

La influencia del orden de las preguntas en los exámenes de primer curso

David López¹, Jordi Cortés², Montse Fernández¹, Joan-Manuel Parcerisa¹, Ruben Tous¹, Jordi Tubella¹

1: Dep. d'Arquitectura de Computadors, 2: Dept. d'Estadística i Investigació Operativa
Universitat Politècnica de Catalunya, UPC – BarcelonaTech

Campus Nord. Jordi Girona 1-3 08034 Barcelona.

{david, montsef, jmanel, rtous, jordit}@ac.upc.edu, jordi.cortes-martinez@upc.edu

Resumen

El orden de las preguntas en un examen no debería tener influencia en sus resultados. Sin embargo, los autores tenemos la sensación de que los estudiantes de primero suelen ser secuenciales a la hora de resolver los exámenes. ¿Lo son realmente?, y si lo son ¿afecta esta manera de contestar los exámenes a los resultados finales? En este artículo analizamos estas cuestiones con un experimento realizado en la asignatura Estructura de Computadores, de primer curso del grado en Ingeniería Informática.

Abstract

The order of the questions on a test should have no influence on the final results. However, the authors had the feeling that students often solve the exam sequentially. Is this assumption true? If so, how does it affect the final results? In this paper we analyze the results of an experiment we designed to answer these questions. The experiment took place in the Computer Organization subject, a first-year course in the Computer Science Degree.

Palabras clave

Organización de los exámenes, costumbres de los estudiantes.

1. Introducción

1.1. Motivación

¿Influye el orden de las preguntas de un examen en las notas de los estudiantes? La respuesta debería ser que no. Sin embargo, la intuición nos dice que los alumnos de primero, que es donde se centra la mayor parte del fracaso en la Ingeniería Informática, suelen resolver el examen de manera secuencial: no planifi-

can el orden en que contestan sus preguntas, sino que empiezan por la primera pregunta, y así continúan hasta llegar a la última.

La mayoría de los profesores experimentados saben, cuando diseñan un enunciado, qué preguntas tendrán más respuestas correctas, bien por la dificultad inherente de los conceptos preguntados, por la dificultad de la pregunta en sí o porque tradicionalmente este es un tema que “se les da mal” a los alumnos (básicamente, porque muchos alumnos no se estudian algunos temas al percibir que no aparecen casi nunca en el examen, o cuando aparecen, valen muy pocos puntos).

Sin embargo, a pesar de que sabemos que en nuestro enunciado hay preguntas que contestarán bien muchos alumnos y otras que sabemos que pocos serán los que la contesten correctamente, muchos profesores no solemos tener en cuenta esta información a la hora de ordenar las preguntas. El orden de las preguntas suele hacerse para una mejor edición del examen (por ejemplo, para evitar que los estudiantes tengan que estar continuamente girando la hoja del enunciado al tener información relevante para una pregunta en ambas caras) o para simplificar el proceso de evaluación (en exámenes que se contestan en el mismo enunciado, que no haya en la misma hoja preguntas que deban corregir dos profesores). No solemos plantearnos el orden de las preguntas en función de otros parámetros porque asumimos que el orden no debería influir en la nota. Sin embargo, ¿podemos estar seguros de ello? Este punto es particularmente importante cuando hablamos de primer curso, donde el nivel de fracaso académico es muy alto.

En este trabajo presentamos un experimento llevado a cabo en una asignatura de primero de Ingeniería Informática, en la que quisimos estudiar si los alumnos tenían una tendencia a contestar los exámenes de manera secuencial y las implicaciones de hacerlo.

1.2. Trabajo relacionado

No hay mucha literatura respecto a la influencia del orden de las preguntas de un examen en un entorno de informática o de ingeniería en general. Básicamente hay trabajos dentro del ámbito de la psicología y con exámenes de tipo respuesta múltiple.

Pettjohn y Sacco [4] estudian la influencia de hacer las preguntas en orden secuencial respecto al temario, orden inverso y orden aleatorio. Respecto a los resultados en la nota no encuentran diferencias, aunque el orden aleatorio es percibido como más difícil y estresante por los alumnos, por lo que aconsejan el orden secuencial para estudiantes con problemas de estrés, de aprendizaje u otras circunstancias especiales. En Tal *et al* [6] se estudia la influencia del orden de las preguntas y del color de las hojas de examen en el resultado. Tampoco encuentran diferencia entre un orden u otro, pero sí demuestran que los estudiantes que contestan en papel blanco tienen mejores resultados que los que contestan en papeles de colores primarios no pastel, dándose los peores resultados en los exámenes contestados en papel azul. Togo [7] afirma que si el temario del curso evoluciona de manera que los últimos temas son los más abstractos y generales, los resultados en aprobados y notas de un examen secuencial superan a los de uno en orden aleatorio. Perlini *et al* [3] afirman que ordenar un examen de fácil a difícil, difícil a fácil o de manera aleatoria no influye en la nota final. Sin embargo, Conejo *et al* [1] demuestran que los profesores no somos buenos jueces de qué es difícil y qué no, por lo que nuestra clasificación fácil / difícil no es fiable.

Todos los estudios anteriores son sobre exámenes de respuesta múltiple, donde difícilmente los estudiantes tendrán problemas de tiempo para acabarlo, como pasa en algunos exámenes de problemas, que son mayoría en los cursos de ingeniería. Jaspers *et al* [2] realizan un estudio sobre problemas de matemáticas en alumnos de primaria, concluyendo que la ordenación más-corto-a-más-largo, más-largo-a-más-corto o aleatorio no influye en la nota, pero sí influye en la percepción de dificultad del examen. Symanzik *et al* [5] realizan un estudio sobre exámenes *on-line* de estadística con preguntas de respuesta múltiple (RM) y preguntas de respuesta escrita corta (REC) concluyendo que hay alumnos secuenciales, alumnos que contestan primero las RM, otros las RC y otros que son aparentemente aleatorios, sin concluir nada respecto a si ello influye en la nota final.

Parece pues que en exámenes de respuesta múltiple el orden no influye. Sin embargo los problemas de los exámenes de ingeniería son diferentes: de desarrollo y más complejos en su respuesta. Además, en ciertas ocasiones no todos los alumnos acaban los exámenes dentro del tiempo asignado, lo que no pasa en los casos estudiados en la bibliografía encontrada. Estamos pues en terreno desconocido.

2. Hipótesis

La pregunta que nos planteamos es: *¿Los estudiantes de primero contestan las preguntas del examen secuencialmente?*

Caso de ser cierto, el orden de las preguntas podría tener una influencia en los resultados finales, aunque hay otros factores que pueden influir:

- La decisión de cómo ordenar. Se nos ocurren dos maneras de ordenar para influir en las notas finales: 1) Ordenar las preguntas según el nivel de dificultad y 2) Ordenar las preguntas según la nota de cada pregunta. Ambas ordenaciones están muy relacionadas, dado que se suele dar más puntuación a las preguntas más difíciles. El problema de ordenar según el nivel de dificultad está en que, como ya se dijo, los profesores no somos buenos juzgando qué es difícil y qué no.
- La cantidad de preguntas contestadas. Si los estudiantes tienen suficiente tiempo para contestar todas las preguntas del examen, entonces la ordenación no debería influir. Sin embargo, un examen donde los alumnos vayan justos de tiempo y no puedan contestar todas las preguntas, el orden sí puede ser un factor clave.
- La puntuación de cada pregunta: si hay gran dispersión en la puntuación de las preguntas, entonces contestar pocas preguntas de muchos puntos podría ser más eficiente para el estudiante que contestar muchas de pocos puntos.

El objetivo primordial de este trabajo es ver si los estudiantes de primero contestan secuencialmente a las preguntas y, en función de ello, presentar una discusión sobre las implicaciones que puede tener.

3. Metodología

3.1. Entorno académico

La asignatura sobre la que se realizó el experimento es “Estructura de Computadores”, ubicada en el segundo semestre del primer curso de Ingeniería Informática de la Facultat d’Informàtica de Barcelona (UPC – BarcelonaTech).

En nuestro plan de estudios, hay 8 asignaturas de 7,5 créditos ECTS en primero. Cuando un estudiante se matricula por primera vez, debe matricularse de cuatro de estas asignaturas (Fundamentos Matemáticos, Física, Programación 1 e Introducción a los Computadores), que consideramos del primer semestre. Las cuatro asignaturas del segundo semestre son: Matemáticas 1, Matemáticas 2, Programación 2 y Estructura de Computadores.

No hay ningún tipo de prerrequisito entre las asignaturas de primero, más allá de la obligación de matricularse de las 4 de primer semestre en la primera

matricula. Las ocho asignaturas se imparten tanto en el semestre de primavera como en el de otoño.

La nota de la asignatura Estructura de Computadores (de ahora en adelante EC) se obtiene a partir del examen final (un 60% de la nota), la nota de laboratorio (un 20%) y la nota del examen parcial (que cuenta un 20%, pero este porcentaje es el máximo entre la nota del parcial y la obtenida en el examen final).

El experimento se llevó a cabo en el semestre de otoño del curso 2012-2013. En la asignatura EC teníamos 135 estudiantes matriculados (119 hombres, 16 mujeres), 67 de los cuales eran repetidores. Se realizaron sendas experiencias en el parcial y en el final. En ambas ocasiones se prepararon dos versiones del enunciado con las mismas preguntas, pero con diferente ordenación. Ambos exámenes podrían considerarse como “tradicionales”, ya que no incluían ninguna pregunta novedosa, sino más bien eran similares a los exámenes de semestres anteriores. Los alumnos disponían tanto del enunciado como de la solución de los exámenes de cursos anteriores.

3.2. Elaboración de los enunciados

Nuestra política a la hora de preparar enunciados es la siguiente: se distribuyen los temas por profesores, y cada profesor genera un número de ejercicios que cubran el temario que se le ha sido asignado (aunque puede cubrir más temario). Los profesores resolvemos todos los enunciados generados y hacemos una reunión para pulir estos enunciados (si los datos son suficientes, si hay algún error, si hay partes del enunciado que pueden mover a confusión a los alumnos, etcétera). Los ejercicios que son aceptados se modifican de acuerdo a los comentarios recibidos. Todos los profesores solucionamos el examen provisional escribiendo la solución completa y cronometrando el tiempo que requiere cada pregunta. En una segunda reunión se decide qué problemas formarán el enunciado definitivo, en función del tiempo asignado al examen y el coste en tiempo de las preguntas. Para calcular el tiempo de resolución de una pregunta por parte de un estudiante, multiplicamos por tres el tiempo medio que nos ha costado a los profesores. Los puntos de cada pregunta se distribuyen proporcionalmente al tiempo que cuesta solucionarla.

Para el parcial (que tiene una duración de 2 horas) se realizó un examen de 8 preguntas, con puntuaciones entre 0,5 y 2,25 puntos. El final (de 3 horas de duración) constó de 10 preguntas, con puntuaciones entre 0,5 y 1,75 puntos.

En el examen parcial se procedió a estimar la dificultad de cada pregunta del enunciado final, basándonos en la experiencia docente de los profesores implicados. Se realizaron dos versiones del examen, con las mismas preguntas, pero ordenadas de manera diferente: en la primera versión las preguntas estaban ordenadas de la más difícil a la más fácil (denomina-

remos a este el enunciado DaF –Difícil a Fácil), mientras que en el segundo enunciado, la ordenación era de fácil a difícil (enunciado FaD).

En el examen final se realizaron dos versiones, pero en este caso la ordenación fue de mayor a menor puntuación (enunciado GaP – Grande a Pequeño) y de menor a mayor puntuación (enunciado PaG)

3.3. Distribución de la población

Para realizar los exámenes, nuestro centro nos proveyó de tres aulas: dos aulas grandes donde deberían examinarse 57 estudiantes en cada una, y una aula pequeña para los 21 estudiantes restantes (denominaremos aulas G1 y G2 a las grandes y P a la pequeña).

Uno de los estudiantes matriculados el semestre del experimento tiene una discapacidad que nos obligaba a que se examinara en el aula P, que está adaptada a este tipo de necesidades especiales. La distribución para el parcial y el final fue diferente, distribuyendo los alumnos alfabéticamente a partir de una posición al azar, con la restricción de que el estudiante con necesidades especiales estuviera en el aula P. A pesar de que la distribución no es aleatoria, creemos que no debería generar sesgos en los resultados obtenidos.

Al tener un aula pequeña y dos grandes y dos enunciados diferentes, decidimos asignar un enunciado en cada aula grande. Faltaba ver si asignábamos un enunciado al aula P al azar, o bien había algún criterio para que P tuviera el mismo enunciado que G1 o G2. Para decidir el emparejamiento del aula P, estudiamos si en el caso de emparejar el aula P con G1 o con G2, una de las combinaciones tenía una mejor distribución de los alumnos por grupos para eliminar el factor grupo en la distribución. Hay 4 grupos (del 10 al 40) y, en nuestra experiencia, los estudiantes del grupo 10 suelen obtener mejores resultados que el resto. Esto es debido a que en nuestro centro los estudiantes se matriculan por orden de nota, y las notas más altas suelen matricularse en el grupo 10. Igualmente, el grupo 40 es el único de la tarde, y suele tener alumnos cuyo nivel de fracaso es tradicionalmente mayor. Las dos opciones pueden verse en el Cuadro 1 para el parcial (Op1P, Op2P) y el final (Op1F, Op2F).

		10	20	30	40
Op1P	G1+P	25,6%	26,6%	24,4%	24,4%
	G2	28,1%	26,2%	21,1%	24,6%
Op2P	G1	29,8%	26,3%	26,3%	17,6%
	G2+P	24,4%	25,6%	20,1%	29,5%
Op1F	G1+P	28,2%	23,1%	24,4%	24,4%
	G2	24,6%	29,8%	21,1%	24,6%
Op2F	G1	26,3%	21,1%	24,6%	28,1%
	G2+P	26,9%	29,5%	21,8%	21,8%

Cuadro 1: Distribución de alumnos de cada grupo de teoría en cada combinación de aulas.

El Cuadro 1 marca, una vez distribuidos los alumnos en las 3 aulas, el porcentaje de estudiantes de cada uno de los 4 grupos de teoría que habría para cada enunciado, suponiendo que se reparte el mismo enunciado en el aula P que en G1 o en G2. Observamos que la opción 1 ofrecía una distribución más equilibrada en ambos casos, y fue la elegida.

Finalmente, se decidió al azar qué versión del examen se distribuiría a cada grupo. En el parcial, el enunciado DaF se repartió en las aulas G1 y P (n=78), y el FaD en el aula G2 (n=57). En el final, el enunciado GaP en las aulas G1 y P (n=78) y el enunciado PaG en el aula G2 (n=57).

Los cálculos estadísticos presentados fueron extraídos con el paquete estadístico R v.2.15.1 a partir de los resultados de las notas de los exámenes.

4. Resultados

4.1. Resultados generales

El parcial no tuvo muy buenas notas, pero los resultados están dentro de lo que, a criterio de los profesores, es habitual en la asignatura. El final fue mucho mejor, aumentando en más de 30 puntos porcentuales el número de aprobados, y en 1,4 puntos la nota media. El Cuadro 2 presenta los resultados.

		Presentados (%)	Aprobados (%)	Nota media (STD)
Parc.	Total	119/135 (88)	32 (27)	3,8 (1,65)
	FaD	51/57 (89)	13 (25)	3,8 (1,48)
	DaF	68/78 (87)	19 (28)	3,9 (1,81)
Final	Total	113/135 (84)	69 (61)	5,2 (1,8)
	GaP	70/78 (90)	41 (59)	5,1 (1,87)
	PaG	43/57 (75)	28 (65)	5,4 (1,73)

Cuadro 2. Resultados de los exámenes (parcial y final), totales y subdivididos entre las dos versiones de cada uno.

No se observan diferencias significativas en las notas obtenidas en ambas versiones de cada examen (analizaremos las notas con más profundidad en la Sección 4.4.). Queremos destacar que observamos una gran diferencia entre la cantidad de estudiantes que tuvieron tiempo de contestar todas las preguntas en el parcial y en el final. En el parcial (de 2 horas) observamos cómo la práctica totalidad de los estudiantes estuvieron contestando preguntas hasta el final del tiempo estipulado, mientras que en el final (de 3 horas) la mayoría entregó el examen antes del tiempo estipulado. En la sección de discusión comentaremos este comportamiento y la influencia que puede haber tenido en los resultados finales.

4.2. Apreciación de la dificultad de los problemas por parte del profesor

Como se ha dicho, en el examen parcial las dos versiones del examen se diferenciaban sólo en la ordenación de las preguntas, de fácil a difícil y viceversa. Aunque es complicado evaluar con exactitud si nuestro criterio a la hora de decidir si un ejercicio es o no fácil, hay datos que vale la pena remarcar:

- La segunda pregunta más fácil (según nuestro criterio) ha sido la que ha tenido un número mayor de no presentados, tanto en la ordenación FaD (segunda pregunta de 8) como en la DaF (séptima de 8).
- De las 3 preguntas donde los alumnos han sacado mejor nota media, ninguna correspondía a las dos consideradas más fáciles, independientemente de la versión de examen. Este cálculo lo hemos realizado sobre las notas ponderadas sobre 10 puntos, no sobre la media absoluta, ya que las preguntas tenían diferentes puntuaciones.
- La segunda pregunta más difícil ha sido la que ha sacado la segunda mejor nota media ponderada en ambas versiones del examen.

De estos datos concluimos que nuestra apreciación de qué es fácil o difícil no se corresponde con la realidad. En parte fue por ello que decidimos cambiar el sistema de ordenación en el examen final.

4.3. ¿Los alumnos responden de manera secuencial?

Hay una serie de datos que indican que esta hipótesis se cumple. Analizaremos por separado los resultados del parcial y del final.

Primero estudiaremos la relación de ejercicios presentados en el parcial. La Figura 1 muestra la diferencia en porcentaje de presentados para cada problema, separados para las dos versiones. Los problemas están ordenados de difícil a fácil (ordenación del grupo DaF). Por ejemplo, el problema MULT¹ lo presentaron el 88.2% (60/68) del grupo DaF, (que lo tenían como el primer problema del examen), y el 76.5% (39/51) del grupo FaD (que lo tenían el último), siendo la diferencia de 11.8% y el intervalo de Confianza IC al 95% entre -3,9 y 27,4 (que es lo que se muestra en el *forest plot* de la Figura 1). Se observa que existe una tendencia en cada grupo a contestar los problemas que tenían en primera posición, aunque únicamente en el problema supuestamente más sencillo (MAT) hubo diferencias estadísticamente significativas entre ambos grupos.

¹ Para no confundir, los problemas no han sido numerados pues la numeración implica un orden, sino que tienen un nombre. Así por ejemplo, el problema MULT se refiere a un problema relacionado con multiplicadores hardware. Los nombres son: MULT, REND, CF, MEM, SUBR, COND, PRG y MAT.

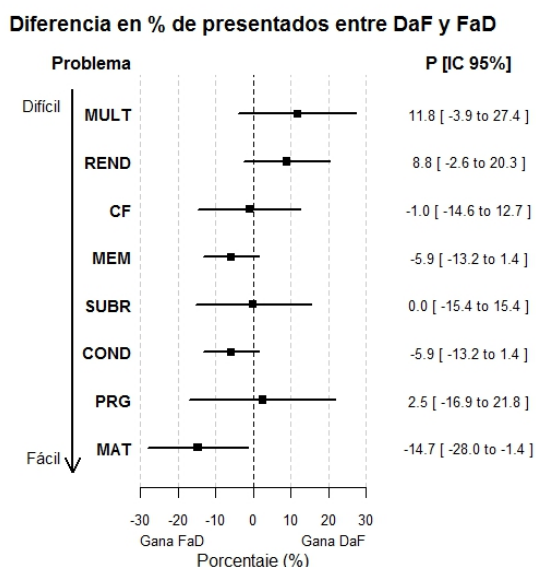


Figura 1: Forest plot de la diferencia entre presentados en cada pregunta entre las dos versiones del examen parcial.

Para un mejor análisis, teniendo en cuenta que contestar unas preguntas u otras no es un factor independiente dentro de cada alumno, debemos estudiar las preguntas en blanco a través de un indicador de secuencialidad. Así, cada ejercicio recibió una puntuación según su posición en el examen (Cuadro 3).

MULT	-7
REND	-5
CF	-3
MEM	-1
SUBR	1
COND	3
PRG	5
MAT	7

Cuadro 3: Score de secuencialidad para los problemas del parcial.

Para cada alumno, su score será la suma de los valores correspondiente a los ejercicios presentados. Un estudiante que conteste a todas las preguntas tendrá un score de 0, independientemente de su nota. Un estudiante que conteste más preguntas fáciles (parte inferior) tendrá a tener valores positivos, mientras que uno que conteste más preguntas difíciles tenderá a valores negativos. Si se cumple la hipótesis de que los estudiantes responden de manera secuencial, entonces esta puntuación debería ser más negativa en el grupo DaF que en el FaD.

La descriptiva de este indicador para ambos grupos puede verse en el Cuadro 4.

	DaF	FaD
Media	-2,22	0,04
Desviación Estándar	5,36	4,38
Mínimo	-15	-13
Q1	-5	-1
Mediana	0	0
Q3	0	2
Max	12	12

Cuadro 4: descriptiva del indicador score.

La Figura 2 muestra la distribución de este indicador para ambos grupos. La cola inferior alargada del grupo DaF sugiere que muchos estudiantes de este grupo se dejaron de contestar problemas fáciles que les hubiesen dado valoraciones positivas en este indicador. Mediante la prueba de la t-Student se estima que la diferencia de este indicador entre ambos grupos es de -2,26, IC95%: [-4,03 a -0,49], lo que es estadísticamente significativo ($p=0,01$). Concluimos pues que observando el número de ejercicios presentados, los alumnos contestan, efectivamente, de manera secuencial y por tanto se cumple la hipótesis inicial.

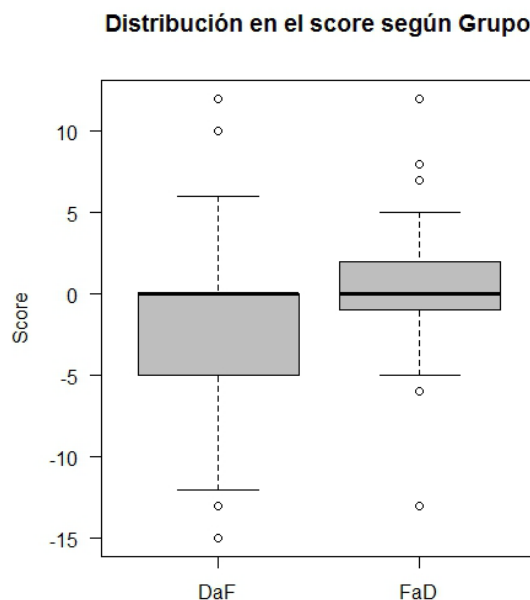


Figura 2: distribución del indicador de score.

Si realizamos este mismo análisis para el examen final, descubrimos que no hay una tendencia a dejar en blanco las últimas preguntas del examen. No hubo diferencias significativas entre ambos grupos para

ningún problema. Elaboramos un indicador de secuencialidad similar al usado para el parcial, adaptado a 10 problemas (se añadieron los valores -9 y 9) y nos encontramos con distribuciones extraordinariamente parecidas. Al aplicar la t-Student estimamos que la diferencia media de este indicador entre ambos grupos es de 0,58, con IC95%: [-0,96 a 2,11], estadísticamente no significativa ($p=0,46$). Al no haber diferencias significativas hemos decidido no incluir las gráficas en el artículo, pues aportan poco. Esta diferencia entre el parcial y el final será analizada en la sección de discusión.

Para reforzar la idea de que los estudiantes contestaban de manera secuencial, analizamos las notas del examen parcial agrupando las notas de las 4 primeras y las 4 últimas preguntas (recordamos que las 4 primeras de una versión son las 4 últimas de la otra). Estos datos se pueden observar en el Cuadro 5, en el que se presenta la nota media sobre 10 y la desviación estándar de las 4 preguntas difíciles y las 4 fáciles del examen para las dos variantes.

	Media de las cuatro preguntas difíciles	Media de las cuatro preguntas fáciles
DaF	3,92 (1,98) - primeras	3,75 (0,61) - últimas
FaD	3,54 (1,67) - últimas	4,2 (2,09) - primeras

Cuadro 5: Media de las 4 preguntas fáciles y las 4 preguntas difíciles (entre paréntesis, la desviación estándar).

Usamos los datos del Cuadro 5 podemos realizar ciertas comparaciones:

- Comparación horizontal: dentro de cada versión, la nota media de las 4 preguntas que se contestan en primer lugar es superior a la media de las que se contestan en segundo lugar (3,92 en las 4 primeras contra 3,75 en DaF, 4,20 en las 4 primeras contra 3,54 en FaD).
- Comparación vertical: la media de las 4 preguntas difíciles es superior en la versión donde eran las 4 primeras que donde eran las 4 últimas (3,92 contra 3,54). Igualmente, la media de las 4 fáciles es superior en el examen donde eran las 4 primeras (4,20 contra 3,75).
- Comparación cruzada: la media de las 4 primeras preguntas es superior cuando son las fáciles que cuando son las difíciles (4,2 contra 3,92). Igualmente, la media de las 4 últimas preguntas es superior cuando son las fáciles que cuando son las difíciles (3,75 contra 3,54).

Ninguna de estas comparaciones ofrece una diferencia estadística significativa, aunque en algún caso están bastante cerca. Sin embargo, refuerzan la hipótesis de que los alumnos contestan secuencialmente.

4.4. ¿Afecta este comportamiento a las notas del examen?

Hemos analizado las notas del parcial para ambas versiones del examen. La media en DaF es 3,85 y en FaD 3,81. La diferencia estimada mediante una prueba t-Student es 0,05 con un IC95%: [-0,56 a 0,66] ($p=0,88$). Por tanto, se deduce que no hay diferencia entre las notas obtenidas en ambos grupos. La Figura 3 muestra gráficamente la distribución de las notas. Las posibles razones para que no haya diferencias se discutirán en la Sección 5.

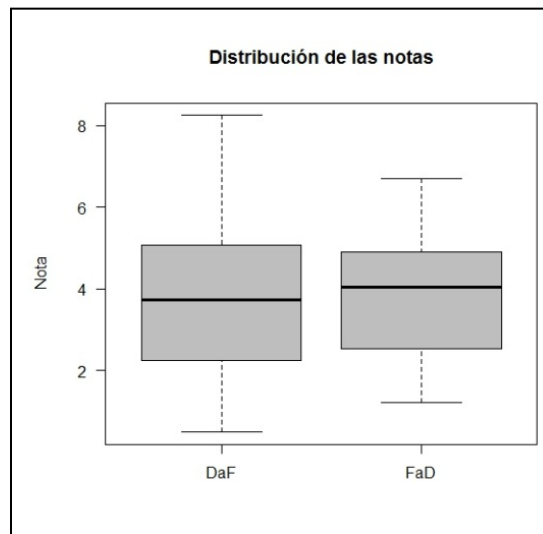


Figura 3: Distribución de las notas del parcial.

En punto 4.3 se demostraba que existía una secuencialidad en el parcial, pero no así en el final. Por tanto no presentamos el análisis de las notas del examen final, aunque lo realizamos y descubrimos que tampoco había diferencias significativas.

5. Discusión

En los resultados presentados en la sección anterior hemos visto una gran diferencia entre los resultados obtenidos en el parcial y los obtenidos en el final. Sin embargo, como también se adelantaba en la Sección 4.1, en el examen parcial la mayoría de los estudiantes no tuvieron tiempo de contestar a todas las preguntas, a diferencia del final, en el que la mayoría entregó antes de finalizar el tiempo asignado. ¿A qué puede ser debida esta diferencia? Podríamos pensar que es simplemente cuestión de la duración del examen, ya que el final dura una hora más que el parcial. Sin embargo, este parámetro está contemplado en el diseño del enunciado, ya que aproximamos el tiempo necesario por el alumno a base de multiplicar por 3 el tiempo medio usado por los profesores para resolver el examen, es decir, menos de 40 minutos para el parcial de 2 horas, y menos de 1 hora para el final de

3 horas. Obviamente podemos equivocarnos en nuestros cálculos, pero nuestra experiencia dice que suele ser habitual que los estudiantes vayan más justos de tiempo en el parcial que en el final.

Una teoría que lo explica (pero difícil de demostrar más allá de la intuición de los profesores) es que los estudiantes van mejor preparados a los finales que a los parciales. El parcial tiene lugar un día con clases, donde tienen que compatibilizar el estudio del mismo con la asistencia a clase, las prácticas y el estudio de otras asignaturas, mientras que el final se realiza en un periodo sin clases ni entregas de prácticas, dedicado únicamente al estudio. Además, el factor madurez de los conocimientos también es importante, ya que después del parcial han dispuesto de un mes y medio para madurar sus conocimientos y obtener una visión global de la asignatura. Tampoco es desdeñable el hecho que sacar mala nota en un parcial (por poca preparación) ponga al estudiante en la situación de espabilar o suspender. Hemos observado que algunos estudiantes se centran en estudiar menos asignaturas, pero las estudian más a fondo.

Sea cual sea la razón, la secuencialidad en la resolución afecta cuando no se tiene tiempo de contestar la totalidad del examen. Si, como en el caso del final, la mayoría de los estudiantes han tenido tiempo de contestar todas las preguntas, seguir un orden u otro no afecta al resultado final. Por tanto, una primera conclusión es que responder secuencialmente afectará únicamente si no se pueden contestar todas las preguntas.

Centrándonos pues en nuestro parcial, hemos concluido que, efectivamente, los estudiantes contestan secuencialmente. Sin embargo, esto no afectó a los resultados finales por lo que a notas se refiere. Al haber hecho una ordenación de fácil a difícil (y viceversa) ¿cómo es posible que si los alumnos contestan secuencialmente no tengan un beneficio si empiezan por las preguntas más sencillas? En el caso de que el estudiante no tenga tiempo de contestar todas las preguntas de un examen, si las ordenamos de fácil a difícil, ¿no les ayudamos?

La respuesta es que hay dos factores que impiden que los alumnos se beneficien de la ordenación fácil a difícil. La primera es que los mismos profesores no somos infalibles a la hora de calificar un problema en fácil o difícil (como hemos visto en el punto 4.2.). La segunda es que los problemas difíciles suelen valer más puntos, por lo que penalizan mucho a los alumnos que no los contestan.

En los exámenes de problemas no es habitual que todas las preguntas valgan los mismos puntos. La puntuación de cada pregunta puede variar dependiendo de un número de factores: la dificultad de la pregunta, el tiempo estimado de respuesta, la importancia del concepto que se pregunta. Hay tantos métodos para decidir el valor de la pregunta como profesores.

Sin embargo, como norma general las preguntas más complejas suelen valer más puntos.

Si, para nuestro parcial, suponemos un caso extremo de un estudiante totalmente secuencial, que empieza por el primer ejercicio y no pasa al segundo hasta que se da por satisfecho con la respuesta al primero, este alumno estaría en cierta ventaja en el examen DaF ya que, como muestra el Cuadro 6, en nuestro parcial las primeras preguntas en esta versión valen más que en la versión FaD.

	1	2	3	4	5	6	7	8
DaF	1	2,75	4	6	8,25	9	9,5	10
FaD	0,5	1	1,75	4	6	7,25	9	10

Cuadro 6: Nota acumulada en cada una de las 8 preguntas de ambas versiones del examen.

Así por ejemplo, un alumno con el examen DaF que ha podido contestar las 5 primeras preguntas podría obtener una nota máxima de 8,25, mientras que si consideramos otro alumno que en el examen FaD ha conseguido contestar una pregunta más (hasta la 6), este último ha contestado más preguntas, pero la nota máxima a la que puede aspirar es menor, un 7,25. Por tanto, el hecho de contestar menos preguntas, pero de más puntos (versión DaF) contrarresta el efecto de tener primero las preguntas fáciles (FaD) donde se contestan más preguntas, pero de menor puntuación.

Nos planteamos probar en el final una ordenación de más a menos puntos (y viceversa), lo que sería independiente de la dificultad de la pregunta, ya que los puntos asignados a cada pregunta dependen del tiempo calculado para su resolución y no de la dificultad de la misma. Pero al ser un examen donde la mayoría de estudiantes tuvieron tiempo de contestar todo el examen, no pudimos extraer conclusiones. Queremos, como trabajo futuro, probar en el próximo parcial la ordenación de más a menos nota.

Una pregunta interesante es: ¿Por qué son secuenciales? En este caso realizamos un estudio cuantitativo, con una población de 21 estudiantes a los que les fuimos preguntando en entrevistas informales sobre sus costumbres a la hora de resolver un examen². Sólo dos afirmaron no empezar por la primera pregunta, pero no por hacer una planificación, sino por buscar una pregunta que supieran hacer “muy bien”; en cuanto encontraban una se ponían con ella, sin mirar el resto del enunciado. El resto de los estudiantes contestaban secuencialmente porque nunca (o casi nunca) se habían encontrado antes de entrar en la Universidad ante un examen que no tuvieran tiempo

² Básicamente charlas de pasillo en los descansos de clase. Quizá no el mejor método de investigación, pero sí uno en el que se reciben respuestas espontáneas y sinceras.

de contestar al completo. Incluso los mejores estudiantes (en la apreciación del profesor que les daba clase) afirmaron estar acostumbrados a no acabar con una pregunta y pasar a la siguiente hasta que no estaban satisfechos con la respuesta de la primera. Nadie afirmó haber recibido nunca el consejo de dedicar los primeros minutos a leer el enunciado al completo y planificar su tiempo. Como trabajo futuro nos planteamos un estudio cualitativo más completo respecto a las costumbres en los exámenes de los estudiantes de primero.

Para finalizar la discusión, nos planteamos la siguiente pregunta: si sabemos que los alumnos son secuenciales ¿qué queremos hacer con esta información? Primero, deberíamos estudiar si hay ordenaciones que influyen en la nota. Hemos visto que en exámenes donde no tienen tiempo de responder a todas las preguntas, la ordenación fácil a difícil no influye en las notas, pero quizá una ordenación de más a menos puntos si tuviera una influencia, o quizá otra ordenación que ahora no somos capaces de ver. En cualquier caso, una conclusión de este trabajo es que deberíamos explicar a nuestros alumnos que el primer paso ante un examen es dedicar un tiempo a planificar en qué orden contestar las preguntas del examen. Se puede argumentar que esto es algo que todos los estudiantes acaban aprendiendo, pero muchos de nuestros estudiantes no se han encontrado nunca ante exámenes donde han ido justos de tiempo. Precisamente nuestros mejores estudiantes nunca habían suspendido nada antes de entrar en la Universidad, siendo los que están más acostumbrados a contestar secuencialmente y pasar a la pregunta dos cuando están satisfechos de la respuesta de la pregunta uno. Estos estudiantes se pueden encontrar inseguros al entrar en la Universidad, pues están en un entorno más exigente. Psicológicamente hablando, es precisamente cuando te encuentras más inseguro, cuando más te aferras a los métodos que te llevaron al éxito en el pasado. Por ello es difícil que se les ocurra espontáneamente cambiar sus creencias respecto a cómo contestar un examen y, por tanto, necesitan nuestra ayuda.

6. Conclusiones

A la hora de elaborar un examen, los profesores no solemos tener en cuenta cómo puede afectar el orden de las preguntas en los resultados finales. Sin embargo, en exámenes donde los estudiantes no tienen tiempo de contestar a todas las preguntas, el orden con el que respondan marcará el número de preguntas respondidas y su puntuación en el examen.

En este trabajo hemos realizado un experimento en una asignatura de primero, donde hemos demostrado

que los estudiantes tienden a contestar secuencialmente, aunque en nuestro estudio no hemos encontrado una ordenación de las preguntas que ofrezca un cambio en los resultados del examen, respecto al número de aprobados o a la nota media.

Saber cómo actúan los estudiantes de primero ante un examen podría ser interesante para ayudar a reducir el fracaso escolar en la Ingeniería Informática, dado que se centra principalmente en primer curso. Una actuación inmediata sería explicar a los estudiantes las ventajas de planificar sus exámenes, pero se requiere más investigación en buscar ordenaciones en las preguntas de los exámenes que sí influyan en el rendimiento.

Este trabajo se ha llevado a cabo con el apoyo del proyecto MiPLE del Ministerio de Ciencia e Innovación español, código TIN2010-21695-C02-02.8 y el proyecto TRAILER de la Comisión Europea (<http://grial.usal.es/agora/trailerproject>).

Referencias

- [1] R. Conejo, E. Guzmán de los Riscos y J.L. Pérez de la Cruz. *Un estudio sobre la dificultad de los ítems en los tests de Informática*. Jenui 2008, pp 242-248, Granada, Julio de 2008.
- [2] Kathryn E. Jaspers, Christopher H. Skinner, Robert L. Williams y Lee B. Saecker. *Effects of problem order on accuracy, preference, and choice of multiplication assignments*. The Behavior Analyst Today 8(3): 347:359. 2007
- [3] Arthur H. Perlini, David L. Lind y Bruno D. Zumbo. *Context effects on examinations: The effects of time, item order and item difficulty*. Canadian Psychology 39(4): 299-307. 1998.
- [4] Terry F. Pettjohn, y Matthew F. Sacco. *Multiple-choice exam question order influences on student performance, completion time, and perceptions*. Journal of Instructional Psychology 34(3). 2007
- [5] Jürgen Symanzik, Robert Erbacher, Palyne Gaenir, Natasha Vakusanivic, Alex Cheng Ki Wun, *On the effect of the ordering of questions in Exams – A visual analysis*. Proceedings of the American Statistical Association. 2004
- [6] Lianit R. Tal, Katherine G. Akers y Gordon K. Hodge. *Effect of paper color and question order on exam performance*. Teaching of Psychology 35(1): 26-28. 2008.
- [7] Dennis F. Togo. *Topical sequencing of questions and advance organizers impacting on student's examination performance*. Accounting Education 11(3). 2002.