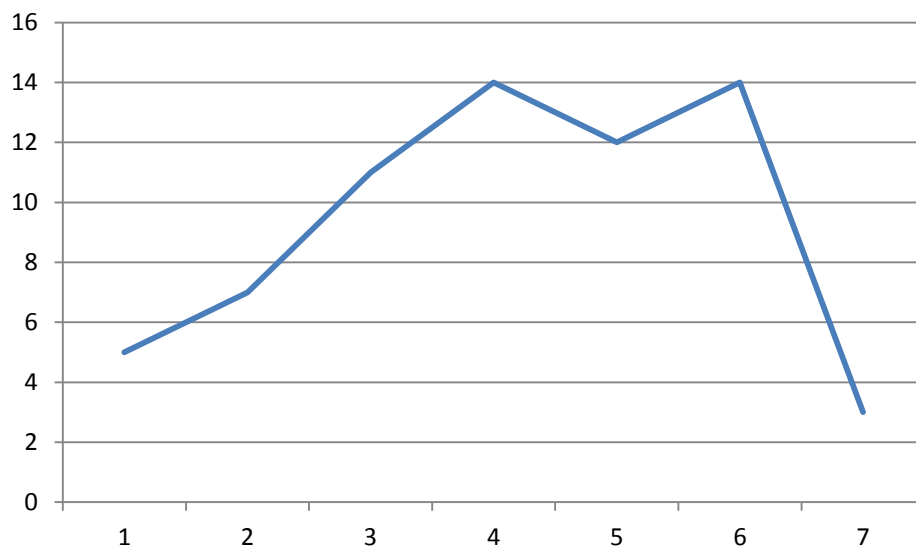


Gráfico 8: *Prefiero el método relativo por evitar el problema de usar distintos modelos de examen que pueden tener distinto grado de dificultad en una misma evaluación.* Fuente: elaboración propia

Cuando la calificación obtenida depende de la posición relativa entre los diversos evaluados, hay una tendencia a que se produzcan conflictos competitivos entre ellos (Berger et al., 2013). Este inconveniente podría llegar a ser grave, especialmente en la actualidad, ya que pese a la fuerte presión competitiva cada vez más tanto las empresas como la sociedad

demandan personas con una elevada capacidad para el trabajo en equipo y la cooperación. Este inconveniente podría incluso ser visto como una ventaja el caso de empresas y sociedades individualistas donde prima la competencia sobre la solidaridad.

Preguntados al respecto, los alumnos no mostraron una preferencia clara, como se ve en el Gráfico 9. Destaca el valor 6 con un total de 18 repuestas, estando el resto muy repartidas, obteniendo cinco de las otras respuestas un valor entre 8 y 11, siendo desviación típica la más elevada de las obtenidas en los diversos ítems (1,78).

"Prefiero el método absoluto por la menor competencia entre alumnos"

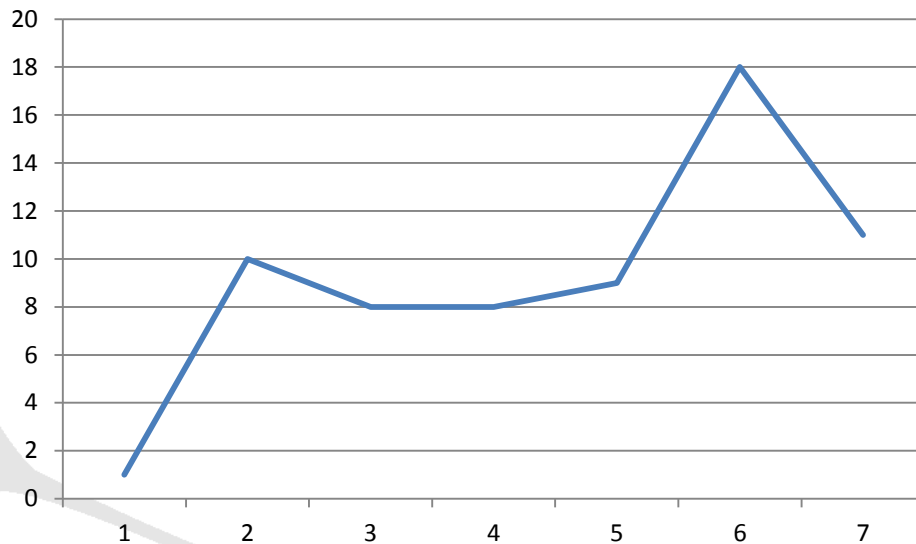


Gráfico 9: *Prefiero el método absoluto por la menor competencia entre alumnos.* Fuente elaboración propia

Otra variable analizada es la fiabilidad entre evaluadores. Ésta es importante cuando la evaluación del alumnado se lleva a cabo por diversos evaluadores y se requiere del juicio de los mismos a la hora de establecer las calificaciones. Cada profesor emplea sus propios juicios de valor, siendo necesario que se establezcan unos criterios comunes si se desea que los resultados de un alumno concreto no se vean afectados por cuál sea el profesor que lo evalúe.

"Prefiero el método relativo por evitar diferencias de notas en función del profesor que me evalúe"

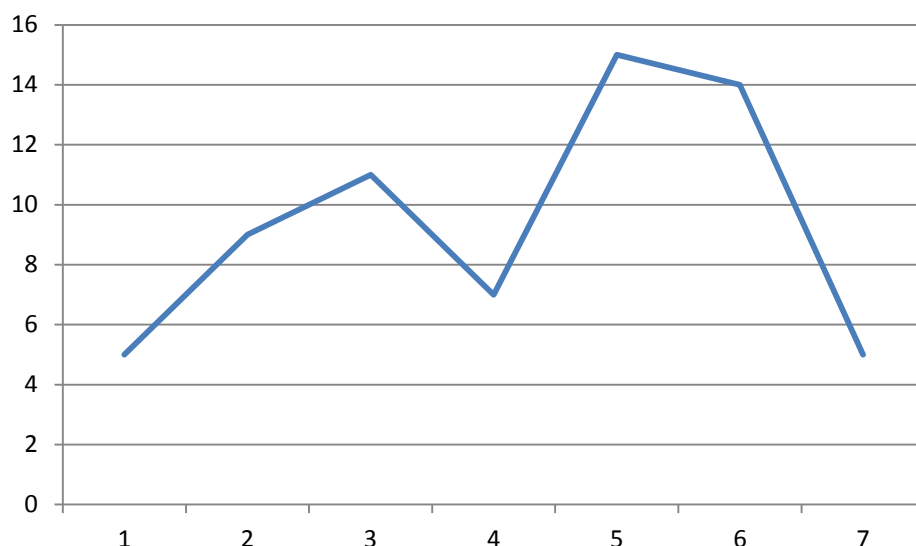


Gráfico 10: *Prefiero el método relativo por evitar diferencias de notas en función del profesor que me evalúe.*
Fuente: elaboración propia

La fiabilidad entre evaluadores se mide correlacionando las valoraciones emitidas por cada uno de ellos respecto a una misma y única prueba de evaluación realizada por cada alumno (Alcaide y González, 1997). Una baja fiabilidad entre evaluadores lleva a una falta de homogeneidad en las calificaciones otorgadas por unos y otros. Los métodos de evaluación relativa como la campana de Gauss tienden a eliminar este problema.

Los resultados de la encuesta, representados en el gráfico 10, muestran una cierta preferencia por los métodos de evaluación relativos, con una mediana y una moda coincidentes (valor 5), pero con una elevada dispersión. La desviación estándar es la segunda más elevada entre las diversas preguntas analizadas, con un valor de 1,78.

Otro aspecto a tener en cuenta al decantarse por un método u otro es la percepción de justicia que los evaluados puedan tener sobre cada uno de ellos. Según diversos autores (Roch et al., 2007 o Schleicher et al., 2009) los evaluados tienden a calificar cómo más justos los métodos basados en medidas absolutas que los basados en medidas absolutas.

Al manifestar su opinión sobre el tipo de método más justo los alumnos se decantan claramente por el empleo de métodos absolutos. Como muestra el Gráfico 11, 45 de los encuestados se decantan por los valores 6 o 7, siendo la mediana 6.

"Prefiero el método absoluto por considerarlo más justo"

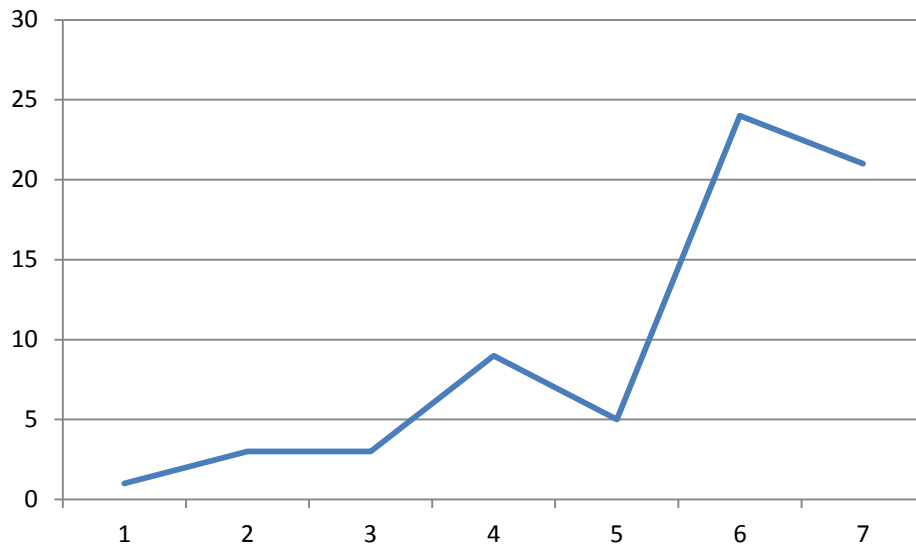


Gráfico 11: Prefiero el método absoluto por considerarlo más justo." Fuente: elaboración propia

Por último, hay opiniones contrapuestas sobre si el empleo de un método u otro afecta a los resultados de la evaluación. Algunos autores como Berger et al. (2013) creen que los métodos de distribución forzada como la campana de Gauss refuerzan los resultados, con mejoras medias de entre un 6% y un 12%, siempre y cuando no se den conflictos competitivos. Por el contrario, Brown (2011) cree que estos métodos pueden reducir el rendimiento de los sujetos con peores capacidades. Scullen et al. (2005), por su parte, afirman que, en conjunto, el rendimiento de los evaluados tiende a aumentar. La mejora de la productividad podría estar ligada a que aquellas organizaciones que recurren a sistemas de evaluación relativos tienden a atraer a sujetos con altas habilidades cognitivas (Blume et al., 2013).

Con el fin de testar la hipótesis de Blume et al. (2013) se ha realizado un análisis de la correlación entre la preferencia por uno u otro modelo de evaluación y las notas obtenidas por los alumnos. Se ha calculado la correlación de Spearman entre las notas obtenidas y las preferencias directas de los alumnos por uno u otro método analizadas en el epígrafe tercero. Dado los diferentes conocimientos, sistemas de evaluación y contenidos de las pruebas realizadas en las dos asignaturas y tres grupos a los que pertenecen los encuestados se ha considerado necesario segmentar las encuestas. Así, se han creado tres grupos formados por los alumnos de Gestión de Recursos Humanos y por los de los grupos 2 y 4 de Dirección de Recursos Humanos. No se precieron correlaciones significativas en ninguno de los casos.

Análisis de Correlaciones (Rho de Spearman)

Grupo/variable		a1	a4	a6	a7
GRH nota	Coeficiente de correlación	0,123	0,147	0,044	-0,265
	Sig. (bilateral)	0,606	0,536	0,853	0,258
DRH2 nota	Coeficiente de correlación	-0,081	-0,104	-0,225	-0,136
	Sig. (bilateral)	0,715	0,637	0,302	0,537
DRH4 nota	Coeficiente de correlación	-0,019	-0,013	0,240	-0,068
	Sig. (bilateral)	0,935	0,956	0,296	0,776

Tabla 3: Análisis de Correlaciones (Rho de Spearman). Fuente: elaboración propia

5. TIPOLOGÍA DE LOS ALUMNOS. ANÁLISIS CLUSTER

Con el fin de analizar la tipología de los alumnos se ha procedido a su agrupación mediante un análisis cluster. Esta técnica permite la clasificación de una serie de individuos en grupos con homogeneidad intragrupo y heterogeneidad intergrupos mediante el empleo de técnicas multivariantes (Scott y Knott, 1974). Más concretamente, se ha considerado adecuado emplear el análisis de clústeres jerárquicos con el método de agrupación de enlace entre grupos, empleando la distancia euclídea al cuadrado dadas las características de las variables a analizar (Quirós, 2015).

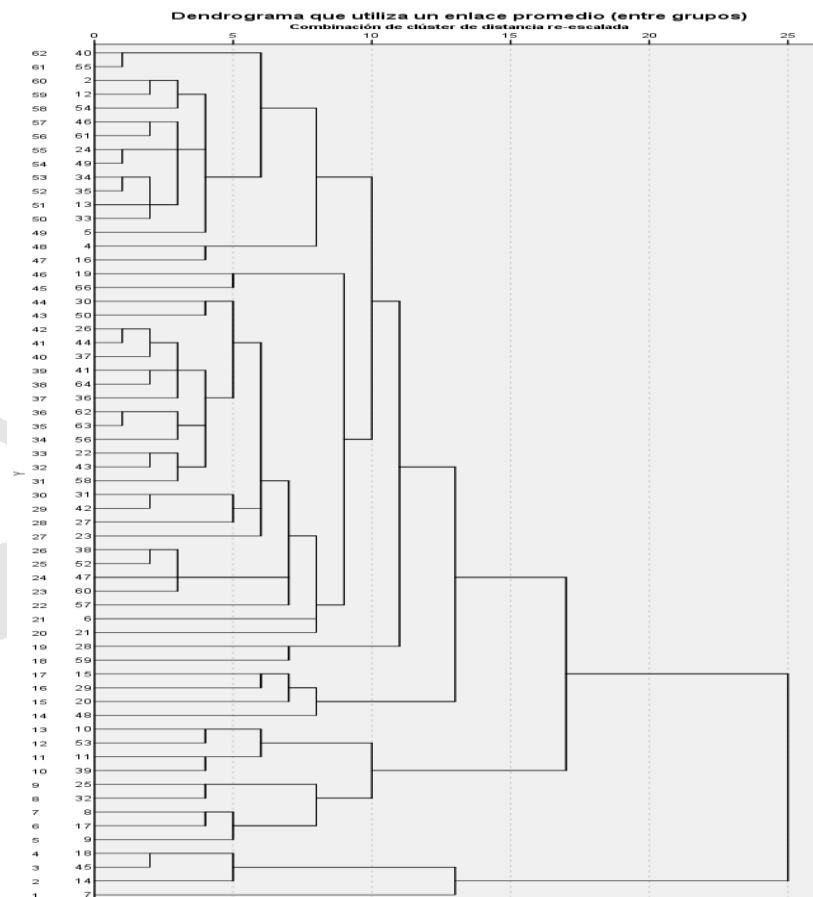


Figura 1: Análisis cluster. Fuente: elaboración propia

Del total de 66 encuestas, se han utilizado para esta parte del trabajo únicamente 62, al faltar en las cuatro restantes información sobre alguno de los ítems analizados. Analizando los 4 clusters principales se puede observar una clara descompensación en el número de estudiantes clasificados en cada uno de ellos. El cluster con más elementos (cluster A) incluye un 73% del total (45 casos), el siguiente en tamaño (Cluster C) cuenta con el 14,5% de los casos analizados, en tanto que los dos restantes incluyen cada uno un 6,5% (4 observaciones).

Los clusters A y D incluyen a alumnos que prefieren claramente ser evaluados mediante métodos de medición absoluta, en tanto que el C comprende a aquéllos que prefieren métodos de medición relativa como la campana de Gauss. Los del cluster B no se decantan claramente por una u otra opción. Analizando el conjunto de los clusters, la amplia mayoría de los encuestados (79%) prefieren métodos de medición absoluta para su evaluación, que es la forma comúnmente empleada en el sistema educativo, siendo una clara

minoría los que escogen la medición relativa (14,5%).

Una segunda diferencia entre los miembros de cada cluster proviene de sus respuestas a las ocho preguntas relativas a los motivos para preferir uno u otro método.

Los alumnos del cluster más amplio (A) son el grupo que está más de acuerdo con los motivos indicados para referir los métodos de medición absoluta, siendo el segundo con una mediana más baja respecto a los de la relativa. Los del cluster B, por su parte, pese a su indiferencia relativa al empleo de uno u otro método son el grupo que menos valoran los motivos para usar métodos relativos.

Los alumnos agrupados en el cluster C se muestran indiferentes respecto a las afirmaciones sobre los métodos absolutos, en tanto que se muestran de acuerdo con las referidas a los métodos relativos, con unos valores medianos de 4 y 5 respectivamente. Prefiriendo en su conjunto ser evaluados mediante métodos relativos.

Por último, el cluster D agrupa a un conjunto de alumnos cuyas respuestas son inconsistentes. Pese a estar de acuerdo con los motivos para el empleo de métodos relativos y ser los que menos valoran el recurso a los absolutos en función de sus características prefieren ser evaluados siguiendo métodos absolutos.

6. CONCLUSIONES

Existen dos tipos de métodos para evaluar el aprendizaje, los basados en mediciones absolutas y los basados en mediciones relativas, siendo el método de la campana de Gauss uno de estos últimos. Cada uno de ellos tiene sus ventajas e inconvenientes que hace que evaluadores y evaluados prefieran unos u otros.

La encuesta realizada a alumnos de dos Grados de la Universidad de Sevilla busca estudiar sus opiniones al respecto. Los resultados muestran una clara preferencia por los métodos absolutos, tanto en general como en el caso particular de su uso en la asignatura concreta en la que se les explicaron las características de ambos métodos. En una escala Likert de 7 niveles, donde los valores inferiores representaban la preferencia por métodos absolutos y los más elevados la preferencia por métodos relativos, se obtuvieron a una batería de cuatro preguntas unas valoraciones con una mediana y moda de 3, estando la mayoría de las respuestas en los niveles 2 y 3.

La encuesta también midió la preferencia por unos u otros métodos en función de las características diferenciales entre ambos sistemas. Las respuestas mostraron una preferencia por los métodos absolutos y sus ventajas frente a los relativos. Las preguntas sobre las ventajas de los absolutos, como la independencia en la valoración de cada alumno, depender del esfuerzo desplegado individualmente, evitar problemas de heterogeneidad intergrupos, una menor competitividad entre alumnos o un mayor sentimiento de justicia conllevaban una fuerte preferencia por métodos absolutos, con valores medianos de entre 5 y 6, con modas en todos los casos de 6. Encuestados sobre las ventajas de los métodos relativos, como mantener la normalidad de los resultados, no verse afectados por la metodología de evaluación ni por la variable dificultad las pruebas a que se someten los diferentes grupos (fiabilidad de formas paralelas) o evitar problemas de fiabilidad de juicio, las alumnos no mostraron una preferencia tan clara. Se obtuvieron valores medianos entre 3 y 5 con modas entre 2 y 5, siendo 4 la indiferencia entre uno y otro tipo de métodos.

Por último, se analizó la posible existencia de grupos de alumnos con preferencias diferentes mediante un análisis cluster jerárquico. Se encontraron cuatro grupos de alumnos, uno formado por la mayoría de los mismos (73%) mostraba una clara preferencia por los métodos absolutos, tanto al ser preguntados por ellos directamente como a la hora de tener en

cuenta sus ventajas e inconvenientes. Otro grupo (14,5% de los encuestados) se encontraba en el caso opuesto, prefiriendo métodos relativos. Otros dos grupos minoritarios (6,5% ambos) incluían a alumnos que no se decantaban por uno u otro método aun estando más de acuerdo con las ventajas de los métodos absolutos y a aquellos que valorando más las ventajas de los métodos relativos que las de los absolutos preferían ser evaluados mediante estos últimos.

REFERENCIAS

- Alcaide M., González, M. & Florez, I. (2011). *Dirección de Recursos Humanos I: La función de personal en la empresa* 2011. Sevilla: Editorial Atril 97, S.L
- Alcaide, M. & González, M. (1997) *Temas actuales de Dirección de Recursos Humanos*. Sevilla: Editorial Atril 97, S.L.
- Arredondo, S. C., Diago, J. C. & Cañizal, A. (2010). *Evaluación educativa de aprendizajes y competencias*. Pearson Educación.
- Berger, J., Harbring, C. & Sliwka, D. (2013). Performance appraisals and the impact of forced distribution—An experimental investigation. *Management Science*, 59(1), 54-68.
- Biggs, J. B. (1987). *Student approaches to learning and studying*. Hawthorn, Vic.: Australian Council for Educational Research.
- Blume, B. D., Rubin, R. S. y Baldwin, T. T. (2013). Who is attracted to an organisation using a forced distribution performance management system?. *Human Resource Management Journal*, 23(4), 360-378.
- Brown, J. (2011). Quitters never win: The (adverse) incentive effects of competing with superstars. *Journal of Political Economy*, 119(5), 982-1013.
- Díaz, C., Batanero, C. & Cobo, B. (2003). Fiabilidad y generalizabilidad. *Aplicaciones en evaluación educativa*. Número 54, 3-21.
- Giumetti, G. W., Schroeder, A. N., & Switzer III, F. S. (2015). Forced distribution rating systems: When does “rank and yank” lead to adverse impact?. *Journal of Applied Psychology*, 100(1), 180.
- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, 263-291.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87(1), 72.
- López, B. G., Pérez, C. P., Carbonell, B. S., Peris, F. S. I. & Ros, I. R. (2007). Actitudes ante el aprendizaje y rendimiento académico en los estudiantes universitarios. *Revista Iberoamericana de Educación*, 42(1), 6.
- Martínez, R. M., Tuya, L. C., Martínez, M., Pérez, A., & Cánovas, A. M. (2009). El coeficiente de correlación de los rangos de Spearman caracterización. *Revista Habanera de Ciencias Médicas*, 8(2), 0-0.
- Osborne, T., & McCann, L. A. (2004). Forced ranking and age-related employment discrimination. *Hum. Rts.*, 31, 6.
- Quirós, F. J. (2015). Análisis de las Tendencias en Gestión de los Recursos Humanos desde una Perspectiva Académica y Empresarial (Tesis doctoral no publicada). Universidad de Sevilla
- Quirós, F. J. (2017). Evaluación del alumnado mediante la medida relativa de sus resultados: el método de la campana de Gauss. *VIII Jornada de Innovación e Investigación Docente (2017)*(2017), p 66-81, 66-81.
- Roch, S. G., Sternburgh, A. M., & Caputo, P. M. (2007). Absolute vs relative performance rating formats: Implications for fairness and organizational

- justice. *International Journal of Selection and Assessment*, 15(3), 302-316.
- Schleicher, D. J., Bull, R. A. y Green, S. G. (2009). Rater reactions to forced distribution rating systems. *Journal of Management*, 35(4), 899-927.
- Schunk, D. H. (1997). *Teorías del aprendizaje*. Pearson educación.
- Scullen, S. E., Bergey, P. K., & Aiman-Smith, L. (2005). Forced distribution rating systems and the improvement of workforce potential: A baseline simulation. *Personnel Psychology*, 58(1), 1-32.
- Scott, A. J. & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 507-512.
- Stewart, S. M., Gruys, M. L., & Storm, M. (2010). Forced distribution performance evaluation systems: Advantages, disadvantages and keys to implementation. *Journal of Management & Organization*, 16(1), 168-179.
- Wherry, R. J., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology*, 35(3), 521-551.

