

EVALUACION DEL RENDIMIENTO EN LA UNIVERSIDAD NACIONAL DE EDUCACION A DISTANCIA *

por CARMEN JIMÉNEZ FERNÁNDEZ
*Universidad Nacional de Educación
a Distancia*

Introducción

Ciertamente, la eficacia de una Universidad se juzga por sus graduados [1]. La Universidad Nacional de Educación a Distancia (UNED) tiene ya diez años de existencia y aunque no se ha realizado un estudio que evalúe la totalidad del sistema [2], no cabe defender *a priori* su inferior calidad [3]. Sarramona afirma que «tan sólo la ignorancia puede hoy poner en duda la necesidad y eficacia de la educación a distancia» [4], afirmación que compartimos en su totalidad.

Y es que el problema real de la UNED, como el de todo nuevo sistema, no es el cambio técnico que lleva consigo, sino los cambios humanos que necesariamente deben acompañar toda innovación [5].

Surgen los problemas cuando a un sistema innovador se responde aplicando el modelo convencional: sustituir el texto oral del profesor por el escrito, querer convertir el centro asociado en *campus* universitario, suplir el examen clásico por la prueba presencial. Las funciones que con mayor o menor bondad cumplen estos medios y estructuras en la universidad clásica no responden de igual modo en la educación abierta. La experiencia no se improvisa y el cambio humano es lento y hasta conflictivo. Los profesores de la UNED lo sabemos hoy, aunque en la práctica no se ensaye o se atine con la necesaria estrategia innovadora.

La teoría general de sistemas nos muestra que la evaluación del estudiante no es sino un subsistema del macrosistema UNED. Para la educa-

* Esta investigación fue presentada al *Congreso Internacional de Universidades Abiertas y a Distancia*, celebrado en Madrid, del 24 al 28 de octubre de 1983, sobre el tema «Evaluación del rendimiento de la enseñanza superior a distancia».

ción superior a distancia, Escotet [6] propone los siguientes subsistemas que interactúan entre sí: gestión, planificación, diseños, producción, mediatización, distribución, utilización, evaluación y mantenimiento. Todos ellos habría que evaluarlos.

Sabemos, además, que la evaluación cobra sentido al relacionarla con la concepción del objeto a evaluar. Es distinto querer evaluar sólo la enseñanza-aprendizaje (objetivos cognitivos de Bloom) que añadir a ello la evaluación del contexto u objetivos afectivos. Tampoco nos preguntamos con Gagné sobre el grado de pertinencia de la educación que se imparte, es decir, si responde a las metas individuales y sociales de individuos y grupos. Nos ceñimos a investigar parte del subsistema evaluación.

Para nosotros este subsistema implica evaluar al estudiante, al diseño curricular y a los objetivos y procesos en general. Los datos que aportamos ahora se centran en la evaluación del rendimiento del estudiante, aunque también abordan la eficacia del diseño curricular y del proceso.

La investigación se centra en una asignatura, Pedagogía Experimental II, de segundo curso de Ciencias de la Educación de la UNED, y al curso académico 1982-83. Pretende arrojar luz sobre dos puntos fundamentalmente: 1) qué grado de acuerdo existe entre los criterios de evaluación de profesores, tutores y alumnos, y 2) qué medios podemos emplear para aproximar nuestras orientaciones y, en consecuencia, estimular más adecuadamente el autoaprendizaje y la formación de los alumnos.

1. El problema

En términos generales, podemos resumirlo como sigue. Parece existir cierto malestar e insatisfacción entre profesores, tutores y alumnos por los criterios que emplean unos y otros al evaluar el rendimiento de éstos. Quizá se eche de menos una mayor claridad en las funciones de los diversos estamentos, así como en el modelo evaluativo de la UNED.

Al evaluar los trabajos y pruebas de los alumnos se cumplen, al menos, dos funciones: educativa o de orientación y social o de sanción, equivalentes a la evaluación formativa y sumativa de Scriven [7], siempre que como él indica se evalúe el proceso y logro de metas valiosas. En la práctica, los tutores realizan la evaluación formativa y los profesores, la sumativa, si bien existen interrelaciones. Así, hay profesores que consideran los informes y sugerencias de los tutores a la vez que orientan a los alumnos a lo largo del curso, así como a través de las unidades que elaboran y tras la realización de las pruebas presenciales. Pero no queda claro «el grado de influencia de la opinión del tutor sobre los alumnos a la hora de emitir la calificación final. Ha sido éste —y sigue siendo, añadimos nosotros— un tema de permanente controversia en la UNED» [8].

En efecto, para Bercovitz «el tutor no es quien decide lo que ha de aprender el alumno y tampoco lo controla, puesto que la corrección de los exámenes es responsabilidad del profesorado de la Sede Central» [9]. Mientras que Tiana Ferrer [10] afirma que la calificación final se determina teniendo en cuenta las calificaciones de las dos pruebas presenciales, las cuatro pruebas de evaluación a distancia (corregidas por el tutor) y las valoraciones contenidas en el informe tutorial. Y añade después que debido al descontento de los tutores sobre la falta de peso real de sus evaluaciones en la nota final, se acordó que la calificación otorgada por el profesor a cada prueba presencial puede ser modificada positiva o negativamente hasta en un 20 por 100 por la nota media obtenida en las correspondientes pruebas de evaluación a distancia [11].

Lo anterior indica lo delicado del tema que consideramos. La autora de estas líneas no ha conocido de modo oficial el acuerdo de que habla Tiana Ferrer a pesar de ser profesora de la UNED desde hace varios cursos. En su trayectoria profesional ha ensayado, por acuerdos de departamento, distintas formas de ponderar las calificaciones de los tutores: desde darles un peso del 25 por 100 siempre que no excedieran en más o menos 2'5 puntos la nota del profesor, a darles un peso discrecional aunque siempre consideradas.

Esta experiencia personal se realiza cuando, en el mismo trabajo, Tiana Ferrer afirma que en una encuesta realizada a los profesores de la Sede Central en 1979, el 40 por 100 manifiesta no concederles ninguna o poca importancia a las calificaciones de los tutores y el 51 por 100 expresa que las considera sólo en casos de alumnos dudosos [12].

Se puede o no admitir el valor de las pruebas de evaluación a distancia para la evaluación sumativa, pero sí hay que admitir la necesidad y posibilidad de tener criterios más comunes de evaluación. De este modo, el tutor no sólo ayudará psicológicamente al alumno, sino que orientará su aprendizaje de acuerdo con las pautas reales que marcan las programaciones de la guía del curso y las hasta ahora casi decisorias pruebas presenciales.

En este contexto se enmarca nuestra investigación.

¿Qué ocurre con los alumnos? Suscriben en muchos casos la queja de los tutores, a la vez que se lamentan de no poder participar formal y realmente en su propia evaluación. Añaden, además, otras quejas o sugerencias relacionadas con dificultades para recibir el material, calidad del mismo, etc.

Aún ignorando los modelos y principios que existen sobre evaluación, la UNED debería reflexionar sobre el siguiente tópico. Si la educación superior a distancia hace al alumno responsable de la conducción de su aprendizaje, ¿no habría que hacerlo también partícipe de su proceso de evaluación? Ciertamente aparece la dificultad ética del fraude [13] y la insuficiencia académica de muchos alumnos. Pese a ello, habría que tener presente que la evaluación debe satisfacer la exigencia de cooperación, pues evaluadores y evaluados persiguen metas comunes y en

distintos momentos del proceso se pueden cambiar los roles [14], en la evaluación institucional o del curriculum, por ejemplo.

Nuestra investigación se propone comprobar de modo empírico si profesores, tutores y alumnos somos capaces de aproximar nuestros criterios de evaluación al evaluar el mismo producto. Profesor, tutores y alumnos hemos evaluado por separado la actividad obligatoria * teniendo siempre como referencia los mismos objetivos y pautas de evaluación. Alumnos y profesor hemos calificado las pruebas presenciales de febrero y junio. El estudio del tema y nuestra experiencia personal nos llevaron a formular las siguientes hipótesis.

2. *Hipótesis*

2.1. Hay diferencias en las calificaciones que otorgamos profesor, tutores y alumnos en la puntuación global de la actividad obligatoria. Los tutores otorgan las calificaciones más altas, seguidos por los alumnos y el profesor.

2.2. Profesor y tutores otorgamos calificaciones distintas, siendo más altas las de éstos, tanto si se considera la evaluación global cuanto si se consideran aspectos parciales y más concretos de la actividad obligatoria. Cuanto más concretas y objetivas sean las reglas de evaluación, menores serán las diferencias. Así, las diferencias en interpretación de la prueba estadística serán menores que en la elección de dicha prueba, y éstas menores que las existentes en la puntuación global.

2.3. Los medios empleados por el profesor para orientar la evaluación de los tutores son eficaces. Así, los tutores asistentes al seminario y/o a las convivencias en los Centros Asociados se acercarán más a las calificaciones del profesor que los no asistentes.

2.4. En las pruebas presenciales de febrero y junio los alumnos se autocalificarán más alto que los califica el profesor. Esta diferencia será mayor en las pruebas de junio que en las de febrero, por haber habido en ésta información adicional.

3. *Variables*

La variable independiente es en todos los casos la interpretación que cada grupo de jueces hace de los mismos criterios de evaluación: los objetivos y pautas de evaluación de la disciplina contenidos en la guía del curso 1982-83. Esta variable sufre modificaciones.

Para la primera hipótesis comparamos la interpretación que hacen de la guía los tres grupos de jueces: profesor, tutores y alumnos al otor-

* Consiste en diseñar una sencilla investigación de dos o más grupos con una sola variable independiente y en realizarla, interpretarla y valorarla.

gar una valoración global. En la segunda nos ceñimos a los dos primeros grupos que otorgan, junto a la evaluación global, dos valoraciones parciales. En la tercera hipótesis la variable independiente cambia, ya que un grupo de tutores asiste al seminario y/o a las convivencias y otro no. Y en la cuarta cambia también, ya que en febrero se pidió la autoevaluación después de informar a los alumnos sobre los aciertos y fallos de su prueba, y en junio, antes de suministrarles esta información.

La variable dependiente es las calificaciones finales que otorga cada grupo.

4. Población y muestra

La población la componen los 142 alumnos matriculados. De ellos, 70 se presentaron a la primera prueba y 56 a la segunda. De estos últimos, sólo 31 respondieron a nuestra invitación de colaborar en la investigación. En cuanto a los tutores, nos dirigimos a todos a través de los Centros Asociados, habiendo recibido un total de 41 calificaciones válidas. Evidentemente, el profesor corrigió la totalidad de los exámenes y trabajos.

5. Diseño

Se trata de un diseño de un factor con dos o tres grupos independientes según el tipo de hipótesis. En siete ocasiones se ha procedido con el diseño de tres grupos, y en cuatro con el de dos.

6. Procedimiento

El proyecto se diseñó antes de comenzar el curso y no se les comunicó a tutores y alumnos. En octubre de 1982 se citó a todos los tutores de la disciplina al seminario que anualmente convoca la UNED. Acudieron los tutores de Málaga, Melilla, Ponferrada, Madrid (Móstoles), Las Palmas, Pontevedra y Segovia, descartándose al final las calificaciones de este último por tener sólo alumnos del curso de adaptación. En la hora y media que duró el seminario comentamos los contenidos y objetivos de la materia, insistiendo en la necesidad de fijarnos en los objetivos y pautas de evaluación citados en la guía del curso a la hora de calificar a los alumnos. Hubo preguntas y respuestas, pero no se descendió a ejemplos reales.

Entre enero y mayo de 1983, la autora se desplazó a los Centros Asociados de Albacete, Burgos, Palencia, La Coruña y Ponferrada. El trabajo en ellos se desarrolló del modo siguiente: tras unos minutos para escuchar sugerencias y dudas de tutores y alumnos, ponía un ejemplo prác-

tico que, con ligeras modificaciones, permitía resumir los conceptos, pruebas estadísticas y estrategias de trabajo de la disciplina. Conviene aclarar que Pedagogía Experimental II tiene un contenido a veces difícil para nuestros alumnos, pero que es muy objetivo y se puede dominar si se capta la estrategia empírico-experimental. Estas convivencias duraban dos horas como mínimo y a lo largo de ellas se ejemplificaban aciertos y fallos de los distintos planteamientos y respuestas de los alumnos de cursos anteriores. Los tutores asistieron, salvo en un caso, por razones de fuerza mayor.

En la primera quincena de mayo se dirigió escrito a todos los tutores pidiéndoles su colaboración para el presente estudio. Se les informaba del objetivo general del mismo y se les indicaba la posibilidad de enviárselo una vez concluido si así lo deseaban. Acompañaba a la carta una ficha-modelo para que puntuaran la actividad obligatoria de sus tutelados sobre una escala de 1 a 10. Se les recordaban objetivos y normas de la guía del curso, pidiéndoseles otorgar: 1) una puntuación global; 2) una puntuación a la elección de la prueba estadística contenida en el trabajo, y 3) otra puntuación a la interpretación de los resultados. En noviembre de 1982 se les envió a todos un modelo de la actividad que pudieron consultar los alumnos que lo desearon, que fueron casi todos. Aparte del modelo enviado, la guía del curso contenía indicaciones precisas que sirvieran de orientación a todos y de autoevaluación al alumno. El contacto se estableció a primeros de mayo para que no se les olvidara, pues es a finales de mayo cuando los tutores realizan los informes finales del alumno.

En cuanto a los alumnos, se les entregó una carta y un modelo de evaluación parecido al de los tutores junto con la segunda prueba presencial con el ruego de que me remitieran al Departamento la información pedida. Les pedía se autoevaluaran en una escala de 1 a 10 tanto en la actividad obligatoria como en cada una de las dos pruebas realizadas, teniendo como referencia los objetivos y pautas de la guía del curso. No les pedí en este caso puntuaciones parciales de la actividad por considerar que ello requiere una cualificación mayor de la que muchos podían tener. Al enviarles las orientaciones y notas de las pruebas de junio, les recordaba me remitieran la autoevaluación si no lo habían hecho, enviándoles para ello una nueva ficha-modelo.

Finalmente, el profesor corrigió todos los exámenes y trabajos, pues aunque éstos los suelen corregir los tutores, les había indicado en la guía del curso —que llega a todos con el sobre de la matrícula— que deberían enviarlo necesariamente al profesor y al tutor. Observé los mismos criterios y tipo de escala.

El examen de febrero ha consistido en una prueba objetiva y en la resolución de un problema acerca del cual se hacen preguntas concretas de interpretación, y la de junio en la resolución de un problema y con base en él se le hacían varias preguntas de interpretación y de definición de conceptos estadísticos precisos.

7. Análisis estadístico

Se supone que se cumplen los requisitos del análisis de varianza (Anova) y se ha aplicado éste en los once casos. Cuando F ha resultado significativa, en diez de ellos, se ha procedido a comparaciones a posteriori aplicando el método de Sheffé en los casos en que el tamaño de los grupos es desigual y el de Tuckey, en aquellos otros en que n es igual. El nivel de significación es de 0'05.

8. Resultados

8.1. Referidos a la primera hipótesis: Hay diferencias en la *calificación global* que otorgan los alumnos ($n_1 = 31$), los tutores ($n_2 = 41$) y el profesor ($n_3 = 40$) a la actividad obligatoria. La siguiente tabla refleja los resultados del Anova.

Tabla 1: Resultados del ANOVA. Calificación global

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	42	2	21	20'38	xxx (*)
Dentro	113	109	1'03		

$$\bar{X}_1 = 6'38; \bar{X}_2 = 7'30; \bar{X}_3 = 5'9$$

Se confirma la primera parte de la hipótesis a un nivel de confianza muy superior al fijado. Al aplicar el método de Sdiefé para las comparaciones a posteriori tenemos lo siguiente:

Tabla 2: Comparaciones a posteriori por el método S.

Contraste	Resultado	Valor crítico	Significación
$ 6'38 - 7'30 (**)$	1'95	2'80	
$\sqrt{1'03 \left(\frac{1}{31} + \frac{1}{41} \right)}$			
$\mu_1 - \mu_3$	3'88	2'80	xxx
$\mu_2 - \mu_3$	6'38	2'80	xxx

(*) xxx = significativa al 1 por mil; xx = significativa al 1 por cien; x = significativa al 5 por cien; en blanco = no significativa.

(**) Para no alargarnos, sólo indicaremos en lo sucesivo los grupos que se comparan. Los simbolizaremos con los parámetros y equivalen en el ejemplo desarrollado a $\mu_1 - \mu_2$.

Empíricamente, el orden aparecido en la calificación promedio es el sostenido por la hipótesis de investigación. Sin embargo, la diferencia encontrada entre las autoevaluaciones y las puntuaciones de los tutores no alcanzan el rango de significación científica. Mientras sí alcanzan significación superior a la fijada las diferencias en las evaluaciones hechas por los alumnos y el profesor y las realizadas por los tutores y el profesor.

8.2. La hipótesis segunda, existen diferencias entre las calificaciones globales y parciales de tutores ($n_1 = 41$) y profesor ($n_2 = 49$), también queda confirmada. En la puntuación global acabamos de verlo al comparar los grupos 2 y 3. Ahora presentamos primero los resultados referidos a la adecuación de la *prueba estadística* y en segundo lugar los referidos a la *interpretación* que se hace de sus resultados.

Tabla 3: *Resultados del ANOVA. Prueba estadística*

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	42'62	1	42'62	25'89	xxx
Dentro	144'63	88	1'64		

$$\bar{X}_1 = 7'40; \bar{X}_2 = 6'02$$

Tabla 4: *Resultados del ANOVA. Interpretación*

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	25'5	1	25'5	17'11	xxx
Dentro	131'25	88	1'49		

$$\bar{X}_1 = 7'23; \bar{X}_2 = 6'16$$

En los tres tipos de puntuaciones hay diferencias significativas al 1 por 1.000. La diferencia empírica mayor es la encontrada entre la puntuación global, 1'40 puntos, y la menor en la interpretación, 1'07, alterándose el orden predicho en la hipótesis. Quizá se deba a que en la carta se especificaba interpretación de la prueba estadística, es decir, se acepta o no la hipótesis de nulidad y en la ficha-modelo se resumía en interpretación. Han podido incluir en ella la validez del diseño que es ya menos objetiva.

8.3. La hipótesis tercera se refiere a la eficacia de dos medios, seminario y convivencias, para aproximar los criterios de evaluación. Presentamos primero los resultados referidos al *seminario*. Tutores asistentes al mismo ($n_1 = 15$), tutores no asistentes ($n_2 = 15$) y profesor ($n_3 = 15$)

son ahora los grupos en comparación, tanto para la puntuación global cuanto para la elección de la prueba estadística y su interpretación. La muestra del profesor se obtuvo tomando al azar la mitad de los alumnos de cada uno de los dos grupos. Referidos a la *puntuación global* tenemos:

Tabla 5: *Resultados del ANOVA. Seminario*

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	64'96	2	32'48	5'19	xx
Dentro	262'77	42	6'26		

$$\bar{X}_1 = 7'56; \bar{X}_2 = 7'43; \bar{X}_3 = 4'89$$

Tabla 6: *Comparaciones a posteriori por el método T.*

C.	R.	V.C.	Significación
$\mu_1 - \mu_2$	0'20	3'44	
$\mu_1 - \mu_3$	4'09	3'44	x
$\mu_2 - \mu_3$	3'89	3'44	x

No se confirma la hipótesis de investigación. La asistencia al seminario no aproximó los criterios de evaluación de tutores y profesor en la puntuación global.

En cuanto a la *elección de la prueba* tenemos:

Tabla 7: *Resultados del ANOVA. Seminario*

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	34	2	17	10'75	xxx
Dentro	66'5	42	1'58		

$$\bar{X}_1 = 7'6; \bar{X}_2 = 7'9; \bar{X}_3 = 5'9$$

Tabla 8: *Comparaciones a posteriori por el método T.*

C.	R.	V.C.	Significación
$\mu_1 - \mu_2$	0'94	3'44	
$\mu_1 - \mu_3$	5'31	3'44	xx
$\mu_2 - \mu_3$	6'25	3'44	xx

También se acepta la hipótesis de no diferencia.

Finalmente, en cuanto a la *interpretación*, estos son los resultados obtenidos.

Tabla 9: *Resultados del ANOVA. Seminario*

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	20'2	2	10'1	6'35	xx
Dentro	67	42	1'59		

$$\bar{X}_1 = 7'4; \bar{X}_2 = 7'7; \bar{X}_3 = 6'2$$

Tabla 10: *Comparaciones a posteriori por el método T.*

C.	R.	V.C.	Significación
$\mu_1 - \mu_2$	0'91	3'44	
$\mu_2 - \mu_3$	3'64	3'44	x
$\mu_1 - \mu_3$	4'55	3'44	xx

Se acepta la hipótesis de nulidad. El seminario no logró en ningún caso establecer diferencias entre los tutores más allá de las esperadas por azar. En los tres casos, la media empírica de los asistentes al seminario es algo menor que la de los no asistentes, pero las diferencias no alcanzan los niveles exigidos por el quehacer científico.

Veamos ahora el posible efecto de las *convivencias* en el mismo orden que acabamos de exponer ($n_1 = 8$; $n_2 = 16$ y $n_3 = 20$). En cuanto a la *puntuación global* tenemos:

Tabla 11: *Resultados del ANOVA. Convivencia*

F.V.	M.C.	g.l.	S.C.	F.	Significación
Entre	52'8	2	26'9	19'55	xxx
Dentro	56'95	41	1'39		

$$\bar{X}_1 = 7; \bar{X}_2 = 8'3; \bar{X}_3 = 5'8$$

Tabla 12: *Comparaciones a posteriori por el método S.*

C.	R.	V.C.	Significación
$\mu_1 - \mu_2$	2'54	2'54	x
$\mu_1 - \mu_3$	2'45	2'54	
$\mu_2 - \mu_3$	6'25	2'54	xx

Se confirma la hipótesis de investigación en todos sus puntos. Los

tutores asistentes a las convivencias y los no asistentes han puntuado de forma distinta sin que la diferencia se explique por azar en más del 5 por 100 de los casos. También se confirma que entre tutores asistentes a las convivencias y profesor las diferencias no son significativas y sí en el caso contrario.

La elección de la prueba estadística arroja los siguientes resultados:

Tabla 13: Resultados del ANOVA. Convivencia

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	46'27	2	23'13	16'64	xxx
Dentro	58'33	41	1'42		

$$\bar{X}_1 = 7'37; \bar{X}_2 = 8'31; \bar{X}_3 = 6'05$$

Tabla 14: Comparaciones a posteriori por el método S.

C.	R.	V.C.	Significación
$\mu_1 - \mu_2$	1'81	2'54	
$\mu_1 - \mu_3$	2'69	2'54	x
$\mu_2 - \mu_3$	5'79	2'54	xx

No se confirma la hipótesis de investigación, aunque los datos empíricos siguen la dirección marcada por ella.

Por último, presentamos los resultados referidos a la interpretación:

Tabla 15: Resultados del ANOVA. Convivencia

F.V.	M.C.	g.l.	S.C.	F.	Significación
Entre	70'73	2	35'86	8'79	xxx
Dentro	167'43	41	4'08		

$$\bar{X}_1 = 6'37; \bar{X}_2 = 8'43; \bar{X}_3 = 5'65$$

Tabla 16: Comparaciones a posteriori por el método S.

C.	R.	V.C.	Significación
$\mu_1 - \mu_2$	2'37	2'54	
$\mu_1 - \mu_3$	0'86	2'54	
$\mu_2 - \mu_3$	4'09	2'54	xx

Se confirma parcialmente la hipótesis de investigación. No existen diferencias significativas entre tutores asistentes a las convivencias y pro-

fesor y sí entre éste y los no asistentes. Sin embargo, las diferencias entre ambos grupos de tutores quedan explicadas por el azar a pesar de lo marcado de la diferencia empírica.

8.4. La última hipótesis sostiene la existencia de diferencias mayores en la prueba de junio que en la de febrero entre las calificaciones de alumnos ($n_1 = 32$) y profesor ($n_2 = 32$). Estos son los resultados de la *prueba de febrero*:

Tabla 17: *Resultados del ANAVA. Prueba de febrero*

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	7'84	1	7'84	3'96	
Dentro	122'86	62	1'98		

$$\bar{X}_1 = 5'69; \bar{X}_2 = 5'1$$

Y éstos en la de junio:

Tabla 18: *Resultados del ANOVA. Prueba de junio*

F.V.	S.C.	g.l.	M.C.	F.	Significación
Entre	6'6	1	6'6	7'41	xx
Dentro	54'4	61	0'89		

$$\bar{X}_1 = 7'14; \bar{X}_2 = 6'5$$

En la prueba de febrero no se confirma la hipótesis de investigación. Es decir, profesor y alumnos otorgaron a la prueba puntuación similar. Recuérdese que se les pidió la autoevaluación casi tres meses después de haberles enviado un modelo de examen ya corregido. Este modelo y en alguna medida el tiempo transcurrido, han logrado aproximar la calificación de ambos jueces. Pero en junio, se han autoevaluado de forma significativamente superior a como lo ha hecho el profesor con lo que sí se confirma esta parte de nuestra hipótesis [15].

9. A modo de conclusión

En cuanto a planteamiento general y experiencia, sugerimos lo siguiente:

1. Sería deseable clarificar las funciones de profesor, tutores y alumnos en el rendimiento de éstos y mejorar e innovar los canales de información mutua. El tema de la autoevaluación de los estudiantes y de las evaluaciones del tutor exige cambios y datos que debemos buscar coordinadamente.

2. La guía del curso debe contener todo cuanto sea de interés para la preparación de la disciplina además de la programación detallada de la misma. Es la única información de primera mano que reciben con puntualidad y seguridad los alumnos y Centros Asociados. Evita malentendidos y consultas innecesarias.

En cuanto a los resultados del estudio, se pueden hacer las siguientes observaciones:

3. Si se mantienen a lo largo del curso las pruebas y trabajos tradicionales, sin medios ni refuerzos específicos, profesor, tutores y alumnos puntúan de modo distinto. En todos los casos considerados, los tutores puntúan más alto que los profesores y que los propios alumnos. Estas diferencias son siempre significativas a un alto nivel de confianza cuando la comparación se establece entre profesor y tutores. Si los grupos en comparación son tutores y alumnos, existen diferencias empíricas a favor de los tutores, pero sólo en un caso de siete es significativa. Parece como si el tutor tuviera «tendencia a mejorar la imagen de sus alumnos ante él» [16].

4. El pedir a los tutores una evaluación global junto a una evaluación analítica, no contribuye a aproximar las calificaciones finales de éstos y profesor si no existe algún tipo de entrenamiento previo [17]. En general, no discriminan entre los diversos aspectos a juzgar.

5. Por su duración y estructura, el seminario no logró unificar o aproximar los criterios de evaluación. Concebidos como hasta ahora, logran si acaso una toma de contacto. Para lograr efectividad, deberían tratar un tema monográfico, estar bien preparados en cuanto a lugar, tiempo y material y tener un título con garra. La invitación cursarla con el envío del programa para que se aprecie los objetivos y el interés del tema. Unas dietas más holgadas podría contribuir a aumentar la asistencia.

6. Las convivencias aparecen como *un medio eficaz* de orientar a tutores —y alumnos— y de aproximar nuestras formas de evaluar. Ejemplos cuasiperfectos, con fallos graves y con fallos poco importantes, aclaran bastante el porqué de determinada nota. De las tres notas otorgadas, en dos casos no ha habido diferencias significativas entre las dadas por el tutor asistente a la convivencia y la dada por el profesor. Y en una ocasión ha habido diferencias significativas entre los dos grupos de tutores, siendo más discriminativos los asistentes a las convivencias.

Estas convivencias podrían lograr mayor eficacia si se aumentara su duración, aunque respetando los necesarios períodos de descanso. En disciplinas como la nuestra se cumple que «un ejemplo vale más que mil palabras».

7. Aquí se confirma que el alumno tiene tendencia a autoevaluarse por encima del valor objetivo de su producto. Necesariamente no existe en ello mala voluntad, sino que más bien alcanza el techo de la información que posee y mezcla, además, factores personales y/o profesionales

que no considera el profesor que evalúa su prueba. Esta afirmación se apoya en los datos. En las pruebas de junio las diferencias entre auto-evaluaciones y evaluación del profesor sólo se explican por azar en menos del 1 por 100 de los casos. Mientras que en las de febrero, en las que ya cotejaron lo hecho con lo que habrían deseado hacer, las diferencias son aleatorias, es decir, puntúan ambos grupos de modo similar.

Parece, por lo tanto, que ejemplos de pruebas ya corregidas en poder de los alumnos o alguna otra técnica, podría aumentar su realismo a la hora de convertirse en juez y parte. En cualquier caso demuestra lo conveniente de enviar la información pertinente junto con la calificación. Y ello no sólo por el grado de conformidad que se logra como acabamos de ver, sino también, y sobre todo, por el efecto que produce el trabajar con información fidedigna, el *feed-back* de los ingleses.

Dirección del autor: Carmen Jiménez Fernández, Secretaria del Departamento de Pedagogía Experimental y Orientación, Facultad de Filosofía y Ciencias de la Educación. UNED. Ciudad Universitaria. Madrid-3.

NOTAS

- [1] FERGUSON, J. (1975) *The Open University from within*, p. 149 (London, University of London Press Ltd.).
- [2] VARIOS *La UNED y sus alumnos* (Madrid, UNED). Con este título, el ICE de la UNED publica anualmente una obra que ofrece los porcentajes de aprobados, suspensos, abandonos, etc., opiniones de sus estudiantes y características de su población.
- [3] JIMÉNEZ FERNÁNDEZ, C. y OTRAS (en prensa) Evaluación de algunas características de los alumnos y de la metodología de la UNED, p. 8, en *Actas del I Congreso de Tecnología Educativa* (Madrid, Sociedad Española de Pedagogía).
- [4] SARRAMONA, J. (1981) Teoría y práctica de la educación a distancia, p. 1. Ponencia presentada al *Seminario Iberoamericano sobre diseño y evaluación de programas de educación superior a distancia* (Santiago, Chile, O.E.A.-C.P.E. E.I.P., Mineo).
- [5] LAWRENCE, P. (1981) How to deal with resistance to change, *Harvard Business Review*, february.
- [6] ESCOTET, M. (1980) *Métodos y técnicas de la Educación superior a Distancia* (San José, Costa Rica, UNED).
- [7] SCRIVEN, M. (1976) The methodology of evaluation, en STAKE, R. E. (Ed.) *Curriculum Evaluation* (Chicago, American Educational Research Association, Monograph Series on Evaluation, n.º 1, Rand MacNally).
- [8] TIANA FERRER, A. (1981) El sistema de evaluación en la UNED, p. 139, en *Seminario Iberoamericano de Educación a Distancia y Educación de adultos* (Madrid, UNED).
- [9] BERCOVITZ, A. (1981) Las estructuras de apoyo: los Centros Asociados, p. 263, en *Seminario Iberoamericano de Educación a Distancia y Educación de Adultos* (Madrid, UNED).
- [10] TIANA FERRER, A. *O.c.*, p. 136.
- [11] Idem, p. 139.
- [12] Idem, pp. 142-143.

- [13] SARRAMONA, J. *O.c.*, p. 11.
- [14] GARZÓN, A. (1981) Una aproximación al problema de la evaluación en la educación a distancia: un modelo sistémico estructural, p. 23, en *Seminario Iberoamericano sobre diseño...*, *o.c.*
- [15] La autora impartió en la UNED una asignatura teórico-práctica de 5.º curso del plan antiguo de Ciencias de la Educación y en una encuesta anónima contestada por 100 alumnos de los 110 que se presentaron a examen, el 80 por cien consideró justa la calificación final obtenida mientras el 10 por cien no la consideró así. Cfr. JIMÉNEZ FERNÁNDEZ, C. y OTRAS, *o.c.*, p. 15.
- [16] TIANA FERRER, A. *O.c.*, p. 139.
- [17] Curiosamente, algunos tutores han querido colaborar en la investigación, pero no han respondido a la ficha modelo, sino a otra más analítica confeccionada por ellos. Al final indicaban que juzgara personalmente qué calificación merecían los alumnos apoyándose en sus apreciaciones.

BIBLIOGRAFIA

- BERCOVITZ, A. (1981): Las estructuras de apoyo: los centros asociados, en *I Seminario Iberoamericano de Educación a Distancia y Educación de Adultos* (Madrid, UNED).
- ESCOTER, M. (1980): *Métodos y técnicas de la educación superior a Distancia* (San José, Costa Rica, UNED).
- FERGUSON, J. (1975): *The open University from within* (London, University of London Press ltd.).
- FERNÁNDEZ HUERTA, J. (1982): Autodidactismo en la educación de adultos a distancia, *Anuario*, 2 (Sección de Ciencias de la Educación. Universidad Autónoma de Barcelona).
- GARZÓN, A. (1981): Una aproximación al problema de la evaluación en la educación a distancia: un modelo sistemático estructural. Ponencia presentada al *Seminario Iberoamericano sobre Diseño y Evaluación de Programas de Educación Superior a Distancia* (Santiago, Chile, O.E.A.-C.E.E.I.P.), Mineo.
- JIMÉNEZ FERNÁNDEZ, C. y OTRAS (en prensa) Evaluación de algunas características de los alumnos y de la metodología de la UNED, Comunicación presentada en el *I Congreso de Tecnología Educativa* (Madrid, Sociedad Española de Pedagogía).
- LAWRENCE, P. (1971) How to deal with resistance to change, *Harvard Business Review*, february.
- MACKENZIE, M. y OTROS (1977) *Etudes ouvertes (systèmes d'instruction postsecondaire à distance)* (París, UNESCO).
- SARRAMONA, J. (1975) *Tecnología de la enseñanza a distancia* (Barcelona, C.E.A.C.).
- SARRAMONA, J. (1980) Problemas y posibilidades de la enseñanza a distancia, *Medios Audiovisuales* (Madrid), 91.
- SARRAMONA, J. (1981) Teoría y práctica de la educación a distancia. Ponencia presentada al *Seminario Iberoamericano sobre Diseño y Evaluación de Programas de Educación Superior a Distancia* (Santiago, Chile, O.E.A.-C.P.E.I.P.). Mimeo.
- SCRIVEN, M. (1967) The Methodology of Evaluation, en STACKE, R. E. (ed.), *Curriculum Evaluation* (Chicago, American Educational Research Association, Monograph Series on Evaluation, n.º 1, Rand MacNally).
- TIANA FERRER, A. (1981) El sistema de evaluación en la UNED. Ponencia presentada en el *I Seminario Iberoamericano de Educación a Distancia y Educación de Adultos* (Madrid, UNED).
- VILLARROEL, A. (1980) Tendencias actuales en la educación a distancia, *Revista Interamericana de Educación de Adultos*, O.E.A., 3: 1-2, pp. 149-177.

SUMARIO: Se trata de estudiar, primero, el grado de eficacia del seminario y la convivencia a la hora de aproximar los criterios de evaluación de profesor y tutores en la UNED. Se concluye que el seminario no logra aproximarlos, pero que ocurre lo contrario con la convivencia. En segundo lugar se estudia la concordancia existente entre las calificaciones otorgadas por profesor, tutores y alumnos. El primero califica significativamente más bajo que los otros dos grupos. Finalmente se analiza el efecto que produce en la autoevaluación el conocimiento formal de los aciertos y errores. Con este conocimiento desaparecen las diferencias existentes entre alumnos y profesor.

Descriptores: Open University, Evaluation of the students, Autoevaluation, Analysis of variance.