

ESTUDIO DE LA APLICACIÓN DEL SISTEMA DE EVALUACIÓN POR PARES A UNA ASIGNATURA DE EXPERIMENTACIÓN¹

José Muñoz García
M^a del Carmen Alfaro Rodríguez
Juan Cota Galán
*Departamento de Ingeniería Química
Facultad de Química
Universidad de Sevilla*

Resumen

En este trabajo se evalúa la eficacia del sistema de evaluación por pares en una asignatura de experimentación de tercer año universitario. El estudio se prolonga a lo largo de dos cursos académicos consecutivos, utilizando los mismos tres pares de correctores. Los resultados obtenidos se comparan con los que se habrían alcanzado de haber utilizado tres correctores. Se presentan diversas estrategias de análisis estadístico de las distintas calificaciones propuestas. La distribución de calificaciones por debajo o por encima de la línea recta de pendiente uno, en la representación lineal de las calificaciones de uno de los miembros de la pareja frente al otro, aporta una herramienta cómoda para un primer análisis de la consistencia de las calificaciones propuestas. La representación de las calificaciones normalizadas respecto a las medias permite también una rápida evaluación de la eficacia de una triple corrección frente a la corrección por pares.

Palabras clave: corrección de exámenes, corrección por pares, calificaciones, evaluación

Abstract

This article deals with the assessment of the efficiency of the pairwise exam marking system applied to a laboratory subject offered in the third university year. This project was carried out over two consecutive academic years with the help of the same three pairs of examiners. The results obtained are compared with those of three examiners (triple marking). Several strategies for the statistical analysis of marks proposed by examiners are considered. The distribution of the marks given by each examiner against those of a second examiner provides a useful tool to perform an initial analysis of results, insofar as they can scatter below or above the straight line of slope 1. The plot of normalised marks versus average marks enables clear conclusions to be drawn on the efficiency of triple versus pairwise marking.

Keywords: exam marking, pairwise marking, marks, evaluation

1. Este trabajo presenta los resultados de un proyecto aceptado en la Convocatoria de Proyectos de Innovación y Mejora Docente 2008/2009 en el marco del I Plan Propio de Docencia de la Universidad de Sevilla.

1. INTRODUCCIÓN. ANTECEDENTES

La asignatura Experimentación en Ingeniería Química I se puede considerar como un ejemplo de materia cuya enseñanza se centra en el trabajo en laboratorio. Es de esperar que las conclusiones a las que se llega sean extrapolables a otras asignaturas de carácter experimental. Se ha de destacar que la enseñanza de materias asociadas a trabajo en laboratorios y plantas piloto está afortunadamente muy extendida en titulaciones de Ciencias, Ciencias de la Salud, Ingenierías Técnicas e Ingenierías Superiores. Por tanto, se puede ser optimista respecto a la rentabilidad de las conclusiones generales de este proyecto ya que se podrían aplicar a una gran variedad de asignaturas de experimentación. La asignatura elegida, se imparte en 3º de la Licenciatura en Química en la Universidad de Sevilla, siendo materia obligatoria para todos los alumnos que cursan dicha Licenciatura. Se divide en cuatro módulos, impartidos por profesores especializados en cada uno de ellos: a) Experimentación en Flujo de Fluidos, b) Experimentación en Transmisión de Calor, c) Experimentación en Transferencia de Materia y d) Experimentación en Cinética Aplicada de Reacciones Químicas. El presente trabajo se aplica sobre el módulo d) y en él han participado los tres profesores que lo imparten, si bien todos los alumnos realizan prácticas de laboratorio de los cuatro módulos.

El sistema de evaluación estándar abarca dos pasos. El primero consiste en el análisis y evaluación por parte del profesor del trabajo realizado en el laboratorio y del informe presentado por la pareja de trabajo. Cada profesor, realiza un análisis crítico por escrito del informe que recibe, y tras entrevistar a los alumnos, lo devuelve para su estudio o modificación. El segundo, consiste

en la realización de un examen global sobre la materia, que es idéntico para cada grupo. Considerando que existen distintas prácticas de laboratorio en el módulo d), que cada práctica se puede realizar bajo diferentes condiciones experimentales y que participan diferentes profesores, la aplicación de un sistema de evaluación que garantice una valoración idéntica del trabajo de los alumnos constituye todo un reto, sobre todo teniendo en cuenta que existe un riesgo de que la calificación del examen por parte del profesor esté influida por su opinión respecto al rendimiento de los alumnos en el laboratorio.

2. OBJETIVOS

El objetivo que se plantea en este trabajo es el de cuantificar la eficacia de la evaluación por pares de una muestra estadísticamente significativa de exámenes globales de la materia impartida en el módulo d) Experimentación en Cinética Química Aplicada. Se compara también el efecto de una triple corrección, para saber si la corrección por pares se puede validar. Debe tenerse en cuenta que la evaluación por pares es una práctica común tanto en la evaluación de proyectos de investigación, como de los sistemas de revisión de trabajos científicos profesionales y que se realiza en algunas universidades del entorno europeo (Grau, Rosa y otros, 2002).

3. METODOLOGÍA

Se han seleccionado, al azar, una muestra de quince exámenes correspondientes al módulo de Cinética Aplicada de Reacciones Químicas y han sido evaluados cada uno de ellos por tres profesores, entre los que se in-

cluye como corrector el coordinador de la asignatura a la que pertenece este módulo. Se ha realizado primero un estudio de regresión lineal por mínimos cuadrados para correlacionar las calificaciones de uno de los profesores frente a las de un segundo profesor. Además, se han llevado a cabo ajustes de regresión de las calificaciones respecto a los valores medios y análisis de los valores residuales de las calificaciones propuestas por cada profesor frente a las notas medias.

4. EVALUACIÓN

La aplicación del estudio estadístico anteriormente mencionado permitirá concluir si la aplicación del sistema de evaluación por pares introduce cambios significativos en las calificaciones que obtienen los alumnos y si ofrece suficientes garantías. Este estudio ha

abarcado los cursos 2008/2009 y 2009/2010. La validez del sistema de corrección por pares se evaluará comparando las desviaciones en las calificaciones de tres parejas posibles de profesores, teniendo como referencia las notas medias calculadas a partir de las propuestas de tres profesores (evaluación triple).

5. RESULTADOS

En las figuras 1 a 3 se representan las calificaciones propuestas por uno de los profesores correctores frente a las dadas por otro de ellos para una población de quince alumnos correspondientes al curso académico 08/09. Estas gráficas ilustran la evaluación por pares realizadas por profesores diferentes (P1, P2 y P3). Además, se muestran los ajustes efectuados por regresión lineal y los parámetros resultantes de este ajuste (Tabla 1). Es evi-

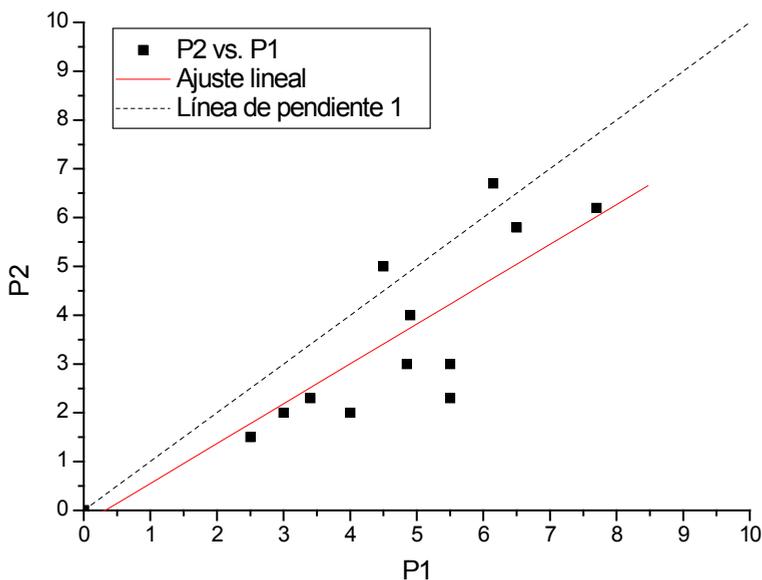


Figura 1. Representación de calificaciones propuestas por el profesor P2 frente al P1 para una población de $N = 15$ alumnos (escogidos aleatoriamente en lo que se refiere a notas y sexo. Curso 08/09).

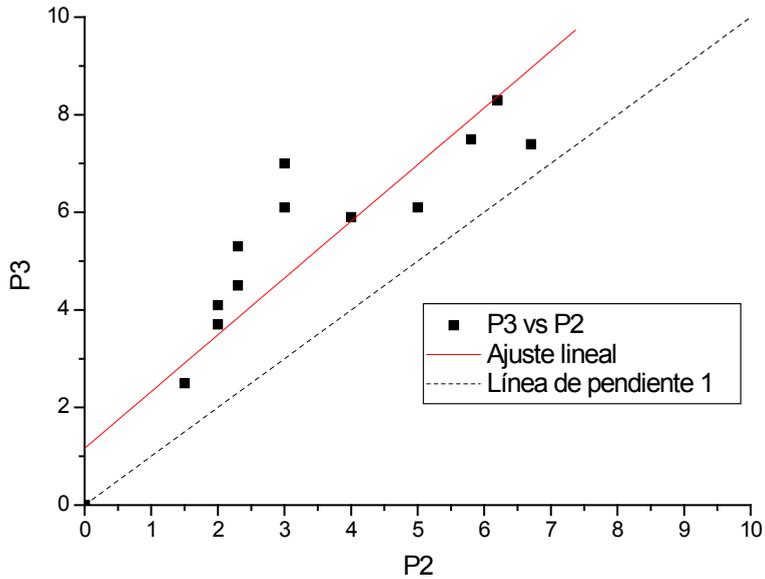


Figura 2. Representación de calificaciones propuestas por el profesor P3 frente al P2 para la misma población de $N = 15$ alumnos utilizada para confeccionar la Fig. 1. Curso 08/09.

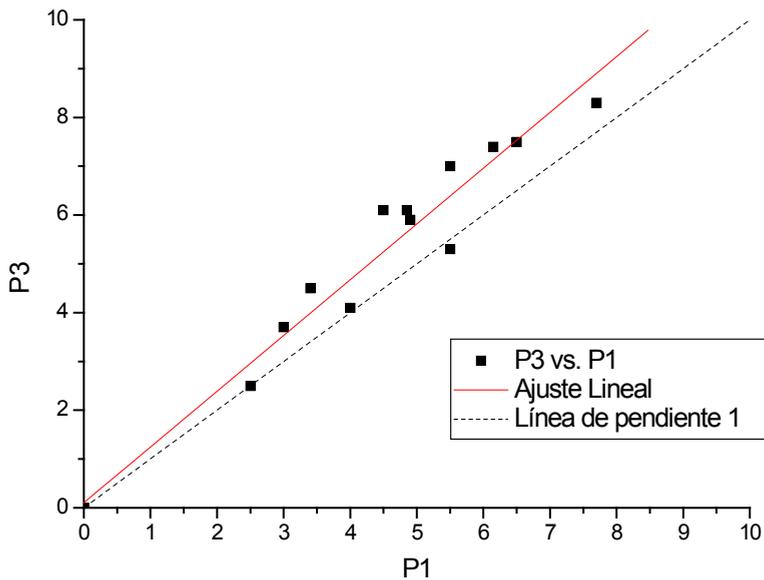


Figura 3. Representación de calificaciones propuestas por el profesor P3 frente al P1 para la misma población de $N = 15$ alumnos utilizada para confeccionar la Fig. 1. Curso 08/09.

Tabla 1. Parámetros resultantes del ajuste lineal realizado a las curvas representadas en las Figuras 1, 2 y 3. Curso 08/09

	PENDIENTE (M)	ERROR ESTÁNDAR DE M	R
P2 vs. P1	0,82	0,11	0,89
P3 vs. P1	1,16	0,14	0,91
P3 vs. P2	1,14	0,06	0,98

dente que cuanto más se acerque a 1 el valor de la pendiente más similar será la velocidad de incremento de las notas propuestas por los profesores implicados en cada representación. Al mismo tiempo, en discontinuo se representa la línea de pendiente 1, que describiría la situación en que ambos profesores hubieran propuesto calificaciones idénticas. Si los puntos representados se encuentran por debajo de esta línea significa que las calificaciones proporcionadas por el profesor representado en el eje de abscisas son superiores a las del profesor representado en el eje de ordenadas. Si la evaluación por pares hubiera sido realizada por los profesores P1 y P2, hubiera habido un 20% de discrepancia entre aptos y no aptos, tomando como referencia la nota estricta de 5,0. Las calificaciones propuestas se encuentran en su mayor parte por debajo de la línea de pendiente 1, lo cual significa que el profesor P2 ha proporcionado calificaciones inferiores a las del profesor P1 (fig 1). Las calificaciones de P2 son superiores a las de P1 en solo un 13,3% de los casos. Si P2 y P3 hubieran realizado la evaluación por pares, la disparidad respecto al porcentaje de aptos-no aptos hubiera alcanzado un 26,7%. Las calificaciones de P3 han sido superiores a las de P2 en todos los casos. Entre P1 y P3 la disparidad de aptos estrictos es de nuevo del 20%.

Entre P1 y P3 la disparidad de aptos estrictos es de nuevo del 20% y las calificaciones de P3 son superiores a las de P1 en seis de cada diez casos. No obstante, hay coinci-

dencia absoluta en un 26,7% de los casos y si se admite una variación de dos décimas, el porcentaje de notas equivalentes alcanzaría el 40% de los casos.

Un análisis de los resultados anteriores pone de manifiesto la existencia de discrepancias entre las calificaciones propuestas por los distintos profesores, siendo los profesores P1 y P3 los que más se aproximan en sus calificaciones. Finalmente indicar que las diferencia en los valores absolutos de las pendientes respecto a 1 son similares en los tres casos (Tabla 1).

En la figura 4 se compara la calificación propuesta por cada profesor frente a la nota media obtenida por cada alumno. Los resultados se han ajustado por regresión lineal.

De la figura 4 se puede concluir que es el profesor P1 quien propone calificaciones más próximas a la nota media obtenida por el alumno, como se deduce de la línea de referencia de pendiente 1. Es de destacar que las calificaciones propuestas por el profesor P3 en todos los casos se localizan por encima de los valores medios. Por el contrario las calificaciones propuestas por el profesor P2 siempre se localizan por debajo de la media. Las pendientes correspondientes a la regresión lineal muestran que, la velocidad de incremento de las calificaciones aportadas por el profesor P3 es superior a la de P1 y éste a su vez, a la de P2. (Tabla 2).

La dispersión según nota, respecto a la media se muestra en la figura 5. En ella se

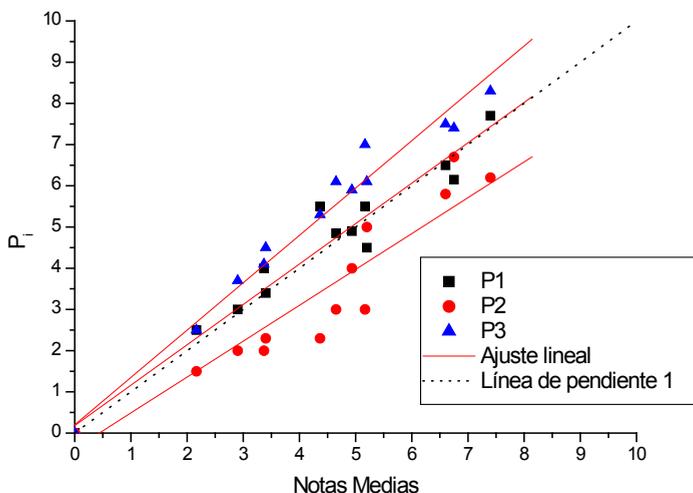


Figura 4. Calificaciones propuestas por los profesores P1, P2 y P3 frente a la nota media de cada alumno.

Tabla 2. Parámetros resultantes del ajuste lineal realizado a las curvas representadas en la Figura 4. Curso 08/09

	PENDIENTE (M)	ERROR ESTÁNDAR DE M	R
P1 vs. Nota Media	0,98	0,05	0,98
P2 vs. Nota Media	0,87	0,08	0,95
P3 vs. Nota Media	1,15	0,04	0,99

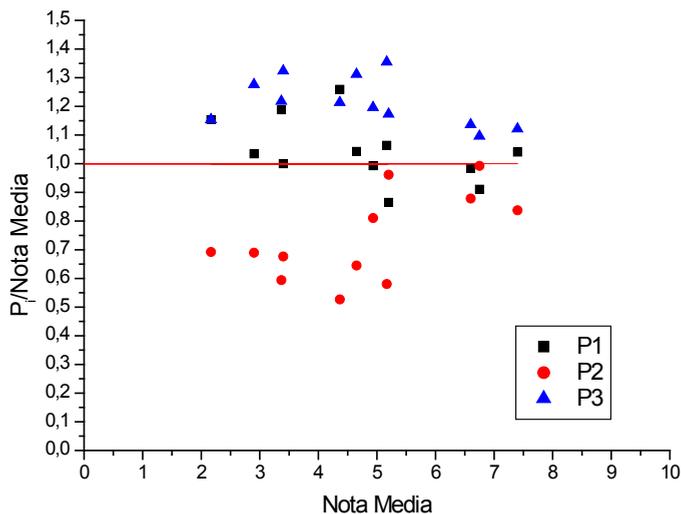


Figura 5. Cociente entre la nota propuesta por cada profesor para cada alumno y su nota media, frente a la nota media. La dispersión respecto a la línea ($y = 1$) muestra el grado de alejamiento de las notas propuestas por cada profesor respecto a la media.

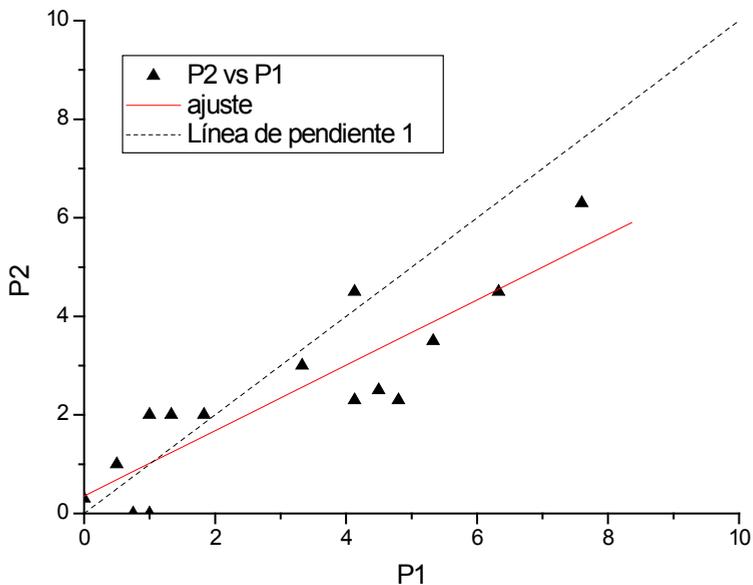


Figura 6. Representación de calificaciones propuestas por el profesor P2 frente al P1 para una población de $N = 15$ alumnos (escogidos aleatoriamente en lo que se refiere a notas y sexo). Curso 09/10

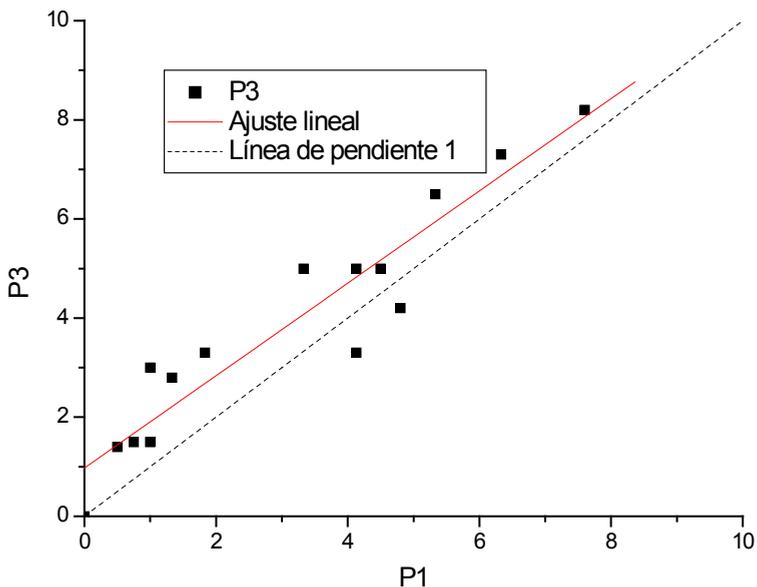


Figura 7. Representación de calificaciones propuestas por el profesor P3 frente al P1 para la misma población de $N = 15$ alumnos utilizada para confeccionar la Fig. 6. Curso 09/10.

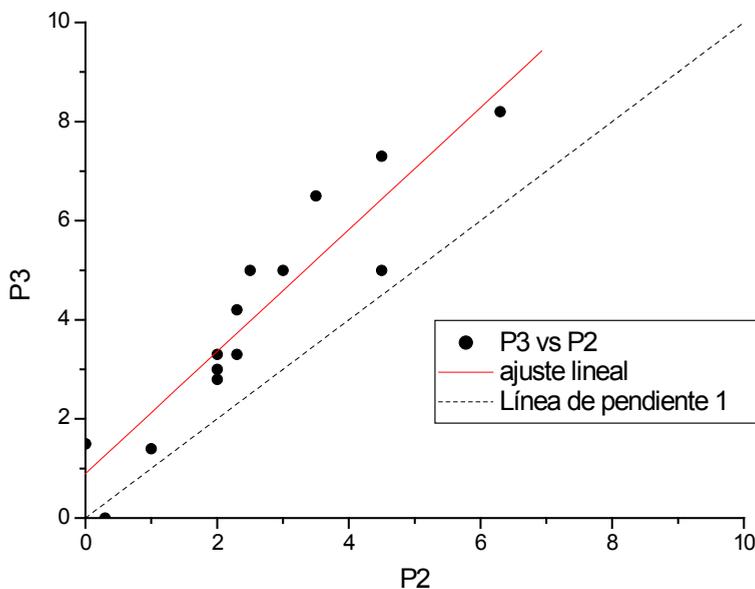


Figura 8. Representación de calificaciones propuestas por el profesor P3 frente al P2 para la misma población de $N = 15$ alumnos utilizada para confeccionar la Fig. 6 .Curso 09/10.

observa la gran disparidad de los resultados presentados por los profesores P2 y P3 respecto a las notas medias. Por el contrario, como era esperable del análisis anterior de los resultados, el profesor P1 se ajusta bastante bien al valor medio de las calificaciones.

Para confirmar la consistencia de los resultados obtenidos en el curso 08/09 se ha repetido el estudio y el análisis para los resultados obtenidos en el curso académico 09/10.

En las figuras 6, 7 y 8 se repite el análisis presentado en las figuras 1-3 para el curso 09/10. Hasta notas inferiores a 3,5 se observa bastante coincidencia en las calificaciones presentadas por los profesores P2 y P1, pero a partir de ella P2 puntúa claramente por debajo de P1 (fig 6). Si la evaluación por pares hubiera sido realizada por los profesores P1 y P2, la reducción de discrepancias entre aptos y no aptos habría descendido del 20 % en el curso 08/09 al 13,3% en el curso

09/10. El porcentaje de casos en que P2 propone calificaciones superiores a P1 asciende a un 40% de los casos.

Respecto al par formado por P2 y P3, decir que son los que aportan mayores discrepancias en calificaciones. Las discrepancias entre aptos y no aptos alcanzan el 33,3% de la muestra, valor superior al calculado en el curso anterior, 26,7%. P2 sólo evalúa por encima de las calificaciones de P3 en un 6,7%, situación muy similar a la del curso 08/09.

Si la evaluación por pares la hubieran realizado P1 y P3, las discrepancias entre aptos y no aptos habrían sido del 20%, valor idéntico al alcanzado en el curso anterior. No obstante, el porcentaje de casos en que las calificaciones de P3 son superiores a las de P1 asciende del 60% en el curso 08/09 al 80% en el 09/10. Por tanto, en general el profesor P3 se mantiene puntuando por encima de P1 (fig 7)

	Pendiente (m)	Error estándar de m	R
P2 vs. P1	0,66	0,10	0,884
P3 vs. P1	0,93	0,09	0,943
P3 vs. P2	1,23	0,13	0,935

Tabla 3. Parámetros resultantes del ajuste lineal realizado a las curvas representadas en las Figuras 6,7 y 8. Curso 09/10.

La correlación lineal entre los profesores P1 y P2 es muy deficiente y muestra una velocidad de incremento muy dispar. La mayor aproximación en estas velocidades de incremento en las calificaciones se observa entre los profesores P1 y P3 ya que considerando el error estándar de la pendiente el valor de pendiente 1 entra dentro del margen de error de la correlación. Esto significa que las valoraciones relativas de P1 y P3 son similares. Es decir, que si un profesor evaluó con un 5 y otro con un 6, luego calificaron con un 8 y un 9 a un alumno más brillante, incremento de +3 en ambos casos. La correlación de notas entre los profesores P3 y P2 es sólo aceptable, consideran-

do la proximidad a pendiente 1 y un valor del coeficiente de correlación muy similar al observado con la correlación P3 vs P1.

En la Figura 9 de nuevo se compara la calificación propuesta por cada profesor corrector frente a la nota media obtenida por el alumno y se muestran los ajustes por regresión lineal correspondientes.

En líneas generales la figura anterior revela el mismo resultado que en el curso académico 08/09, es decir, el profesor P1 es el que más se aproxima a la nota media mientras que el profesor P2 puntúa por debajo y el profesor P3 por encima de la media, independientemente del grupo de estudian-

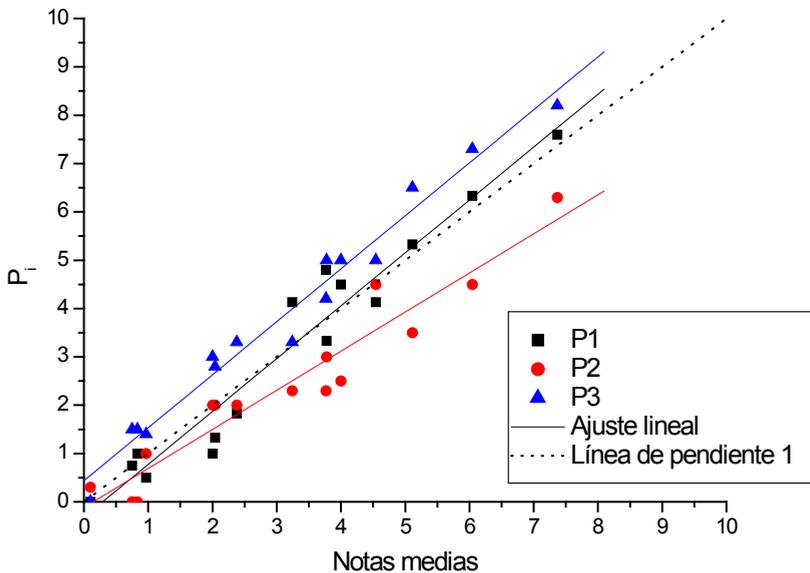


Figura 9. Calificaciones propuestas por los profesores P1, P2 y P3 frente a la nota media de cada alumno.

tes considerado. Los valores de las pendientes correspondientes a la regresión lineal para los profesores P1 y P3 ($1,094 \pm 0,07$ y $1,096 \pm 0,05$, respectivamente) no son significativamente diferentes indicando que ambos presentan una velocidad de incremento de calificaciones muy similar.

La dispersión según nota en relación a la nota media mostrada en la figura 10 revela una tendencia a la disminución de la disparidad en las calificaciones propuestas a medida que éstas son más altas. Si comparamos con el curso anterior, la dispersión es similar excepto para notas inferiores a 2,5. En ambos cursos se debe prestar atención a la disparidad respecto a la media para el entorno de la calificación de 4,0.

5. CONCLUSIONES

Los resultados indican que el resultado de la corrección por pares mejora la evaluación del alumno. A pesar de corresponder a exámenes de una asignatura experimental, no se ha detectado influencia en las calificaciones de los profesores debido a la evaluación continua en el laboratorio, ni por la evaluación de los informes de la labor experimental. Por tanto, las conclusiones alcanzadas se pueden extrapolar a una asignatura teórica, acompañada o no con resolución de problemas. Se demuestra que es necesario incluir un tercer corrector para calificar con la precisión exigida actualmente. A pesar de que las concesiones de becas y premios se pueden decidir por diferencias muy exiguas,

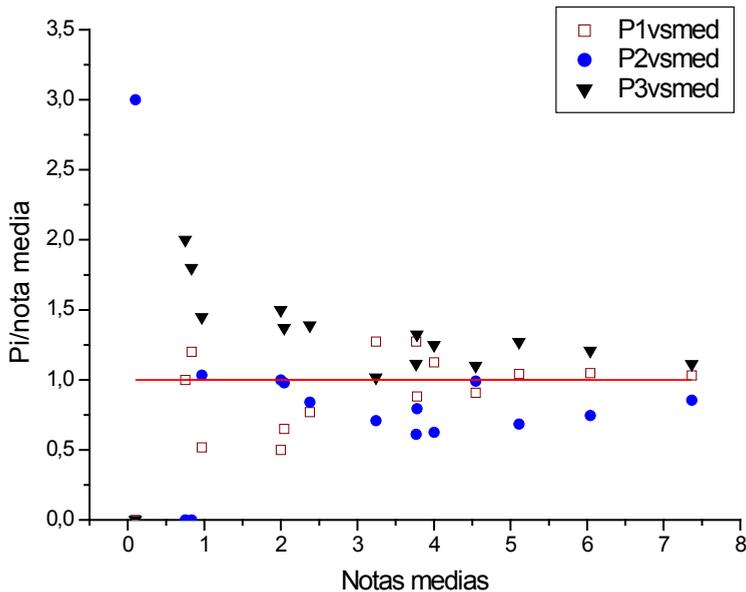


Figura 10. Cociente entre la nota propuesta por cada profesor para cada alumno y su nota media, frente a la nota media. La dispersión respecto a la línea ($y = 1$) muestra el grado de alejamiento de las notas propuestas por cada profesor respecto a la media C.curso 09/10.

no parece necesario proponer el sistema de corrección triple, ni siquiera el de pares para las calificaciones de cursos ordinarios a lo largo de la carrera universitaria, debido al coste asociado y a que la ponderación de calificaciones a lo largo de varios cursos amortigua los errores. Por el contrario, sí que se puede concluir claramente que la corrección de las pruebas de acceso a la Universidad debería implicar a tres profesores (triple corrección), considerando que la elección de los estudios a realizar por los alumnos, o de la universidad donde cursarlos, puede depender de la centésima de la nota final. Si por razones de coste no fuera posible, se debería extender la corrección por pares, aceptando la nota media. Sólo si las desviaciones entre las dos calificaciones excedieran de un

porcentaje a fijar, se recurriría de oficio a una tercera corrección.

6. BIBLIOGRAFÍA

Grau, R, y Cuxart, A. (2002). La calidad en el proceso de corrección de las pruebas de acceso a la Universidad: variabilidad y factores. *Revista de Investigación Educativa* , 20, 1, 209-223.

AGRADECIMIENTOS

Los autores quieren expresar su agradecimiento a sus respectivas familias, por su comprensión en relación con el tiempo dedicado al desarrollo de este trabajo.