

**enseñanza  
de la  
física  
en la universidad**

Tomás Escudero Escorza

Premio Nacional a la  
Investigación Educativa 1977



ESTUDIOS de EDUCACION

---





**MINISTERIO DE EDUCACION  
INSTITUTO NACIONAL DE CIENCIAS DE LA EDUCACION**

**Enseñanza de la física  
en la Universidad**



# **ENSEÑANZA DE LA FISICA EN LA UNIVERSIDAD**

## **LA EVALUACION PERIODICA COMO ESTIMULO DIDACTICO**

**Por Tomás Escudero Escorza**

Premio Nacional a la  
Investigación Educativa 1977

© EDITA: SERVICIO DE PUBLICACIONES DEL MINISTERIO DE EDUCACION  
IMPRIME: R. GARCIA BLANCO - AVDA. PEDRO DIEZ, 3 - MADRID - 19  
DEPOSITO LEGAL: M. 23.884 - 1979  
I. S. B. N. 84 - 369 - 0711 - 6

## INDICE

	Págs.
Prólogo ... ..	11
<b>1. La enseñanza de las ciencias: campo abierto a la investigación ... ..</b>	<b>15</b>
1.1. Apuntes históricos ... ..	16
1.2. Definición, objetivos, criterios y características peculiares de la investigación en didáctica de las ciencias ... ..	17
1.3. Los modelos en las ciencias y la investigación didáctica ... ..	19
1.4. El proceso de la investigación ... ..	21
1.4.1. Enfoques ... ..	22
1.4.2. Tipos de investigación ... ..	24
1.4.3. Métodos ... ..	27
1.5. Areas de investigación ... ..	30
1.6. Mirando el futuro ... ..	36
1.7. Consideraciones sobre la investigación en didáctica de las ciencias en España ... ..	37
<b>2. Consideraciones sobre la enseñanza de la física en la Universidad ... ..</b>	<b>41</b>
2.1. La resistencia a la innovación ... ..	41
2.2. Variedad y flexibilidad en el currículum ... ..	42



	Págs.
2.3. Algunos aspectos de la enseñanza de la física en la Universidad ... ..	46
2.3.1. Física para no-físicos ... ..	47
2.3.2. Matemáticas para físicos ... ..	49
2.4. El proceso didáctico ... ..	50
2.4.1. Estado inicial de los alumnos ... ..	52
2.4.2. Metodología didáctica ... ..	54
2.4.3. Evaluación de la actuación del profesorado.	58
2.5. Experimentación ... ..	59
2.6. Formación de profesores ... ..	61
2.7. La estructura departamental ... ..	64
<b>3. La evaluación periódica como estímulo didáctico: un análisis histórico ... ..</b>	<b>65</b>
3.1. Fundamentaciones teóricas ... ..	66
3.2. Periodicidad de la evaluación ... ..	69
3.3. Influencia del tipo de instrumento de evaluación ...	71
3.4. Influencia del conocimiento de los resultados ... ..	73
3.5. Influencia sobre el rendimiento escolar ... ..	74
3.6. Comentario final ... ..	76
<b>4. Descripción del estudio experimental ... ..</b>	<b>79</b>
4.1. Planteamiento del problema: propósitos del trabajo.	79
4.2. Niveles de tratamiento experimental ... ..	84
4.3. Sujetos base de nuestro estudio ... ..	86
4.4. Diseño experimental ... ..	89
4.5. Control experimental ... ..	94
4.6. Comentario sobre la validez interna y externa de nuestros experimentos ... ..	105
<b>5. Fundamentos del análisis multivariante ... ..</b>	<b>107</b>
5.1. Modelo general de análisis multivariante lineal (efectos fijos) ... ..	108
5.2. Contraste de hipótesis lineales multivariantes ... ..	114

	Págs.
5.3. Procedimientos de contraste de hipótesis ... ..	118
5.4. El análisis de covarianza ... ..	123
<b>6. Análisis de datos ... ..</b>	<b>125</b>
6.1. Planteamiento inicial ... ..	125
6.2. Descripción del programa (BMD 11V) ... ..	125
6.3. Naturaleza de los datos de entrada ... ..	128
6.4. Hipótesis contrastadas ... ..	134
6.5. Interacción de la evaluación periódica con el contexto educativo ... ..	148
6.6. Otros resultados de nuestros experimentos ... ..	151
6.7. Límites de nuestro trabajo ... ..	152
<b>7. Conclusiones ... ..</b>	<b>155</b>
<b>Bibliografía ... ..</b>	<b>158</b>



## PROLOGO

La presente obra recoge una investigación de carácter experimental sobre didáctica de la Física a nivel universitario. La planificación general, recogida de referencias y puesta a punto del tema, se llevó a cabo durante el curso 1973-74 en la Escuela de Educación de la Universidad de Stanford (California). El posterior diseño experimental, su aplicación y análisis, se realizó en el Departamento de Física Fundamental (Cátedra de Optica) de la Universidad de Zaragoza, durante los cursos 1974-1976 y principios del 1976-1977.

El contenido del estudio refleja la constante preocupación de que estamos trabajando con un amplio tema poco tratado en nuestro país; de ahí la necesidad de incluir dos capítulos de síntesis, enfocados de una parte a poner a punto las ideas para una mejor comprensión de este trabajo y de otra para servir de base a futuras investigaciones que puedan realizarse en este campo.

En nuestra labor sintetizadora hemos intentado en todo momento abstraer de la temática general aquello que nos aparecía como relevante en nuestro sistema educativo y situación particular, ya que hablar de Ciencias de la Educación sin ceñirse a un determinado contexto social es de escasa operatividad.

El primer capítulo pretende ofrecer de forma compacta el qué, por qué, cómo y dónde de la investigación en didáctica de las ciencias. Dentro de nuestro empeño, consideramos que este capítulo viene a ser un punto de partida necesario porque centramos el campo de trabajo y las líneas alternativas de acción posibles. La labor de síntesis ha sido dificultosa en algunos aspectos, sobre todo en lo relativo a la metodología de investigación, por tener que poner en concordancia distintos tratamientos que partían de puntos de vista muy diversos.

\* Restringiendo el tema a una facultad, sección o departamento de Física en la Universidad, nuestro segundo capítulo ofrece un análisis de la problemática didáctica que nos aparece como más peculiar. Nuestro proyecto se mueve en la esfera universitaria, que presenta ciertas particularidades específicas que conviene detectar y señalar. La problemática tratada es en su mayoría generalizable a nivel internacional; sin embargo, hemos puesto particular énfasis en aquello con más clara incidencia en nuestra Universidad, debiendo dejar constancia de que la visión que ofrecemos es el fruto de un trabajo de síntesis reposado con base en un gran número de referencias, la mayoría trabajos cortos sobre un tema muy concreto, y que dicha síntesis bien pudiera haber tenido otra estructuración y otros puntos relevantes. El tema es delicado de tratar por la gran diversidad de aspectos que intervienen y que pueden ser manipulados, consciente o inconscientemente, en apoyo de ciertas ideas prefijadas. En nuestro caso hemos pretendido minimizar enfoques personales, aunque estamos seguros de que la síntesis responde a un planteamiento concreto de entre los varios válidos con los que se hubiera podido abordar el tema.

Los siguientes capítulos recogen nuestro estudio experimental propiamente dicho, en el que hemos analizado los efectos estimulantes en el rendimiento escolar de ciertos tratamientos de evaluación frecuente. Los experimentos están centrados en la disciplina física para estudiantes universitarios con una especialización científica distinta de la Física.

El tercer capítulo es un análisis histórico y puesta a punto del tema de la evaluación frecuente como estímulo didáctico, que sirve para centrar y justificar nuestro proyecto dentro de una línea concreta de investigación en metodología didáctica que viene siendo abordada desde principios de siglo. El desarrollo metodológico y didáctico ha trasladado el interés de los investigadores a la búsqueda de interacciones significativas del tratamiento con otras variables, rendimiento previo, actitudes, aptitudes, etcétera. Nuestro trabajo experimental intenta ofrecer ciertos aportes a la problemática de las interacciones del tratamiento elegido, campo que todavía carece de una exploración a fondo.

La descripción de los objetivos, motivaciones y metodología del trabajo experimental aparece en el capítulo cuarto, en el que hacemos especial hincapié en lo relativo a diseño y control experimental. Al igual que en la mayoría de los capítulos de la memoria, hemos añadido a la redacción puramente descriptiva elemen-

tos formativos e informativos que considerábamos interesantes en función de los objetivos básicos de trabajo en su conjunto.

Los capítulos quinto y sexto recogen el modelo de análisis multivariante empleado y el consiguiente análisis de datos y de efectos significativos. Las motivaciones que hemos tenido para desarrollar el modelo multivariante han sido las siguientes:

- a) Porque entendíamos que entroncaba con un trabajo con ciertas orientaciones formativas.
- b) Porque presentamos el modelo perfectamente ajustado al modo en que hemos llevado a cabo el análisis de datos, esto es, el enfoque general del *modelo de regresión*, en el que las *matrices-diseño* de las variables independientes juegan el papel principal.
- c) Porque viene a ser una aportación para el investigador aplicado, frecuentemente cercado por textos especializados que presentan el análisis multivariante a un nivel teórico difícil de abordar y textos puramente aplicados que no ofrecen apenas fundamentación teórica. Entendemos que el enfoque que hemos desarrollado es de gran facilidad de aplicación general por la simplicidad de las matrices-diseño que intervienen.

El último capítulo recoge las conclusiones generales de nuestro estudio y su significado para trabajos posteriores.

No cabe duda de que en una investigación didáctica experimental, gran parte de las dificultades prácticas surgen de la necesidad de colaboración de muchas personas implicadas en el propio contexto experimental. En nuestro caso hemos de señalar el apoyo, aliento y facilidades que en todo momento hemos encontrado en las personas y estamentos a los que hemos solicitado colaboración, tanto en la Facultad de Ciencias de la Universidad de Zaragoza como en la Academia General Militar.

Mi especial agradecimiento debe dirigirse al profesor J. Casas Peláez, impulsor y orientador del trabajo; a los profesores de ambos Centros, J. Aporta, E. Bernabéu, J. R. Blasco, J. L. Brun, R. Corbalán, M. A. Rebolledo, C. Santamaría, J. Savirón, J. A. Turégano y M. J. Yzuel, que participaron en la aplicación de los tratamientos experimentales, y a mis compañeros del Instituto de Ciencias de la Educación, que me proporcionaron la ayuda cotidiana y el ambiente propicio necesarios.

Zaragoza, abril 1978

Tomás ESCUDERO ESCORZA



## 1. LA ENSEÑANZA DE LAS CIENCIAS: CAMPO ABIERTO A LA INVESTIGACION

Es un hecho evidente que la ciencia y la tecnología juegan un papel cada vez más preponderante en el desarrollo de la humanidad, de ahí que la formación en ciencias sea un aspecto de gran relevancia dentro de la formación básica de todo individuo. Los educadores en ciencias adquieren en este contexto una enorme responsabilidad social, que obligatoriamente les conduce a la búsqueda de alternativas óptimas de transmisión del conocimiento científico y, en última instancia, de formación de futuros científicos y de desarrollo armonioso de la propia ciencia. En definitiva, el profesor de ciencias se encara abiertamente con una necesaria acción investigadora que sistemáticamente analice, y en lo posible mida, cómo es y cómo debe ser su acción educativa. Acción que a la postre supone adoptar también en el plano didáctico algo universalmente aceptado por todos los científicos en su tarea investigadora, esto es, la vieja y conocida idea cuya paternidad se atribuye a lord Kelvin (1) de que "cuando no podemos medir aquello de lo que estamos hablando, nuestro conocimiento es confuso e insatisfactorio".

El presente capítulo intenta ofrecer una visión sucinta de la situación y problemática general de la investigación en didáctica de las ciencias, siguiendo un enfoque *integrador*, esto es, sin distinguir entre las didácticas de cada disciplina científica de forma sistemática. Este supuesto no lo fundamentamos solamente en consideraciones de espacio y alcance del trabajo, sino que responde a motivos más profundos: nuestra idea de que hay muchos

---

(1) La presente cita es recogida por KUHN, T., en "The function of measurement in modern physical science", en *Quantification*, W. Harry (ed.), Indianapolis, Bobbs-Merril, 1961, 31.



más aspectos que unen a las distintas ciencias en el plano didáctico que los que las separan.

Las ideas que aparecen en nuestro análisis se encuadran en el marco de la investigación en didáctica de las ciencias, pero son aplicables en gran parte a la investigación en didáctica en general; algo lógico cuando aquélla no es sino una parcela de ésta.

### 1.1. Apuntes históricos

La investigación en ciencias sociales y del comportamiento es fundamentalmente un fenómeno del siglo XX. No puede sorprendernos, por tanto, que la investigación en educación, campo aplicado de las ciencias del comportamiento, sea un área de trabajo relativamente nueva, cuyo desarrollo ha debido ser precedido por el establecimiento como campos de investigación de las ciencias en que se basa: psicología, sociología, antropología, etc.

Remontándonos a la época en que la educación no era un fenómeno de masas, hecho que se produce entrado el presente siglo, los programas didácticos se basan en una concepción elitista de la educación. Así, la enseñanza de las ciencias venía a basarse en la acción tutorial de un profesor que enfatizaba primordialmente la observación de fenómenos naturales. Posteriormente, las ideas democratizadoras de ciertos pensadores como Bacon y Locke estimularon progresivamente el diseño y adaptación de programas para el uso escolar (2).

El comienzo de la investigación sistemática en didáctica de las ciencias nos atrevemos a situarlo en los años veinte, con la preponderancia y liderazgo del "Teacher College" de la Universidad de Columbia. En dicha institución se desarrollaron gran número de trabajos y tesis doctorales sobre el tema, aunque no siempre se tratara de problemas relevantes, y los logros fueran acordes con el esfuerzo dedicado. La principal dificultad parecía centrarse en la escasa orientación de investigadores-profesionales de muchos de los que realizaban los trabajos, siendo sus estudios "ad hoc" y no un puente de comunicación con otros investigadores. Ante esta situación, la consecuente escasez de resultados valiosos iba a producir una atmósfera de cierta desconfianza ante la investigación didáctica, dentro de la comunidad educativa.

Sin embargo, uno de aquellos trabajos, publicado en 1927 por

---

(2) SMITH, H. A.: "Educational research related to science instruction for the elementary and junior high school: A review and commentary", *Journal of Research in Science Teaching*, 1 (1963), 200.

Gerald S. Craig tras tres años de trabajo en su escuela-laboratorio (3), supuso una aportación decisiva en el campo de la enseñanza de las ciencias, viniendo a justificar por sí solo el gran esfuerzo de la Universidad de Columbia en los años veinte y treinta. Craig introduce ideas que están plenamente vigentes medio siglo después, concibiendo a la ciencia como una parcela decisiva y necesaria dentro del contexto general de la educación de la persona. Las leyes, generalizaciones y principios de la ciencia adquieren relevancia y son aplicables en numerosos aspectos de la vida con los que todo individuo va a enfrentarse. Asimismo, se señala expresamente la necesidad de considerar no sólo los aspectos cognoscitivos de la instrucción científica, sino también las dimensiones afectivas, actitudes, apreciación e intereses.

Sin embargo, es después de la segunda guerra mundial, cuando el desarrollo amplio de la metodología y de los procedimientos de análisis de datos permite dar un carácter más científico y riguroso a la investigación didáctica, proporcionando la posibilidad de analizar la realidad didáctica tal como se presenta, con muchas variables interaccionando.

En el caso de la enseñanza de las ciencias, el segundo gran esfuerzo comienza, a nuestro entender, tras la llegada de la era espacial, incidiendo fundamentalmente en el diseño curricular y en la elaboración de "proyectos" integrados o diferenciados por disciplinas. El desarrollo en cuanto a enfoques y métodos es notable, apareciendo algunos buenos ejemplos de proyectos extensos (PSSC, Harvard Project Physics, CHEM, Nuffield Project, ESCP, BSCS, etc.).

## **1.2. Definición, objetivos, criterios y características peculiares de la investigación en didáctica de las ciencias**

Todo intento sistemático para definir, estudiar y buscar soluciones alternativas del proceso de enseñanza y aprendizaje, con vistas a una planificación y desarrollo lógico de los programas de educación en ciencias, se encuadra en la investigación que nos concierne.

La búsqueda y definición de problemas propios de investigación surge como en otras disciplinas desde la perspectiva teórica y desde la práctica. En el primer caso, cuando existe vaguedad

---

(3) CRAIG, G. S.: "Certain techniques used in developing a course study in science for the Horance Mann elementary school", *Teacher College Contributions of Education*, Columbia University, New York, Teacher College Press, 1927.

o inconsistencia por parte de la teoría para explicar el proceso de enseñanza y aprendizaje, en el segundo, cuando se detectan deficiencias al planificar programas o interaccionar con los alumnos. Parece claro que los problemas que, en general, deben ofrecer más interés para el investigador, son aquellos con la doble dimensión, teórica y práctica.

La investigación didáctica presenta por el momento un nivel de desarrollo relativamente escaso, lo que es causa a menudo de resultados algo indefinidos en vez de teorías bien organizadas. Sin embargo, se evidencia un gran paralelismo en cuanto a concepción y proceso con la que se realiza en otras ciencias más paradigmáticas. Las diferencias observables provienen más del nivel de desarrollo que del planteamiento (4).

Los criterios son los mismos que los de toda investigación en cualquier campo o disciplina, esto es: a) relevancia, b) conceptualización adecuada, y c) metodología robusta. Criterios que son todos igualmente importantes y necesarios, pero que presentan una problemática particular en el caso de la enseñanza de las ciencias (5).

La *relevancia*, trivial a primera vista, es a veces conflictiva en materia curricular. Un "proyecto" válido desde el punto de vista interno puede perder fácilmente su relevancia e interés ante dificultades de validez externa, esto es, de aplicabilidad real, bien por carencia del material exigido o por inexistencia de profesorado con unas características concretas.

La *conceptualización* no presenta atributos y dificultades diferentes que en la investigación en otros campos en cuanto a su nivel de profundidad; sin embargo, las diferencias aparecen en materia de extensión. El buen investigador en didáctica de las ciencias va a necesitar una sólida preparación en áreas aparentemente disjuntas, tales como: a) ciencias de la naturaleza, generalmente con énfasis en una de ellas; b) educación en general (problemas, teorías, métodos, tendencias, etc.); c) psicología del aprendizaje; d) diseños experimentales y métodos estadísticos (hay que notar que gran parte de la investigación didáctica es experimental); e) diseño y evaluación del curriculum (encuadre interdisciplinar de las ciencias), y f) sociología de la educación.

Ciertamente, la posibilidad de encontrar un investigador con

---

(4) CRONBACH, L. J., y SUPPES, P.: *Research for Tomorrow's Schools*, National Academy of Education, London, MacMillan, 1969, 129-130.

(5) TYLER, R. W.: "Analysis of strengths and weakness in current research in science education", *Journal of Research in Science Teaching*, 5 (1967-68), 58.

tales características es escasa, lo que nos reafirma en la idea de que el futuro de esta investigación se encuentra en el trabajo en equipo, que agrupe a especialistas en las distintas áreas. Sin embargo, siempre será necesario que los especialistas en ciencias implicados posean una amplia base en psicología y metodología.

La cuestión *metodológica*, que ha venido a ser el gran caballo de batalla durante mucho tiempo, ofrece unas perspectivas esperanzadoras no sólo por el gran desarrollo de la estadística aplicada, diseños experimentales, análisis multivariante, etc., y de los procedimientos de análisis de datos que se está produciendo en los últimos años, sino también porque se han robustecido, a través del trabajo de buen número de investigadores dedicados a dicha tarea, las vías de absorción de dichos métodos y tecnología por los equipos que investigan en proyectos didácticos.

Por último, debemos señalar un claro aspecto diferenciante de la investigación didáctica con relación a otras más modélicas, como por ejemplo la investigación en ciencias físicas. En esta última, la abstracción no es a menudo problemática, viniendo dictada por el éxito de una práctica científica, bien por un experimento brillante, bien por un modelo hipotético-deductivo bien desarrollado. En definitiva, el físico investiga generalmente dentro de paradigmas establecidos y aceptados por el éxito anterior. Sin embargo, el investigador no cuenta siempre, por el momento, con esos paradigmas universalmente aceptados en los que fundamentar su investigación, viéndose generalmente obligado a crear sus propias abstracciones y proposiciones contrastables, lo que dificulta el proceso de globalización a partir de fenómenos concretos (6).

### **1.3. Los modelos en las ciencias y la investigación didáctica**

Un aspecto paralelo y de alguna forma entroncado con el problema de paradigmas apuntado en el apartado anterior es el del uso de modelos científicos bien establecidos y contrastados dentro de la investigación didáctica. La aplicación sistemática de tales modelos no solamente supone la vía más eficaz de desarrollo, sino que supondrá, seguramente, la superación gradual de defectos de rigor provenientes de la inflación de ideas moralistas, políticas y doctrinarias que se observa dentro de las teorías educati-

---

(6) COHEN, E. G.: *A New Approach to Applied Research: Race and Education*, Columbus, Ohio, Merrill, 1970, 14-15.

vas. Ideas que a menudo no tienen una clara fundamentación sociológica o psicológica, sino que se apoyan en modas sociales con un alto grado de transitoriedad.

Weinstock (7) aborda este tema, definiendo lo que entiende por modelos científicos, clasificándolos y discutiendo sucintamente su grado de aplicabilidad. El ve al *modelo científico* como algo fundamentalmente filosófico y siempre en términos de su función lógica en interpretación de una teoría, que pudiera definirse como un sistema deductivo del que siguen consecuencias lógicas al conjuntar hechos observados con el bloque de hipótesis fundamentales. Sin embargo, señala que, en general, establecer un modelo no es suficiente para deducir una ley experimental, ya que las nociones teóricas están coordinadas frecuentemente por reglas de correspondencia con varios conceptos experimentales.

En la actualidad, el nivel de aplicación a la investigación didáctica de los ocho tipos de modelos científicos que presenta Weinstock es muy diverso, en algunos por razones de índole técnica, esto es, dificultades provenientes de que su uso conlleva complicaciones metodológicas y matemáticas que no todo investigador puede resolver, y en otros, porque el propio carácter o estructura del modelo no ofrece un entronque claro con la naturaleza de los fenómenos comportamentales y educativos.

Los modelos *analógicos*, aunque van siendo cada vez más aplicados a medida que avanza la tecnología en que se sustentan, encuentran dificultades en lo relativo a reproducción fiel de fenómenos comportamentales, generalmente complejos y condicionados por muchas variables y factores intervinientes.

Ciertos enfoques que modelan procesos psico-sociales con relaciones entre identidades matemáticas, encuentran un gran apoyo en los modelos *matemáticos* en los que se sigue el proceso de identificación de variables, establecimiento de hipótesis y ecuaciones, resolución, extrapolación, etc.

Los modelos *estadísticos*, con aplicación específica al muestreo e inferencia, son y han sido uno de los pilares fundamentales para el desarrollo de las ciencias del comportamiento. Su fundamentación probabilística les hace criticables desde un enfoque determinista, pero este enfoque no parece desprenderse claramente de un análisis sereno de la realidad educativa.

A pesar de su incapacidad para descubrir relaciones de causa-

---

(7) WEINSTOCK, H. R.: "The concept of 'model' and educational research", *Journal of Research in Science Teaching*, 4 (1966), 45-51.

efecto, ya que dan igual peso a la concurrencia de situaciones  $A \rightarrow B$  y  $B \rightarrow A$ , los modelos de *correlación* son de uso habitual y de gran utilidad en el proceso de detección de problemas relevantes y predicción de probabilidades de respuesta en gran número de procesos y fenómenos educativos.

La aplicación de modelos *de escala*, procesos reales o imaginarios que preservan las proporciones relativas, presenta el problema, a pesar de su uso frecuente, de que el cambio de escala introduzca distorsión e irrelevancia.

Los modelos *intuitivos* juegan un papel similar al de los modelos de correlación, cuando se buscan parámetros y se realizan estimaciones referentes a nociones, conceptos o constructos de tipo intuitivo tales como inteligencia, habilidad, etc.

Los modelos de estimación *bayesianos* que tratan con el problema de encontrar, a priori, la mejor distribución de determinadas variables, van a jugar un papel decisivo en el futuro desarrollo de la investigación en ciencias del comportamiento. Su relativamente escasa aplicación se debe a que introducen un tratamiento matemático extenso y, a veces, complejo.

Los modelos *conceptuales*, capaces de ilustrar entramados de ideas más que de interpretar teorías, van a seguir jugando su tradicional papel de apoyo para describir procesos de naturaleza compleja que se dan a menudo en el terreno didáctico.

Con todos sus pros y contras, en la actualidad, y posiblemente en un futuro inmediato, el desarrollo de la investigación en didáctica se apoya principalmente en los modelos analógicos y en los tres de fundamentación probabilística, estadísticos, de correlación y bayesianos.

#### **1.4. El proceso de la investigación**

La investigación en didáctica de las ciencias presenta unas características modélicas y metodológicas similares a las de otras áreas educativas, aunque, debido a sus peculiaridades específicas, en los últimos años se haya podido ver un enfoque más marcado hacia el desarrollo curricular.

En este apartado vamos a analizar el proceso de investigar bajo tres puntos de vista: a) *enfoques* de partida, b) *tipos* de investigación, y c) *métodos* empleados para llevar a cabo la investigación. La conjugación adecuada de los diversos componentes que aparecen en estos tres aspectos nos ofrece una clasificación

global capaz de representar a la gran mayoría de trabajos de investigación en el terreno didáctico.

### 1.4.1. Enfoques

Los estilos o enfoques aplicados a la investigación didáctica tienen su fundamento en la propia forma de entender el proceso de enseñanza y aprendizaje.

Atkin (8) nos habla de la predominancia de dos de estos enfoques, el comportamental o psicológico y el de ingeniería.

El enfoque *comportamental* está basado en los modelos de investigación de las ciencias del comportamiento, principalmente en la psicología, de forma que sus técnicas y puntos de vista teóricos no parten del concepto globalizado de educación, sino que lo hacen de las ciencias en que se fundamenta: psicología, sociología, antropología, etc. En definitiva, se parte de la idea de que el proceso educativo se compone de procesos sociológicos, psicológicos, etc., y de que aquél puede llegar a ser entendido y explicado a partir del estudio aislado de éstos.

Este enfoque es mucho más válido para la sociología y psicología, la clase es un excelente medio para investigar en estos campos, que para estudiar el proceso didáctico en su conjunto, ya que en éste se dan ciertas interacciones que no se detectan en un estudio fragmentado. A la postre, parece necesario un enfoque más global, es decir, más educativo.

El enfoque de *ingeniería*, utilizado últimamente con profusión, tiene un carácter tecnológico y econométrico en razón de su propio origen. Su fundamento está en identificar primero los objetivos a los que se dirige el sistema, y segundo, en diseñar el sistema capaz de alcanzarlos. Este enfoque, que es totalmente compatible con el comportamental, aunque presenta atributos muy diferentes, engloba al análisis de sistemas, análisis operativo, PPBS, análisis coste-beneficio, etc.

Este segundo enfoque se encara con ciertos problemas relevantes, especialmente cuando se trata de definir los objetivos del proceso educativo. Resulta complejo avanzar más allá del campo cognoscitivo a corto plazo, porque inmediatamente se topa con distintos sistemas de valores y productos sociales en competencia, por otra parte difíciles de cuantificar. Ante esta profunda proble-

---

(8) ATKIN, J. Myron: "Research styles in science education", *Journal of Research in Science Teaching*, 5 (1967-68), 338-345.

mática, los sistemas operativos pierden gran parte de su validez general.

A pesar de las deficiencias apuntadas en los dos bloques tratados, no debe olvidarse que la gran mayoría de los logros didácticos han provenido de trabajos con uno u otro estilo y que no es aconsejable su sustitución mientras no se cuente con vías alternativas más eficaces.

Ante esta problemática, el propio Atkin sugiere la introducción de un nuevo enfoque que denomina *fenomenológico*, basado en la *teoría educativa* y en el *contexto clase*. Añade que, un buen principio sería el introducir durante períodos largos en distintas clases a especialistas provenientes de los más diversos campos, pedagogía, sociología, psicología, filosofía, ciencias, derecho, economía, etc., a fin de que intentaran buscar entronques y relaciones entre sus campos para describir las actividades de la clase y ofrecer explicaciones a los procesos que se producen. En suma, sustituir el enfoque fragmentario por uno conjunto, aunque se presente la dificultad de aunar la labor de especialistas en distintos campos orientados a estudiar el proceso didáctico de forma integrada.

Estas ideas nos recuerdan la necesidad de que en didáctica, al igual que se hace en otras ciencias, se investigue con una micro-perspectiva y una macro-perspectiva. Algunos físicos investigan la estructura de los núcleos atómicos y otros la del universo, algunos biólogos trabajan a nivel molecular y otros en problemas de ecología y medio ambiente. En didáctica el micro-nivel es y seguirá siendo abordado por psicólogos, sociólogos, antropólogos, etc., y el macro-nivel debe serlo por los "educadores" o equipos que integren distintos especialistas.

Aunque no puede afirmarse que el enfoque fenomenológico u otros similares hayan venido a cubrir el campo de investigación inabordable por los trabajos comportamentales o de ingeniería, sí puede decirse que en los últimos años se está conduciendo la investigación didáctica de las ciencias, dentro de un contexto más amplio. Este hecho se refleja en las opiniones esperanzadas emitidas últimamente por los críticos de la investigación, en contraste con posiciones de cierto desaliento observadas años atrás (9).

---

(9) (TYLER, R. W.: "Research in science teaching in a larger context", *Journal of Research in Science Teaching*, 11 (1974), 138.



### 1.4.2. Tipos de investigación

En cualquier disciplina y campo del saber aparecen tres tipos de trabajos de investigación, los teóricos, los empíricos y los aplicados. A veces se denomina a la investigación teórica como fundamental, pero el añadido no nos parece acertado porque vemos un sentido más amplio en la denominación fundamental que lo meramente teórico. Ciertamente, la gran mayoría de la investigación teórica es fundamental, pero muchos de los trabajos empíricos lo son igualmente. Del mismo modo, usamos la denominación "empírica" para el segundo tipo en vez de experimental, por ser un término más amplio y englobante.

En didáctica de las ciencias aparecen estos tres tipos de trabajos, pero que nosotros vamos a presentar como cuatro, siguiendo a Jacobson (10). Estos cuatro tipos son: a) *filosóficos o teóricos*, b) *empíricos*, c) *de decisión* y d) *de desarrollo* (estudios formativos). Esta clasificación supone el desdoblamiento de la investigación típicamente aplicada en dos tipos que presentan caracteres diferentes en el terreno didáctico.

a) Los estudios *filosóficos* se centran en el análisis de presupuestos básicos, de la naturaleza de las situaciones problemáticas y de las posibles consecuencias de las acciones propuestas, todo ello con base a las recomendaciones de un proceso mental sistemático. En el caso de la didáctica de las ciencias no son este tipo de estudios los más abundantes, sin embargo, son la vía adecuada para buscar estructuras didácticas de las disciplinas científicas que sean eficaces en los distintos niveles de la enseñanza, teniendo presente que se trata con individuos y grupos que forman conjuntos que son, generalmente, algo más que la suma de sus componentes.

b) Los trabajos *empíricos* suponen la recolección de datos de comportamiento de estudiantes, profesores, grupos de clase, centros, etc., con vistas a, tras el consiguiente análisis, aceptar o rechazar ciertas hipótesis establecidas. A veces el estudio se centra en el descubrimiento de hechos o relaciones entre variables sin perseguir directamente las relaciones causa-efecto (sondeos o estudios de correlación), y otras se manipulan las variables en busca de la causalidad (estudios puramente experimentales). En este caso, como apunta Von Weizsäcker (11), a diferencia de la

---

(10) JACOBSON, W. J.: "Approaches to science education research: Analysis and criticism", *Journal of Research in Science Teaching*, 7 (1970), 217-225.

(11) VON WEIZSACKER, D. F.: *La Importancia de la Ciencia*, Barcelona, Ed. Labor, 1968, 95.

experiencia cotidiana, "con un experimento controlado se transforma la naturaleza dada en realidad manejable y con el apoyo de una teoría matemática se interpreta la respuesta a una pregunta planteada".

Contrariamente al caso de los estudios filosóficos, los trabajos de investigación empíricos en didáctica de las ciencias han sido muy frecuentes. Sin embargo, los frutos no siempre han sido abundantes ni, por supuesto, en consonancia con los esfuerzos realizados. La principal causa de este hecho podría estar en la desmedida tendencia a comparar métodos didácticos aislados cuando quizá fuera necesario buscar su complementariedad. El uso del laboratorio puede ser el ejemplo más esclarecedor. Entendemos que no tiene mayor sentido dedicar esfuerzos, tal y como se ha hecho muy frecuentemente, a comparar la eficacia de las clases basadas en el laboratorio con las clases expositivas, cuando lo razonable sería buscar cómo deben conjuntarse entre ellas y con las clases de problemas, lecturas, proyecciones, etc.

Asimismo, la tendencia a establecer en los trabajos experimentales el nivel de significación estadística de una forma mecánica a  $\alpha=0,05$ , ha supuesto posiblemente la paralización de vías de trabajo que pudieran haber sido prometedoras. Este nivel de riesgo, aceptable para estudios de drogas, resistencia de materiales, etc., pudiera no ser necesariamente lógico para el estudio de innovaciones didácticas. Hay que tener presente que en el proceso educativo intervienen muchas variables condicionando y modelando el comportamiento del discente durante períodos largos de su vida. Por ello, es difícil encontrar factores aislados que al manipularlos por períodos relativamente cortos, la duración de un tratamiento experimental suele variar entre varas semanas y un curso académico, van a afectar al comportamiento de los alumnos o sujetos base del estudio de forma significativa a un nivel de riesgo muy bajo.

Una mayor flexibilidad en los niveles de riesgo aumentaría la capacidad y audacia de los experimentadores didácticos. A veces los hallazgos experimentales se producen por casualidad y parece necesario incrementar, sin perder el rigor excesivamente, la probabilidad de que la casualidad del hallazgo se produzca. La historia de la física nos ofrece un ejemplo estimulante en este sentido en la persona de Oersted, que buscó la casualidad del hallazgo hasta en su propia actividad didáctica. Es así como, en una experiencia ilustrativa ante la clase con corrientes e imanes, observó por primera vez ciertas interacciones que supusieron el

punto de arranque del área de la física conocida como electromagnetismo (12).

En el terreno didáctico, el contexto es tan variado y sorprendente que se debe estar preparado para detectar algún fenómeno insospechado que se produzca.

c) Los estudios *de decisión* tienen por objetivo prever las consecuencias de acciones didácticas futuras, tomando como base resultados de experiencias e investigaciones didácticas ya realizadas.

Es evidente que todo trabajo didáctico va a fundamentarse en trabajos de otro tipo, pero en el caso de los estudios de decisión es imprescindible el apoyo de análisis filosóficos, empíricos y estudios de desarrollo que muestren posibilidades y consecuencias de aplicación a corto, medio y largo plazo.

Los estudios históricos y de educación comparada se encuadran perfectamente en este tipo de trabajos, siempre que se planteen con un carácter analítico objetivo que proporcione elementos para un futuro desarrollo. Lo mismo puede decirse de los estudios predictivos enfocados a abrir caminos y escenarios didácticos, bien para establecerlos, bien para recibirlos con capacidad de adaptación.

d) Los estudios *de desarrollo* se centran en la preparación y evaluación de nuevos materiales didácticos, enfoques, procedimientos, programas, etc., con el fin de mejorar y facilitar el rendimiento escolar.

Todo estudio de desarrollo sigue, generalmente, un proceso que podríamos sintetizar en los siguientes puntos:

- Exploración de ideas en la literatura y experiencias anteriores.
- Planificación preliminar.
- Desarrollo de material, programas, técnicas, etc.
- Aplicación a pequeña escala.
- Revisión del material, técnicas, etc., para aplicación a gran escala.
- Aplicación a gran escala.
- Evaluación sumativa de resultados.
- Preparación del programa, técnicas, etc., para edición masiva.
- Análisis general de la experiencia.

---

(12) Véase GAMOW, G.: *Biography of Physics*, New York, Harper-Torchbooks, 1964, 135-136.

La realización de estos estudios entraña comúnmente costos elevados y muchos recursos humanos, sin embargo, los intentos a escala reducida requieren pocos medios en ambos sentidos y por otra parte son tremendamente valiosos como elemento de perfeccionamiento y autocritica del profesorado que los lleva a cabo. Por otra parte, si estos estudios a pequeña escala se encuadran en un marco filosófico común, pueden suponer el fundamento de programas de gran alcance.

Los trabajos de estructuración de programas de formación de profesorado deben encuadrarse en general en este tipo, porque suponen un proceso similar al señalado.

### 1.4.3. Métodos

Hasta mitades del presente siglo la investigación educativa ha estado de alguna forma asimilada a la investigación psicológica, lo que supuso que enfoque y métodos psicológicos fueran preponderantes casi de forma exclusiva. Sin embargo, el desarrollo de la investigación psicológica con la aparición del conductismo y gestaltismo, así como la aparición del fenómeno curricular y de enfoques más amplios en la investigación educativa, fueron la base para el establecimiento definitivo de la investigación educativa como campo distinto de la psicología, aunque en ciertos aspectos cercano y generalmente necesitado de su apoyo.

El problema metodológico ha sido grave durante mucho tiempo. Tanto es así que se creó el llamado "mito de la metodología", basado en la hipótesis de que "las dificultades más serias con las que se enfrenta la investigación en ciencias del comportamiento son metodológicas y que solamente cuando tengamos la metodología sólida, el progreso será rápido y seguro". Bridgham (13) comenta este problema y dice que la metodología es de suma importancia pero no el único requisito, además, el "mito" tiende a un desvanecimiento paulatino con el progreso de los últimos años. Este autor opina que "el primer requerimiento para una línea productiva de investigación en didáctica de las ciencias es una idea o conjunto de ideas que puedan dar cuenta de los fenómenos educativos significativos, que el buen método es aquel que nos ayuda a clarificar la conexión entre la idea y los fenómenos sin distorsionarlos indebidamente, y que, por lo tanto, habrá tantos bue-

---

(13) BRIDGHAM, R. G.: "Methods in research in science education". *Journal of Research in Science Teaching*, 11 (1974), 169.

nos métodos como grupos relevantes de fenómenos educativos e ideas valiosas para usar en investigación”.

La metodología se entronca con los enfoques y tipos de investigación, pero se diferencia en que mientras éstos tienen una relación inmediata con la base filosófica o fines perseguidos, aquella se centra en los procedimientos utilizados. Sin embargo, la separación clara resulta a veces dificultosa, porque el método puede venir delimitado por uno de los otros aspectos. Así, un *enfoque molecular* como el comportamental lleva casi irremediablemente al *método molecular*.

Sin entrar en discusiones detalladas sobre técnicas concretas, la investigación didáctica nos ofrece cinco métodos globales como los más usados comúnmente.

a) El método de *investigación en acción* es un procedimiento probatorio similar al usado por el inventor de aparatos. Como en este caso, el estudio científico crítico se sustituye por el “tanteo” hasta conseguir un resultado apetecido, y las soluciones a las que se llega son arbitrarias, dificultándose el proceso de generalización. La línea seguida viene a responder a un enfoque de tipo fenomenológico, pero con la grave simplificación de su falta de apoyo en bases teóricas.

El procedimiento estuvo en boga entre los años cincuenta y sesenta como consecuencia de la frustración existente por los escasos resultados de la investigación didáctica que hemos apuntado con anterioridad. Un buen número de profesores, sobre todo de ciencias y matemáticas, decidieron elevar el rendimiento de su enseñanza probando métodos y técnicas hasta elegir las más productivas.

b) El método *naturalista* es un refinamiento del anterior, ya que la investigación en acción se restringe a aquellas variables que aparecen como más críticas en la clase, a medida que se analiza la situación problemática. El procedimiento es de reciente aplicación y resulta prematuro analizarlo críticamente, sin embargo, cuenta con defensores de prestigio como Piaget, Bruner, Gagné, Suppes y Watson (14).

c) El método *molecular* supone el tratamiento separado de las distintas variables que influyen en el proceso didáctico, respondiendo perfectamente a un enfoque como el que hemos denominado comportamental o psicológico. La idea subyacente es la

---

(14) SCANDURA, J. M. y ANDERSON, J.: “Educational research and the science educator”, *Science Education*, 52 (1968), 357.

de construir una teoría del aprendizaje estudiando separadamente el efecto de variables como refuerzo, motivación, inteligencia, capacidad de expresión, capacidad de socialización, etc., en vez de hacerlo a partir del propio proceso en su conjunto. Es decir, algo similar al método seguido por la teoría cinética para explicar la presión de un gas dentro de un recipiente, en vez de usar la ley de los gases  $PV=KT$ .

La mayor parte de los trabajos que siguen este método se incluirían en los de tipo empírico o empírico-experimental, apoyándose generalmente en técnicas estadísticas univariadas.

d) El método *multi-molecular* es una extensión del anterior que permite el estudio conjunto de los efectos de varias variables y sus interacciones. En términos físico-químicos podríamos decir que hemos pasado de las moléculas más simples a los compuestos complejos.

El procedimiento es más poderoso que el molecular y se ajusta mejor a la idea de que la educación es un fenómeno complejo en el que intervienen interaccionando un elevado número de variables y factores, pudiendo suceder que algunas de las interacciones sean más significativas que los propios efectos principales aislados.

El método implica naturalmente algunas complicaciones técnicas, generalmente estadísticas, pues supone tratamientos multivariados de cierta complejidad. Esta problemática apoya abiertamente el acercamiento de especialistas en estadística a la investigación didáctica.

Otro problema a resaltar, dentro de este procedimiento, es que la introducción de muchas variables puede oscurecer la interpretación de resultados y dificultar la generalización teórica si dichas variables son irrelevantes. Para evitar este hecho, se deben detectar previamente las variables más cruciales dentro de un enfoque acumulativo y no fragmentario.

e) El método *del replanteamiento*, recientemente sugerido por Bridgham (15), toma el estudio de casos como vía de desarrollo, de forma que el énfasis está más en la selección que en la generación de teorías.

El procedimiento parte de la construcción de una "teoría sobre caso concreto" por varios investigadores independientemente. Posteriormente, cada uno de ellos presenta al grupo su teoría par-

---

(15) BRIDGHAM, R. G.: "Methods in research in science education" (o. c.), 169-174.

particular, seleccionándose la o las más convincentes, que se muestran a su vez a un grupo más amplio de especialistas. A cada uno de estos últimos se les pide: a) decidir qué teoría de las presentadas es satisfactoria, o b) proponer una alternativa más satisfactoria, o c) indicar si no encuentra una teoría satisfactoria. El proceso puede repetirse hasta encontrar un acuerdo en un sentido u otro.

Este método se nos presenta como muy laborioso y de difícil extensión porque supone un ingente trabajo de coordinación y la colaboración de un elevado número de especialistas. Sin embargo, lo presentamos como algo sugestivo y que pudiera ser eficaz en el futuro.

## 1.5. Areas de investigación

En términos generales, temas relevantes para el investigador en didáctica de las ciencias son los relacionados con la materia a enseñar, su estructura y secuenciación, los objetivos que hacen necesaria y justifican la formación científica, los métodos didácticos y su sustentación teórica, el profesorado, su formación y actuación en el proceso didáctico y el alumnado, sus actitudes, aptitudes y rendimientos.

Nosotros vamos a parcelar todo este terreno de investigación en didáctica de las ciencias en siete grandes áreas, apoyándonos en el excelente análisis sobre el tema de Shulman y Tamir (16).

### A. Estructura disciplinar

Area que ha aglutinado el esfuerzo de muchos investigadores en los últimos años, particularmente en lo relativo a contenidos, organización y orientación de los programas. Los resultados han supuesto el paso paulatino de cursos centrados en el conocimiento de hechos, leyes, teorías científicas y aplicaciones tecnológicas, a cursos que ponen el énfasis en la naturaleza, estructura y unidad de la ciencia y en el proceso de "indagación" científica.

Esta nueva orientación se ha apoyado en las ideas de algunos famosos pensadores en este campo. Bruner (17) aboga por la en-

---

(16) SHULMAN, L. S. y TAMIR, P.: "Research on teaching in the natural sciences", en *Second Handbook of Research on Teaching*, Travers, A. (ed.), Chicago, Rand McNally, 1973.

(17) BRUNER, J. S.: *The Process of Education*, Cambridge, Mass., Harvard University Press, 1960.

señanza de la estructura disciplinar como defensa contra la excesiva información y como garantía de relevancia, apoyándose en la idea, más psicológica que filosófica, de que las generalizaciones son retenidas y transferidas más fácilmente que los hechos. Schwab (18) llega más lejos diciendo que la estructura no solamente precede a los hechos sino que determina cuáles serán considerados como tales. Por último, Gagné (19) y Ausubel (20) propugnan, apoyados en enfoques típicamente psicológicos, que la organización del material de instrucción debe ser aquella que más facilite su integración con la estructura de conocimiento organizada del discente, aquél a través de *jerarquías de aprendizaje* y éste alrededor del concepto de *estructuras cognoscitivas*.

A nuestro entender, estas visiones y enfoques resaltan una cuestión crucial para los educadores en ciencias, esto es, cómo transmitir una concepción particular o estructura del conocimiento científico a los estudiantes, de forma que se convierta en componente permanente de su propia estructura cognoscitiva.

#### B. *Objetivos de la educación científica*

Los objetivos educativos han variado y varían a lo largo de los tiempos a medida que se presentan nuevas posibilidades y nuevas demandas sociales. Sin embargo, como apunta Tyler (21), es posible que los nuevos objetivos se acepten sin considerar críticamente sus ventajas y solidez frente a las alternativas que se rechazan.

Las importantes connotaciones sociales envueltas en los distintos aspectos de la ciencia y su enseñanza dan relevancia al tema de los objetivos de la educación científica como área de investigación. Sin embargo, dichas connotaciones son tan variadas que no puede pensarse en que sea tarea exclusiva de los científicos su tratamiento. Estos deben ser apoyados por expertos en otros campos del conocimiento.

---

(18) SCHWAB, J. J.: "The teaching of science as enquiry", en *The Teaching of Science*, Schwab, J. J., y Brandwein, P. F., Cambridge, Mass., Harvard University Press, 1962.

(19) GAGNE, R. L.: *The Conditions of Learning*, New York, Holt, Rinehart & Winston, 1965.

(20) AUSUBEL, D. P.: *Educational Psychology: A Cognitive View*, New York, Holt, Rinehart & Winston, 1968.

(21) TYLER, R. W.: "Resources, models and theory in the improvement of research in science education", *Journal of Research in Science Teaching*, 5 (1967-68), 44.



Shulman y Tamir (22) señalan tres niveles de discusión dentro de los objetivos de la enseñanza científica: a) normativo, b) cognoscitivo y c) técnico.

El nivel *normativo* se centra en el “qué aprender”, surgiendo cuestiones profundas como ¿se debe aprender a ser científico? o ¿se debe aprender a entender la ciencia? Inclinarsé abiertamente por una u otra dirección podría ser peligroso, porque posiblemente la dirección apropiada varía de un nivel educativo a otro.

El nivel *cognoscitivo* se centra en el “a través de qué”, esto es, en la estructura global del proceso de enseñar la ciencia. Tiene una relación inmediata tanto con las “teorías psicológico-evolutivas” como con la “estructura disciplinar”.

El nivel *técnico* viene a ser la operativización de los dos niveles anteriores en el “cómo” y en el “qué hay que conseguir en cada momento”. Su relación es directa con los aspectos formales del proceso de definición de objetivos de la instrucción. Mientras Gagné (23) y Mager (24) se apoyan en la psicología comportamental para proponer que los objetivos deben ser claramente expresados en términos de conducta cuantificable con mención de respuesta, condiciones, contenido y criterio, Ebel (25) y Eisner (26) proponen unos objetivos menos rígidos en su formalización, objetivos expresivos o de expresión, no necesariamente basados en una conducta medible del alumno, apoyados en la idea de que el fin último del proceso educativo no es cuantificable de forma objetiva, ni debe circunscribirse a unas condiciones prefijadas.

Dentro de este área es reseñable el problema de la acomodación de objetivos a los distintos niveles de la educación científica. La abstracción requerida en cada escalón es un punto crítico para la secuenciación lógica de todo proyecto curricular.

### C. Métodos pedagógicos

El desarrollo de la metodología didáctica ha estado íntimamente ligado con el de las teorías psicológicas del aprendizaje,

---

(22) SHULMAN, L. S. y TAMIR, P.: “Research on teaching...” (q. c.).

(23) GAGNE, R. L.: *The Conditions...* (o. c.).

(24) MAGER, D. F.: *Preparing Instructional Objectives*, Palo Alto, Cal., Fearon Publishers, 1962.

(25) EBEL, R. L.: *Measuring Educational Achievement*, Englewood Cliffs, N. J., Prentice Hall, 1965.

(26) EISNER, E. W.: “Instructional and expressive educational objectives. Their formulation and use in curriculum”, en *Instructional Objectives*, Popham, W. J.; Eisner, E. W.; Sullivan, H. J., y Tyler, L. L., AERA Monograph Series on Curriculum Evaluation, n.º 3, Chicago, Rand Mac Nally, 1969.

al tiempo que los programas se apoyaban cada vez más en los conceptos de "interrogación o indagación" y "descubrimiento". Los tres enfoques metodológicos más tratados en la práctica han sido:

a) El *tutorial*, es el que el profesor reduce su papel de informador y aumenta el de guía del alumno para resolver problemas que le impiden seguir adelante.

b) El de *secuencias de instrucción* inductivas y deductivas establecidas previamente como variables manipulables por el profesor.

c) El de Suchman (27), en el que el alumno es expuesto ante un fenómeno concreto por medio de un film, experiencia, etc., y posteriormente formula preguntas al profesor que se limita a contestarlas. Este método es similar al socrático aunque mucho menos tutorial.

Estos tres métodos no se nos presentan como excluyentes, sino más bien como complementarios, sobre todo los dos primeros, en los que el profesor juega un papel más activo en la dirección del proceso de aprendizaje. En cuanto al tercero, la iniciativa corresponde fundamentalmente al alumno, tomando el profesor un papel casi exclusivamente receptivo y de moderador.

#### D. Estudios didácticos comparativos

Los distintos enfoques psicológicos y metodológicos abren el camino a ciertos estudios empíricos, conducentes a demostrar la eficacia o ineficacia relativa entre distintos métodos y técnicas didácticas.

El laboratorio, que es el elemento más distintivo de la educación científica, ha solido jugar un papel preponderante en los trabajos de este área. Con los nuevos enfoques centrados en la estructura de la ciencia y en el proceso científico más que en el aprendizaje de teorías, hechos y técnicas, dicho papel adquiere más relevancia si cabe, pasando de ser un soporte para demostraciones y confirmaciones a ser el auténtico centro del proceso de enseñanza de las ciencias.

Ya hemos apuntado la relativa ineficacia en cuanto a resultados de los estudios centrados en la mera comparación de métodos y procedimientos didácticos, por la razón sencilla de dificultar

---

(27) SUCHMAN, J. R.: *The Elementary School Training Program in Scientific Enquiry*, Project n.º 216, National Defense Education Act., University of Illinois, 1962.

tades de generalización una vez que nos movemos a otros contextos, niveles, profesores y estudiantes. Máxime cuando pueden aparecer interacciones significativas entre las características individuales de los estudiantes y profesores y los métodos instructivos. Así, por ejemplo, las actitudes o aptitudes de los estudiantes pueden modular la eficacia de un procedimiento didáctico. El buscar hasta qué punto esto es así, es un tema abierto a los investigadores en el que Cronbach y Snow vienen haciendo hincapié hace años y sobre el que han publicado recientemente un excelente tratado (28).

Con relación a los trabajos en este área, todo lo apuntado nos inclina a tomar la postura de apoyar los enfoques de complementación y coordinación de métodos y abandonar los meramente comparativos de escasa operatividad práctica.

### E. Desarrollo y evaluación del currículum

Gran parte de los trabajos en las cuatro áreas anteriores se plantean con la idea de encontrar bases sólidas para la elaboración del currículum de formación en ciencias. Esto es perfectamente lógico, porque el verdadero interés del profesor que tiene que enseñar ciencia está en tener las ideas claras sobre lo que tiene que enseñar, como puede y debe hacerlo, con qué secuenciación, por qué, para qué, cómo evaluar, etc., y esto es en definitiva el *currículum*. En esta perspectiva, no cabe duda de que el currículum es el auténtico y definitivo caballo de batalla del enseñante, pero el diseño curricular, si se quiere que sea eficaz, tiene unas connotaciones tan amplias y profundas que da por sí solo relevancia a la investigación en todas las restantes áreas que señalamos.

En los últimos años, el desarrollo y evaluación del currículum ha llevado el peso fundamental de la investigación en didáctica de las ciencias, pero aunque se han diseñado algunos proyectos valiosos, el impacto en la enseñanza escolar ha sido débil. La causa habría que verla seguramente en las dificultades de centros y profesores, con una determinada estructura y actitudes, para adaptarse o adaptar un currículum que posiblemente requiera ciertos cambios estructurales y actitudinales profundos.

El análisis y desarrollo de sistemas y técnicas de evaluación del rendimiento de los alumnos, así como el de otros estamentos

---

(28) CRONBACH, L. J. y SNOW, R. E.: *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*, New York, Irvington Publishers, 1977.

del sistema educativo, entra de lleno en la investigación curricular ya que ésta implica procesos globales de evaluación tanto formativa como sumativa.

#### F. Formación del profesorado de ciencias

El buscar métodos y procedimientos que proporcionen capacidades didáctico-profesionales, teóricas y prácticas, a los futuros profesores de ciencias de todos los niveles, como complemento lógico a su formación académica, es un área de investigación cuya importancia no es necesario resaltar. Antes hemos visto cómo el desarrollo curricular venía a ser un objetivo central para todo investigador en didáctica de las ciencias. Pues bien, una vez que contamos con un buen "proyecto", lo que necesitamos son profesores con capacidad y actitud para introducirlo en la práctica escolar.

Esta formación didáctica del profesorado plantea una serie de problemas y debe cubrir una serie de aspectos que atañen e implican en el tema tanto a los departamentos científicos universitarios, responsables directos de la formación académica, como a los servicios pedagógicos de apoyo, los Institutos de Ciencias de la Educación en el caso español. El estructurar debidamente la formación del profesorado no podrá hacerse, seguramente, si no se hace bajo planteamientos de estrecha colaboración entre estos estamentos implicados (29).

Dentro de este tema, las dos líneas más diferenciadas son aquellas consecuencia de enfoques conductistas y humanistas (30). En la primera, la preocupación principal es el comportamiento que el profesor debe observar, y en la segunda, su desarrollo y maduración profesional. También aquí, al igual que en lo relativo a enfoques de la investigación didáctica, es previsible que la conjunción de planteamientos ofrezca nuevas y eficaces vías de formación del profesorado.

#### G. Desarrollo científico del estudiante

La programación didáctica de cualquier disciplina científica

---

(29) Para un tratamiento más amplio de este tema consultar: ESCUDERO, T. y FERNANDEZ URÍA, E.: "La formación del profesorado de ciencias, un análisis de la problemática", *Bordón*, 214 (1976), 301-315.

(30) Un análisis de ambos enfoques aparece en: ESCUDERO, T. y FERNANDEZ URÍA, E.: "Enfoques conductistas y humanistas en la formación práctica del profesorado", *Educadores*, 90 (1976), 641-656.

debe ser acorde con las posibilidades de asimilación y comprensión de los estudiantes a los que va a ir dirigida. El diseño de unidades didácticas comienza con el análisis del nivel de conocimiento y capacidades cognoscitivas de los estudiantes a los que se va a enseñar. Por lo tanto, paralelamente a los estudios sobre estructura disciplinar para enseñar ciencia con el máximo de eficacia, se necesitan pautas claras y orientaciones sobre los niveles de desarrollo mental de los docentes en cada uno de los estadios de la educación. En otras palabras, cuál es el grado de abstracción científica que es capaz de absorber la mente de un alumno de un determinado nivel.

La investigación en este área entronca la epistemología genética con las estructuras científicas, en un proceso de búsqueda de momentos adecuados para exponer al estudiante ante nuevos hechos, generalizaciones y procesos científicos, para que los asimile y pasen a formar parte de su estructura mental en desarrollo.

La didáctica de la ciencia se encuentra en este punto con dos temas de investigación de suma importancia y que deben abordarse en coordinación, el papel de la ciencia como vehículo para el desarrollo mental del individuo y el desarrollo mental como condición para comprender las distintas etapas, técnicas, generalizaciones y aspectos de la ciencia.

## **1.6. Mirando al futuro**

La panorámica que acabamos de ofrecer nos muestra a una investigación que parece haber superado casi definitivamente el proceso de puesta en marcha, encontrándose en una clara situación de consolidación y aproximación al rigor científico que hace concebir el futuro de la misma con alto grado de confianza. Sin embargo, no se nos escapa que todavía son muchos y variados los problemas que quedan por resolver y el camino a andar, hasta que la investigación en didáctica de las ciencias adquiera el nivel paradigmático que tiene la que se realiza en otros campos, como la física por ejemplo.

Aunque muchos de estos problemas y posibles vías de abordaje de los mismos ya han sido señalados a lo largo de nuestro análisis, parece conveniente ofrecer una síntesis en la que resaltemos alguno de los aspectos que a nuestro entender son más relevantes con vistas al futuro.

a) La investigación didáctica ha mostrado cierta tendencia a dicotimizarse en aspectos teóricos, relacionados con las teorías

de aprendizaje o, en aspectos prácticos, relativos a problemas concretos del profesor en el aula, descuidándose un poco la conjunción de ambas perspectivas. Es evidente que el auténtico desarrollo se facilita con el tratamiento de problemas que tienen un entronque doble, teórico y práctico, y que es esta línea de trabajo la que deben perseguir la mayoría de los investigadores.

b) El desarrollo de paradigmas en la investigación didáctica, problema de gran trascendencia, debe apoyarse en el uso y aplicación más frecuente de modelos científicos bien establecidos.

c) La complejidad del proceso educativo dificulta su comprensión y análisis a partir de enfoques unilaterales. Se necesitan enfoques amplios y fenomenológicos capaces de entender el proceso de enseñanza y aprendizaje desde muy distintas perspectivas en coordinación. La figura del investigador en didáctica aislado debe dejar paso al equipo aglutinador de especialistas en distintas disciplinas.

d) La experimentación didáctica en el aula debe ser apoyada y generalizada a todos los niveles, aun a riesgo de no obtener siempre los resultados deseados, porque es un medio claro de perfeccionamiento continuo del profesorado y una fuente potencial inagotable de ideas y problemas para el investigador. En este contexto, los niveles de riesgo estadístico deben ser tratados con flexibilidad y conocimiento de causa para que, sin perder rigor en los resultados, no se cierren vías de trabajo que con tratamientos más adecuados puedan ser productivas.

e) Los trabajos en todas las áreas de investigación deben ser coordinados convenientemente, porque no puede pensarse en un desarrollo didáctico eficaz si el de las distintas áreas no está debidamente compensado. Buenos profesores no pueden llevar a cabo una labor adecuada si no cuentan con buenos proyectos curriculares y viceversa, pero buenos proyectos curriculares no pueden ser desarrollados si no existen buenos modelos básicos de enseñanza-aprendizaje y de estructura disciplinar.

### **1.7. Consideraciones sobre la investigación en didáctica de las ciencias en España**

Hasta este momento hemos estudiado en un plano general la problemática, presente y futura, de la investigación en didáctica de las ciencias. Sin embargo, nuestra vinculación profesional a este campo nos impulsa a concluir este análisis, tratando someramente ciertos problemas de infraestructura peculiares que encon-

tramos en nuestro país, para que la enseñanza de las ciencias se convierta realmente en "campo abierto a la investigación".

Nuestra discusión no vamos a centrarla en los problemas de infraestructura de la investigación en España, planificación y escasez de recursos (que evidentemente afectan a todo tipo de investigación), sino que nos remitiremos a aquello más característico del tema que nos ocupa, desligándolo asimismo en la manera de lo posible, de la investigación en ciencias de la educación y del comportamiento que se realiza en las secciones de sociología, psicología y pedagogía universitarias, aunque seamos conscientes de que dichos trabajos, en general circunscritos a problemáticas más globales, sean un gran apoyo para la didáctica específica de las ciencias.

En este campo concreto, vemos directamente implicados a tres estamentos del sistema educativo español: a) los centros universitarios de formación de futuros profesores de ciencias, Facultades de Ciencias y Escuelas Universitarias de E.G.B.; b) los Institutos de Ciencias de la Educación, y c) los centros de enseñanza básica y secundaria.

Las **Facultades de Ciencias Universitarias** deben ser el primer y uno de los fundamentales puntos de apoyo de la investigación en didáctica de las ciencias. Máxime cuando se trate de proyectos centrados en estructura disciplinar, objetivos de la educación científica, currículum y formación de profesores. Sin embargo, tenemos fundadas sospechas de que en dichas facultades, o no se considera este tipo de investigación como responsabilidad propia, o bien se considera que su relevancia e importancia es de segundo orden. Por nuestra parte, es evidente que no podemos pensar así, sobre todo si pensamos que nuestra ciencia en los años venideros, y esto sí que es responsabilidad no cuestionada de las facultades científicas, será tanto mejor cuanto mejor sean las formas y objetivos de su enseñanza en la actualidad.

Esta frecuente actitud de indeferencia ante la didáctica científica, dificulta sobremanera la realización de tesis doctorales en el tema, algo que todavía se agrava más por la anacrónica parcelación del grado de doctor en nuestro país según especialidades. Este hecho supone, de facto, una pantalla para muchos trabajos interdisciplinarios, y ciertamente, la investigación en didáctica, aunque sea didáctica de las ciencias, tiene mucho de interdisciplinaria.

Las **Escuelas Universitarias de Formación del Profesorado de E.G.B.** tienen ante esta investigación una problemática específica

muy distinta, y es que se encuentran en una marcada situación de inferioridad de recursos de todo tipo. En contradicción con el enunciado de su rutilante nombre, su contexto actual no responde a un centro en el que la investigación didáctica, si se quiere aplicada al nivel educativo básico, sea una parte importante de su quehacer, antes bien, razones internas y externas de muy diverso tipo las han relegado casi totalmente a la labor formativa.

Esta situación es de cierta gravedad porque resulta difícil pensar en una buena formación de profesores si dicha formación no se apoya en trabajos paralelos de investigación didáctica. Un centro de formación de profesores con estos planteamientos se nos presenta tan anómalo como un centro de formación de científicos sin investigación en ciencia.

Ocho años después de su puesta en marcha, los **Institutos de Ciencias de la Educación** siguen teniendo ciertos problemas graves que dificultan su labor, consistente en el cumplimiento de una de las funciones previstas en el decreto de creación (31), esto es, la investigación didáctica. Estos problemas son fundamentalmente de infraestructura interna de tipo administrativo. La política ministerial habida en nuestro país desde 1969 no ha considerado necesario dotar a estos Institutos de una estructura y de unas plantillas básicas en las que apoyar su funcionamiento. Esto es evidentemente un error de política educativa, sobre todo cuando los I. C. E. s. se ven cada vez más abrumados con competencias de todo tipo, que en estas circunstancias resultan difíciles de afrontar debidamente.

En este contexto, la acción de los I. C. E. s. en materia de investigación didáctica ha sido desigual y discontinua. La falta de una mínima estructura básica ha sido causante de diferencias notables entre un I. C. E. y otro en cuanto a funcionamiento, según las facilidades y habilidad de cada uno de ellos para afrontar sus innumerables problemas administrativos. La ausencia de una plantilla básica dificulta los trabajos de investigación con orientación de futuro y continuidad, ya que los equipos que la realizan están en una clara situación de inestabilidad que tarde o temprano desemboca en su desintegración. Solamente investigadores estables pueden afrontar una investigación productiva, de lo contrario la rueda de formación y desintegración de equipos se repite periódicamente, convirtiendo en permanente el proceso de puesta en marcha.

---

(31) Decreto 1678/1969, de 24 de julio, sobre creación de los Institutos de Ciencias de la Educación ("BOE" de 15 de agosto de 1969).



Tampoco la distribución presupuestaria de los fondos dedicados a los I. C. E. s. es favorable para la investigación, capítulo que supone un porcentaje mínimo del presupuesto total.

Los Centros educativos de Enseñanza Básica y Secundaria presentan un contexto general desfavorable para la experimentación didáctica. Falta de recursos de todo tipo y, sobre todo, falta real de estímulo para que el profesorado la realice. Dentro de los sistemas de promoción del profesorado de estos niveles es mínimo o inexistente el papel de los trabajos orientados al perfeccionamiento didáctico. Bien es cierto que existe un ambiente más o menos generalizado de actitud pasiva ante el tema, pero un cambio de actitudes no puede conseguirse sin estímulos eficaces.

El problema de los recursos humanos es evidente que es general para todos los estamentos que acabamos de analizar. En nuestro país tenemos carencia de investigadores en ciencias de la educación y en concreto en didáctica de las ciencias, pero también es cierto que este problema tendería a resolverse si se solucionan estos problemas de infraestructura y política educativa apuntados.

Con una estructura adecuada, el progreso de la investigación en didáctica de las ciencias en nuestro país no sería real si no se concibiera basado en la coordinación y colaboración permanente entre Facultades, Escuelas Universitarias, I. C. E. s. y Centros Escolares, así como con los departamentos universitarios especializados en ciencias del comportamiento.

## **2. CONSIDERACIONES SOBRE LA ENSEÑANZA DE LA FÍSICA EN LA UNIVERSIDAD**

En el contexto general de la enseñanza de la Física, el nivel universitario presenta una problemática similar a la que aparece en otros niveles, aunque se particulariza y adquiere un énfasis especial en determinados puntos concretos.

El hecho de que nuestro estudio experimental se mueva en la esfera universitaria nos ha llevado a señalar de forma expresa esa problemática de la enseñanza de la Física más peculiar del nivel superior, tal como queda reflejada en un amplio número de referencias revisadas, pero teniendo siempre presente que nuestra labor principal ha sido la de extraer de contextos, en general, más amplios y de poner en cierto orden secuencial aquellos aspectos y temas que tienen, a nuestro entender, una más clara relevancia en el ámbito de nuestro más directo interés.

### **2.1. La resistencia a la innovación**

La resistencia al cambio parece ser un fenómeno muy generalizado entre el profesorado de todos los niveles con experiencia docente.

Las innovaciones se enfrentan siempre con cierta inercia contra los cambios estructurales y metodológicos entre educadores y organismos docentes. Nosotros mismos hemos analizado algunos ejemplos del ámbito universitario en otros países, relativos al problema de integración de especialistas en didáctica en departamentos científicos tradicionales, constatando dicha inercia y problemas de aislamiento e incompreensión (32).

---

(32) ESCUDERO, T. y FERNANDEZ URÍA, E.: *Proyecto de Instituto de Ciencias de la Educación. Estructura y funciones*. II-III Plan Nacional de Investigación del INCIE, ICE de la Universidad de Zaragoza, 1975, 22.

Lafourcade (33) señala tres características que se dan en parte del profesorado universitario y que a su entender pudieran ser causa de dificultades para la innovación pedagógica: a) personalidad contraria a innovaciones profundas, b) escaso interés docente, y c) indecisión a aceptar nuevos modelos si no se ha contrastado perfectamente su superioridad sobre los vigentes.

Por otra parte, Eraut (34) viene a centrar más el problema, apuntando que la renovación didáctica universitaria se enfrenta a menudo con las tres dificultades siguientes: a) falta de recursos dirigidos a la innovación didáctica, b) falta de asesoramiento para desarrollar las innovaciones, y c) ignorancia acerca de innovaciones posibles.

Por nuestra parte, pensamos que quizás un aspecto que contribuye igualmente a esta inercia contra la innovación es la escasez de estímulos que recibe el profesor universitario, al igual que el de otros niveles, para perfeccionar su actuación docente. En muchos casos el ser mejor o peor profesor sólo reporta estímulos morales. Creemos, además, que ésta bien puede ser la causa, al menos en parte, que justifique la característica b) de las señaladas por Lafourcade.

## 2.2. Variedad y flexibilidad en el currículum

El problema de los cambios e innovaciones curriculares está ligado con los cambios frecuentes que se producen acerca de los tópicos de interés para los estudiantes de Física y los físicos mismos.

Así, por ejemplo, está próximo el tremendo interés producido por la física nuclear tras sus éxitos de aplicación en fisión y fusión y ciertos hallazgos espectaculares en el campo de las partículas elementales. Sin embargo, la investigación en estos temas suponía tales recursos humanos y económicos que sólo podía llevarse a cabo en laboratorios estatales o universitarios especialmente dotados. De esta forma, los avances de la física quedaban íntimamente ligados con los avances en investigación pura, que iba a absorber la imagen académica de la física (35).

Este hecho y otros similares tuvieron incidencia en la estruc-

---

(33) LAFOURCADE, P. D.: *Planeamiento, Conducción y Evaluación en la Enseñanza Superior*, Buenos Aires, Kapelusz, 1974, 17-19.

(34) ERAUT, M.: "Promoting innovation in teaching and learning: Problems, processes and institutional mechanisms", *Higher Education*, 4 (1975), 14 y 15.

(35) VICK, F. A.: "The case for teaching physics". *Physics Education*, 3 (1968), 338.

turación de los cursos de formación de físicos, enfatizando aquellos aspectos básicos que promovieran la continuación del desarrollo de la física fundamental.

En la actualidad se ha producido un cierto cambio en los tópicos de interés de los físicos, caso del estado sólido y su relación con la tecnología, lo que vuelve a plantear el tema de las innovaciones curriculares.

Sobre el tema parece haber acuerdo en dos puntos de arranque concretos: la necesidad de estructurar un curriculum capaz de cubrir un rango más amplio de actividades profesionales, como apunta Jossem (36), y la necesidad de fortalecer el papel de plataforma de comprensión de principios básicos que tiene la física en los estudios de especialización de otras ciencias, como señala Ingram (37).

En la III Conferencia Internacional sobre Enseñanza de la Física, Lord Beeching (38) indicaba que para posibilitar la variedad curricular, las Universidades debían hacer un primer esfuerzo en evitar que sus alumnos creyeran que la más alta y mejor carrera para un físico es la investigación fundamental, ahondando en la idea de que su formación supone o debe suponer, más que un entrenamiento profesional específico, un bloque de conocimiento que le capacitan y hacen útil para un amplio número de tareas profesionales.

Estas observaciones refuerzan nuestra creencia de que no parece apropiado considerar al físico, y por consiguiente estructurar su formación, como un profesional al estilo tradicional. Es difícil encontrar a físicos que trabajen como profesionales independientes tal como hacen muchos ingenieros o arquitectos.

Bajo este enfoque de opciones multiprofesionales de los estudios de física, al futuro graduado debe hacérsele sentir que una sólida formación en física y en fundamentos de filosofía natural le proporciona una visión panorámica que le capacita para adaptarse a muchas profesiones, incluso a algunas no relacionadas íntimamente con la física (39).

Este enfoque va a repercutir sobre los currícula, debiendo ser lo suficientemente flexibles para seguir formando personas orien-

---

(36) JOSSEM, E. L.: "More options for the physics major", *American Journal of Physics*, 40 (1972), 1889.

(37) INGRAM, D. J. E.: "The changing face of the physics", *Physics Education*, 10 (1975), 136.

(38) BEECHING, Lord: "Opening address to the 3<sup>rd</sup> international conference on physics education", en *The Education of a Physicist*, Brown, S. C. y Clarke, N. (ed.), London, Oliver & Boyd, 1966, 5.

(39) INGRAM, D. J. E.: "The changing face..." (o. c.), 136.

tadas hacia la física fundamental, y al mismo tiempo ofrecer la posibilidad de orientarse a otras áreas tradicionalmente poco tratadas. A este respecto, Bates (40) ofrece una extensa gama de posibles áreas de especialización del físico poco tenidas en cuenta en la actualidad.

En este mismo nivel de discusión, entendemos que es necesario posibilitar, asimismo, el que estudiantes orientados hacia otras profesiones menos ligadas con la física y las ciencias naturales, sociólogos, periodistas, médicos, filósofos, etc., puedan adquirir, si lo consideran conveniente para su quehacer profesional, como parte integrante de su curriculum de formación, una sólida base de conocimientos físicos y científicos en general.

Aunque el interés de nuestro análisis se centre en la asignatura física y en los departamentos de física, la búsqueda de soluciones alternativas al problema planteado se escapa del mero ámbito de los departamentos de física. El problema es interdisciplinario e interdepartamental y creemos que únicamente desde esa perspectiva puede llegarse a una solución eficaz. Es claro que para los físicos el tema es importante, dado el gran alcance interdisciplinario de su disciplina.

Parte del profesorado universitario es consciente de esta problemática, habiendo empezado a dar los pasos necesarios para resolverla. En agosto de 1974 se celebró en la Universidad del Estado de Pennsylvania la Conferencia sobre Tradición y Cambio en la Educación de Graduados en Física, llegándose a la conclusión de que o se introducían ciertos cambios en los currícula o en poco tiempo se llegaría a una crisis de superpoblación de graduados en física. Para abordar el problema se indicaron las siguientes líneas de acción:

a) Modificar el sentido del doctorado hacia una base con muchos canales ocupacionales posibles en vez de un entrenamiento para una futura línea específica de trabajo.

b) Ampliar los currícula hacia unas nuevas especialidades con gran demanda social potencial, tales como física médica, problemas energéticos, etc., al tiempo que se emprendía una campaña dirigida a los estudiantes que tuviera por lema *applied is beautiful*.

c) Tratar los problemas planteados por la tecnología no controlada, ahora que la campaña mundial contra la tecnología pare-

---

(40) BATES, C. A.: "Physics degree courses", *Physics Education*, 8 (1973), 135-141.

(41) GREEN, A. E. S.: "Report on the conference on tradition and change in physics graduate education", *American Journal of Physics*, 43 (1975), 214-222.

ce en decadencia, e iniciar un gran movimiento, en el que los físicos debieran ser actores principales, para llevar a cabo nuevos e importantes avances tecnológicos con fines pacíficos en todas las áreas.

d) Ofrecer posibilidades de formación profesional de acuerdo con su formación académica para los futuros graduados con orientación profesional a la docencia. Este punto se consideró de gran importancia, basándose en el hecho de que una gran parte de los graduados en física van a tener relación con la docencia en su actividad profesional.

El tema de la interdisciplinaridad fue asimismo analizado, viéndose las grandes dificultades que plantea en el momento. Dificultades que Green (42) achaca principalmente a tres leyes que él denomina *acadinámicas*:

Ley 1.<sup>a</sup> Cada profesor quiere hacer su propio trabajo con un mínimo de control y dirección, buscando especialidades de investigación que minimicen la necesidad de una competición interna y externa, esto es, un nicho exclusivo.

Ley 2.<sup>a</sup> Cada departamento siente que su papel establece un amplio dominio territorial que guarda celosamente contra toda invasión en potencia, tomando decisiones que maximicen sus operaciones microscópicas.

Ley 3.<sup>a</sup> Las Universidades están organizadas para promover el crecimiento y evolución de personas y departamentos individuales con un mínimo de conflictos. Esto es, la estructura de poder de una Universidad típica refuerza los departamentos disciplinarios.

Ante esta panorámica, entendemos que un cambio sustancial en la orientación de los programas no puede llevarse a cabo desordenadamente y de inmediato, ya que los resultados serían, seguramente, desastrosos. Sin embargo, es necesario introducir cambios e innovaciones paulatinamente, debiéndose tener presente que la introducción de una opción curricular es un proceso complejo, que requiere una labor sistemática en equipo profunda, tras haber llegado a un acuerdo, tal como apunta Lafourcade (43), en los siguientes puntos:

a) Los objetivos que definen la naturaleza y orientación de la carrera.

b) La identificación de los múltiples sectores del saber que focalizarán contenidos y prácticas desde distintos ángulos, coordi-

---

(42) GREEN, A. E. S.: "Report on the conference..." (o. c.), 220.

(43) LAFOURCADE, P. D.: *Planeamiento, Conducción y...* (o. c.), 50-52.

nados por las características y orientación general de la carrera.

- c) La validez científico-social de los contenidos.
- d) Las necesidades esenciales del futuro graduado para iniciarse satisfactoriamente en la especialidad elegida.
- e) Las posibles alternativas de elección de los alumnos que determinan una adecuada flexibilidad al curriculum.
- f) La naturaleza de los distintos ciclos que se establecen y su ponderación en relación a criterios finales bien definidos.
- g) La necesaria organicidad y coherencia de la estructura global que evite repeticiones y exigencias disímiles.

Por nuestra parte entendemos que las nuevas opciones curriculares deben atender a cuestiones de más impacto social y cubrir, en principio, las demandas más prioritarias. Así, como ejemplo típico que se ajuste a esta idea, pensamos en la opción *física del ambiente* como algo de gran interés. Interés que queda reflejado en algunos ejemplos como el que ofrece Cowan (44), que ha diseñado un curriculum en este campo, con expresión de tópicos, bibliografía y las subespecialidades siguientes:

1. Física del ambiente en general.
2. Polución del aire.
3. Fuentes y conversión de la energía.
4. Radiación ambiental.
5. Ruido.
6. Polución térmica.
7. Polución del agua.
8. Sistemas de transporte.

### **2.3. Algunos aspectos de la enseñanza de la física en la Universidad**

No cabe duda que son muchos los aspectos y casos concretos de la enseñanza de la física en el nivel superior que tienen importancia como para ser tratados y analizados de forma aislada. Sin embargo, un análisis exhaustivo de casos y aspectos alargaría en demasía esta síntesis, dificultando sensiblemente su unidad estructural. Por todo ello, hemos considerado conveniente tocar aquí dos aspectos concretos: la física para no físicos y las matemáticas para físicos, por la importancia cuantitativa, esto es, de número de estudiantes, que tiene el primero, y por la trascendencia que tiene el segundo en la formación integral del graduado en física.

---

(44) COWAN, D. J.: "Environmental topics in an undergraduate physics curriculum", *American Journal of Physics*, 40 (1972), 1748-1756.

### 2.3.1. Física para no físicos

Paralelo al problema de variedad curricular para los futuros físicos, en el sentido tradicional de la palabra, nos aparece otro sumamente importante como es el de la enseñanza de la física a otro tipo de científicos y también a otro tipo de profesionales.

La enseñanza de las ciencias, y en particular de la física, a estudiantes universitarios de disciplinas no científicas o tecnológicas es un fenómeno que va alcanzando gran relevancia, sobre todo en países con sistemas curriculares universitarios de opciones abiertas. El hecho es positivo porque, a la vez que se va mejorando la actitud del estudiantado en general hacia la ciencia en sí (45), se pueden demostrar la gran cantidad de conexiones que tiene la ciencia con otros campos del saber.

Los beneficios de este nuevo enfoque para la ciencia y la sociedad parecen evidentes; sin embargo, es importante resaltar que ello complica sensiblemente la estructura curricular universitaria a gran escala, siendo necesario el tomar como presupuesto de partida las opciones interdisciplinarias. Por dicha razón, no parece fácil la generalización de este enfoque de forma rápida, aunque en los últimos años se han llevado a cabo ciertas experiencias en dicho sentido con un alto grado de éxito, tal como la del Beloit College (USA) (46), con un curso de física de seis meses para no científicos, en el que se procura evitar el predominio de los intereses del profesor de física que lo imparte.

El fundamento del curso anterior está en hacer que el estudiante descubra por sí mismo que la física puede ser relevante para su campo de especialización, lo que se consigue haciéndole escribir un trabajo sobre un tópico que conecte dicho campo con la física o la tecnología.

El primer día de clase, cada estudiante recibe un documento de referencias de textos y artículos titulado *Physics and Everything*, que relacionan la física con catorce áreas del conocimiento (antropología y arqueología, comunicación y lingüística, economía y geografía, educación, lenguas y cultura extranjera, derecho internacional, historia, música, filosofía, psicología, sociología, religión y teología, seres vivos y ecología). Al cabo de una semana, cada estudiante ha debido seleccionar el campo en el que quiere trabajar, indicando la bibliografía consultada y el por qué de su

---

(45) COTHERN, C. R.: "Teaching science to nonscience majors. Some attitudes, ideas and approaches", *American Journal of Physics*, 41 (1973), 8.

(46) BAYLEY, J. M.: "Physics and everything: A bibliography", *American Journal of Physics*, 39 (1971), 1347-1352.



elección. Cada estudiante será orientado a través de acción tutorial y conferencias generales, al tiempo que periódicamente deberá presentar memorias de la situación de su trabajo, con vistas a evitar la demora en su iniciación. Un mes antes de terminar el curso, todos los trabajos deberán estar concluidos y presentados, examinándose y discutiéndose por la clase. Paralelamente a la realización del trabajo, los estudiantes siguen un ritmo de tres horas de clase semanales, más dos horas y media de laboratorio o proyección filmada, en las que se hace hincapié en fundamentos de física, enfatizando principalmente cuestiones epistemológicas y todas aquellas conectadas con los tópicos seleccionados por los alumnos.

Más relevante desde el punto de vista didáctico que el tema analizado, se nos presenta el de la enseñanza de la física para otro tipo de científicos y técnicos que no sean físicos puros, químicos, ingenieros, etc.

La enseñanza de la física a científicos no físicos, plantea unos problemas diferentes, en general de rango actitudinal, que deben ser estudiados como tales. Nosotros mismos hemos considerado oportuno centrar nuestro trabajo experimental, si bien en el área de metodología didáctica y no en la puramente curricular, que es el punto central de nuestra actual discusión, en este campo, no solamente buscando soluciones alternativas a un problema concreto, sino conscientes de la importancia de este campo de trabajo.

Hoy en día se acepta a la física como una constante, dentro del rango disciplinario, en la formación de cualquier científico y tecnólogo tanto como en la formación de los físicos mismos. La razón es clara y sobradamente conocida: el papel fundamental de la física como punto de arranque básico para cualquier desarrollo científico y tecnológico. Sin embargo, es obvio pensar que no todo científico necesita el mismo tipo de conocimientos de física, ni adquirirlos bajo un mismo enfoque y énfasis.

Así como se planteaba la necesidad de flexibilizar y dar variedad al curriculum de formación del físico, igualmente será necesario adecuar convenientemente los cursos de física dirigidos a formar científicos con un énfasis de especialización diferente.

La adecuación a que nos referimos en el punto anterior puede verse a menudo dificultada por problemas de comunicación entre físicos y otros científicos y tecnólogos. Alley (47) realizó en 1972 un sondeo en los Estados Unidos y Canadá en 132 departamentos

---

(47) ALLEY, R. E.: "Physics in undergraduate engineering education. Report of a survey", *American Journal of Physics*, 40 (1972), 1063-1069.

de física y escuelas de ingenieros universitarios, constatando el hecho de una pobre comunicación entre ellos, así como que muchos de los desacuerdos en lo relativo al papel de la física en la formación del ingeniero son atribuibles a la falta de interés de cualquiera de los dos grupos por aprender lo que los otros están haciendo, y a la falta de cooperación en la planificación de currícula que aborden más eficazmente las necesidades de los estudiantes.

También el mundo de la medicina va adquiriendo conciencia de la necesidad que tiene de basar su desarrollo en la física. Tanto es así que la física aplicada a la medicina es una de las especialidades más sugerentes en la actualidad tanto para físicos como médicos. Son muchas las Universidades que están potenciando departamentos de física aplicada a la medicina, bien circunscritos a las facultades de ciencias tradicionales o a las de medicina (48).

Un buen ejemplo de desarrollo interdisciplinar es el curso de *física médica* ofrecido por el Departamento de Física de la Universidad de Delaware (49), en el que tras revisar los fundamentos de la estructura de la materia, electromagnetismo y radiaciones, y al mismo tiempo que se abordan, desde los prismas físico y médico, once áreas físico-médicas relevantes, cada estudiante debe seguir dieciocho horas de prácticas en un hospital, seis de ellas en visita de reconocimiento por los distintos departamentos en que se usan técnicas físicas, y las restantes realizando experiencias en dos departamentos elegidos por el mismo, sobre las que deberá presentar la correspondiente memoria.

### 2.3.2. *Matemáticas para físicos*

Hemos señalado cómo la física es un soporte importante para el quehacer de cualquier científico; asimismo, la matemática es el lenguaje imprescindible para entender y desarrollar la física. La causa es obvia, no sólo por cuestiones de instrumentación, sino porque los procesos mentales en el campo científico incluyen aspectos de naturaleza lógico-deductiva, similares a los que tienen lugar en los procesos de razonamiento matemático. Sin embargo, en las ciencias aparecen otros componentes muy importantes, como son las estructuras de razonamiento inductivo y experimental.

---

(48) Puede consultarse el número monográfico "Physics courses in higher education", *Physics Education*, 10 (1975).

(49) ONN, D. G.: "A medical physics course based upon hospital field experience", *American Journal of Physics*, 40 (1972), 1147-1152.

Al hablar de matemáticas y didáctica de la física pueden plantearse dos problemas concretos que entendemos deben ser evitados en lo posible. Primero, una tendencia a matematizar excesivamente la enseñanza de la física, centrándose primordialmente en deducciones de fórmulas, desarrollos y aplicaciones numéricas y dejando en un segundo término las estructuras de razonamiento propiamente físico, y segundo, una tendencia a enseñar matemáticas a los futuros físicos de forma muy abstracta, como si de futuros matemáticos se tratara (50).

Por lo que respecta al primer punto, vemos necesario plantear la enseñanza de la física basándose en la propia estructura de la disciplina, esto es, conjugando en cada momento los razonamientos lógico-deductivos y los inductivos y experimentales.

Con relación a las matemáticas que necesita y debe aprender un físico, Friedman (51) opina que ningún futuro físico debe ser expuesto a un profesor de matemáticas que piense que el hecho de que las matemáticas puedan aplicarse al mundo real es irrelevante y no tiene nada que ver con lo que enseña, por la sencilla razón de que las matemáticas que el físico necesita y debe conocer tienen que ser aplicables al mundo real.

Una formación matemática purista, que en ningún caso debe confundirse con rigurosa, podría ser la causa de que a menudo el estudiante tienda a matematizar la física y solicite física matematizada para encontrarse más cómodo en ella.

En nuestra opinión, y en previsión de esta problemática, los profesores de matemáticas para físicos, bien matemáticos o físico-matemáticos, deben disminuir el énfasis en la enseñanza de técnicas matemáticas detalladas, ahondando en los conceptos más que en los hechos. El estudiante necesita capacidad de comprender más que de hacer, la que puede adquirirse fácilmente cuando se necesite si se tiene la primera.

#### **2.4. El proceso didáctico**

Es incuestionable que el profesor de física universitaria debe poseer un sólido conocimiento de su disciplina. Sin embargo, esta condición, que es absolutamente necesaria, no es suficiente para tener capacidad de dirigir el proceso didáctico de forma eficaz.

---

(50) FRIEDMAN, B.: "Mathematics in the training of a physicist", en *The Education of a Physicist*, Brown, S. C. y Clarke, N. (ed.), London, Oliver & Boyd, 1966, 67.

(51) FRIEDMAN, B.: "Mathematics in the training..." (o. c.), 64.

Son muchos otros los aspectos que debe tener en cuenta el profesor, además de la temática y contenido disciplinar, antes de comenzar un nuevo curso y durante su desarrollo. En este sentido es claro el párrafo de Oppenheimer (52) que ofrecemos a continuación, que bien pudiera ser un marco de reflexión previo a todo curso.

“Es cierto que no se puede aprender física memorizando fórmulas y terminología únicamente, pero también es cierto que la memorización juega un papel muy importante en el desarrollo intelectual. Es cierto que los estudiantes deben contar con libertad para descubrir por sus propios medios, pero también es cierto que se pueden beneficiar mucho de una acción tutorial bien dirigida. Es cierto que las experiencias empíricas con fenómenos naturales son esenciales para desarrollar una comprensión intuitiva de los mismos, pero también es cierto que, para muchos, la representación matemática de dichos fenómenos les proporciona una intuición aún más aguda que la experiencia. Es cierto que los estudiantes deben ser envueltos de forma activa en un proceso de aprendizaje, pero también es cierto que pueden darse bellos procesos de aprendizaje mientras se escucha una conferencia o se atiende a un programa de televisión. Es cierto que los estudiantes quieren aprender aquello que es relevante para su futura acción profesional, pero también es cierto que les gusta acercarse a lo fantástico y llegar al límite de lo desconocido. Es cierto que los profesores deben ser objetivos en sus representaciones disciplinares y que no deben forzar a sus estudiantes a aceptar su propio enfoque y filosofía, pero también es cierto que inevitablemente la enseñanza es una aventura de defensa de ciertas ideas con la consiguiente exclusión de otras.”

Es evidente que la tarea de conjugar adecuadamente todas esas ideas, ninguna de ellas excluye a otra, no es algo que pueda hacerse con poca imaginación y conocimiento, ni con poco esfuerzo y dedicación.

El proceso de la enseñanza requiere del profesorado la capacidad para manipular muchas más variables que los propios contenidos disciplinares. Para que los contenidos sean transmitidos

---

(52) OPPENHEIMER, F.: “Teaching and learning”, *American Journal of Physics*, 41 (1973), 1310-1311.

con eficacia, necesitan de un ambiente y situaciones educativas propicias, así como ser dirigidos a unos estudiantes emocionalmente serenos que estén convenientemente motivados profesional y científicamente. El profesorado juega un papel preponderante en este proceso, y en todo momento debe sentirse con parte de la responsabilidad ante situaciones didácticamente poco eficaces.

Este punto de vista se apoya en la idea, para nosotros básica, de que el proceso didáctico, estructúrese como se quiera, se fundamenta en la acción de un profesor o grupo de profesores, o si se amplían conceptos, en la acción de un educador o grupo de educadores. No entendemos una buena educación sin un buen profesor. Esto nos lleva a pensar en el valor relativo que va a tener todo desarrollo curricular y metodológico si no va acompañado de actitudes positivamente comprometidas del profesorado.

Continuando nuestra discusión del proceso didáctico, vamos a analizar algunos puntos concretos que hemos considerado más relevantes.

#### 2.4.1. *Estado inicial de los alumnos*

Es casi imposible prever, sobre criterios objetivos, el régimen de organización que exigirá el desarrollo de un curso, sin un claro conocimiento del estado inicial de preparación, capacidad, interés y expectativas de los alumnos a quien va dirigido. El disponer de datos sobre los aspectos apuntados ayudará a orientar las estrategias para lograr el éxito, diferenciar exigencias y establecer el grado de eficacia del esfuerzo realizado (53).

Entendemos, asimismo, que el profesor no debe estructurar el curso pensando en un alumno promedio en las citadas características porque se le pueden presentar alguno o algunos de los siguientes problemas:

- a) El curso está estructurado para un alumno teórico que no existe.
- b) Los alumnos más capaces superan las exigencias requeridas con poco esfuerzo y a veces llegan a aburrirse.
- c) Los alumnos menos capaces tienen excesivas dificultades y algunos optan por el abandono.

Hablando en términos estadísticos, un profesor no puede contentarse con tener en cuenta la tendencia central de la distribución de los alumnos de la clase, la dispersión existente también es importante y, matizando más, la propia forma de la distribución.

---

(53) LAFOURCADE, P. D.: *Planeamiento, Conducción y...* (o. c.), 60.

A veces la distribución tiene unas características tales que dificultan la labor del profesor de cara a conseguir una estructuración del curso eficaz.

Este hecho apuntado convierte al proceso de educar en algo complicado y emocionante. Si todos los educandos respondiesen a un tipo prefijado, su educación y comportamiento respondería a un modelo determinista de aplicación más o menos mecánica y escasos alicientes. Es posible que la aplicación de modelos deterministas al proceso educativo sea causa de falta de estímulo didáctico en parte del profesorado. Este fenómeno puede ser paralelo en todos los campos del conocimiento. Gran parte de la belleza de la física radica en sus dificultades y en su demanda de considerables esfuerzos mentales.

A nivel de cursos universitarios de física, puede ser erróneo el considerar como constante didáctica el contenido y nivel de exigencias de la disciplina o curso, en vez de considerar constante algo que paradójicamente varía de curso a curso: el nivel de los alumnos. Entendemos que mantener la filosofía básica de exigencias adecuadas al nivel universitario no es incompatible con el criterio de flexibilidad dentro de ciertos límites.

El problema de niveles de exigencia no lo sería tal con una estructuración vertical racional de todo el proceso de formación de un futuro graduado en ciencias, con unos sistemas de evaluación y selección periódicos que permitieran el acceso a los estudios superiores, únicamente a aquellos capaces de desenvolverse bien dentro de dicho nivel. Entendemos que la Universidad no debe tomar la postura de rigidez o pasividad ante defectos notorios de dicha estructuración vertical de la formación. Ella es una parte primordial en el organigrama educativo y debe tomar una postura activa de responsabilidad a corregir deficiencias en vez de una postura pasiva de espera para ejercer rígidamente su labor de barrera.

Desde un punto de vista didáctico y formativo se nos presentan como ineficaces: a) cursos con nivel de contenidos y exigencias excesivos porque promueven abandonos y frustraciones, y b) cursos con nivel de contenidos y exigencias bajos porque pierden posibilidades de formación en parte del alumnado. Los contenidos y exigencias deben estar al límite razonable de capacidades y esfuerzos de la distribución de alumnos.

#### 2.4.2. Metodología didáctica

En nuestro primer capítulo hemos señalado cómo se iban abandonando los estudios dedicados a buscar la superioridad relativa de métodos didácticos aislados, en una línea de búsqueda del *método ideal*. Hoy se buscan métodos apropiados a determinados tiempos didácticos, para determinados alumnos, profesores, disciplinas, etc. Esto es, las interacciones entre la metodología e individuos, disciplinas, contextos, etc., que son parte del proceso de enseñanza y aprendizaje.

Este enfoque viene a resaltar más, si cabe, el papel del profesor como agente directo y a veces único conocedor de las distintas variables que van a incidir en la estructuración metodológica del curso que imparte.

También se ha apuntado anteriormente algo que parece ser fundamental en la formación en cualquier disciplina y particularmente en las ciencias: la necesidad de que el alumno participe activamente en su proceso de aprendizaje. Somos conscientes de la imposibilidad de aplicar en todo momento una metodología activa, pero esto no es grave si los métodos pasivos no adquieren la preponderancia y se conjugan convenientemente con los activos.

Las características estructurales de la disciplina física le ofrecen al profesor grandes posibilidades de utilización de metodología variada. No es difícil embarcar a los alumnos en actividades formadoras bien a nivel individual, en pequeño o gran grupo. Si se acepta la vía del descubrimiento escalonado con tutoría del profesor, la estructura del curso se basará en problemas, experiencias y trabajos, dosificados de abajo a arriba, que van a mantener en actividad constante a los alumnos, promoviendo su creatividad, posibilitando diferentes enfoques y líneas de ataque y proporcionando una flexibilidad temporal adecuada para cada alumno.

Los métodos tradicionalmente denominados pasivos, conferencias, proyecciones y clases magistrales, pueden jugar un papel importante en momentos didácticos en los que a las vez de económicos son eficaces. Momentos tales como:

- a) Analizar el contenido general de un curso, sus objetivos, sus métodos, su alcance dentro del curriculum global.
- b) Introducir un tema, situarlo en contexto, interrelacionarlo con otros, motivarlo, comunicar fenomenología y experiencias, explicar el desarrollo histórico, comunicar entusiasmo científico hacia él.

c) Desarrollar conceptualmente un tema a través de un esquema global, insistiendo en los conceptos que aparecen y sus relaciones.

d) Aclarar de forma sintética la estructura de una parte estudiada de la materia.

e) Proveer de información acerca de nuevos logros científicos.

f) Proveer de bibliografía y documentación comentada.

g) Proporcionar ilustraciones experimentales y tecnológicas.

Las deducciones formales, resolución de problemas y la aplicación experimental, que van a constituir necesariamente la mayor parte de los tiempos didácticos en el desarrollo de cualquier curso de física en la Universidad, requieren generalmente la aplicación de métodos activos, trabajo individual escalonado, método de proyectos, seminarios de discusión, laboratorio, etc., actuando el profesor de *manager* de la formación de los alumnos y no de mero transmisor de información.

La preponderancia de los métodos activos se hace evidente cuando se analizan las dificultades para conseguir niveles de comportamiento deseados de los alumnos, tales como asimilación de conceptos, desarrollo de destrezas manuales y mentales, desarrollo de capacidad de abstracción y síntesis, desarrollo de la creatividad y capacidad de interrelacionar conceptos, etc., a través de un proceso de transmisión de información, por muy bien estructurada y bella que sea.

El compromiso eficacia-economía didáctica debe ser críticamente evaluado en toda acción docente. Cualquier esfuerzo del profesor es solamente válido desde el punto de vista pedagógico si se transforma en aumento de eficacia del proceso formador de los alumnos. Con este criterio, podría ser en ciertas ocasiones un desajuste didáctico el reproducir en clase temas, lecciones, demostraciones, problemas, etc., que están contenidos en textos, notas y apuntes editados o que pudieran ser entregados a los alumnos en forma de notas expresamente preparadas por el profesor, ya que la preparación exhaustiva de dichos temas, de cara a la exposición en la clase, supone un gran esfuerzo que pudiera ser a menudo más rentable si se dedicara a aclarar puntos oscuros encontrados por los alumnos en el estudio de esos temas y a preparar actividades complementarias que amplíen los conocimientos y su comprensión. Por otra parte, el desmenuzar los contenidos disciplinares en clase refuerza la actitud pasiva de los alumnos, algo pedagógicamente no deseable.

A veces será aconsejable el desarrollo exhaustivo de un tema



en la clase, pero la norma general debe ser la contraria. Comprenderemos por otra parte que aspectos como masificación de las clases, jornadas docentes sobrecargadas, etc., dificultan en ocasiones la racionalización del proceso didáctico; sin embargo, en muchos cursos de los últimos años de licenciatura, el número de alumnos es suficientemente manejable como para generalizar los métodos activos.

La preponderancia de métodos pasivos puede ser la causa del énfasis de los procesos lógicos-deductivos en ciertos cursos de física, ya que encuentran más posibilidades de ajuste. La introducción de métodos activos vendría a ser una necesidad si aceptamos como buena la idea de Schoene (54) de que la metodología inductiva y deductiva son igualmente importantes en la enseñanza de la física, apoyándose una en otra y no compitiendo.

Otro aspecto que debe tenerse en cuenta en la estructuración metodológica de un curso de física es el relativo a recursos didácticos y en concreto a tecnología educativa. La comprensión de fenómenos físicos se ve facilitada comúnmente con el uso de ilustraciones a través de retroproyector, proyectores de diapositivas y cine, paneles, etc. El escaso uso de estas técnicas didácticas no se basa tanto en la carencia de la tecnología, sino en la de material de paso eficaz. Sin embargo, parte de dicho material de paso, especialmente láminas de retroproyector, diapositivas y paneles, son fácilmente producibles por el propio profesorado.

A pesar de ser conscientes de que el modo de entender la enseñanza de la física en la Universidad requiere en estos momentos un análisis crítico profundo y, quizás, ciertos cambios sustanciales en lo relativo a metodología didáctica, no consideramos como vía óptima el que los distintos profesores se embarquen en el proceso de cambiar abruptamente su forma de desarrollar los cursos, porque se producirían desajustes considerables que conducirían a una situación poco menos que caótica. Las innovaciones metodológicas deben introducirse paso a paso, contrastando continuamente la eficacia de los cambios. Si se quiere promover un cambio profundo en un período corto, es necesario estructurar un plan maestro y lógico por medio de un equipo compuesto por especialistas en física, educadores, psicólogos, etc., con base a estudios de eficacia realizados previamente, pero no implantarlo mientras los profesores que van a usarlo no se hayan integrado y adaptado perfectamente a la nueva estructura.

---

(54) SCHOENE, H.: "Essentials for a minimum course in physics", en *Teaching Physics To-day. Some important topics*, Paris, O. C. D. E., 1965, 17.

El estudio experimental que presentamos es un ejemplo de innovación sencilla que puede suponer un escalón dentro de un proceso de renovación más profunda. Las revistas especializadas en didáctica de la física están llenas de magníficos ejemplos ilustrativos en este sentido sobre aplicación de enseñanza en equipo, aprendizaje en equipo, sistema tutorial por parte de estudiantes en frente de toda la clase, puesta a punto de temas, etc.

Lo señalado hasta este punto ofrece una amplia gama de recursos metodológicos que el profesor de física puede y debe emplear en el desarrollo de sus clases. La física por su estructura es muy rica en matices didácticos, presentando además ciertas características que facilitan más que en otras disciplinas la aplicabilidad de muchos de tales métodos y técnicas. La labor del profesor es, una vez conocida esa amplia gama de posibilidades, buscar los tiempos de aplicación de cada una de ellas, no necesariamente todas, teniendo presente la estructura y el nivel temático, el tipo de alumnos, sus características personales como profesor y los recursos con que cuenta.

Como punto final de nuestra discusión sobre metodología debemos detenernos en un aspecto fundamental. Es el problema de la coordinación organizativa entre las clases llamadas de teoría, problemas y prácticas de una asignatura. A veces la situación es de separación clara en cuanto a profesorado y estructuración. Este hecho, que puede tener justificaciones económico-administrativas, no las tiene desde un punto de vista metodológico y didáctico. La separación de teoría, problemas y prácticas es didácticamente poco aconsejable y bajo ningún punto de vista viene impuesta por la estructuración de la física, que es un cuerpo de conocimiento compacto en el que se conjugan aspectos teóricos y prácticos. Por estas razones abogamos por la unificación de los tres tipos de clases en una sola, la clase de la asignatura, lo que no significa que tenga que ser necesariamente un solo profesor el encargado de desarrollarla siempre y dirigir todos los tiempos didácticos. En la mayoría de los casos, será un equipo de dos o tres profesores bien coordinados el encargado de hacerlo.

La problemática resultante de la separación de las clases se agrava cuando esa separación es discriminatoria, esto es, que el peso relativo de una de ellas, por ejemplo problemas, sea inferior al de las otras. Una situación como la apuntada estaría en conflicto con la idea de considerar a la resolución de problemas como factor determinante a la hora de evaluar el rendimiento de los alumnos en la disciplina, porque presentaría problemas de

*validez de contenido* a la evaluación. En todo momento, el ajuste entre interacción-didáctica y evaluación debe ser total.

#### 2.4.3. *Evaluación de la actuación del profesorado*

Es evidente que algunos de los puntos tratados en el presente capítulo no son específicos de la enseñanza de la física a nivel superior, pero aunque son de aplicabilidad más general, se recogen por su relevancia. Este es el caso concreto de la evaluación del profesorado, aspecto que tiene implicaciones profundas en todo proceso de reforma e innovación didáctica (55).

Los puntos críticos en todo proceso de evaluación de profesores están en la búsqueda de instrumentos adecuados y en el uso que se le dé a la información obtenida. En cuanto a los instrumentos de evaluación es lógico pensar en los colegas y superiores del profesor y en sus alumnos, en cuanto al uso de la información, el proceso será diferente si el análisis tiene fines correctivos o si tiene fiscalizadores.

Entendemos que la evaluación de un determinado profesor por medio de sus colegas y superiores no debe ser algo institucionalizado y rígido, porque tarde o temprano conduciría seguramente al conflicto. Este proceso debe estructurarse de forma que el propio profesor decida de entre diversas opciones de evaluación establecidas cuál es la que quiere adoptar en su caso por considerarla más conveniente, más eficaz y menos perturbadora para su desenvolvimiento docente normal.

Los propios alumnos son un natural instrumento de evaluación del profesorado, tanto mejor cuanto el nivel de madurez de dichos alumnos es más elevado. Es obvio que para que esto sea válido debe contarse con cuestionarios bien contruidos, fiables y que no pretendan introducir sesgos de ningún tipo, así como unas condiciones ambientales relajadas que no distorsionen externamente la natural interacción didáctica entre alumnos y profesores.

En todo caso, debe complementarse toda la información sobre la acción del profesor obtenida a través de los distintos posibles instrumentos. El proceso debe ser racional y basado en criterios de prudencia porque el tema ofrece el riesgo de caer con cierta facilidad en resultados poco fiables y, por lo tanto, poco válidos.

El uso de los resultados de la evaluación del profesor con carácter fundamentalmente de fiscalización puede presentar riesgos, a menudo, difíciles de controlar. Si el peso de la evaluación se

---

(55) LAFOURCADE, P. D.: *Planeamiento, Conducción y...* (o. c.), 243.

basa principalmente en lo que dicen los alumnos, puede que el profesor tienda a hacerles ciertas concesiones gratuitas, negativas desde un punto de vista de eficacia formadora, para asegurarse buenos juicios. Si el peso de la evaluación se basa principalmente en colegas y superiores, puede que el profesor tienda desmesuradamente al perfeccionamiento en exposiciones y orden didáctico, descuidando en algún sentido la marcha real, las motivaciones, etc., de los alumnos.

## **2.5. Experimentación**

Aunque entendemos que las demostraciones y prácticas de laboratorio debieran haberse tratado en el apartado anterior, aislamos su discusión por su importancia dentro del proceso de enseñar física, así como por ser un tema conflictivo y a veces mal enfocado en el proceso de formación de físicos.

Pensamos que ningún profesional de la física discutiría la importancia del laboratorio en el proceso de formación de un físico, cualquiera que vaya a ser su orientación profesional y área de especialización. Sin embargo, no es arriesgado señalar el poco peso específico que tienen a menudo las prácticas y demostraciones de laboratorio en dicho proceso de formación.

Ahondando en las ideas anteriores, entendemos que el tópico de que no se cuenta con material e instalaciones de laboratorio adecuadas es un factor mucho menos decisivo que la ausencia de programaciones compactas y bien estructuradas en las que el laboratorio sea un aspecto relevante. Es necesario decidirse a operacionalizar la idea apuntada en el párrafo anterior y sacar a las prácticas de su tradicional "status" de acompañamiento de las clases de problemas, que a veces lo son a su vez de las de teoría.

Unas prácticas mal estructuradas y desenfocadas pueden suponer una pérdida de tiempo para el alumno, pero esto no da pie a pensar que las prácticas no deben existir en el curriculum.

Repetimos que el punto de partida básico debe ser la ruptura de la separación artificial que pueda existir de las prácticas con el resto del contexto de la asignatura, unificándolo en un solo cuerpo bien organizado y adecuadamente coordinado en tiempos, con la idea clara de que ningún tema está completo sin los problemas y las prácticas de laboratorio relativas al mismo.

Por lo que respecta al desarrollo de las prácticas, es importante señalar que exponer al alumno a material complicado para la realización de las mismas es a menudo penoso, porque no aca-

ba de entender las complicadas explicaciones de los folletos y guiones, y poco eficaz desde el punto de vista formativo, porque le hace pensar mucho en el aparato y poco en el fenómeno físico que se trata, que es en definitiva lo que se pretende (56). El manejo de aparatos complejos es algo sustancial, en primera aproximación, y debe entenderse como opcional en un buen número de casos.

El laboratorio es un elemento en la formación del físico con un carácter fundamentalmente activo y de descubrimiento personal. La repetición de experiencias de acuerdo con lo señalado en guiones o libros de recetas, tiene un escaso valor formativo en sí mismo. El diseño de programas y material de instrucción con base en el laboratorio por parte del profesor, debe hacerse de forma que proporcione al estudiante lo mínimo posible para que eche a andar y vaya pensando en lo que puede hacer y en el significado de lo que hace en cada momento de la experiencia (57). En definitiva, exponerlo, a escala reducida, a una situación similar a la que va a encontrarse continuamente a lo largo de su vida profesional. El alumno debe recibir ayuda únicamente cuando no tenga recursos para seguir adelante, algo que naturalmente va a ser diferente para cada alumno o grupo de alumnos.

Las demostraciones de laboratorio realizadas por el propio profesor es otro elemento didáctico muy importante que se encuentra con una problemática similar a la ya apuntada para las prácticas de laboratorio.

Una buena demostración de laboratorio requiere tiempo y programación adecuada por parte del profesor, así como las características que aparecen a continuación (58):

- a) El aparato debe ser a escala real.
- b) La experiencia debe ser visible para todos los alumnos.
- c) El profesor debe exponer claramente lo que se pretende, lo que se hace y lo que pasa en todo momento.
- d) Las operaciones deben ser dramatizadas y realizadas con suspense.
- e) Los resultados inesperados deben resaltarse.
- f) Las paradojas suelen ser importantes para mantener el interés.

---

(56) GOLDBERG, H. S.: "An introductory mechanics laboratory at U.I.C.C.", *American Journal of Physics*, 41 (1973), 1321.

(57) PRICE, R. M. y BRANDT, D.: "Walk-in laboratory: A laboratory for introductory physics", *American Journal of Physics*, 42 (1974), 126.

(58) MILLER, J. S.: "Matters of importance in the teaching of physics", en *Teaching Physics To-day. Some important topics*, París, O. C. D. E., 1965, 53-54.

Estas consideraciones acerca del uso del laboratorio en el proceso de formación del físico y en la enseñanza de la física en general, nos refuerzan la idea de que es necesario convertirlo de verdad en el centro, o al menos en uno de los centros más importantes, del proceso.

## 2.6. Formación de profesores

La importancia de este tema ya ha quedado señalada al aislar en el capítulo anterior a la formación de profesorado como una de las áreas relevantes de investigación en didáctica de las ciencias. Para los departamentos de física universitarios esta importancia se hace más patente, porque el tema les compete de forma muy directa.

Es cierto que un departamento de física no debe ser responsable único de la formación profesional de un futuro profesor de física, pero también es cierto que muchas de las necesidades didácticas de ese futuro profesor se pueden cubrir más fácilmente por las secciones o departamentos de física que por cualquier otra institución (59). Por ello, los departamentos de física deben tomar parte activa, en colaboración con los servicios o departamentos pedagógicos de apoyo, en el proceso de buscar el grado máximo de efectividad en la formación y perfeccionamiento de profesores.

Son muchos los problemas que rodean esta temática de la formación del profesorado íntimamente ligada con la enseñanza de la física en niveles previos al universitario. Así, por ejemplo, no es extraño encontrar ciertos errores en textos y documentos de física elemental, lo que pudiera ser achacado a que la física elemental es difícil y además ha habido muy poco estudio sistemático sobre el tema. A este respecto, el profesor Krans (MIT) defiende que la *física escolar* debiera ser una rama de investigación de las secciones de física universitarias, pues supondría una valiosa contribución a la formación de los futuros profesores, a quien Krans ve con amplio interés en la física y suficiente capacidad intelectual para enseñarla, pero sin la disposición psicológica, creatividad, concentración y perseverancia necesarias para investigar sobre el tema (60).

---

(59) TAWNEY, D. A.: "The preparation of physics teachers for secondary schools: The role of university physics departments", *Physics Education*, 8 (1973), 377.

(60) Recogido por TAWNEY, D. A.: "The preparation..." (o. c.), 378.

Una especialidad como ésta complementaría los componentes didácticos generales aportados por el servicio educativo correspondiente (el Instituto de Ciencias de la Educación, por ejemplo), centrándose en la búsqueda de ideas y métodos claros para estructurar la física elemental, en el desarrollo de métodos experimentales sencillos de alto nivel formativo y en la investigación sobre la historia del desarrollo científico.

Otro problema planteado en la enseñanza de la física a niveles primario y secundario es la cada vez mayor tendencia a estructurar los currícula en ciencias a estos niveles de forma integrada, en vez de como suma de varias asignaturas aisladas. Esto plantea a las secciones científicas universitarias la necesidad de formar profesores de ciencia integrada, con lo que será necesario la colaboración entre ellas, la readaptación del profesorado existente y la reorientación de los cursos de formación de futuros profesores de ciencias (61).

En conexión con el tercer ciclo universitario aparece el tema de la formación de profesorado de física universitario, en la que, además de la tradicional componente de especialización e investigación en una determinada parcela de la física, deben introducirse como componentes de formación aspectos relativos a la didáctica, historia, filosofía y entorno de la física. Esta idea se ve apoyada por la opinión del profesorado actual, que ve la necesidad de una formación más amplia que la puramente en física (62).

Así, por ejemplo, el departamento de física de la Universidad de Kansas nos ofrece un modelo de programa de formación de jóvenes profesores universitarios que parte de los siguientes su- puestos básicos (63):

1. El profesor encuentra a los estudiantes donde estén, personal e intelectualmente.
2. Para enseñar se debe comprender cómo aprenden los estudiantes.
3. Los materiales y conceptos desarrollados por los educadores pueden ser una gran ayuda en el desarrollo del estilo personal de enseñar.

---

(61) KEOHANE, K. W.: "Teaching physics in the future", *Physics Education*, 3 (1968), 339.

(62) STRASSENBURG, A. A.: "College physics teacher preparation. How to do it", *American Journal of Physics*, 39 (1971), 1309.

(63) SPEARS, J. y ZOLLMAN, D.: "Orientation for the new teaching assistant. A laboratory based program", *American Journal of Physics*, 42 (1974), 1062-1066.

4. Muchos estudiantes no aprenden porque el profesor siempre contesta sus cuestiones.

Este programa tiene una orientación muy experimental y de interacción permanente dentro del círculo de los tres tiempos didácticos: a) objetivos y programación, b) interacción didáctica, y c) evaluación. Los contenidos del programa son variados, basándose en todo momento en los cuatro supuestos básicos anteriores y ciñéndose las prácticas didácticas a las clases prácticas de laboratorio. La duración del programa es de seis meses.

En una línea similar, la Universidad de California (64) tiene en marcha un programa trimestral para familiarizar a los profesores ayudantes con técnicas distintas de la clase magistral tradicional, tales como resolución de dudas, clases de discusión, resolución de problemas en grupo, etc., y para demostrar cómo afecta al aprendizaje de los alumnos el comportamiento de los profesores en clase.

El procedimiento seguido es el siguiente:

1. Se graba en video-registro la actuación de cada profesor ayudante.
2. La grabación es analizada individualmente y con ayuda del tutor.
3. Se mantienen seminarios semanales de discusión.

Por el momento, se señalan como frutos más importantes del programa los siguientes:

- a) Los profesores ayudantes se sienten más libres a experimentar con distintas técnicas didácticas.
- b) La interacción verbal entre los estudiantes y los profesores ayudantes se ha incrementado.
- c) El comportamiento de los profesores ayudantes cambia de forma favorable hacia el alumnado.

Este programa es voluntario, pero en dos años de funcionamiento lo han realizado la tercera parte de los profesores ayudantes de física (100 profesores aproximadamente) de la Universidad de California de Berkeley.

Estos ejemplos y la panorámica expuesta anteriormente nos

---

(64) MOZER, F. S. y NAPELL, S. M.: "Instant replay and the graduate teaching assistant", *American Journal of Physics*, 43 (1975), 242-244.



muestran que en materia de formación de profesores es mucha e importante la labor a realizar por los departamentos de física universitarios, máxime cuando gran parte de la producción de graduados en física va a desembocar de una u otra forma en la enseñanza.

## **2.7. La estructura departamental**

Somos conscientes que el desarrollo de muchas de las ideas y sugerencias que aparecen en este capítulo presupone un análisis crítico previo de la estructura y funcionamiento más eficaz de los departamentos que componen una Sección de Físicas. Estructura que por otra parte desborda el campo puramente didáctico y por consiguientes nuestro centro de interés y posibilidades.

Sin considerar aspectos de tipo económico-administrativo, entendemos que una excesiva departamentalización puede dificultar el desarrollo curricular variado y flexible a través de una aplicación incontrolada de las leyes acadinámicas de Green (65). Asimismo, una concentración mal entendida podría sofocar la posibilidad de aparición de nuevas vías desconocidas de trabajo y formación.

Creemos que en uno u otro enfoque, el punto de partida clave está en la debida coordinación, sin relaciones de dependencia prefijadas, entre las distintas opciones disciplinares e interdisciplinares. A este respecto es de señalar el esfuerzo en este sentido de algunos departamentos en Universidades españolas y extranjeras, que tienden a aglutinar parcelas variadas de la física bajo un mismo organigrama estructural. Esto facilita, por una parte, la realización de proyectos de investigación en los que concurren aspectos de las distintas subáreas, lo que aúna esfuerzos y amplía el campo de investigación. Y, por otra parte, repercute en la propia capacidad departamental para ofrecer opciones curriculares variadas y flexibles.

Nosotros estamos en la creencia que este proceso de coordinación de departamentos de física no es más que el inicio de un proceso más ambicioso, del cual ya hay algún precedente, en el que la coordinación se extienda a departamentos interdisciplinarios científicos y hasta interfacultativos.

---

(65) Véase nota (42).

### **3. LA EVALUACION PERIODICA COMO ESTIMULO DIDACTICO: UN ANALISIS HISTORICO**

En nuestro análisis de la problemática de la investigación didáctica apuntábamos la necesidad de entroncar los estudios de tipo experimental con las teorías psicológico-educativas, ya que es el único medio para que los hallazgos tengan validez operativa y puedan ser generalizados a contextos más amplios que aquellos en los que se ha llevado a cabo el experimento.

Dado que el trabajo experimental que hemos realizado, y que analizamos en los capítulos siguientes, está centrado en el estudio de efectos de tratamientos basados en distintas formas de evaluación periódica del alumnado como medio de estimulación del aprendizaje, consideramos de importancia para el conjunto de nuestro estudio llevar a cabo un análisis histórico de la situación de este tipo de estímulos que enmarque nuestro trabajo experimental y fundamente teóricamente el por qué de su realización. En definitiva, intenta ser una aportación nueva a una línea de desarrollo didáctico con la que los investigadores teóricos y experimentales vienen trabajando desde principios de siglo.

El tema es antiguo, pero no ha sido abordado en todo momento con regularidad. Han existido momentos, coincidentes con aquellos en que los sistemas de evaluación y diagnosis psicológicos eran más criticados, en los que fue prácticamente abandonado. La racionalización del concepto de evaluar producida en los últimos años ha desenterrado el tema que nos ocupa como un medio para el desarrollo didáctico.

La mayor parte de los estudios sobre el tema que hemos revisado no tuvieron como base disciplinaria las ciencias; asimismo, muchos de ellos no fueron llevados a cabo en el marco universitario, pero ello no les quita valor dentro de un contexto didáctico general. Al fin y al cabo la didáctica de la física en la Uni-

versidad es una parcela interior a la más general a la que nos estamos refiriendo en este análisis.

### 3.1. Fundamentaciones teóricas

En los últimos setenta años ha habido gran cantidad de discusiones teóricas sobre la naturaleza, extensión y consecuencias del efecto de la evaluación en el proceso de aprendizaje. Los intentos investigadores para encontrar respuestas válidas han ido secuenciándose de forma desordenada, y, por desgracia, los hallazgos no han sido todavía lo contundentes que era de desear para decir que contamos con la solución al problema y que el tema está concluido (66).

Jones (67) hablaba en el primer cuarto de siglo del doble valor de los exámenes de clase como estímulo para el aprendizaje, proporcionando al estudiante elementos de información sobre su situación de aprendizaje y motivando al estudiante hacia una participación más activa ante la prueba de evaluación. Asimismo, Colvin (68) señalaba en 1913, al finalizar el proceso de aprendizaje, la necesidad de enfatizar más el uso de exámenes en clase, no como meras pruebas de evaluación, sino como incentivos para mantener en mente el trabajo abordado a lo largo de períodos largos de tiempo.

El que el examen motiva el trabajo de los estudiantes es algo generalmente aceptado por todos los educadores y psicólogos de la educación; sin embargo, es evidente que existen muchos puntos oscuros sobre el valor concreto de dicho tipo de estímulo. Brown y Holtzman (69) concluyen que la *motivación* parece ser un factor más importante en el rendimiento que cualquier rutina sistemática de estudio. Esto nos convierte al nivel actitudinal en un estrato de vital importancia de cara a promover el rendimiento académico. ¿Cuáles son las actitudes que elevan el rendimiento? Peterson (70) encontraba en 1916 que el aprendizaje de los estudiantes

---

(66) BALCH, J.: "The influence of the evaluating instrument on students' learning", *American Educational Research Journal*, 1 (1964), 176.

(67) JONES, H. E.: "Experimental studies of college teaching", *Archives of Psychology*, núm. 68, 1923.

(68) COLVIN, S. S.: *The Learning Process*, New York, The Macmillan Co., 1913, 176.

(69) BROWN, W. F. y HOLTZMAN, W. H.: "A Study-Attitudes questionnaire for predicting academic success", *Journal of Educational Psychology*, 46 (1955), 75-84.

(70) PETERSON, J.: "The effect of attitude on immediate and delayed retention: A class experiment", *Journal of Educational Psychology*, 7 (1916), 523-532.

era influenciado positivamente ante la presencia de un examen, ya que tomaban una actitud más activa en el proceso de aprendizaje. Reforzando estas ideas, Pease (71) encontró que los estudiantes de enseñanza media y universitarios rendían mucho mejor ante pruebas anunciadas previamente que ante pruebas no anunciadas. Resultados similares obtuvieron Tyler y Chalmers unos años después (72).

Existe otro aspecto al que todavía no hemos hecho alusión, y es el de que los estudiantes aprenden algo al examinarse y, sobre todo, cuando reciben los resultados y soluciones del mismo, hecho que está profusamente señalado en la literatura psicológico-educativa (73). En este punto debemos reconocer que este efecto no será tan importante en los exámenes de estructura más abierta, esto es, aquellos del tipo ensayo tradicional sobre un tema concreto. Sin embargo, no es este tipo de evaluación el que ha sido más considerado por los investigadores como medio de estimular el aprendizaje, sino aquella más cerrada, y en general más objetiva, tal como tests. Las críticas hacia este tipo de evaluación-estímulo, por considerar que no proporcionan al estudiante cauce para organizar y expresar sus ideas, pueden ser desechadas dentro de un análisis global, porque enfrentan al estudiante con gran cantidad de información específica relevante que podría escapar fácilmente de su atención (74). Además, las pruebas de tipo objetivo bien construidas son un medio óptimo para abordar aspectos de comprensión y aplicación que son, en definitiva, los dos niveles de la taxonomía educativa más básicos de cara a desarrollar procesos mentales de orden superior en cualquier disciplina.

Standlee y Popham (75) señalan que aunque es generalmente aceptado el hecho de que la evaluación periódica influye positivamente en el aprendizaje, la razón por la que esto sucede no es clara. Mientras Deputy (76) cree que el factor decisivo es el co-

---

(71) PEASE, G. R.: "Should teachers give warning of tests and examinations", *Journal of Educational Psychology*, 21 (1930), 273-277.

(72) TYLER, F. T. y CHALMERS, T. M.: "The effect on scores of warning junior high school pupils of coming tests", *Journal of Educational Research*, 37 (1943), 290-296.

(73) NOLL, U. H.: "The effect of written tests upon achievement in college classes: An experiment and a summary of evidence", *Journal of Educational Research*, 32 (1939), 345.

(74) KEYS, N.: "The influence of true-false items on specific learning", *Journal of Educational Psychology*, 27 (1936), 511.

(75) STANDLEE, LL. S. y POPHAM, W. J.: "Quizzes contribution to learning", *Journal of Educational Psychology*, 51 (1960), 322-325.

(76) 'DEPUTY, E. C.: "Knowledge of success as a motivating influence in college work", *Journal of Educational Psychology*, 20 (1929), 327-334.

nocimiento del progreso, Curtis y Woods (77) especulan que las influencias en el aprendizaje pueden estar en la forma de manejar los resultados de los tests. Standlee y Popham ofrecen cuatro razones o explicaciones posibles al hecho que nos concierne, y aunque ven posibilidades en todas ellas, sus preferencias se centran en la primera. Las alternativas son las siguientes:

- a) Los exámenes proporcionan *motivación extrínseca*, de forma que los estudiantes trabajan más duramente con el fin de obtener mejores calificaciones.
- b) *El conocimiento de los resultados* proporciona al estudiante la oportunidad de conocer mejor sus puntos débiles y sus puntos sólidos, de forma que pueden centrarse en aquéllos, con lo que elevan su rendimiento final.
- c) El uso frecuente de tests *estructura* el curso, ya que el profesor que emplea este estímulo didáctico está diciendo de alguna forma a los estudiantes: "Estos son los hechos y principios que yo creo que son importantes; recuérdalos."
- d) Los tests o exámenes *refuerzan la actividad* en la disciplina, por el mero hecho de hacerlos.

Rothkopf (78 y 79) hipotetiza que los sucesos tipo test, tales como cuestiones, tienen un efecto general facilitador en el aprendizaje a partir de material escrito. En un principio identificó estos efectos con una clase de actividades conocidas como *inspection behavior* (comportamiento de inspección), pero posteriormente prefirió el término comportamiento o respuestas *matemagénicas* (matema = aprendizaje, ginestai = nacer), ya que *inspection behavior* suele ser entendido en un sentido más restringido que el deseado (80).

Davis (81) ve muy deseable evaluar después de completar cada unidad didáctica porque cada una de éstas, al tiempo que cubre

---

(77) CURTIS, F. D. y WOODS, G. G.: "A study of the relative teaching value of four common practices in correcting examination papers", *School Review*, 37 (1929), 615-623.

(78) ROTHKOPF, F. Z.: "Some conjectures about inspection behavior in learning from written sentences and the response mode problem in programmed self-instruction", *Journal of Programmed Instruction*, 2 (1963), 31-46.

(79) ROTHKOPF, F. Z.: "Some theoretical and experimental approaches to problems in written instruction", en *Learning and the Educational Process*, Krumboltz, J. D. (ed.), Chicago, Rand McNally, 1965, 193 y siguientes.

(80) ROTHKOPF, F. Z.: "Learning from written instructive materials: An exploratory of the control of inspection behavior by test-like events", *American Educational Research Journal*, 3 (1966), 241-249.

(81) DAVIS, A.: "Testing and the course of classroom learning", *Journal of Educational Psychology*, 34 (1943), 526.

aspectos relativos a los objetivos y organización psicológica, representa una de las bases más significativas para la evaluación durante el proceso de aprendizaje. En esta misma línea se enmarca la sugerencia que ofrecen Monk y Stallings (82) a los profesores, al decir que debieran usarse más regularmente las evaluaciones, no solamente para estimular el estudio de los alumnos o medir su aprendizaje, sino como diagnóstico de las dificultades del estudiante con la materia y como medio de contrastar la efectividad de su labor docente.

En el caso del aprendizaje de las matemáticas, lo que nosotros extenderíamos a la mayoría de los aspectos del aprendizaje de la física, Glucksman (83) señala que dicho aprendizaje depende en gran parte de la base adquirida en conceptos o habilidades enseñadas anteriormente, de forma que, siendo importante ampliar los conocimientos del alumno, es igualmente importante asegurar que el estudiante ha obtenido previamente un nivel mínimo de conocimientos y habilidades, hecho que fundamenta la necesidad de llevar a cabo un proceso de evaluación continua del proceso de aprendizaje de los alumnos.

### 3.2. Periodicidad de la evaluación

Este capítulo nos muestra cómo la literatura ofrece bastantes muestras empíricas que fundamentan la idea teórica de que el conocimiento del propio progreso es un factor importante de motivación para el aprendizaje posterior. Igualmente, los estudios empíricos demuestran la superioridad del aprendizaje distribuido, esto es, siguiendo un proceso escalonado de asimilación y de comprobación de dicha asimilación sobre el concentrado, en el que el proceso se compone de un solo escalón (84). Obviamente, un medio de asegurar la operación de estos dos factores en una situación escolar típica será la administración de exámenes con más frecuencia que la ordinaria.

Ante una situación como la planteada han sido varios los investigadores que han buscado la frecuencia óptima de administración de exámenes con vistas a estimular el aprendizaje de los

---

(82) MONK, J. J. y STALLINGS, W. M.: "Another look at the relationship between frequency of testing and learning", *Science Education*, 55 (1971), 187.

(83) GLUCKSMAN, M. C.: "The use of retesting as a teaching device in an elementary algebra course", *School Science and Mathematics*, 73 (1973), 725.

(84) KEYS, N.: "The influence on learning and retention of weekly as opposed to monthly tests", *Journal of Educational Psychology*, 25 (1934), 427-436.

alumnos. Scott (85) llega a la conclusión de que cuantos más exámenes-estímulo son administrados el efecto es más positivo. Keys (86) observa que el rendimiento global al final del período de tratamiento era un 8 por 100 superior en los alumnos que habían tenido tests cada semana que en los que habían tenido tests mensualmente. A similares resultados llegan Fitch, Drucker y Norton (87) cuando comparan la periodicidad semanal con la mensual. Observan, sin embargo, Standlee y Popham (88) que el uso muy frecuente de exámenes no tiene efectos muy significativos a la larga, coincidiendo con algo que ya había sido observado por Keys (89). Standlee and Popham concluyen que el efecto positivo de administrar tests con cierta frecuencia puede ser contrarrestado si la frecuencia se hace excesiva.

A la vista de estos resultados es difícil llegar a conclusiones definitivas sobre la periodicidad óptima de la administración de exámenes para estimular el aprendizaje. Es probable que este aspecto no sea de importancia crítica dentro del tema general que estamos tratando y que como apunta Balch (90) la frecuencia óptima dependa de la naturaleza de la disciplina que se imparta y de la estructuración del curso que haya establecido el profesor correspondiente. Nosotros nos hemos movido en esta línea en nuestro trabajo experimental, de forma que la periodicidad ha venido en todo momento marcada por la duración de la unidad didáctica que se estaba explicando en cada momento y del tipo de examen-estímulo que era usado como tratamiento. La frecuencia ha sido mayor cuando el estímulo eran tests objetivos que cuando eran resúmenes de la materia explicada por el profesor.

Con relación al tiempo que debe pasar entre el período de explicación por parte del profesor y el uso del examen como estímulo para el aprendizaje de la materia, Sones y Stroud (91) encontraron que para el caso de pruebas de elección múltiple la efectividad era inversamente proporcional a la demora entre el período original de aprendizaje y la administración del test.

---

(85) SCOTT, I. O.: "Use of examination to stimulate learning", *California Journal of Secondary Education*, 13 (1938), 223-225.

(86) KEYS, N.: "The influence on learning..." (o. c.), 427-436.

(87) FITCH, M. L., DRUCKER, A. J. y NORTON, J. A., Jr.: "Frequent testing as a motivating factor in large lecture classes", *Journal of Educational Psychology*, 42 (1951), 1-20.

(88) STANDLEE, LL. S. y POPHAM, W. J.: "Quizzes..." (o. c.), 322-325.

(89) KEYS, N.: "The influence on learning..." (o. c.), 427-436.

(90) BALCH, J.: "The influence..." (o. c.), 175.

(91) SONES, A. M. y STROUD, J. B.: "Review, with special reference to temporal position", *Journal of Educational Psychology*, 31 (1940), 602-609.

### 3.3. Influencia del tipo de instrumento de evaluación

En el estudio de la evaluación frecuente como medio de estimulación del aprendizaje, también han existido controversias acerca del tipo de instrumento que debiera usarse como medio más efectivo. Jersild (92) señala que como medio didáctico, el valor de los tests de dos opciones, o verdadero-falso, es dudoso, ya que los efectos observados en un experimento con alumnos de primero de psicología no fueron significativos. Continúa el propio Jersild indicando que desde un punto de vista teórico, los tests verdadero-falso presentan dos serios inconvenientes:

a) La probabilidad de acierto o error es la misma, aún con testando al azar, pudiendo tener tanto efecto en perpetuar errores como fortaleciendo asociaciones correctas.

b) No requieren demasiado ejercicio mental del que se examina. Puede ser un ejercicio de tipo más pasivo que activo.

Sin embargo —continúa diciendo Jersild— estas críticas no son sostenibles contra preguntas de elección múltiple o cuestiones cortas abiertas.

En una línea argumental similar se mueve Cocks (93) afirmando: "A todos nos es familiar la vieja máxima: Nunca presentes afirmaciones falsas o incorrectas a la clase. La mayoría de los educadores estarían de acuerdo probablemente en que la presentación repetida a un niño de afirmaciones falsas resultaría desgraciadamente en aprendizaje erróneo". Remmers y Remmers (94) introducen un estudio sobre el tema diciendo: "Una de las objeciones que se hace frecuentemente contra las afirmaciones verdadero-falso es la vieja máxima pedagógica de que no se debe nunca presentar al docente una asociación falsa".

Sin embargo, Ross y Pirie (95) indican como resultado de sus trabajos que, aunque los tests verdadero-falso tienen ciertas limitaciones como estímulo didáctico, la persistencia de errores por su administración parece no mantenerse cuando inmediatamente

---

(92) JERSILD, A.: "Examination as an aid to learning", *Journal of Educational Psychology*, 20 (1929), 608.

(93) Cita recogida por ROSS, R. T. y PIRIE, M., en "The persistence of errors in successive true-false tests", *Journal of Educational Psychology*, 25 (1934), 422, del libro de COCKS, A. W.: *The Pedagogical Value of the True-False Examination*, Baltimore, Worwick and York, 1929.

(94) REMMERS, H. H. y REMMERS, E. M.: "The negative suggestion effect of true-false examination question", *Journal of Educational Psychology*, 17 (1926), 52.

(95) ROSS, R. T. y PIRIE, M.: "The persistence of error in successive true-false tests", *Journal of Educational Psychology*, 25 (1934), 422-426.



a la aplicación se dan las respuestas verdaderas, siendo que el efecto de discutir detalladamente el por qué de las respuestas tendía a ser positivo pero no muy significativo sobre la mera indicación de las respuestas verdaderas.

Keys (96) toma una postura de defensa como resultado de sus experimentos con pruebas verdadero-falso, indicando sus efectos positivos sobre el rendimiento y el desarrollo de una actitud crítica hacia afirmaciones no probadas en psicología, disciplina con la que realizó el trabajo experimental.

Hasta el momento hemos centrado nuestro análisis sobre los tipos de instrumentos en los tests de verdadero-falso, sin embargo, la controversia va mucho más lejos, presentando dos frentes límites, los defensores a ultranza de las pruebas tipo ensayo y los que defienden sobre todo las pruebas de tipo objetivo, especialmente las de elección múltiple. Una lectura detallada de la situación histórica de esta controversia presentada por Balch (97) nos confirma la idea de que no parece apropiado inclinarse por uno u otro extremo, ya que los dos tipos de instrumentos se complementan con vistas a cubrir todo el rango comportamental en el que hay que evaluar el aprendizaje del alumno y por supuesto estimularlo, caso de ser usados como estímulo didáctico.

La postura de los ultradefensores de las pruebas tipo ensayo o más bien atacantes de las pruebas objetivas, aparece más radicalizada que la de sus oponentes en el análisis de Balch. Esto puede basarse en que a nadie se le escapa el valor de las pruebas tipo ensayo como medio de valorar la capacidad organizativa y creativa del alumno en su interacción con la materia tratada en el proceso de aprendizaje. Sin embargo, Balch interviene como mediador en su análisis de la controversia señalando que no hay pocos educadores que toman una postura totalmente negativa ante las pruebas de tipo objetivo con base a argumentos hipotéticos, que no ofrecen evidencias que los justifiquen.

La escalada en el uso de pruebas objetivas en una realidad en el mundo didáctico, toda vez que se ha clarificado el panorama de su validez para evaluar ciertos niveles de la taxonomía educativa, en unas disciplinas con menor dificultad técnica que en otras. Este es el caso del trabajo de Nelson (98) relativo a la evaluación de la enseñanza de las ciencias, campo en que gran parte

---

(96) KEYS, N.: "The influence of true-false items..." (o. c.), 511-520.

(97) BALCH, J.: "The influence..." (o. c.), 169-182.

(98) NELSON, C. H.: "Evaluation objectives of science teaching", *Science Education*, 43 (1959), 20-27.

de los aspectos de comprensión de fenómenos y conceptos y aplicación pueden ser eficazmente y de forma muy fiable evaluados con pruebas objetivas, especialmente con pruebas de elección múltiple, aunque se trate de relaciones de inferencia en materia compleja.

### 3.4. Influencia del conocimiento de los resultados

El conocimiento de los resultados aparece como un punto crítico en la eficacia de la evaluación periódica como estímulo didáctico. Tanto es así, que en nuestro análisis de la literatura sobre el tema y si nos centramos en pruebas de tipo objetivo principalmente, no hemos encontrado estudios que hayan producido efectos claramente positivos a no ser que introdujeran como parte del tratamiento el dar los resultados, de una u otra manera, a los estudiantes con posterioridad inmediata a la aplicación del test. Cuando este aspecto es introducido el panorama parece esclarecerse.

Deputy (99) señala que los resultados de muchos trabajos empíricos indican que el aprendizaje tiene lugar de forma más rápida cuando los estudiantes conocen sus éxitos, pero indica asimismo como resultados de sus trabajos, que en situación de clase regular este efecto parece depender de la actitud con que los estudiantes acepten el procedimiento.

Kitch (100) observó con estudiantes de biología de enseñanza secundaria que el hecho de que ellos mismos corrigieran sus propios tests-estímulo producía efectos positivos significativos al medir el rendimiento global.

Angell (101) observó con alumnos de química de primer curso universitario que el conocimiento inmediato de los resultados de sus tests-estímulo producía efectos más positivos que el conocimiento con dos o tres días de retraso, siendo este efecto ligeramente superior en lo relativo a aspectos de aplicación que en materia de información funcional o conocimiento de hechos.

Stone (102) llegó a la conclusión de que el efecto producido

---

(99) DEPUTY, E. C.: "Knowledge of success..." (o. c.), 327-334.

(100) KITCH, L. V.: *An Experiment in Integrating Testing with Learning in High School Biology*, M. A. Thesis, University of Southern California, 1932.

(101) ANGELL, G. W.: "The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry", *Journal of Educational Research*, 42 (1949), 391-394.

(102) STONE, G. R.: "The training function of examinations: Retest performance as a function of the amount and kind of critique information", *Research Report No. AFPTRC-TN-55-8*, San Antonio, Texas, USAT Personnel Training Research Center, Lackland Air Force Base, 1955.

por tests de elección múltiple en los futuros oficiales del Ejército del Aire americano, era más positivo cuanto más información se añadiera al devolvérselos corregidos. Conclusiones similares ofrecen Plowman y Stroud (103) para estudiantes de los últimos años del nivel secundario.

Los hallazgos de Curtis (104) con alumnos de la Universidad de Michigan y de Minnesota refuerzan los de Kitch (105) al ver efectos positivos si los alumnos corrigen sus propios tests o los de algún compañero, siguiendo las indicaciones del profesor.

En un estudio con la asignatura de matemáticas para futuros profesores de enseñanza básica, Beeson (106) concluye que sus hallazgos parecen justificar su recomendación de que los profesores de matemáticas hagan mayor uso de procedimientos de evaluación que dan inmediato conocimiento, pregunta por pregunta, de los resultados a los estudiantes.

Los ejemplos apuntados parecen dejar en claro que los trabajos con el fin de demostrar el valor didáctico de la evaluación periódica deben introducir como parte sustancial del tratamiento la necesidad de informar lo más inmediatamente posible a los estudiantes, sobre las respuestas correctas de las pruebas que se apliquen.

### 3.5. Influencia sobre el rendimiento escolar

En el análisis realizado hasta el momento de los aspectos considerados como más importantes en la problemática del uso didáctico de la evaluación periódica, hemos comprobado repetidamente la influencia de este estímulo en el rendimiento del alumnado. Sin embargo, no queremos concluir este análisis histórico sin señalar explícitamente otras evidencias empíricas sobre el sentido y nivel del efecto de la evaluación periódica sobre el rendimiento escolar de los alumnos.

Hertzberg, Heilman y Leuenberger (107) comprobaron el efecto positivo de los tests-estímulo con estudiantes de psicología uni-

---

(103) PLOWMAN, L., y STROUD, J. B.: "Effect of informing pupils of the correctness of their responses to objective test questions", *Journal of Educational Research*, 36 (1942), 16-20.

(104) CURTIS, F. D.: "Testing as a means of improving instructions", *Science Education*, 28 (1944), 29-31.

(105) KITCH, L. V.: *An Experiment...* (o. c.).

(106) BEESON, R. O.: "Immediate knowledge of results and test performance", *Journal of Educational Research*, 66 (1973), 224-226.

(107) HERTZBERG, O. E., HEILMAN, J. D. y LEUENBERGER, H. W.: "The value of objective tests as teaching devices in educational psychology classes", *Journal of Educational Psychology*, 23 (1932), 371-380.

versitarios en evaluaciones de rendimiento no muy lejanas del período de instrucción, caso de un examen parcial. Al mismo tiempo, comprobaron el efecto de homogeneizar el rendimiento del grupo de clase, lo que desde el punto de vista de manejo didáctico del grupo facilita sensiblemente la labor del profesor. En el caso de evaluación del rendimiento muy alejada del período de instrucción, el efecto de los tests objetivos no aparecía como significativo.

Turney (108) consiguió igualar el rendimiento final de un grupo de estudiantes de psicología de la educación con el de otro equivalente en coeficiente intelectual pero que había rendido un 20 por 100 más de media en un examen previo sobre el tema. Esto lo consiguió con la aplicación durante un semestre de trece tests intermedios. Smeltzer (109) observó un aumento del rendimiento final del 10 por 100 con estudiantes de la misma disciplina y del mismo nivel que los empleados por Turney, viéndose cómo el beneficio parecía acentuarse para los estudiantes de menor rendimiento.

Kirkpatric (110) llegó a conclusiones similares a las apuntadas anteriormente con estudiantes de física de enseñanza media, observando asimismo cómo el rendimiento superior se mantenía para evaluaciones distantes cuando se usaban tests de estructura similar a los empleados en el tratamiento didáctico. Sin embargo, este efecto retardado no era significativo en un test de comprensión especial a pesar de haberse producido en una aplicación inmediata al tratamiento. Igualmente parecían observarse mayores beneficios para los alumnos de rendimiento previo inferior.

Noll (111), sin embargo, concluye sus estudios diciendo que el efecto de tests escritos de forma ocasional es negativo de cara al rendimiento global de la clase, y que sólo se observan efectos positivos en los estudiantes de rendimiento más bajo. La forma en que Noll realizó su trabajo experimental, comparando los resultados de un grupo de estudiantes con los del año anterior, a pesar de homogeneizarlos en cuanto a rendimiento en ciertas pruebas psicológicas y de rendimiento estándar, nos plantea a nivel per-

---

(108) TURNEY, A. H.: "The effect of frequent short objective tests upon the achievement of college students in educational psychology", *School and Society*, 33 (1931), 760-762.

(109) El experimento de SMELTZER, C. H., en su tesis doctoral, *An Experimental Evaluation of Certain Teaching Procedures in Educational Psychology*, Ohio State University, 1931, es recogido y sintetizado por PRESSEY, S. L.: *Psychology and the New Education*, New York, Harper's, 1933, 363-366.

(110) KIRKPATRICK, J. E.: *The Motivating Effect of a Specific Type of Testing Program*, Ph. D. Thesis, Iowa University Research Studies, 1933.

(111) NOLL, V. H.: "The effect of written..." (o. c.), 345-358.

sonal serias dudas metodológicas sobre la fiabilidad de dichas conclusiones.

Selakovich (112) no observó efectos positivos estadísticamente significativos en sus experiencias con tests-estímulo, pero vio cómo este tratamiento estimulaba la discusión entre los alumnos sobre la materia, así como, en opinión del profesor de la asignatura, la interacción de los alumnos hacia el profesor, sus preguntas y observaciones. Sin embargo, se notaba cierta tendencia a que muchas de las discusiones se centraran en detalles de los tests en vez de en problemas de significación más amplia.

Rothkopf (113), por su parte, observó que las preguntas tipo test, administradas tras el aprendizaje a partir de materia en prosa, facilitaban la comprensión específica y general, facilitándose únicamente la específica cuando las cuestiones eran aplicadas previamente a la lectura del material. En todo caso, este efecto era más positivo cuando al estudiante se le presentaba la respuesta correcta.

Pikunas y Mazzota (114) incrementaron en un 10 por 100 el rendimiento de alumnos de química de un centro de enseñanza media, usando un programa de evaluación semanal. Señalando, que el hecho de que la reducción del número de alumnos que rendían menos del 60 por 100 en la prueba criterio fuera de 66 a 17 tiene unas implicaciones importante desde el punto de vista didáctico.

Mckenzie (115) observó con estudiantes del último año de nivel elemental, que los tests de inferencia parecían facilitar un tipo particular de razonamiento mental que envolvía manipulaciones lógicas de conceptos tratados en un texto de historia.

### 3.6. Comentario final

A lo largo del análisis histórico que presentamos, hemos intentado seguir una línea objetiva manifiesta, al mostrar todos los pros y contras que considerábamos más relevantes de los ofrecidos por la literatura revisada, sobre la eficacia de la evalua-

---

(112) SELAKOVICH, D.: "An experiment attempting to determine the effectiveness of frequent testing as an aid to learning in beginning college courses in american government", *Journal of Educational Research*, 55 (1962), 178-180.

(113) ROTHKOPF, F. Z.: "Learning from written..." (o. c.), 241-249.

(114) PIKUNAS, J. y MAZZOTA, J.: "The effect of weekly testing in the teaching of science", *Science Education*, 49 (1965), 373-376.

(115) MCKENZIE, G. R.: "Some effects of frequent quizzes on inferential thinking", *American Educational Research Journal*, 9 (1972), 231-239.

ción frecuente como medio de estimulación didáctica del aprendizaje. Ciertamente, el número de trabajos sobre el tema es abundante aunque, como puede observarse, quedan muchos puntos oscuros sin resolver sobre los que volveremos a centrarnos al introducir y enmarcar el estudio que presentamos a continuación.

Existen ciertos aspectos en los que parece que existe un acuerdo más o menos definitivo:

a) La evaluación frecuente es un estímulo que, en líneas generales, afecta positivamente al proceso didáctico, elevando el rendimiento, homogeneizando los grupos de clase, etc.

b) El conocimiento inmediato de los resultados de las pruebas de evaluación-estímulo por parte de los estudiantes, aparece como un factor decisivo para que este tipo de tratamiento didáctico sea eficaz.

c) La frecuencia de la aplicación de las pruebas-estímulo no parece ser un punto crítico de trabajo, creyéndose que debe venir definida por la estructuración curricular que haya impuesto el profesor para el desarrollo del curso.

d) Las discusiones que enfrentan a las pruebas de evaluación cerradas con las abiertas no han servido para demostrar la superioridad de unas sobre otras, sino su carácter complementario, dentro de un proceso general de evaluación del comportamiento y rendimiento académico de los estudiantes.

Una clara demostración de que el tema abordado en este análisis histórico sigue siendo de sumo interés para los investigadores educativos, es la constante aparición de trabajos sobre el mismo, tanto en su vertiente teórica como en la práctica. Así, por ejemplo, el concepto de "*comportamiento o respuestas matemáticas*" introducido por Rothkoph (116), es un campo de trabajo en plena vigencia, tratado con asiduidad en las revistas especializadas en psicología educativa. Asimismo, no es extraño encontrar nuevos ejemplos prácticos de aplicación de la evaluación como estímulo didáctico para la enseñanza matemático-científica, tales como los que muy recientemente nos ofrecían Stover (117) y Lashier y Wren (118).

---

(116) Véase notas (78), (79) y (80).

(117) STOVER, D. W.: "Testing and grading: Procedures for improved student motivation", *The Mathematics Teacher*, 70 (1977), 498-503.

(118) LASHIER, W. S. Jr. y WREN, E. L.: "Effect of pretest feedback and mathematics skills overview on IPS (Introductory Physical Science) achievement", *Science Education*, 61 (1977), 513-518.



## **4. DESCRIPCION DEL ESTUDIO EXPERIMENTAL**

### **4.1. Planteamiento del problema: propósitos del trabajo**

Hemos señalado varias veces en los capítulos anteriores la necesidad de que los trabajos de investigación en didáctica de las ciencias se planteen desde la doble perspectiva de intentar contribuir a una línea de fundamentación teórica en desarrollo y buscar soluciones alternativas a problemas concretos que se hayan planteado.

En todo momento hemos pretendido diseñar y desarrollar nuestro estudio con esa doble perspectiva antes aludida; por un lado entroncar con la línea de trabajo psicológico-educativa que ve a la evaluación frecuente como un estímulo didáctico de considerable eficacia potencial y, por otro, contrastar la eficacia del método de evaluación frecuente como medio de estimular el rendimiento en los cursos de física universitaria dirigidos a futuros científicos, químicos y militares en nuestro caso, cuyo campo concreto de especialización no es la física.

El análisis presentado en el capítulo anterior muestra cuál es la fundamentación y situación de los trabajos sobre la evaluación frecuente como estímulo didáctico. Hemos podido comprobar que aunque algunas de las interrogantes que se habían planteado parecen tener una contestación más o menos clara, existen otras sobre las que hay que seguir investigando antes de que puedan ser respondidas de forma científica.

Algunos de esos puntos oscuros que se detectan en el análisis histórico citado son los siguientes:

a) ¿Cuál es el efecto de interacción entre la evaluación frecuente y el tipo de alumnos de que se trate? Esto es, ¿los beneficios serán mayores para los alumnos de rendimiento previo alto, medio o bajo? ¿Serán estos beneficios mayores o menores según



sean algunas características psicológicas de los alumnos? Lo mismo podría decirse en lo relativo a interacciones con el tipo de profesorado, centros, etc. Algunos de los trabajos analizados intentan ofrecer algo con relación a alguna de estas cuestiones, pero lo observado por Smeltzer (119) y Kirkpatrick (120), en relación a que los beneficios son mayores para los alumnos de rendimiento inferior, no concuerda perfectamente ni con lo observado por Proger (121), referente a la ausencia de interacción, ni por Mach (122), relativo a que el beneficio es exclusivo de los alumnos de rendimiento previo medio.

b) La interacción entre la evaluación frecuente, el tipo de prueba-estímulo que se use y el tipo de examen final. Es decir, que hay que controlar el efecto de distintos tipos de pruebas de evaluación-estímulo con diferentes tipos de pruebas de evaluación-criterio (123).

c) La interacción entre la evaluación frecuente y distintos métodos didácticos y disciplinas. A este respecto, es posible que los efectos sean diferentes bajo distintas estructuras del desarrollo de un curso, o para distintas disciplinas (124).

d) La interacción entre el tipo de prueba-estímulo que se use, la frecuencia con que se apliquen, y las características personales del estudiante. En este sentido pudiera ser que las reacciones de un individuo a los distintos tipos de pruebas, a las distintas frecuencias de aplicación, etc., dependan de sus capacidades y satisfacción como estudiante (125).

e) La interacción con los distintos niveles educativos. Ello conduce a que el efecto del método que nos ocupa pudiera ser diferente para alumnos de nivel elemental que para alumnos de nivel universitario.

Estos aspectos que acabamos de señalar, o bien no han sido estudiados convenientemente o no se ha llegado a conclusiones definitivas sobre las distintas interrogantes que plantean. El que los problemas a resolver se encuentren generalmente en el terre-

---

(119) Véase nota (109).

(120) Véase nota (110).

(121) PROGER, B. B.: *The Relationship Between Four Testing Programs and the Resultant Achievement and Test Anxiety Levels of High and Low Previous Achievement Sixth Grade Arithmetic Students*, Ed. D. Thesis, Lehigh University, 1968.

(122) MACH, G. R.: *A Comparative Study of Students Performance in an Intermediate Calculus Class as a Result of Different Evaluation Programs*. Ph. D. Thesis, Purdue University, 1963.

(123) MONK, J. J. y STALLINGS, W. M.: "Another look..." (o. c.), 184.

(124) STANDLEE, LL. S. y POPHAM, W. J.: "Quizzes..." (o. c.), 322.

(125) BALCH, J.: "The influence..." (o. c.), 177.

no de las interacciones entre distintas variables, tal como lo plantean los trabajos citados (todos ellos relativamente recientes dentro del contexto general del análisis histórico presentado), es consecuencia de algo ya señalado con anterioridad en nuestro trabajo; la tendencia cada vez más generalizada a entender el proceso didáctico como un problema claro de interacciones entre distintas variables y no como una suma de efectos aislables producidos por las distintas variables en juego. El reciente progreso de la metodología de investigación a que nos hemos referido en el primer capítulo, posibilita operacionalizar este nuevo enfoque, orientando los trabajos experimentales, apoyados en diseños complejos, a la búsqueda de la significación de tales interacciones.

Nuestro trabajo, centrado en el nivel universitario y en la disciplina física, no pretende abordar todos esos puntos oscuros porque eso es algo que se escapa de cualquier trabajo aislado. Sin embargo, sí pretende aportar nuevos datos sobre los niveles de significación de las interacciones tipo de prueba estímulo-tipo de alumno, tipo de prueba-tipo de prueba criterio y tratamiento-tipo de centro, así como sobre los niveles de significación de los efectos principales aislados, intentando contribuir con ello al estudio del valor de este tipo de estímulo para el profesor de ciencias, algo que como dicen Pikunas y Mazzota (126) no ha sido intentado sistemáticamente hasta ahora.

En lo relativo a aquella segunda perspectiva de aplicación concreta de la que hablábamos al principio de este apartado, esto es, su capacidad para responder a un problema real planteado, nuestro estudio ha sido realizado precisamente en clases de física para futuros químicos y militares, con la idea de contrastar la validez de este método para mejorar el rendimiento académico de los alumnos en dichos cursos.

Consideramos que nuestro intento está justificado si, a modo de inspección y sin pretensiones de generalización, examinamos ciertos datos recogidos durante el curso académico 1974-75 en el que fue llevado a cabo nuestro trabajo experimental, con la única finalidad de sondear el contexto educativo general en el que se iba a llevar a cabo nuestro estudio.

a) Por lo que respecta al curso Selectivo de la Academia General Militar (datos proporcionados por el Gabinete-Psicopedagógico), limitándonos al conjunto de los cuatro grupos que partici-

---

(126) PIKUNAS, J y MAZZOTA, J.: "The effect of weekly..." (o. c.), 373.

paron en nuestro trabajo, la situación relativa de aptos después de los primeros *parciales* fue:

<b>Asignatura</b>	<b>Núm. de aptos</b>	<b>% sobre matriculados</b>
Algebra ... ..	48	26,09
Cálculo ... ..	49	26,63
Química ... ..	42	22,83
Física ... ..	6	3,26

b) Por lo que respecta al curso 1.º de Químicas y centrándonos únicamente en los alumnos de los cuatro grupos que siguieron nuestra experiencia (datos proporcionados por los profesores respectivos de dichos cursos), la situación venía a ser la siguiente:

<b>Asignatura</b>	<b>Puntuación media 1.º parcial</b>
Física ... ..	3,58
Matemáticas .. ..	4,76
Química ... ..	6,18
Biología ... ..	5,81
Geología .. ..	5,76

Debemos recordar que estos datos son relativos a un grupo de alumnos determinado, aquéllos que siguieron el curso regularmente. Para el global de los cuatro grupos, las puntuaciones medias hubieran sido posiblemente menores.

c) Por lo que respecta al curso de 3.º de Químicas, la situación relativa era más cercana, aunque la media del primer parcial de la asignatura Electricidad y Optica era inferior en medio punto al de Química Inorgánica II. Con la asignatura de Química Analítica no llevamos a cabo una comparación cuantitativa por seguir un sistema de evaluación diferente.

Sin entrar en detalles de interpretación del significado de estos datos, ya que sólo tienen un valor ilustrativo en este momento, podemos observar que justifican nuestro intento y cualquier otro semejante, con vistas a buscar medios de elevar el rendimiento de este tipo de alumnos en física.

En el capítulo segundo, al hablar de la enseñanza de la física para no-físicos, hemos visto como el tema es de interés den-

tro del contexto general de la enseñanza de la física, ya que a menudo se presentan situaciones similares a las que acabamos de analizar.

La explicación causal de este tipo de problemática debe necesariamente estar en los efectos y en las interacciones correspondientes de los tres aspectos siguientes:

a) *La estructura de la física*, cuyo aprendizaje requiere la conjugación casi perfecta de aspectos de razonamiento lógico-deductivo y experimental, dándole un carácter especial que presupone un tipo de abstracción mental no fácil de proporcionar al estudiante.

b) *La actitud del profesorado de física* que, con cierta frecuencia, diseña sus pruebas de evaluación con vistas a medir ese nivel de abstracción a que nos referimos en el punto a), cuando el proceso didáctico lo ha dirigido centrándose, o al menos dando más peso relativo, a procesos de carácter lógico-deductivo. En definitiva, vemos con más frecuencia de la deseada, ciertas deficiencias en la *validez de contenido* del proceso evaluador.

c) Las *actitudes* de los alumnos hacia la física son decisivas para el desarrollo de los cursos y su aprendizaje. Es lógico pensar que las actitudes de los no-físicos hacia la física deben ser diferentes de las de los físicos o futuros físicos, y que a menudo, aquellos estarán menos positivamente motivados hacia su estudio que éstos. En este punto hay que hacer referencia al hecho de que esta motivación se va a ver seguramente afectada por el grado de aceptación de los contenidos disciplinares a que son expuestos los alumnos como parte importante de su formación para su futura actividad profesional.

Aunque no hemos abordado de forma profunda el estudio de estos tres aspectos, si hemos llevado a cabo un sondeo que además de enmarcarnos el contenido educativo en que se ha movido nuestra experiencia, ofrece ciertos indicadores significativos y refuerza empíricamente algunas de nuestras hipótesis previas. Los resultados de dicho sondeo serán expuestos más adelante.

El planteamiento de nuestro trabajo responde y se apoya en la idea de que es prácticamente inviable llevar a cabo una tarea de innovación didáctica eficaz en la Universidad de forma global y de una vez por todas. Toda innovación eficaz debe llevarse a cabo de forma graduada pero sistemáticamente. Esta idea la fundamentamos en la creencia de que la clave del proceso in-

novador no está, aunque es evidente su gran importancia, en el mero desarrollo curricular y metodológico, sino más bien en las capacidades y sobre todo en las actitudes didácticas del profesorado, pero cambiar actitudes y proporcionar nuevas capacidades no es tarea fácil ni desde luego realizable de inmediato.

Por otra parte, es difícil pensar en un trabajo de investigación realizado por un individuo, dentro del marco curricular o de experiencia a gran escala. Eso es tarea de un equipo bien conjuntado de distintos especialistas, que cuente con amplias capacidades de decisión y recursos para planificar el trabajo y el contexto administrativo en que debe moverse.

Nuestro trabajo no pretende ser sino un punto de arranque para futuros proyectos más ambiciosos en los que se vayan intentando y experimentando reformas más globales y profundas. Igualmente, pensamos ofrecer algo positivo hacia ese cambio gradual de las actitudes y capacidades del profesorado con nuestra aportación empírica de que es factible la investigación didáctica en nuestra Universidad, que existe profesorado sensiblemente motivado por el tema con capacidades para participar en este tipo de trabajos, y que el tema es de gran interés para el propio desarrollo global de la Universidad como máximo estamento formador.

Como señalan Krathwohl y Payne (127); "El papel del profesor es crear un ambiente que aumente la probabilidad de los cambios deseados en el comportamiento de los alumnos. Esto debe hacerlo interaccionando con ellos individualmente o en grupo, o haciendo que interaccionen entre ellos mismos, con material didáctico, etc.". Nosotros hemos intentado contrastar hasta qué punto éramos capaces de ayudar a ciertos profesores a mejorar dicho ambiente. Esa contrastación, para que tenga validez, la hemos planteado en términos científicos, esto es, basados en un experimento controlado.

#### **4.2. Niveles de tratamiento experimental**

Nuestro estudio de los efectos de la evaluación frecuente sobre el rendimiento académico ha sido abordado partiendo del supuesto de que los tres tipos de evaluación-criterio más comúnmente usados en los cursos de física universitaria, son: a) temas

---

(127) KRATWOHL, D. R. y PAYNE, D. A.: "Defining and assessing educational objectives", en *Educational Measurement*, Thorndike, R. L. (ed.), Washington, American Council of Education, 1971, 17.

o exámenes tipo ensayo, b) resolución de problemas y c) preguntas cortas, tradicionalmente conocidas como test, que vienen a ser un punto intermedio entre los exámenes de teoría y de problemas.

Esta situación de base nos hizo pensar en la conveniencia de estudiar los efectos de tres niveles del tratamiento que correspondieran a dichos tres tipos de evaluación-criterio, de forma que pudiéramos observar los efectos de interacción que se produjeran entre el tipo de evaluación-estímulo empleado y el rendimiento según el tipo de evaluación-criterio correspondiente.

En el capítulo anterior hemos observado como la gran mayoría de trabajos sobre el tema se han centrado en el uso de tests objetivos como estímulo didáctico. Aunque otros tipos de evaluación frecuente son usados a menudo como estímulo, tal es el caso de entrega de problemas, trabajos prácticos, etc., no tenemos referencias de que hayan sido medidos los efectos de estos otros niveles del tratamiento de forma controlada.

Las razones expuestas nos llevaron a definir los tres niveles de tratamiento que iban a ser usados en nuestro estudio y que son los siguientes:

*Nivel 1.* Evaluación del nivel de comprensión, inmediatamente después del desarrollo de cada unidad didáctica a través de un test corto de preguntas cerradas (5-10 preguntas), doble alternativa o elección múltiple, sobre la materia contenida en la correspondiente unidad didáctica. Tras la aplicación de cada test, el profesor comentaba cuáles eran y el porqué de las respuestas acertadas con lo que se producía el necesario, tal y como se ha podido observar en el capítulo anterior, "feed-back" hacia los alumnos que habían seguido la prueba.

*Nivel 2.* Evaluación de la capacidad de aplicación de lo desarrollado en cada unidad didáctica, a través de dos problemas sobre dichos contenidos formulados inmediatamente después de desarrollada la unidad y recogidos al comienzo de la clase siguiente. Con vistas a asegurar que todos alumnos trabajarán de una forma u otra cada problema, una vez entregados, el profesor seleccionaba al azar a uno de tales alumnos para que lo desarrollara para toda la clase. De esta forma, al tiempo que se proporcionaba a todos los alumnos el cómo y porqué de la resolución concreta de un determinado problema, se conseguía que todos ellos estudiaran los problemas aunque los hubieran copiado de algún compañero.

El hecho de que los alumnos podían ayudarse en la resolución de los problemas es algo que no podía ser controlado tal y como llevamos a cabo la experiencia. Esta supuesta anomalía está confundida con nuestro *nivel 2* del tratamiento y como tal la señalamos. Por lo tanto, no podemos asegurar que todos los alumnos que siguieron este nivel del tratamiento trabajaron los problemas de forma individual, pero si podemos asegurar que al menos se los estudiaron.

*Nivel 3.* Evaluación de la capacidad de sintetizar una unidad didáctica amplia a través de un resumen (dos páginas como máximo) de los puntos más fundamentales abordados por el profesor, a entregar unos días después del desarrollo de la unidad. En este nivel del tratamiento se consideró oportuno limitar la extensión del resumen-ensayo a dos páginas para evitar la dispersión, no nos interesaba que nos entregaran una copia de los apuntes tomados en clase, y obligar al alumno a que sintetizara al máximo toda la temática tratada.

Al igual que en el nivel anterior, dado que los alumnos llevaban a cabo este trabajo fuera del aula y sin control externo, no podemos asegurar que lo hicieran individualmente sin ayudarse del trabajo de los compañeros.

Por razones de simplicidad, hablaremos del *nivel 0* del tratamiento para aquellos alumnos que hayan servido como sujetos testigo o grupo de control.

### **4.3. Sujetos base de nuestro estudio**

Ya hemos señalado con anterioridad nuestro interés en llevar a cabo el trabajo con estudiantes de física universitaria cuyo campo primordial de especialización no fuera la física. Por razones de volumen de sujetos, consideramos oportuno llevar a cabo nuestro trabajo con estudiantes de la licenciatura de químicas y con estudiantes del curso selectivo para el ingreso en la Academia General Militar. Estos dos bloques de alumnos representan los dos grupos más numerosos de estudiantes de física en la Universidad de Zaragoza y por otra parte, responden a dos tipos de estudiantes de ciencias perfectamente definidos y diferentes del estudiante de física no-físico.

En las experiencias realizadas en la Facultad de Ciencias de la Universidad de Zaragoza (curso 1974-75), la muestra invitada

fueron todos los alumnos de primer curso de ciencias químicas y todos los alumnos de la asignatura Electricidad y Optica correspondiente a tercer curso de ciencias químicas.

En las experiencias en el curso selectivo de la Academia General Militar (curso 1974-75), la muestra invitada fueron los alumnos de las secciones 2, 9, 10 y 14 (40 alumnos en cada una) de entre las dieciséis secciones en que habían sido divididos el total de alumnos de dicho curso.

Los alumnos del primer curso de químicas estaban divididos en cuatro grupos de unos cien alumnos aproximadamente cada uno (grupos A, B, C y D) por riguroso orden alfabético de apellidos. Procedimiento que asegura el agrupamiento aleatorio en términos metodológicos.

Los alumnos de Electricidad y Optica de tercer curso de químicas (170 aproximadamente) había sido divididos en dos grupos de idéntica forma que los de primer curso.

Las dieciséis secciones de alumnos del curso selectivo para el ingreso en la Academia General Militar habían sido confeccionadas por el Gabinete Psico-pedagógico del Centro siguiendo un procedimiento rigurosamente aleatorio para los nuevos aspirantes y distribuyendo de forma similar, por todas las secciones, los aspirantes repetidores de segundo año. La selección de las secciones 2, 9, 10 y 14 para llevar a cabo el experimento se basó en simples razones de facilidad y concordancia de horarios que evitaran filtraciones en los alumnos.

El hecho de partir de grupos de alumnos formados por procedimientos aleatorios facilitó nuestra problemática metodológica en la fase inicial del diseño, proporcionando al mismo tiempo el supuesto básico de asignación de sujetos aleatoria del que debe partir todo estudio experimental con vistas a asegurar la validez interna del mismo.

En resumen, nuestras muestras invitadas, ya que estamos tratando con tres poblaciones diferentes, alumnos de primer curso de ciencias químicas, alumnos de tercer curso de ciencias químicas y alumnos del curso selectivo para el ingreso en la Academia General Militar, eran: a) 400 alumnos de primer curso de químicas, b) 170 alumnos de tercer curso de químicas y c) 160 alumnos del curso selectivo para ingreso en la Academia General Militar.

Sin embargo, nuestras muestras reales, lo que iba a quedar de nuestras muestras invitadas, se redujeron considerablemente bien por abandonos naturales (éstos suponen en junio alrededor



del 60 por 100 de los matriculados) o bien por restricciones metodológicas que introdujimos con vistas a asegurar la fiabilidad de los resultados.

Nuestras muestras reales se redujeron a:

- a) 114 alumnos de primer curso de ciencias químicas (28,5 por 100 de la muestra invitada).
- b) 50 alumnos de tercer curso de ciencias químicas (29,8 por 100 de la muestra invitada).
- c) 75 alumnos del curso selectivo para el ingreso en la Academia General Militar (46,8 por 100 de la muestra invitada).

Como puede suponerse el porcentaje más amplio de reducción de la muestra proviene de los abandonos naturales y no de nuestras restricciones metodológicas que en ningún caso supusieron más del 15 por 100.

Con posterioridad analizaremos con detalle los laboriosos procedimientos seguidos para controlar la homogeneidad de nuestros grupos en juego. Dicha laboriosidad se explica porque en todo momento hemos trabajado en contexto real, cuidando extremadamente no imponer ningún agente externo del control didáctico que pudiera afectar al trabajo, convirtiéndolo en experiencia de laboratorio. En investigación en física, el laboratorio es un elemento mucho más valioso que en investigación didáctica, porque variables perturbadoras que afecten en forma de ruido al fenómeno que queremos estudiar son generalmente aislables o al menos controlables. Sin embargo, en experimentación didáctica es por el momento imposible aislar ciertas variables intervinientes, algunas de las cuales no las tenemos perfectamente definidas. Ante este hecho, como bien señala Tuckman (128), parece más eficaz no ejercer controles externos en la experimentación didáctica, sino internos a través de diseños robustos.

Los controles internos que hemos llevado a cabo para seleccionar nuestras muestras reales han sido llevados a cabo *post-facto* por la razón obvia de tener que esperar a ver cuál era el porcentaje final de abandonos naturales. Sin embargo, esto no significa que hayamos incurrido en las debilidades metodológicas que presentan los experimentos *ex post-facto* (129), sino que robustece internamente nuestro trabajo por la sencilla razón de que:

---

(128) TUCKMAN, B. C.: *Conducting Educational Research*, New York, Harcourt, Brace & Jovanovich, 1972, 8.

(129) GLASS, G. W.: "Design of evaluation studies", en *Educational Evaluation: Theory and Practice*, Worthen, B. R., y Sanders, J. R. (ed.), Worthington, Ohio, Jones, 1973, 229.

a) nuestros grupos eran aleatorios previamente al experimento, b) el diseño del experimento es *pre-facto* y c) las manipulaciones *post-facto* para la debida homogeneización de los grupos lo han sido sobre la base de valores en variables medidas *pre-facto*. En definitiva, el proceso de homogeneización de grupos que hemos seguido nos va a asegurar que comparamos sujetos semejantes en aquellos aspectos que nos aparecen relevantes al trabajo, con la única diferencia de haber sido expuestos o al nivel 0 del tratamiento o a otro de los tres restantes niveles.

#### 4.4. Diseño experimental

En la actualidad, el término diseño experimental es considerado como un concepto más amplio del que pudiera inferirse de nuestra exposición posterior. Kirk (130) habla de su relación con las cinco actividades siguientes:

- a) Formulación de hipótesis estadísticas.
- b) Reglas de decisión para el contraste de hipótesis.
- c) Plan de recogida de datos.
- d) Plan de análisis de datos.
- e) Toma de decisiones sobre la probabilidad de verdad o falsedad de las hipótesis contrastadas.

En definitiva, Kirk nos está hablando del plan completo para llevar a cabo un trabajo experimental.

Winer (131) se mueve en términos similares cuando compara el diseño de un experimento con los planos del proyecto de un arquitecto. Es evidente que distintos proyectos y diseños pueden ser válidos para conseguir unos fines determinados, la clave está en elegir aquel que ofrezca mayores ventajas.

El propio Winer asimila al arquitecto con el diseñador de experimentos y al experimentador con el posible futuro propietario de la vivienda. En muchos casos, en la experimentación científica, el diseñador y el experimentador son una misma persona o equipo o al menos trabajan en estrecha colaboración.

Aunque detrás de nuestro trabajo ha existido un plan global o diseño en este sentido amplio, no consideramos oportuno explicitarlo abiertamente, porque puede inferirse de la lectura del trabajo en su totalidad y porque, como en el caso de la arqui-

---

(130) KIRK, R. E.: *Experimental Design: Procedures for the Behavioral Sciences*, Belmont, California, Brooks/Cole, 1968, capítulo 1.

(131) WINER, B. J.: *Statistical Principles in Experimental Design*, New York, McGraw-Hill, 2.<sup>a</sup> ed., 1971, Introducción.

itectura, la relevancia de un plan previo desaparece ante la realidad de la realización concreta. En la exposición de trabajos experimentales del tipo del que nos concierne, cuando se habla del diseño experimental se suele hacer en el sentido más restringido que apunta Kirk (132), esto es, plan de asignación de sujetos a las condiciones experimentales y el tipo de análisis estadístico asociado con dicho plan. En estos términos vamos a presentar el diseño de nuestros experimentos.

El número total de experimentos distintos llevados a cabo ha sido cuatro, tres en la Facultad de Ciencias y uno en el curso selectivo de la Academia General Militar. Por razones de control experimental que analizaremos más adelante, nos vimos obligados a dividir el trabajo con los alumnos de primer curso de ciencias químicas en dos experimentos distintos.

Con vistas a evitar problemas de identificación y facilitar el seguimiento global del trabajo, hemos denominado a cada experimento con una palabra compuesta por las primeras letras de los nombres o apellidos de los profesores que intervinieron en el correspondiente experimento. Los cuatro experimentos son los siguientes:

1) *Experimento SANBRU*: seguido por 58 alumnos de los grupos A y B del primer curso de ciencias químicas durante el *segundo trimestre* del curso 1974-75. La materia estudiada fue termodinámica y electrostática, siguiéndose el texto de los profesores J. M. Savirón y J. Yarza (133).

El experimento fue llevado a cabo con un diseño factorial cruzado ( $2 \times 2$ ), siendo los dos niveles del factor 1, tratamiento, no tratamiento, y los dos del factor 2, alto rendimiento (puntuación por encima de la media en el primer parcial), bajo rendimiento (puntuación por debajo de la media en el primer parcial).

La distribución de alumnos entre factores y niveles factoriales fue la siguiente:

		Factor 1		
		Tratamiento	No tratamiento	
Factor 2	Alto rendimiento	14	15	29
	Bajo rendimiento	15	14	29
		29	29	58

(132) KIRK, R. E.: *Experimental Design...* (o. c.), capítulo 1.

(133) SAVIRON, J. M. y YARZA, J. C.: *Física General, Termodinámica, Electromagnetismo y Ondas*, Zaragoza, 1974.

El tratamiento seguido por el grupo experimental fue el que hemos denominado anteriormente como *nivel 3*, esto es, entrega de resúmenes, de dos folios como máximo, de lo explicado por el profesor en unidades didácticas amplias.

El grupo experimental (tratamiento), además de llevar a cabo el mismo trabajo que el grupo de control (no-tratamiento), sintetizó el contenido del trimestre en *dos* resúmenes escritos en casa, que entregaron individualmente al profesor unos días después de que hubieran terminado las explicaciones de la unidad didáctica correspondiente y previamente a realizar el *examen-criterio*, esto es, *el segundo parcial*.

2) *Experimento JUSREB*: seguido por 56 alumnos de los grupos C y D, del primer curso de ciencias químicas durante el *tercer trimestre* del curso 1974-75. La materia estudiada fue electricidad y ondas, siguiéndose el mismo texto que en el experimento anterior y el de óptica del profesor J. Casas (134).

El diseño experimental era un factorial cruzado ( $2 \times 2$ ) idéntico al experimento SANBRU.

En este caso el criterio para clasificar a los alumnos en alto rendimiento previo o bajo rendimiento previo fue la nota media de la suma de calificaciones individuales en el primer y segundo examen parcial.

La distribución de alumnos entre factores y niveles factoriales fue la siguiente:

		Factor 1		
		Tratamiento	No tratamiento	
Factor 2	Alto rendimiento	14	14	28
	Bajo rendimiento	14	14	28
		28	28	56

El grupo experimental entregó cuatro o cinco tandas de dos problemas tipo cada una, además de seguir las mismas actividades que el grupo control, con anterioridad a realizar el examen-criterio (tercer parcial). Los alumnos que entregaron menos de cuatro tandas de problemas fueron excluidos de la muestra definitiva.

(134) CASAS, J.: *Optica*, Universidad de Zaragoza, Departamento de Física Fundamental, 1974.

3) *Experimento CORBIZU*: seguido por 50 alumnos de los dos grupos de la asignatura "Electricidad y Optica", correspondientes al tercer curso de la licenciatura de ciencias químicas, durante el *tercer trimestre* del curso 1974-75. La materia estudiada fue electromagnetismo y ondas siguiendo el texto de Reitz y Milford (135) e interferencias y polarización siguiendo el del profesor J. Casas.

El diseño experimental volvía a ser un factorial cruzado ( $2 \times 2$ ) de características similares a los dos anteriores, siendo el criterio para clasificar a los alumnos en bajo y alto rendimiento la nota media de una suma de calificaciones del primer y segundo examen parcial.

La distribución de alumnos entre factores y niveles factoriales fue la siguiente:

		Factor 1		
		Tratamiento	No tratamiento	
Factor 2	Alto rendimiento	12	13	24
	Bajo rendimiento	12	13	26
		25	25	50

El tratamiento seguido por el grupo experimental fue el que hemos denominado anteriormente como *nivel 1*, esto es, realización de un test corto (5-10 ítems) al comienzo de la siguiente clase a la que se terminara la unidad didáctica correspondiente. En este caso, los ítems eran del tipo de doble alternativa y las respuestas verdaderas se comentaban por parte del profesor una vez finalizada la prueba estímulo.

Los componentes del grupo experimental realizaron cuatro o cinco de estas pruebas a lo largo del trimestre, además del trabajo ordinario de la clase. Los alumnos que realizaron menos de cuatro pruebas fueron desechados del estudio.

El criterio (variable dependiente) volvió a ser la calificación del tercer parcial.

(135) REITZ, J. R. y MILFORD, F. J.: *Fundamento de la Teoría Electromagnética*, México D. F., UTEHA, 1969.

4) *Experimento BLATUR*: seguido por 75 alumnos de las secciones, 2, 9, 10 y 14 del curso selectivo para el ingreso en la Academia General Militar, durante el *segundo cuatrimestre* del curso 1974-75. La materia estudiada fue termodinámica y electricidad, siguiendo el texto de los profesores J. M. Savirón y J. Yarza.

Debido a la falta de homogeneidad de los grupos fue imposible usar un diseño factorial cruzado, por lo que se aplicó un análisis de covarianza multivariante en un sentido, con cuatro niveles de tratamiento.

La distribución de alumnos entre los niveles de tratamiento, fue la siguiente:

Tratamiento			
Nivel 1 <sub>1</sub>	Nivel 3	Nivel 0	Nivel 1 <sub>2</sub>
19	17	20	19
75			

El *nivel 0* corresponde al grupo de control y el *nivel 3* corresponde al grupo experimental que presentó tres o cuatro resúmenes individuales de las unidades didácticas que se habían explicado.

Los niveles 1<sub>1</sub> y 1<sub>2</sub> corresponden a los grupos experimentales que realizaron seis tests-estímulo cortos con items del tipo de doble alternativa o elección múltiple. Ambos grupos usaron los mismos tests y al mismo tiempo, siendo comentadas las respuestas por el profesor en ambos casos una vez terminada la prueba. La división en 1<sub>1</sub> y 1<sub>2</sub> la señalamos expresamente por tratarse de dos grupos de alumnos diferentes y porque además el profesor del nivel 1<sub>1</sub> seguía un método de desarrollo de la clase más inductivo que el profesor del nivel 1<sub>2</sub>.

Esta división apuntada podía permitirnos estudiar, caso de que fuera significativa, la interacción entre el *nivel 1* de nuestro tratamiento y el método didáctico general seguido por el profesor.

Podemos comprobar que en los cuatro experimentos realizados el diseño experimental es relativamente simple y fácil de manejar. En todo caso, este hecho es consecuencia de que el contexto educativo en el que realizábamos nuestro estudio era de esa forma y como tal debíamos tratarlo. En todo momento nos apoyamos en el supuesto básico de que nuestra ingerencia en el proceso establecido debía ser mínima, salvo en lo relativo a aplicación de los niveles del tratamiento. Por dicha razón, la mayor

parte del control muestral fue realizada post-facto pero a partir de datos y de un proceso llevado a cabo de forma pre-facto.

En todos los experimentos, el análisis estadístico lo hemos llevado a cabo según un modelo general de análisis multivariante cuyos fundamentos introduciremos en el siguiente capítulo. En los tres primeros experimentos nuestro análisis equivale a un análisis de varianza y en el cuarto a uno de covarianza. Como ya veremos más adelante, esta diferencia se refleja en nuestro modelo general en un diseño diferente en la matriz de variables independientes. Lo mismo podría decirse para el caso de que estuviéramos hablando de análisis de regresión.

Nuestras variables criterios o dependientes eran multivariantes porque en nuestras pruebas-criterio (exámenes parciales) aparecían calificaciones distintas para teoría, problemas, etc., lo que nos ofrecía la posibilidad de contrastar las interacciones entre los distintos tipos de prueba-estímulo y los distintos tipos de prueba-criterio.

El llevar a cabo los análisis de forma univariada para cada tipo de prueba-criterio simplificaba el modelo global del tratamiento estadístico, pero redundaba en una dispersión repetitiva del análisis y en una falta de coherencia en el proceso.

#### 4.5. Control experimental

Algunos autores como Tuckman (136) señalan que la esencia de la investigación experimental es el *control*, de forma que es imposible llegar a una conclusión válida sobre el efecto de una condición o tratamiento si no se elimina (o limita) la influencia de otras condiciones o factores en dicho efecto.

Por otro lado, Calfee (137) habla de las relaciones del control experimental con la planificación del diseño y de los procedimientos a seguir para buscar una contestación clara a la pregunta experimental.

A nosotros nos cuesta trabajo hablar de forma separada de control, diseño, procedimiento, etc., posiblemente porque, en definitiva, estamos hablando de una misma cosa con distintos aspectos íntimamente relacionados. Volvemos a la idea de diseño experimental en el sentido amplio antes apuntado, de forma que para que ese diseño sea capaz de proporcionarnos respuestas

---

(136) TUCKMAN, B. C.: *Conducting...* (o. c.), 72.

(137) CALFEE, R. C.: *Human Experimental Psychology*, New York, Holt, Rinehart and Winston, 1975, 23.

válidas debemos llevar a cabo un minucioso proceso de control de los posibles efectos perturbadores.

Campbell y Stanley (138) hablan de la necesidad de controlar dos tipos de variables: a) aquellas que puedan afectar a la validez interna del experimento, esto es, la seguridad de que los resultados obtenidos son precisos y consecuencia del proceso experimentado, y b) aquellas que afectan a la validez externa, esto es, al grado de representatividad de los resultados del estudio.

En cualquier trabajo de experimentación científica el problema del control es de máxima relevancia. Aislar debidamente las variables significativas y minimizar el ruido que interviene es una de las mayores preocupaciones de todo científico que realiza un trabajo experimental. En los experimentos en física el desarrollo técnico ha contribuido a facilitar los problemas de control, máxime cuando, en muchos casos, el refinar los procedimientos de control interno no supone deterioro en la generalización de los resultados. En el terreno didáctico se llega las más de las veces a una situación de competencia entre la validez interna y la validez externa. En ese momento el investigador debe tomar una decisión y señalar cuál es el óptimo de control para conseguir lo que se pretende.

Volviendo al caso concreto de nuestro trabajo nos encontramos con un ejemplo clarificador del problema que acabamos de plantear.

Hemos señalado cómo en todo momento al plantear nuestro trabajo hemos pretendido evitar al máximo el control externo de forma que las experiencias se desarrollaran en un contexto educativo mínimamente perturbado, es decir, en un contexto similar al que hubiese existido si no hubiéramos llevado a cabo nuestro trabajo, ya que en el grado en que consigamos esto los resultados van a tener validez externa o nivel de aplicabilidad.

Al plantear las experiencias, nos encontramos con el hecho de que, salvo en el curso selectivo de la Academia General Militar, en donde un mismo profesor impartía tres de las cuatro secciones elegidas, cada uno de los grupos elegidos en la Facultad de Ciencias tenía un profesor diferente. Es incuestionable que el hecho de que en cada experimento el grupo experimental y el de control tuvieran el mismo profesor era importante para asegurar la validez interna, sin embargo, el modificar la estructura establecida pensamos que iba a afectar al trabajo en su conjunto de

---

(138) CAMPBELL, D. T. y STANLEY, J. C.: "Experimental and quasi-experimental design for research on teaching", en *Handbook for Research on Teaching*, Gage, N. L. (ed.), Chicago, Rand McNally, 1963.



forma negativa, porque el efecto experimental que posiblemente íbamos a introducir podría tener un efecto más perturbador que el corregido y además modificaríamos el contexto natural, afectando negativamente a la validez externa. La alternativa que se nos presentó fue la de robustecer el control interno de forma que minimizáramos el problema y manteniáramos la necesaria consistencia del experimento y de sus resultados. Al analizar a continuación los procedimientos de control usados, comprobaremos que el hecho de que tratamiento y profesor estén confundidos no va a tener incidencias significativas sobre la validez de los resultados.

A. *Control de la selección: Homogeneización de los grupos experimentales y de control*

El primer paso en la selección de los grupos nos fue proporcionado por la Facultad y la Academia que habían formado sus grupos de forma aleatoria. Con este presupuesto y toda vez que los grupos que iban a participar en cada experimento eran equivalentes, asignamos al azar el tratamiento para uno de ellos, tratamiento que se introdujo en las actividades del curso sin modificar en absoluto el plan de trabajo anterior, que era idéntico al seguido por el grupo de control.

Es evidente que en la época en que se desarrollaron los experimentos, segunda y tercera parte del curso, la mortalidad natural era muy importante y que los grupos se habían reducido a menos de la mitad de los matriculados. Esta mortalidad natural reducía considerablemente nuestra muestra a una representación de la población de alumnos de este tipo de cursos de física que siguen regularmente clases y exámenes. Por lo tanto, nuestros resultados iban a ser válidos y así los constatamos expresamente, para alumnos que siguen de forma regular los cursos.

Sin ejercer ningún control externo en el alumnado, se desarrolló el experimento en el contexto natural y fue a posteriori cuando, a partir de los datos obtenidos, llevamos a cabo un proceso científico de selección de la muestra definitiva y homogeneización de los grupos experimentales y de control.

Ante el supuesto de que la mortalidad natural podía haber afectado de distinta forma a cada grupo, llevamos a cabo un proceso de homogeneización siguiendo la técnica de apareamiento individual (*matched-pair technique*). La variable de control para este apareamiento individual fue la calificación en el parcial o

parciales realizados previamente al experimento en la asignatura correspondiente. Es importante hacer constar que en todos los experimentos, el grupo experimental y el de control realizaban juntos los mismos exámenes-criterio (exámenes parciales).

En el experimento llevado a cabo en la Academia General Militar el apareamiento fue imposible porque los grupos eran, a pesar de la selección aleatoria inicial, claramente distintos. Esto fue la causa de que tuviéramos que llevar a cabo un análisis de covarianza multivariante, en vez de un análisis de varianza multivariante.

Tras el apareamiento y seleccionando únicamente a aquellos alumnos que habían seguido regularmente clases y exámenes, la situación previa al experimento de los grupos experimentales y de control era la siguiente:

#### 1) Experimento SANBRU

	Experimental	Control
Teoría 1. <sup>er</sup> parcial ... ..	(3,07; 1,65)	(3,27; 1,78)
Problemas 1. <sup>er</sup> parcial ... ..	(2,5 ; 2,37)	(2,47; 2,14)
Teoría + Problemas ... ..	(2,78; 1,62)	(2,90; 1,65)

El primer valor de los paréntesis correspondiente a la media aritmética del grupo y el segundo a la desviación típica.

#### 2) Experimento JUSREB

	Experimental	Control
1. <sup>er</sup> parcial ... ..	(3,77; 1,5 )	(4,25; 1,39)
2. <sup>o</sup> parcial ... ..	(4 ; 1,8 )	(3,58; 1,8 )
1. <sup>o</sup> + 2. <sup>o</sup> parciales ... ..	(3,88; 1,45)	(3,91; 1,35)

#### 3) Experimento CORBIZU

	Experimental	Control
1. <sup>er</sup> parcial ... ..	(3,72; 1,82)	(3,57; 1,41)
2. <sup>o</sup> parcial ... ..	(4,42; 1,52)	(4,68; 1,39)
1. <sup>o</sup> + 2. <sup>o</sup> parciales ... ..	(4 ; 1,53)	(4,13; 1,26)

#### 4) Experimento BLATUR

En este caso, repetimos, no se llevó a cabo el apareamiento por ser imposible la homogeneización. La situación de los grupos era la siguiente:

	Experim. 1 <sub>1</sub>	Experim. 3	Control	Experim. 1 <sub>2</sub>
(1. <sup>o</sup> y 2. <sup>o</sup> parciales).	(4,19; 2,14)	(4,31; 1,57)	(4,42; 1,94)	(3,72; 1,71)

La selección de la variable calificación previa en la asignatura de física correspondiente, como criterio para llevar a cabo el apareamiento y homogeneización, se hizo por considerarla la más relevante de todas las posibles. Los educadores suelen usar para estos procesos variables de rendimiento previo como la que hemos usado, o variables de aptitudes, siendo la más utilizada la inteligencia, en general el coeficiente. Nosotros elegimos el rendimiento previo por considerarlo más decisivo para el rendimiento posterior que la capacidad, que sin restarle importancia viene moderada por aspectos de interés y esfuerzo individual. Los educadores no nos preocupamos tanto de las capacidades como de los rendimientos y a esto hemos considerado oportuno remitirnos. Además, el medir el coeficiente intelectual de los sujetos del estudio nos iba a introducir un efecto experimental que hubiera perturbado la validez interna sin aportar grandes beneficios para la selección muestral definitiva.

La misma variable de rendimiento previo, en este caso la media de los dos grupos, experimental y de control, fue el criterio para clasificar a cada individuo en el grupo de rendimiento alto o en el de rendimiento bajo. Como se ha visto con anterioridad, la distribución de los alumnos en cada uno de los niveles de este factor fue prácticamente idéntica para los grupos experimentales y de control.

En el experimento BLATUR resultó imposible la introducción de este factor de rendimiento porque la distribución era muy diferente para cada grupo, hecho que reforzó nuestra idea de llevar a cabo un análisis de covarianza.

#### *B. Control de historia: Homogeneización de las condiciones experimentales y de control*

Aparte del tratamiento es importante que el grupo experimental se mueva en el mismo contexto, esto es que tenga las mismas experiencias e historia que el grupo de control.

En nuestro trabajo no ha habido necesidad de controlar este aspecto a gran escala porque el contexto era prácticamente el mismo para los grupos que intervenían en cada experimento. Únicamente hemos procurado optimizar esta igualdad, cuidando detalles pequeños como ritmos de explicación, materia abordada, mismos ejercicios, textos, etc.

Estos detalles señalados nos inclinaron a realizar dos experimentos con el curso primero de ciencias químicas en vez de uno

como teníamos previsto. El propósito era el de llevar a cabo un experimento con los cuatro grupos con un diseño factorial ( $4 \times 2$ ) en el que aplicáramos los tres niveles del tratamiento. La constatación del hecho de que los profesores seguían una marcha idéntica dos a dos aunque los cuatro grupos realizaban los mismos exámenes al mismo tiempo, hizo que considerásemos conveniente sacrificar el estudio más compacto y elegante ( $4 \times 2$ ) y convertirlo en dos ( $2 \times 2$ ).

En este mismo terreno hay que señalar que en todos los experimentos los dos profesores implicados eran lo más similares posibles en nivel, área de especialización, etc., y que su trabajo y dirección de las clases era en estrecha colaboración, cuidando al máximo la homogeneidad del contexto del experimento y minimizando el efecto del factor profesor. Para facilitar las cosas, el profesor de problemas en los dos experimentos en que no lo era el mismo profesor de teoría era la misma persona para los dos grupos.

El control de la historia es a menudo algo delicado y difícil porque los problemas pueden venir de fuentes a las que el experimentador no tiene acceso. Esto sucedió en uno de nuestros experimentos al plantearse una huelga de profesorado, la huelga era intermitente y afectaba a uno de los grupos del experimento y no a otro. La buena disposición del grupo de profesores implicados hizo que las clases siguieran con toda normalidad.

### C. *Control de la instrumentación*

Al controlar la instrumentación se intenta evitar todos los sesgos que afecten a la medición de los datos del estudio.

Ya hemos señalado cómo en nuestros experimentos todos los sujetos realizaron los mismos exámenes-criterio y en las mismas condiciones. Sin embargo, la corrección de los exámenes la realizaba el profesor concreto encargado de cada grupo, al menos en lo relativo a exámenes de teoría. En dos de los experimentos, JUSREB y CORBIZU, el profesor y calificador de los exámenes de problemas era una misma persona para los dos grupos.

En exámenes abiertos como los que tradicionalmente se realizan en nuestras facultades, el instrumento de evaluación no es la prueba en sí, sino el profesor que corrige.

El sistema de apareamiento y selección de sujetos usado nos resolvía el problema de cualquier sesgo potencial de instrumentación, esto es, de distintos criterios de evaluación, ya que el cri-

terio de apareamiento (rendimiento previo) estaba confundido con el criterio evaluador del profesor que dirigía el grupo. En otras palabras, el criterio de apareamiento inicial era relativo para cada grupo.

Como ese mismo criterio relativo se mantuvo en la evaluación de los exámenes-criterio (variables dependientes), podemos asegurar, a no ser que los profesores cambiaran su criterio evaluador en un trimestre, que el sesgo de instrumentación no se produjo y que los efectos que pudieran producirse lo eran *con* el profesor y no *por* el profesor. Lo que no podemos asegurar con nuestro estudio es que los mismos efectos se hubieran producido con otro profesor cualquiera. Es decir, la interacción profesor-tratamiento se escapa de nuestro trabajo por las razones señaladas; pero los efectos no han sido en ningún caso producidos por el profesor en sí. Sin el tratamiento los efectos no se hubieran producido.

Aunque científicamente deberíamos concluir aquí nuestra discusión sobre este aspecto, debemos señalar que tal como se homogeneizó a los profesores de cada experimento, tal como se coordinó su trabajo y tal como se asignaron de forma aleatoria al grupo experimental o de control, tenemos fundadas razones para inclinarnos abiertamente a la creencia de que el efecto aislado del tratamiento ha sido el determinante principal de los efectos significativos y no la interacción tratamiento-profesor.

A pesar del control introducido contra el sesgo de instrumentación, llevamos a cabo el estudio de la *fiabilidad entre-juicios* de los profesores que participaban en cada experimento. Los dos profesores calificaron separadamente y sin tener conocimiento de las calificaciones otorgadas por el otro, un mismo grupo de exámenes. Un estudio de la media y desviación típica de las dos distribuciones de notas, nos deparó en todos los casos una igualdad casi absoluta y la correlación alcanzó en todos los casos valores por encima de 0,95. Esta innecesaria precaución volvió a confirmarnos en la idea de que el control de instrumentación era prácticamente absoluto.

En definitiva, este riguroso control nos asegura la misma validez de resultados que si hubiéramos contado con un mismo profesor. Lo que no nos resuelve es algo que tampoco nos resolvería el mismo diseño con un único profesor, esto es, el peso que tiene en el efecto global el efecto de interacción profesor-tratamiento. Para solucionar este problema sería necesario replicar un mismo experimento con distintos tipos de profesores dentro de un único

contexto y esto supone un proyecto de investigación con serias dificultades de realización práctica en estos momentos.

#### D. *Control del efecto experimental*

Es un hecho conocido que siempre que se produce una intervención experimental existe probabilidad de que el comportamiento de los sujetos base del estudio y el del propio agente experimentador cambie más por el hecho en sí de una nueva situación que sabe es experimental, que por el propio efecto del tratamiento estudiado. En otras palabras, aparece un *efecto reactivo a la situación experimental*.

El efecto sobre los sujetos experimentales es conocido como el efecto *Hawthorne*, descubierto en Chicago en 1920 con los trabajadores de la General Electric Company (139), viéndose que la productividad de los obreros aumentaba durante su participación en experimentos independientemente del tipo de experimento que se llevase a cabo.

El efecto sobre el agente experimentador íntimamente ligado con sus *expectativas* alcanzó relevancia en el mundo educativo tras la aparición en 1968 del famoso estudio de Rosenthal y Jacobson (140) sobre la influencia de las expectativas del profesor en el rendimiento de sus alumnos.

En nuestro trabajo hemos intentado en todo momento controlar el que ninguno de estos dos efectos se produjera. Para ello, los experimentos se desarrollaron sin la menor publicidad. Los alumnos nunca fueron informados de que se estaba llevando a cabo un trabajo experimental. Muy pocas personas conocían este hecho y para evitar problemas les fue solicitada su discreción, algo que en todo momento fue una realidad. Por otra parte, la recogida de datos sobre los resultados de los grupos en parciales y exámenes finales en otras asignaturas se hizo una vez que el experimento había concluido.

Para reafirmarnos en la no existencia de efecto Hawthorne y evitar efectos de las expectativas de los agentes experimentadores, en este caso los profesores, éstos no conocieron exactamente durante la realización de los experimentos en qué consistía el conjunto del trabajo ni cuáles eran las hipótesis y los procesos que el experimentador iba a contrastar. El diseño global del trabajo

---

(139) Una discusión amplia de este hecho aparece en BROWN, J. A. C.: *The Social Psychology of Industry*, Middlesex, England, Penguin Books, 1954.

(140) ROSENTHAL, R. y JACOBSON, L.: *Pygmalion in the classroom*, New York, Holt, Rinehart and Winston, 1968.

es algo que el experimentador guardó en secreto hasta que los experimentos habían concluido. Estas medidas de control nos dan seguridades de que ninguno de los dos efectos experimentales hayan modificado y distorsionado los resultados del trabajo.

#### E. *Control de rendimiento en otras disciplinas*

Con posterioridad a la realización de los experimentos se llevó a cabo una recogida y análisis de datos sobre los rendimientos antes, durante y después de la realización del experimento, de los alumnos base de nuestro estudio en otras disciplinas.

El análisis de los rendimientos previos en otras asignaturas en los experimentos en la Facultad de Ciencias, nos confirmó en la idea de que la selección y homogeneización había sido correcta. En ningún caso las diferencias en las medias grupales en otras asignaturas superaban el medio punto y las distribuciones de notas y calificaciones eran muy similares. Este hecho se acentuaba cuando el profesor que impartía otra asignatura era el mismo para los dos grupos. En alguna asignatura la comparación cuantitativa era dificultosa por existir distintos grados de mortalidad en la muestra o porque los profesores de cada grupo seguían unos criterios de calificación totalmente diferente. Sin embargo, en el plano cualitativo la similitud se hacía evidente.

Esta homogeneidad previa se mantuvo tras las realización de los experimentos, hecho que lo confirmó el análisis de los rendimientos finales. Podemos afirmar que el experimento en *la asignatura de física no alteró ni positiva ni negativamente el desarrollo y el rendimiento en el resto de las asignaturas del grupo.*

En lo que respecta al experimento en el curso selectivo de la Academia General Militar, también podemos afirmar que el experimento no afectó al resto de asignaturas; pero sin embargo, el análisis de rendimientos nos confirmó la idea de que los grupos no eran homogéneos como ya habíamos detectado con la asignatura de física. Tanto en álgebra, cálculo como en química, las distribuciones en cada grupo presentaban unas diferencias similares a las que habían aparecido en física.

#### F. *Control de las actitudes del alumnado*

Es evidente que las actitudes de los alumnos pueden tener a veces un peso en los rendimientos mayor que las mismas capacidades. Por lo tanto, no podíamos llevar a cabo nuestro trabajo sin tener en cuenta esta faceta.

El primer problema a resolver fue el de introducir el tratamiento de forma natural, sin presiones que cambiaran el contexto y nos introdujeran un efecto experimental potencial. Sin embargo, era absolutamente necesario motivar de alguna manera a los alumnos para que realizaran el trabajo que se les proponía.

Tras discusión abierta del experimentador con los profesores implicados, se llegó a la fórmula de introducir el tratamiento explicándoles a los alumnos que era una técnica para ayudarles a elevar los bajos rendimientos que evidentemente se habían producido en los exámenes anteriores, pero que no era obligatorio de la misma forma que no eran obligatorias, de hecho, las clases. Sin embargo, en caso de duda entre dos calificaciones al final de curso, los tests, problemas o resúmenes entregados podrían ser un factor de ayuda para el alumno.

La realidad es que, esto que apareció en un principio como uno de los problemas más delicados del trabajo, se resolvió sin ninguna dificultad y los alumnos de la Facultad de Ciencias lo aceptaron de forma natural y sin plantear el más mínimo problema.

Los alumnos del curso selectivo de la Academia General Militar lo aceptaron con algún pequeño reproche aunque sin grandes dificultades. Esto es natural dada la multiplicidad de situaciones de control a que están sometidos este tipo de alumnos.

En relación con las actitudes de los alumnos como estudiantes de física, llevamos a cabo un pequeño sondeo, a través de un cuestionario con items del tipo de escala Likert elaborado al respecto, de las actitudes y opiniones de los sujetos base de nuestro trabajo sobre la facilidad de la asignatura de física, su interés, el interés de la evaluación frecuente y el interés del tratamiento para aquellos alumnos que hubieran participado en los grupos experimentales.

Las medias aritméticas que vamos a reproducir tienen un valor de simple sondeo y no pueden tomarse como base para tomar decisiones importantes. Nuestro propósito al recogerlas y reproducirlas es el simple de mostrar que las actitudes no han sido un factor diferencial en el comportamiento de los grupos de nuestros experimentos.

Para evitar el efecto experimental este sondeo fue realizado una vez concluido el experimento.



### 1. Experimento SANBRU

	Experimental	Control
Facilidad (*)	2,32	2,41
Interés por la física	3,09	2,92
Interés evaluación frecuente	3,56	3,23
Interés tratamiento	2,8	

(\*) Facilidad ésta puntuada en una escala de 1 a 5 y los tres restantes aspectos en una escala de 1 a 4.

### 2. Experimento JUSREB

	Experimental	Control
Facilidad	2,32	2,35
Interés física	2,87	2,98
Interés eval. frec.	3,04	3,4
Interés tratamiento	3,11	

### 3. Experimento CORBIZU

	Experimental	Control
Facilidad	2,56	3,62
Interés física	3,13	3,02
Interés eval. frec.	3,6	3,17
Interés tratamiento	3,24	

### 4. Experimento BLATUR

	Trat. 1 <sub>1</sub>	Trat. 3	Control	Trat. 1 <sub>2</sub>
Facilidad	2,36	2,51	2,59	2,47
Int. física	2,08	2,85	3,18	2,72
Int. eva. frec.	3,26	3	3,26	3
Int. trat. ..	3,05	2,65		2,56

Como puede observarse la situación actitudinal en los grupos que componen cada experimento es muy semejante y las diferencias observadas no parecen importantes, máxime cuando hemos usado un instrumento simple y sin contrastar. Sin embargo, este añadido a nuestro trabajo era necesasio para controlar un problema que sinceramente no esperábamos pero que hubiera podido ser importante.

En líneas generales los alumnos piensan que la asignatura tiene ciertas dificultades pero que es interesante tanto para su formación como para su actividad profesional.

Los alumnos piensan que la evaluación frecuente tiene interés y que el tratamiento seguido tiene valor positivo, sobre todo en el caso de tests y entrega de problemas. La entrega de resúmenes (nivel 3) ha sido el tratamiento peor aceptado por el alumno, posiblemente por ser el más novedoso y el que introducía un mayor cambio en la mecánica tradicional de trabajo. Los resultados que analizaremos más adelante confirman nuestra hipótesis previa de que este nivel de tratamiento puede ser muy valioso y tiene una alta capacidad didáctica al nivel universitario.

En este mismo sentido, los datos parecen indicarnos que los alumnos no interpretan la entrega de resúmenes con un método de evaluación frecuente. Los otros dos métodos sí que parecen entenderlos así.

Remontándonos a lo expuesto al comienzo de este capítulo sobre las causas de los bajos rendimientos en física de alumnos con otra especialización científica que la física, el sondeo realizado nos indica que no parece que el tercer factor señalado, esto es, las actitudes contrarias de los alumnos, sean algo determinante. La generalidad de los alumnos expresa su interés por la física, al mismo tiempo que señalan las dificultades que encuentran en ella.

Creemos sinceramente en la necesidad de replantearse este problema con la base expuesta en el punto anterior y buscar vías de solución estudiando los otros factores que indicábamos, la estructura de la disciplina y la actitud del profesorado en cuanto a impartidores de cursos de física dirigidos a no-físicos.

#### **4.6. Comentario sobre la validez interna y externa de nuestros experimentos**

En este capítulo creemos haber dado suficientes pruebas de que los resultados de nuestros experimentos deben tener un alto grado de fiabilidad, o lo que es lo mismo, que los experimentos tal como se han llevado a cabo lo han sido con suficiente validez interna.

En lo relativo a la validez externa o capacidad de generalización, ya hemos señalado asimismo que con nuestro estudio no se puede ir científicamente más allá de los alumnos de ciencias químicas y aspirantes a militares del tipo de los tratados y que regularmente asisten a clases y exámenes, esto es, alumnos que deliberadamente no abandonan la asignatura con anterioridad a que el

curso académico haya concluido. De la misma forma que nuestro trabajo se ha centrado en la asignatura de física y que toda generalización a cualquier otra no está científicamente fundamentada en los resultados de este estudio.

Es esperable que resultados similares se pueden producir en asignaturas cercanas en contextos educativos parecidos y con otro tipo de alumnos de ciencias. Sin embargo, nosotros debemos ser necesariamente honestos al exponer el grado de aplicación que fundamenta científicamente nuestro trabajo, aunque tengamos fundadas esperanzas de que puede tener un valor informativo mucho más amplio. Cualquier nueva hipótesis para otros contextos debe ser contrastada experimentalmente.

También es éste el momento de volver a repetir que en experimentación didáctica es mucho más crítico el problema de replicación que en experimentación en el laboratorio de física, con ser de gran relevancia asimismo. Si tuviéramos que tomar decisiones curriculares importantes en nuestra Universidad con la única base de nuestro trabajo, estaríamos en una situación de gran probabilidad de equivocarnos. Sería necesario replicar nuestro trabajo y ver que es lo permanente.

Con lo expuesto no cuestionamos la validez de nuestro estudio, sino que queremos reforzar lo señalado en el primer capítulo sobre la necesidad de replicar experiencias didácticas ante el hecho concreto de que las interacciones con el contexto educativo pueden ser decisivas. Nosotros hemos controlado el contexto de la segunda parte del curso académico 1974-75 en la Universidad de Zaragoza, no nos atreveríamos a decir que ese mismo contexto permanece hoy y que los resultados en este momento serían idénticos. Ciertamente, suponemos que en líneas generales todo sería muy parecido pero científicamente no podemos asegurarlo. Este tipo de perturbación es intrínseco a la investigación en ciencias del comportamiento humano y debemos ser científicamente prudentes al manejar datos ofrecidos por proyectos en dicho marco.

## 5. FUNDAMENTOS DEL ANALISIS MULTIVARIANTE

Hemos señalado en el capítulo anterior que nuestras variables dependientes no eran variables simples sino multivariantes, ya que están compuestas de calificaciones de teoría y problemas en los exámenes criterio en tres de los experimentos, y de calificaciones de teoría, problemas y cuestiones en uno de ellos. Por lo tanto, para llevar a cabo el análisis de datos en cada uno de nuestros experimentos de forma compacta, sin partir la variable dependiente, necesitamos el uso de un modelo de análisis estadístico multivariante que se ajuste a nuestros diseños experimentales.

En este capítulo no vamos a presentar un estudio riguroso y exhaustivo de las peculiaridades del análisis multivariante por ser algo que desborda los límites de un trabajo experimental aplicado como el nuestro, sin embargo, hemos considerado conveniente desarrollar los fundamentos del modelo estadístico en el que se apoya nuestro análisis de datos con un tratamiento comprensible para el investigador aplicado, pero sin descuidar el nivel mínimo de rigor matemático necesario.

Anderson (141) señala que tal como el análisis de la varianza basado en el modelo de efectos fijos puede obtenerse de la teoría de mínimos cuadrados, el análisis multivariante se puede también obtener directamente a partir de la teoría de los tests de hipótesis lineales de la teoría de la regresión. Sin embargo, en el caso multivariado es más amplio el rango de elección de tests de significación.

Nosotros hemos llevado a cabo la descripción estocástica de diseños experimentales en los que se usa funciones lineales observables de variables aleatorias partiendo de un enfoque general conocido como *modelo de regresión*, en el que el tratamiento de la

---

(141) ANDERSON, T. W.: *An Introduction to Multivariate Statistical Analysis*, New York, Wiley, 1958, 178.

regresión, varianza y covarianza es idéntico, apareciendo diferencias únicamente en la composición de la matriz-diseño de la variable independiente (142).

### 5.1. Modelo general de análisis multivariante lineal (efectos fijos)

En el análisis univariado en un sentido, partimos de un conjunto de observaciones independientes  $x_{ij}$ , a partir de  $k$  poblaciones normales con medias  $\mu_1, \dots, \mu_k$  y varianza común desconocida  $\sigma^2$ .

La observación  $i$  de la población  $j$  viene generada por el modelo lineal

$$x_{ij} = \mu + \gamma_j + e_{ij}$$

donde  $\mu$  es un parámetro general común a todas las observaciones,  $\gamma_j$  se debe al tratamiento asociado con la población  $j$  y  $e_{ij}$  es el error. Los valores  $e_{ij}$  para todos los pares  $(i, j)$  son valores aleatorios distribuidos independientemente de forma normal con media nula y varianza  $\sigma^2$ .

Reparametrizando el modelo podemos escribir

$$x_{ij} = \mu_j + e_{ij}, \text{ con } \mu_j = \mu + \gamma_j.$$

Introduciendo el concepto de matriz-diseño, tenemos un medio para escribir el modelo lineal de todas las observaciones en forma matricial.

Si el número de unidades muestrales de la población  $j$  es  $N_j$  y establecemos  $N = N_1 + N_2 + \dots + N_k$ , podemos escribir las observaciones y errores (residuos) como vectores de  $N$  componentes,

$$\mathbf{x}' = (x_{11} \dots, x_{N_1}, \dots, x_{1k} \dots, x_{N_k}),$$

$$\mathbf{e}' = (e_{11} \dots, e_{N_1}, \dots, e_{1k} \dots, e_{N_k}),$$

y los parámetros como

$$\boldsymbol{\mu}' = (\gamma_1 \gamma_2, \dots, \gamma_k, \mu).$$

---

(142) Para el desarrollo de este capítulo nos hemos apoyado en gran manera en notas internas y comunicaciones directas recibidas de la profesora SITGREAVES, R., en el curso Education 351A-Multivariate Analysis, Stanford University, California, 1973-1974.

La matriz-diseño para este modelo es

$$A = \begin{pmatrix} 1 & 0 & . & . & . & . & 0 & 1 \\ . & . & . & . & . & . & . & 1 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & 0 & . & . & . & . & . & . \\ 0 & 1 & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & 1 & . & . & . & . & . & . \\ . & 0 & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & 0 & . \\ . & . & . & . & . & . & 1 & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & . & 1 & 1 \end{pmatrix}$$

La matriz-diseño  $A$  está compuesta de  $k$  submatrices  $[N, x(k+1)]$  de forma que en la submatriz  $j$ , la columna  $j$  y la última contiene un *uno* en cada posición, mientras el resto son ceros, y, por lo tanto, el rango de  $A$  es  $k$ . Las  $k$  primeras columnas son linealmente independientes y forman una base de la matriz-diseño.

La distribución de unos y ceros en cualquier matriz-diseño va a depender del plan experimental y de su modelo lineal. En el caso del análisis de covarianza contendrá asimismo a los valores de la variable o variables concomitantes, y en el análisis de regresión los unos y ceros serán sustituidos por los valores de las variables independientes, variables que serán en general continuas en vez de discretas como en el análisis de varianza.

Con esta formulación matricial el modelo se expresa

$$X = A\mu + e$$

Si nos movemos al caso multivariante, nuestra variable dependiente es una matrix  $X$  de dimensiones  $(N \times p)$  compuesta de  $N$  conjuntos de observaciones independientes, tal que los elementos de una fila tienen una distribución multinormal con matriz común de covarianza  $\Sigma$  de dimensiones  $(p \times p)$  (143).

---

(143) MORRISON, D. F.: *Multivariate Statistical Methods*, New York, McGraw Hill, 1967, apartado 3.3.

El modelo lineal para la observación  $i$  en la respuesta  $h$  será

$$x_{ih} = a_{i1} \xi_{1h} + \dots + a_{iq} \xi_{qh} + e_{ih}.$$

Cada respuesta en el vector  $i$  tiene los mismos coeficientes  $a_{ij}$ , de forma que la misma matriz  $A$  es válida con vector de medias nulo y matriz de covarianza  $\Sigma$ .

En notación matricial nuestro modelo se expresa

$$[1] \quad X = A\xi + E^*,$$

donde la matriz  $A$  tiene dimensiones  $(N \times q)$ , la matriz de parámetros no conocidos  $\xi$ ,  $(q \times p)$  y la matriz de residuos  $E^*$  tiene obviamente las mismas dimensiones que la matriz  $X$ , esto es  $(N \times p)$ .

Calculando el valor esperado, nuestro modelo se escribe en la forma más usual

$$[2] \quad E(X) = A\xi.$$

Si el rango de  $A$  es  $r < q$  podemos subdividir  $A$  de la forma  $A = (A_1, A_2)$  donde  $A_1$  es una matriz base de  $A$  de dimensiones  $(N \times r)$  y  $A_2$  es una matriz de dimensiones  $N \times (q - r)$ .

Como  $A_1$  es base de  $A$  es posible encontrar una matriz  $C$  de dimensiones  $r \times (q - r)$  tal que  $A_2 = A_1 C$ .

Igualmente podemos dividir  $\xi$  de forma que

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix},$$

con  $\xi_1$  de dimensiones  $(r \times p)$  y  $\xi_2$  de dimensiones  $(q - r) \times p$ .

El modelo inicial puede reparametrizarse en la forma

$$[3] \quad A\xi = A_1\beta,$$

donde  $\beta = \xi_1 + C \xi_2$  es una matriz de parámetros desconocidos de dimensiones  $(r \times p)$ .

El modelo se convierte en

$$[4] \quad E(X) = A_1\beta.$$

La densidad conjunta de todas las observaciones se puede escribir

$$[5] \quad P(X) = \frac{|\Sigma|^{-N/2}}{(2\pi)^{pN/2}} \exp \left[ -\frac{1}{2} \text{tr} \left( (X - A_1 \beta)' \Sigma^{-1} (X - A_1 \beta)' \right) \right],$$

esta fórmula es una extensión para la distribución normal de multivariadas de la que expresa la densidad de una distribución normal simple (144).

$$n(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$

$$\text{Como } \text{tr} \left( (X - A_1 \beta)' \Sigma^{-1} (X - A_1 \beta)' \right) = \text{tr} \left( (X - A_1 \beta)' (X - A_1 \beta)' \Sigma^{-1} \right),$$

tenemos que

$$[6] \quad P(X) = \frac{|\Sigma|^{-N/2}}{(2\pi)^{pN/2}} \exp \left[ -\frac{1}{2} \text{tr} \left( (X - A_1 \beta)' (X - A_1 \beta)' \Sigma^{-1} \right) \right]$$

Sin introducir restricciones sobre los parámetros, los estimadores de máxima verosimilitud de  $\beta$  y  $\Sigma$  son (145)

$$[7] \quad \hat{\beta} = (A_1' A_1)^{-1} A_1' X,$$

$$[8] \quad \hat{\Sigma} = \frac{1}{N} (X - A_1 \hat{\beta})' (X - A_1 \hat{\beta}) = \frac{1}{N} (X'X - X'A_1 (A_1' A_1)^{-1} A_1' X)$$

Transformando el término exponencial de [6] obtenemos

$$[9] \quad P(X) = \frac{|\Sigma|^{-N/2}}{(2\pi)^{pN/2}} \exp \left[ -\frac{1}{2} \text{tr} \left\{ \left( (X - A_1 \hat{\beta})' (X - A_1 \hat{\beta}) + (\hat{\beta} - \beta)' A_1' A_1 (\hat{\beta} - \beta) \right) \Sigma^{-1} \right\} \right]$$

$A_1' A_1$  es una matriz definida positiva, por lo que podemos realizar la transformación ortogonal

$$A_1' A_1 = T T', \text{ donde } T \text{ es no singular,}$$

$$\text{cumpliéndose que } (A_1' A_1)^{-1} = (T')^{-1} T^{-1}.$$

(144) MOOD, A. M. y GRAYBILL, F. A.: *Introduction to Theory of Statistics*, New York, McGraw Hill, 1963, 123.

(145) Una discusión de este tema puede consultarse en ROY, S. N.: *Some aspects of Multivariate Analysis*, New York, Wiley, 1957.



Supongamos que

$$[10] \quad T^{-1} A_1 = O_2,$$

donde  $T$  tiene dimensiones  $(r \times r)$  y  $O_2$ ,  $(r \times N)$ , entonces

$$O_2 O_2' = T^{-1} A_1 A_1 (T')^{-1} = T^{-1} T T' (T')^{-1} = I.$$

Podemos construir una matriz  $O_1$  de dimensiones  $(N-r) \times N$  tal que

$$[11] \quad \begin{cases} O_1 O_1' = I \\ O_1 O_2' = O \end{cases}$$

Establezcamos  $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$  de dimensiones  $(N \times p)$  tal que

$$[12] \quad \begin{cases} Y_1 = O_2 X \\ Y_2 = O_1 X \end{cases}$$

Con lo que a partir de [10] obtenemos

$$[13] \quad E(Y_1) = O_1 A_1 \beta = O_1 O_2' T' \beta = O,$$

$$[14] \quad E(Y_2) = O_2 A_1 \beta = T^{-1} A_1 A_1 \beta = T' \beta.$$

La transformación ortogonal conserva la independencia de las filas de  $Y$  y la matriz de covarianza para los  $p$  componentes de una fila dada de  $Y$  sigue siendo  $\Sigma$ .

La densidad de los componentes de  $Y$  es

$$[15] \quad P(Y) = \frac{|\Sigma|^{-N/2}}{(2\pi)^{pN/2}} \exp \left[ -\frac{1}{2} \operatorname{tr} \left\{ Y_1' Y_1 + (Y_2 - T' \beta)' (Y_2 - T' \beta) \Sigma^{-1} \right\} \right]$$

$$= \frac{|\Sigma|^{-N/2}}{(2\pi)^{pN/2}} \exp \left[ -\frac{1}{2} \operatorname{tr} \left( (Y_1' Y_1) \Sigma^{-1} \right) \right]$$

$$\times \frac{|\Sigma|^{-r/2}}{(2\pi)^{pr/2}} \exp \left[ -\frac{1}{2} \operatorname{tr} \left( (Y_2 - T' \beta)' (Y_2 - T' \beta) \Sigma^{-1} \right) \right]$$

$Y'Y$  es una matriz aleatoria de dimensiones  $(p \times p)$  con una distribución Wishart, con matriz de covarianza  $\Sigma$  y  $(N - r)$  grados de libertad (146, 147). En rigor, es necesario un tercer parámetro, la dimensionalidad de la normal, para definir la distribución Wishart (148), sin embargo, nosotros la especificaremos por dos, siguiendo a Anderson y Morrison.

Si escribimos  $O_* = \begin{bmatrix} O_1 \\ O_2 \end{bmatrix}$  en la que  $O_* O_*' = O_*' O_* = I$ ; tenemos que

$$[16] \quad Y'Y = X' O_*' O_* X = X' X,$$

pero como

$$Y'Y = Y_1' Y_1 + Y_2' Y_2$$

tenemos que

$$Y_1' Y_1 = X' X - Y_2' Y_2.$$

Por otra parte

$$[17] \quad Y_2' Y_2 = X' O_2' O_2 X = X' A_1 (T')^{-1} T^{-1} A_1 X = \\ = X' A_1 (A_1' A_1)^{-1} A_1 X,$$

con lo que

$$[18] \quad Y_1' Y_1 = X' X - X' A_1 (A_1' A_1)^{-1} A_1 X$$

y por lo tanto  $X' X - X' A_1 (A_1' A_1)^{-1} A_1 X$  tiene una distribución Wishart con matriz de covarianza  $\Sigma$  y  $(N - r)$  grados de libertad.

Por transformaciones algebraicas llegamos asimismo a que

$$[19] \quad (X - A_1 \hat{\beta})' (X - A_1 \hat{\beta}) = X' X - X' A_1 (A_1' A_1)^{-1} A_1 X$$

con lo que  $(X - A_1 \hat{\beta})' (X - A_1 \hat{\beta})$  es asimismo una matriz aleatoria  $(p \times p)$  con una distribución Wishart con matriz de covarianza  $\Sigma$  y  $(N - r)$  grados de libertad.

Aplicando [17] obtenemos

$$[20] \quad (Y_2 - T' \hat{\beta})' (Y_2 - T' \hat{\beta}) = (\hat{\beta} - \beta)' A_1' A_1 (\hat{\beta} - \beta), \text{ y por lo tanto}$$

$$[21] \quad Y_2 = T' \hat{\beta}.$$

(146) ANDERSON, T. W.: *An Introduction...* (o. c.), capítulo 7.

(147) MORRISON, D. F.: *Multivariate...* (o. c.), 98.

(148) WINER, B. J.: *Statistical Principles...* (o. c.), 93.

Tras este análisis contamos con suficientes conocimientos sobre la forma de las distribuciones que entran en juego como para intentar abordar el problema de contrastar hipótesis lineales.

## 5.2. Contraste de hipótesis lineales multivariantes

Volvamos a nuestro modelo  $X = A_1\beta + E$  y consideremos el problema de contrastar una hipótesis lineal multivariante.

La hipótesis es contrastable si puede representarse en la forma (149).

$$[22] \quad H_0 : C_1 \beta M = O,$$

donde  $C_1$  y  $M$  son matrices especificadas por el investigador de dimensiones  $(g \times r)$  y  $(p \times u)$ , respectivamente.

Recordemos que no estamos trabajando con el modelo más general [1] sino que hemos introducido una partición en la matriz  $A$  y que  $C_1$  representa la partición correspondiente de una matriz  $C$  aplicable a dicho modelo general.

La hipótesis alternativa es

$$[23] \quad H_1 : C_1 \beta M \neq O.$$

Consideremos en primer lugar el caso en que  $M$  es la matriz identidad de dimensiones  $(p \times p)$ . En este caso, cuando  $H_0$  es cierta se cumple que

$$C_1 \beta = O,$$

existiendo una matriz  $G$  de dimensiones  $(g \times g)$  tal que

$$[24] \quad G C_1 (T')^{-1} = O_1^*,$$

donde  $O_1^*$  es una matriz de dimensiones  $(g \times r)$  tal que

$$[25] \quad O_1^* O_1^{*'} = I.$$

Asimismo, podemos construir una matriz  $O_2^*$  de dimensiones  $(r - g) \times r$  tal que se cumpla

---

(149) Véase BOSE, R. C.: "The fundamental theorem of linear estimation", *Proceedings of 31<sup>st</sup> Indian Scientific Congress*, 1944.

$$[26] \quad O_1^* O_2^{*'} = O.$$

$$O_2^* O_2^{*'} = I.$$

$$\text{De esta forma, la matriz } O^* = \begin{bmatrix} O_1^* \\ O_2^* \end{bmatrix}$$

es ortogonal.

Volviendo al conjunto de variables  $Y_2$  de dimensiones  $(r \times p)$ , bajo la transformación

$$[27] \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} O_1^* \\ O_2^* \end{bmatrix} Y_2$$

y usando [14] se llega a

$$[28] \quad E(Z_2) = O_2^* T' \beta,$$

y cuando  $H_0$  es cierta a partir de [24] a

$$[29] \quad E(Z_1) = O_1^* T' \beta = G C_1 (T')^{-1} T' \beta = G C_1 \beta = O.$$

En estas condiciones con  $H_0$  cierta, la densidad multivariante original [15] se transforma en

$$[30] \quad P(Y_1, Z_1, Z_2) = \frac{|\Sigma|^{-\frac{(N-r)}{2}}}{(2\pi)^{\frac{p(N-r)}{2}}} \exp \left[ -\frac{1}{2} \text{tr} \left( (Y_1' Y_1) \Sigma^{-1} \right) \right] \times \\ \times \frac{|\Sigma|^{-g/2}}{(2\pi)^{pg/2}} \exp \left[ -\frac{1}{2} \text{tr} \left( (Z_1' Z_1) \Sigma^{-1} \right) \right] \times \\ \frac{|\Sigma|^{-\frac{(r-g)}{2}}}{(2\pi)^{\frac{p(r-g)}{2}}} \exp \left[ -\frac{1}{2} \text{tr} \left( (Z_2 - O_2^* T' \beta)' (Z_2 - O_2^* T' \beta) \Sigma^{-1} \right) \right]$$

Por lo tanto, cuando  $H_0$  es cierta, la matriz aleatoria  $Z_1' Z_1$  de dimensiones  $(p \times p)$  tiene una distribución Wishart con matriz de covarianza  $\Sigma$  y  $g$  de grados de libertad, siendo dicha distribución independiente de la de  $Y_1' Y_1$ .

En este caso,  $Z_1' Z_1$  representa la suma de cuadrados y productos cruzados de  $g$  observaciones independientes de un vector  $p$ -dimensional cuyos componentes tienen una distribución normal multivariante con medias nulas y matriz de covarianza  $\Sigma$ .

Como  $Y_1' Y_1$  y  $Z_1' Z_1$  son independientes y tienen ambos una distribución Wishart, entonces (150)  $Y_1' Y_1 + Z_1' Z_1$  es una matriz  $(p \times p)$  distribuida de forma Wishart con matriz de covarianza  $\Sigma$  y  $(N - r + g)$  grados de libertad, cumpliéndose que  $Z_1' Z_1 = Y_2' O_1^* O_1^* Y_2$

Volvamos al caso más general de hipótesis

$$H_0 : C_1 \beta M = O$$

$$H_1 : C_1 \beta M \neq O.$$

Consideremos la variable  $Z_1$  definida en [27] de la forma  $Z_1 = O_1^* Y_2$ . Las  $g$  filas de  $Z_1$  están distribuidas independientemente unas de otras y los  $p$  componentes de cada una de ellas tienen una distribución normal multivariante con matriz de covarianza  $\Sigma$ .

Sin restricciones de hipótesis se cumple que

$$[31] \quad E(Z_1) = G C_1 \beta.$$

Llevando a cabo la transformación

$$[32] \quad Z_1^* = Z_1 M,$$

como los componentes de cada fila de  $Z_1^*$  son combinación lineal de una fila de  $Z_1$  las  $g$  filas de  $Z_1^*$  están distribuidas independientemente unas de otras y las  $u$  componentes de cada una de ellas tienen una distribución normal multivariante con matriz de covarianza  $M' \Sigma M$  (151).

Cuando  $H_0$  es cierta, a partir de [29] se cumple

$$[33] \quad E(Z_1^*) = E(Z_1 M) = G C_1 \beta M = O.$$

(150) ANDERSON, T. W.: *An Introduction...* (o. c.), 161.

(151) MORRISON, D. F.: *Multivariate...* (o. c.), 86.

Llevando a cabo una transformación similar con  $Y_1$  tenemos

$$[34] \quad Y_1^* = Y_1 M.$$

Al igual que con  $Z_1^*$ , las  $(N - r)$  filas de  $Y_1^*$  están distribuidas independientemente unas de otras y los  $u$  componentes de cada una de ellas tienen una distribución normal multivariante con matriz de covarianza  $M' \Sigma M$  y valor esperado cero de acuerdo con [13].

La distribución conjunta de  $Y_1^*$  y  $Z_1^*$  será de la forma

$$[35] \quad P(Y_1^*, Z_1^*) = \frac{|M' \Sigma M|^{-\frac{(N-r)}{2}}}{(2\pi)^{\frac{u(N-r)}{2}}} \exp \left[ \frac{1}{2} \text{tr} [Y_1^{*'} Y_1^* (M^{-1} \Sigma^{-1} (M')^{-1})] \right] \\ \times \frac{|M' \Sigma M|^{-u/2}}{(2\pi)^{u/2}} \exp \left[ \frac{1}{2} \text{tr} [(Z_1^{*'} - GC_1 \hat{\beta} M)' (Z_1^* - GC_1 \hat{\beta} M) (M^{-1} \Sigma^{-1} (M')^{-1})] \right]$$

Cuando  $H_0$  es cierta, usando [33], tenemos que

$$[36] \quad P(Z_1^*) = \frac{|M' \Sigma M|^{-u/2}}{(2\pi)^{u/2}} \exp \left[ \frac{1}{2} \text{tr} [Z_1^{*'} Z_1^* (M^{-1} \Sigma^{-1} (M')^{-1})] \right]$$

Volviendo a los datos originales  $X$ , tenemos

$$[37] \quad Y_1^{*'} Y_1^* = M' [X' X - \hat{\beta}' (A_1' A_1) \hat{\beta}] M = \\ = M' [X' X - X' A_1 (A_1' A_1)^{-1} A_1' X] M = E$$

$$[38] \quad Z_1^{*'} Z_1^* = M' \hat{\beta}' C_1' [C_1 (A_1' A_1)^{-1} C_1'] C_1 \hat{\beta} M = \\ = M' X' A_1 (A_1' A_1)^{-1} C_1' [C_1 (A_1' A_1)^{-1} C_1']^{-1} (A_1' A_1)^{-1} A_1' X M = H.$$

Las matrices simétricas  $E$  y  $H$  son generalizaciones de las sumas de cuadrados y productos cruzados *intra* y *entre* grupos respectivamente, de forma que sus términos diagonales son las formas cuadráticas del denominador y numerador respectivamente del estadístico  $F$  calculado por Roy (152) para cada una de las  $u$

(152) Véase ROY, S. N.: "On a heuristic method of test contribution and its uses in multivariate analysis", *Annals of Mathematical Statistics*, 24 (1953), 22-230.

variables. Por lo tanto, cualquier hipótesis univariada puede generalizarse a la situación multivariante si reemplazamos las sumas de productos por sus extensiones matriciales (153).

Del tipo de distribución de  $Y_1^*$  se desprende que, por razonamientos análogos a los desarrollados con anterioridad, los elementos de E tienen una distribución Wishart con matriz de covarianza  $M'\Sigma M$  y  $(N - r)$  grados de libertad.

Asimismo, las filas de  $Z_1^*$  están distribuidas independientemente unas de otras y de la matriz E y sus componentes tienen una densidad multinormal con matriz de covarianza  $M'\Sigma M$  y valores esperados nulos cuando  $H_0$  es cierta. Por lo tanto, cuando  $H_0$  es cierta, los elementos de H representan las sumas de productos de  $g$  observaciones independientes de un vector  $u$  - dimensional con valores esperados nulos y matriz de covarianza  $M'\Sigma M$ . El rango de H es  $s$ , donde:

$$[39] \quad s = \min(g, u).$$

### 5.3. Procedimientos de contraste de hipótesis

Hasta el momento hemos examinado los efectos de una hipótesis nula del tipo  $H_0: C_1\beta M = 0$  sobre la densidad de probabilidad de las observaciones.

Por medio de transformaciones sucesivas hemos llegado a las matrices muestrales E y H que son estadísticos suficientes para abordar el problema de contrastar hipótesis lineales multivariantes.

La búsqueda de tests estadísticos para el contraste de  $H_0$  ha sido enfocada de forma diversa por distintos investigadores. Por otra parte, este tema que presenta ciertas dificultades debido a la complejidad de las distribuciones de los estadísticos en juego parece ser un campo de investigación del máximo interés de la investigación actual, apareciendo continuamente valiosas aportaciones.

Sitgreaves (154) habla de dos enfoques generales: 1) los proce-

---

(153) MORRISON, D. F.: *Multivariate...* (o. c.), 166.

(154) Véase nota (142).

dimientos de "análisis canónico" que tratan con funciones de las raíces de  $HE^{-1}$ , y 2) los procedimientos de *step-down*, que consisten en series de tests univariados.

Nosotros nos limitaremos a discutir brevemente los tests que más aparecen en la literatura y que están dentro del enfoque que Sitgreaves denomina *análisis canónico*.

#### A. El criterio de la razón de verosimilitud generalizada

Este criterio fue introducido en primer lugar por Wilks (155) para el contraste de medias de vectores de varias poblaciones, siendo extendido posteriormente para el caso de coeficientes de regresión por el propio Wilks (156) y por Bartlett (157).

En el sustrato de este procedimiento se encuentra la idea de que la generalización multivariante del análisis univariante que Rao (158) denomina *análisis de dispersión*, supone la partición de la matriz de suma de productos de forma idéntica al caso univariado y el test de contraste resultante viene a ser una generalización del test F para dicho caso.

El test estadístico resultante, según este criterio, es (159):

$$[40] \quad \Lambda = \frac{|\hat{\Sigma}_{\Omega}|^{N/2}}{|\hat{\Sigma}_{\omega}|^{N/2}} = \frac{|E|^{N/2}}{|H + E|^{N/2}}$$

de donde

$$[41] \quad \Lambda^{2/N} = \frac{|E|}{|H + E|} = \frac{1}{|HE^{-1} + I|} = U.$$

El estadístico U es la razón de dos determinantes o varianzas generalizadas, donde E es la matriz (promediada) de suma de

(155) WILKS, S. S.: "Certain generalizations in the analysis of variance", *Biometrika*, 24 (1932), 471-494.

(156) WILKS, S. S.: "Moment-generating operators for determinants of product moments in samples from a normal system", *Annals of Mathematics*, 135 (1935), 312-340.

(157) BARTLETT, M. S.: "The vector representation of a sample", *Proceedings of the Cambridge Philosophical Society*, 30 (1934), 327-340.

(158) Véase RAO, C. R.: *Advanced Statistical Methods in Biometric Research*, New York, Wiley, 1952.

(159) ANDERSON, T. W.: *An Introduction...* (o. c.), 188.



productos intra-grupos y H la matriz de suma de productos entre-grupos.

Si se concibe el valor de un determinante como el índice escalar de la varianza multivariante generalizada, el estadístico U viene a ser la razón entre la varianza intra-grupos y la varianza total (160), de forma que U será tanto menor cuanto mayor diferencia real exista entre grupos.

El cálculo del estadístico U es sencillo pero su distribución es compleja, por eso se trabaja con distribuciones alternativas de más fácil manejo.

Rao (161) ha demostrado que cuando  $H_0$  es cierta, el estadístico  $\frac{1 - U^{1/u}}{U^{1/u}} \cdot \frac{h}{g \cdot u}$  tiene una distribución aproximada F con  $(g \times u)$  y  $(h)$  grados de libertad, donde

$$t = \begin{cases} \sqrt{\frac{u^2 g^2 - 4}{u^2 + g^2 - 5}}, & \text{si } g^2 + u^2 \neq 5. \\ 1, & \text{si } g^2 + u^2 = 5, \end{cases}$$

y

$$h = \left[ N - r - \frac{(u - g + 1)}{2} \right] t - \frac{u \cdot g}{2} + 1.$$

Este estadístico tiene una distribución exacta F si

$$s = \min(u, g) = 1, 2.$$

Recordemos que  $g$  y  $u$  son el número de filas y columnas, respectivamente, de la matriz nula de la hipótesis [22].

## B. Test basado en la mayor raíz de $HE^{-1}$

Si se parte del principio de unión-intersección de Roy (162) en vez del de la razón de verosimilitud, se obtiene un test basado en la mayor raíz de  $HE^{-1}$ . Al usar este principio se acepta  $H_0$  si

(160) OVERALL, J. E. y KLETT, C. J.: *Applied Multivariate Analysis*, New York, McGraw Hill, 1972, 316.

(161) RAO, C. R.: *Linear Statistical Inference and its Applications*, New York, Wiley, 1965, 471.

(162) ROY, S. N.: "On a heuristic..." (o. c.), 220-230.

y sólo si se aceptaría toda hipótesis univariada  $H_0: C_1 \beta M a = 0$  para todo vector no nulo  $(a)$  de  $u$  componentes (163).

Para una hipótesis univariada del tipo anterior el test estadístico es

$$[42] \quad F(a) = \frac{N-r}{g} \cdot \frac{a' Ha}{a' Ea}, \quad \text{de forma que}$$

$H_0$  es aceptada si  $F(a) \leq$  valor crítico y rechazada si  $F(a) >$  valor crítico.

Para la hipótesis multivariante  $H_0: C_1 \beta M = 0$ , la región de aceptación es equivalente a la definida por

$$\max_a F(a) \leq \text{valor crítico}$$

$HE^{-1}$  tiene  $s = \min(g, u)$  raíces no nulas, que en orden creciente denominamos  $c_1, c_2 \dots c_s$ .

El valor máximo de  $F(a)$  es proporcional a  $c_s$  y la hipótesis multivariante es aceptada si

$$c_s \leq \text{valor crítico.}$$

Cuando  $u$  o  $g$  son la unidad,  $HE^{-1}$  sólo tiene una raíz no nula, llegándose a un test equivalente al de la razón de verosimilitud.

En la práctica no se trabaja con  $c_s$  sino con  $\theta_s$ , mayor raíz de  $H(H + E)^{-1}$ , debido a la existencia de gráficas tabuladas de la distribución de  $\theta_s$  elaboradas por Heck (164).

$$[43] \quad \theta_s = \frac{c_s}{1 + c_s}$$

Si la hipótesis multivariante nula es cierta, los únicos parámetros de la distribución de  $\theta_s$  son (165)

(163) MORRISON, D. F.: *Multivariate...* (o. c.), 165.

(164) HECK, D. L.: "Charts of some upper percentage points of the distribution of the largest characteristic root", *Annals of Mathematical Statistics*, 31 (1960), 625-642.

(165) MORRISON, D. F.: *Multivariate...* (o. c.), 109.

$$s = \min (g, u),$$

$$m = \frac{|g - u| - 1}{2},$$

$$n = \frac{N - r - u - 1}{2}.$$

$H_0$  es aceptada si  $\theta_s \leq$  valor crítico, donde dicho valor crítico se obtiene de las gráficas de Heck.

### C. Test basados en la suma de raíces

Lawley (166) y Hoteiling (167) introdujeron la suma de raíces de  $HE^{-1}$  como el test estadístico de la hipótesis lineal multivariante. Este es

$$[44] \quad \sum_{i=1}^s c_i = \text{tr } HE^{-1}.$$

La distribución de este estadístico es compleja pero cuando  $H_0$  es cierta, la distribución límite de  $N \text{tr } HE^{-1}$  es la  $\chi^2$  con  $(g \cdot u)$  grados de libertad (168).

Anderson (169) presenta una extensión de este criterio a la suma de las raíces de  $H(H + E)^{-1}$  llegando al estadístico:

$$[45] \quad \sum_{i=1}^s \theta_i = \sum_{i=1}^s \frac{c_i}{1 + c_i} = \text{tr } H(H + E)^{-1}$$

donde la distribución límite de  $N \text{tr } H(H + E)^{-1}$  cuando  $H_0$  es cierta, es asimismo una  $\chi^2$  con  $(g \cdot u)$  grados de libertad.

---

(166) LAWLEY, D. M.: "A generalization of Fisher's Z-test", *Biometrika*, 30 (1938), 180-187.

(167) HOTEILING, H.: "Multivariate quality control, illustrated by the air testing of sample bombrights", en *Selected Techniques of Statistical Analysis*, Eisenhart, C. y al. (ed.), New York, McGraw Hill, 1947, 11-184.

(168) MORRISON, D. F.: *Multivariate...* (o. c.), 198.

(169) ANDERSON, T. W.: *An Introduction...* (o. c.), 224.

#### 5.4. El análisis de covarianza

El análisis presentado nos ofrece un modelo general de análisis multivariante aplicable directamente al estudio de la varianza, covarianza y regresión. Hemos apuntado como la única diferencia en uno u otro caso estará en la forma de la matriz-diseño  $A_1$  y en el correspondiente significado de los elementos de la matriz de parámetros de regresión.

A modo de ilustración diremos que cuando trabajamos con covariables, nuestro modelo sin ellas  $E(X) = A_1\beta$ , se convierte en

$$[46] \quad E(X) = A_1\beta + P\beta^*, \text{ esto es}$$

$$E(X) = (A_1, P) \begin{pmatrix} \beta \\ \beta^* \end{pmatrix} = k\Omega,$$

donde

$$k = (A_1, P)$$

y

$$\Omega = \begin{pmatrix} \beta \\ \beta^* \end{pmatrix} = (M' k)^{-1} k' X.$$

En la práctica, y así se recoge en nuestro análisis de datos, todo se reduce a introducir en la matriz  $A_1$  las covariables como columnas añadidas y llevar a cabo el análisis de forma idéntica.

En otras palabras, nuestra matriz-diseño compuesta de variables vectoriales discretas se ve aumentada con una o varias variables vectoriales continuas (este es el caso más general). Cuando se trate de regresión, todas las columnas de  $A_1$  serán generalmente variables vectoriales continuas.



## 6. ANALISIS DE DATOS

### 6.1. Planteamiento inicial

Las características de nuestros diseños experimentales, tipos de variables en juego y naturaleza de los efectos que pretendíamos analizar, nos presentaban al modelo de análisis multivariante que acabamos de introducir en el capítulo anterior, como el más idóneo y poderoso para tomar decisiones con base científica sobre los efectos y resultados de nuestros experimentos. Tengamos presente que la variable rendimiento escolar tiene una distribución teórica que concuerda con el presupuesto básico del modelo estudiado.

La generalidad del modelo de análisis facilitaba nuestra labor que iba a estar centrada casi exclusivamente en la búsqueda de la matriz-diseño de cada uno de los cuatro análisis.

Para diseñar nuestras matrices  $A_i$  nos apoyamos en el trabajo que previamente habíamos realizado tutelados por la profesora Sitgreaves, al analizar los resultados de un estudio piloto sobre *Memoria de reconocimiento de dibujos y nombres* (170).

El análisis de datos fue llevado a cabo en el Centro de Cálculo de la Facultad de Ciencias de la Universidad de Zaragoza, con el programa BMD11V (BMDX63 en la antigua denominación) (171) que estudia las hipótesis lineales generales multivariantes.

### 6.2. Descripción del programa (BMD 11V)

Este programa estudia la regresión múltiple para multivaria-

---

(170) ESCUDERO, T.: *Final Project*, Education, 493A, Stanford University, California, 1974.

(171) DIXON, W. J. (ed.): *BMD-Biomedical Computer Programs*, Los Angeles, University of California Press, 1973.

bles dependientes según el modelo  $X = A_1\beta + E$ . Calcula el estadístico U y estadístico aproximado F, siguiendo el enfoque de Rao (172) que se apoya en el criterio de la razón de verosimilitud, expuesto en el capítulo anterior, para el contraste de hipótesis de la forma  $C_1\beta M = D$ .

Hasta el momento nosotros hemos identificado siempre la matriz D con la matriz nula, ya que esta es la situación en la mayor parte de los casos, y además D es siempre nula en las hipótesis que contrastamos en nuestros experimentos. Sin embargo, el programa trabaja con matrices D, diseñadas por el investigador, no necesariamente nulas.

Las matrices  $C_1$ , M y D deben ser diseñadas y especificadas para cada una de las hipótesis que se desee contrastar.

El out-put del programa incluye:

1. La matriz de suma de productos

$$(A_1, X)' (A_1, X) = \begin{bmatrix} A_1' A_1 & A_1' X \\ X' A_1 & X' X \end{bmatrix}$$

2. La matriz de coeficientes de regresión  $\hat{\beta} = (A_1' A_1)^{-1} A_1' X$ .

3. La matriz de suma de productos de residuos

$$R = X' X - X' A_1 \hat{\beta} = X' X - X' A_1 (A_1' A_1)^{-1} A_1' X.$$

La matriz R ha sido denominada en el capítulo anterior como  $Y_1' Y_1$ .

4. Para cada hipótesis, el programa especifica las matrices  $C_1$ , M y D seguidas de

a)  $L = C_1 \hat{\beta} M - D$ .

b)  $V = C_1 (A_1' A_1)^{-1} C_1'$ .

c)  $E = M' R M = M' [X' X - X' A_1 (A_1' A_1)^{-1} A_1' X] M$ .

d)  $H = L' V^{-1} L = (C_1 \hat{\beta} M - D)' [C_1 (A_1' A_1)^{-1} C_1']^{-1} (C_1 \hat{\beta} M - D)$ .

Cuando D es nula H toma la forma que aparece en la fórmula [38] del capítulo anterior.

---

(172) Véase nota (161).

e) determinante  $|E| = d_1$ .

f) determinante  $|E + H| = d_2$ .

g) estadístico  $U = \frac{d_1}{d_2}$  con grados de libertad  $(u, g, N - r)$ ,

en donde

$u = n.$ º de columnas de  $M$

$g = n.$ º de filas de  $C_1$

$r = n.$ º de variables independientes (columnas de  $A_1$ )

$N = n.$ º de observaciones.

h) estadístico  $F$  (aproximado)  $= \frac{1 - U^{1/t}}{U^{1/t}} \cdot \frac{h}{g u}$ , con grados de libertad  $[(g \cdot u), h]$ , donde

$$h = \left[ N - r - \frac{(u - g + 1)}{2} \right] t - \frac{u g}{2} + 1 \quad y$$

$$t = \begin{cases} \left\lceil \frac{\sqrt{u^2 g^2 - 4}}{u^2 + g^2 - 5} \right\rceil, & \text{si } u^2 + g^2 \neq 5. \\ 1, & \text{si } u^2 + g^2 = 5. \end{cases}$$

Debemos recordar que el estadístico  $F$  calculado en todas las hipótesis contrastadas es exacto y no aproximado porque siempre se ha dado el caso que  $u$  o  $g$  era la unidad.

La decisión de aceptar o rechazar nuestras hipótesis nulas ha sido siempre con base al criterio de comparar el estadístico  $F$  que nos calculaba el programa con el valor crítico correspondiente a partir de las tablas de la distribución  $F$  (173).

La hipótesis era aceptada si

$$F \leq \text{valor crítico}$$

y rechazada si

$$F > \text{valor crítico.}$$

Ya hemos señalado que todas nuestras hipótesis  $H_0$  han sido planteadas con matriz  $D = 0$ , de forma que nuestra hipótesis de partida es la *no existencia* de efecto. En los casos en que la decisión ha sido la de rechazar nuestra hipótesis nula, algún efecto se ha producido de forma estadísticamente significativa.

---

(173) MORRISON, D. F.: *Multivariate...* (o. c.), 310.



### 6.3. Naturaleza de los datos de entrada

En nuestros cuatro experimentos, la multivariable dependiente de entrada es una matriz  $X$  de dimensiones  $(N \times p)$ , donde  $N$  es el número de observaciones independientes y  $p$  es el número de univariabes que componen la multivariable.

Los valores de  $N$  y  $p$  para los experimentos han sido:

	N	p
SANBRU	58	2
JUSREB	56	2
CORBIZU	50	3
BLATUR	75	2

En los tres experimentos donde  $p$  es dos, las  $N$  observaciones incluyen las calificaciones del examen-criterio en teoría y problemas y en el que  $p$  es tres se incluye además las calificaciones de cuestiones cortas (test).

Para los tres experimentos realizados en la Facultad de Ciencias (SANBRU, JUSREB y CORBIZU) y, que responden a un diseño factorial  $(2 \times 2)$ , un elemento genérico de la matriz  $X$  es de la forma  $X_{ijkl}$ , donde

$i = 1$  (tratamiento),  $2$  (no-tratamiento)

$j = 1$  (rendimiento previo alto),  $2$  (rendimiento previo bajo)

$k = 1$  (teoría),  $2$  (problemas),  $3$  (test)

$l = 1, 2, \dots, n_{ij}$  (número de sujetos en cada celda).

Con esta base, la matriz de entrada del experimento CORBIZU es la siguiente

$$X = \begin{pmatrix} X_{1111} & X_{1121} & X_{1131} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ X_{111n_{11}} & X_{112n_{11}} & X_{113n_{11}} \\ X_{1211} & X_{1221} & X_{1231} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ X_{121n_{12}} & X_{122n_{12}} & X_{123n_{12}} \\ X_{2111} & X_{2121} & X_{2131} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ X_{211n_{21}} & X_{212n_{21}} & X_{213n_{21}} \\ X_{2211} & X_{2221} & X_{2231} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ X_{221n_{22}} & X_{222n_{22}} & X_{223n_{22}} \end{pmatrix}$$

En los experimentos SANBRU y JUSREB, X sólo se compone de las dos primeras columnas por no tener examen-criterio de cuestiones.

Los valores de  $n_{ij}$  para los tres experimentos sectoriales ( $2 \times 2$ ) son las siguientes:

	$n_{11}$	$n_{12}$	$n_{21}$	$n_{22}$	N
SANBRU	14	15	15	14	58
JUSREB	14	14	14	14	56
CORBIZU	12	13	12	13	50

La matriz-diseño  $A_1$  para estos tres experimentos factoriales ( $2 \times 2$ ) es

$$A_1 = \begin{array}{cccc|l} 1 & 0 & 0 & 0 & \left. \vphantom{\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array}} \right\} n_{11} \\ \hline \hline \hline \hline 1 & 0 & 0 & 0 & \left. \vphantom{\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array}} \right\} n_{12} \\ 0 & 1 & 0 & 0 & \\ \hline \hline \hline \hline 0 & 1 & 0 & 0 & \left. \vphantom{\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array}} \right\} n_{21} \\ 0 & 0 & 1 & 0 & \\ \hline \hline \hline \hline 0 & 0 & 1 & 0 & \left. \vphantom{\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array}} \right\} n_{22} \\ 0 & 0 & 0 & 1 & \\ \hline \hline \hline \hline 0 & 0 & 0 & 1 & \end{array}$$

Por lo tanto se cumple que

$$A_1' A_1 = \begin{vmatrix} n_{11} & 0 & 0 & 0 \\ 0 & n_{12} & 0 & 0 \\ 0 & 0 & n_{21} & 0 \\ 0 & 0 & 0 & n_{22} \end{vmatrix}, (A_1' A_1)^{-1} = \begin{vmatrix} 1/n_{11} & 0 & 0 & 0 \\ 0 & 1/n_{12} & 0 & 0 \\ 0 & 0 & 1/n_{21} & 0 \\ 0 & 0 & 0 & 1/n_{22} \end{vmatrix}$$

$$A_1' X = \left( \sum_{l=1}^{N_{ij}} x_{ijkl} \right)_{ijk}, \quad \hat{\beta} = (x_{ijk} \dots)_{ijk}$$

$X'X$  es la matriz ( $p \times p$ ) de suma de productos y  $X'X - X'A_1\hat{\beta}$  es la matriz ( $p \times p$ ) de residuos.

En el experimento realizado en la Academia General Militar (BLATUR), el análisis es de covarianza en un sentido, siendo el elemento genérico de la matriz  $x_{ikl}$ , donde

$i = 1, 2, 3, 4$  distintos tratamientos ( $i = 3$  grupo control).

$k = 1$  (teoría),  $2$  (problemas).

$l = 1, 2, \dots, n_i$ .

La matriz X tiene por tanto la forma

$$X = \begin{pmatrix} x_{111} & x_{121} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ x_{11n_1} & x_{12n_1} \\ x_{211} & x_{221} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ x_{21n_2} & x_{22n_2} \\ x_{311} & x_{321} \\ \text{---} & \text{---} \\ \text{---} & \text{---} \\ x_{31n_3} & x_{32n_3} \\ x_{411} & x_{421} \\ \text{---} & \\ \text{---} & \\ x_{41n_4} & x_{42n_4} \end{pmatrix}$$

Los valores de  $n_i$  son los siguientes:

	$n_1$	$n_2$	$n_3$	$n_4$	N
BLATUR	19	17	20	19	75

La matriz-diseño  $A_1$  para este experimento es la siguiente:

$$A_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & u_{11} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ 1 & 0 & 0 & 0 & u_{1n_1} \\ 0 & 1 & 0 & 0 & u_{21} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ 0 & 1 & 0 & 0 & u_{2n_2} \\ 0 & 0 & 1 & 0 & u_{31} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ 0 & 0 & 1 & 0 & u_{3n_3} \\ 0 & 0 & 0 & 1 & u_{41} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ 0 & 0 & 0 & 1 & u_{4n_4} \end{pmatrix}$$

apareciendo la covariable rendimiento previo como quinta columna de dicha matriz-diseño ( $u_{ij}$ ).

En este caso se cumple que  $A_1' A_1$  es una matriz simétrica de la forma

$$A_1' A_1 = \begin{pmatrix} n_1 & 0 & 0 & 0 & \sum_{i=1}^{n_1} u_{i1} & u_{11} \\ & n_2 & 0 & 0 & \sum_{i=1}^{n_2} u_{i2} & u_{21} \\ & & n_3 & 0 & \sum_{i=1}^{n_3} u_{i3} & u_{31} \\ & & & n_4 & \sum_{i=1}^{n_4} u_{i4} & u_{41} \\ & & & & 4 & \sum_{i=1}^{n_i} u_{ii}^2 \end{pmatrix}$$

y su inversa es

$$(A_1' A_1)^{-1} = \begin{vmatrix} 1/n_1 & 0 & 0 & 0 & 0 \\ & 1/n_2 & 0 & 0 & 0 \\ & & 1/n_3 & 0 & 0 \\ & & & 1/n_4 & 0 \\ & & & & 0 \end{vmatrix} + \frac{1}{W} \begin{vmatrix} u_{1.}^2 & u_{1.}u_{2.} & u_{1.}u_{3.} & u_{1.}u_{4.} & -u_{1.} \\ & u_{2.}^2 & u_{2.}u_{3.} & u_{2.}u_{4.} & -u_{2.} \\ & & u_{3.}^2 & u_{3.}u_{4.} & -u_{3.} \\ & & & u_{4.}^2 & -u_{4.} \\ & & & & 1 \end{vmatrix}$$

donde

$$u_{i.} = \frac{\sum_{l=1}^4 n_l u_{il}}{n_i}$$

$$W = \left[ \sum_{i=1}^4 \sum_{l=1}^4 n_l u_{il}^2 - \sum_{i=1}^4 \frac{\left( \sum_{l=1}^4 n_l u_{il} \right)^2}{n_i} \right]$$

La matriz  $A_1' X$  tiene por lo tanto la forma

$$A_1' X = \begin{vmatrix} \left( \sum_{l=1}^4 n_l x_{ikl} \right)_{ik} \\ \left( \sum_{l=1}^4 n_l u_{il} x_{ikl} \right)_k \end{vmatrix}$$

La matriz de los parámetros estimados  $\hat{\beta}$  es de la forma

$$\hat{\beta} = (A_1' A_1)^{-1} A_1' X = \begin{vmatrix} x_{11.} & x_{12.} \\ x_{21.} & x_{22.} \\ x_{31.} & x_{32.} \\ x_{41.} & x_{42.} \\ 0 & 0 \end{vmatrix} - \begin{vmatrix} b_1 u_{1.} & b_2 u_{1.} \\ b_1 u_{2.} & b_2 u_{2.} \\ b_1 u_{3.} & b_2 u_{3.} \\ b_1 u_{4.} & b_2 u_{4.} \\ -b_1 & -b_2 \end{vmatrix}$$

donde

$$b_k = \frac{W_k}{W} \quad y$$

$$W_k = \sum_{i=1}^4 \sum_{l=1}^{n_i} [(u_{il} - u_{i.}) (x_{ikl} - x_{ik.})].$$

$X'X$  es la matriz ( $p \times p$ ) de suma de productos con la misma forma que en los experimentos anteriores y en la matriz de residuos  $X'X - X'A_1\hat{\beta}$  de dimensiones ( $p \times p$ ) vuelve a tener incidencia la existencia de una variable concomitante. Esta matriz es simétrica, siendo sus elementos funciones de los valores  $W_k$  y  $W$  explicitados con anterioridad.

El programa BMD 11 V escribe directamente los valores de todas estas matrices cuya forma para nuestros experimentos hemos querido detallar, a fin de resaltar la facilidad en el desarrollo analítico que ofrece una matriz-diseño  $A_1$  como la usada en nuestros experimentos y al mismo tiempo mostrar, cómo la introducción de una variable concomitante en dicha matriz diseño complica de forma significativa el cálculo posterior. Sin embargo, en cuanto al uso del programa BMD 11 V se trabaja de la misma forma con o sin variables concomitantes una vez diseñada  $A_1$ , aunque naturalmente debe tenerse presente siempre la forma que tiene  $\hat{\beta}$  en cada caso, para diseñar las hipótesis que se quieren contrastar.

#### 6.4. Hipótesis contrastadas

En los cuatro experimentos las hipótesis contrastadas han sido de la forma:  $H_0: C_1\hat{\beta}M = 0$ , donde  $C_1$  y  $M$  son matrices expresamente diseñadas para cada contraste, de acuerdo con las características del diseño empleado, forma de  $\hat{\beta}$  y tipo de efecto que se pretendía estudiar su significación.

Ya hemos apuntado que la decisión de aceptar o rechazar  $H_0$  ha estado basada en todos los casos en la comparación del valor del estadístico F calculado por el programa BMD 11 V con el valor crítico correspondiente a partir de las tablas de la distribución F. En todas nuestras hipótesis el valor estadístico F sigue una distribución F exacta porque o bien el número de filas de  $C_1$  o el de columnas de  $M$  es la unidad.

## A. Experimento SANBRU

En este caso la matriz  $X(x_{ijkl})$  (variable dependiente) contenía las calificaciones de *teoría* ( $k = 1$ ) y *problemas* ( $k = 2$ ) del examen crítico (2.º parcial).

El *tratamiento* ( $i = 1$ ) había sido la entrega de resúmenes al final de una unidad didáctica larga (nivel 3 del tratamiento), y dentro del grupo experimental y de control los alumnos habían sido divididos según su *rendimiento* previo alto ( $j = 1$ ) o bajo ( $j = 2$ ). El subíndice  $l$  representaba el valor de orden de cada observación dentro de las cuatro celdas del diseño ( $2 \times 2$ ).

La siguiente tabla nos muestra los resultados obtenidos en los diez contrastes realizados del tipo  $H_0: C_1 \hat{\beta} M = 0$ , con expresión de las matrices  $C_1$ ,  $M$  y  $L$  (valor del contraste  $C_1 \hat{\beta} M$ ), efecto contrastado, valor del estadístico  $F$  calculado, grados de libertad y decisión con expresión del nivel de riesgo ( $\alpha$ ).

La figura 1 nos ilustra los resultados del tratamiento experimental en la calificación global, teoría y problemas.

De la tabla anterior se observa que el nivel 3 del tratamiento ha producido efectos altamente significativos, fundamentalmente en el examen criterio de teoría y para todos los alumnos que lo han seguido, aunque el efecto haya sido algo superior para los alumnos de rendimiento previo bajo.

El efecto global del tratamiento es significativo al nivel ( $\alpha = 0,01$ ), aunque se debe principalmente al efecto global en teoría, ya que el efecto global en problemas sólo es significativo al nivel ( $\alpha = 0,1$ ).

Las interacciones estudiadas no ofrecen efectos significativos salvo en el caso de que aislemos dicha interacción en teoría para alumnos de rendimiento previo alto o bajo. (Caso que en rigor no supone una interacción, sino un efecto principal aislado para un tipo u otro de alumnos.)

Es evidente que el nivel de significación de los efectos producidos es ciertamente sorprendente, sin embargo, los efectos en sí podrían entrar dentro de cualquier previsión lógica. Un tratamiento que incide sobre la capacidad de sintetizar información por parte de los alumnos ha producido efectos positivos sobre la parte de la prueba criterio en la que el alumno debe hacer más uso de su capacidad de síntesis, esto es, en el examen de teoría tipo ensayo. Sin embargo, el efecto también ha sido positivo, aunque a un nivel de significación mucho menor, cuando al alumno se le ha pedido demostrar su capacidad de aplicación.



TABLA 1  
*Contrastes en el experimento SANBRU*

Hipótesis	Efecto	F	(n <sub>1</sub> , n <sub>2</sub> )	Decisión
1. C <sub>i</sub> = (11 — 1 — 1) M' = (11) L = 6,77	Global del tratamiento (TG)	15,99	(1,54)	Rechazada (***)
2. C <sub>i</sub> = (11 — 1 — 1) M' = (10) L = 1,71	Tratamiento en teoría (TT)	26,98	"	Rechazada (***)
3. C <sub>i</sub> = (11 — 1 — 1) M' = (01) L = 1,71	Tratamiento en problemas (TP)	3,03	"	Rechazada (*)
4. C <sub>i</sub> = (1 — 1 — 11) M' = (10) L = 0,46	Global de la interacción (tratamiento × rendimiento) (TRG)	0,07	"	Aceptada
5. C <sub>i</sub> = (1 — 1 — 11) M' = (01) L = 0,21	En teoría de la interacción (tratamiento × rendimiento) (TRT)	0,48	"	Aceptada
6. C <sub>i</sub> = (1 — 1 — 11) M' = (01) L = 0,21	En problemas de la interacción (tratamiento × rendimiento) (TRP)	0,05	"	Aceptada
7. C <sub>i</sub> = (10 — 10) M' = (10) L = 2,18	En teoría de la interacción (tratamiento × rendimiento alto) (TR1T)	10,13	"	Rechazada (***)
8. C <sub>i</sub> = (010 — 1) M' = (10) L = 2,87	En teoría de la interacción (tratamiento × rendimiento bajo) (TR2T)	17,34	"	Rechazada (***)
9. C <sub>i</sub> = (10 — 10) M' = (01) L = 0,96	En problemas de la interacción (tratamiento × rendimiento alto) (TR1P)	1,92	"	Aceptada
10. C <sub>i</sub> = (010 — 1) M' = (01) L = 0,75	En problemas de la interacción (tratamiento × rendimiento bajo) (TR2P)	1,16	"	Aceptada

(\*\*\*) F<sub>0,01</sub> (1,54) = 7,15.

(\*\*) F<sub>0,05</sub> (1,54) = 4,03.

(\*) F<sub>0,1</sub> (1,54) = 2,8.

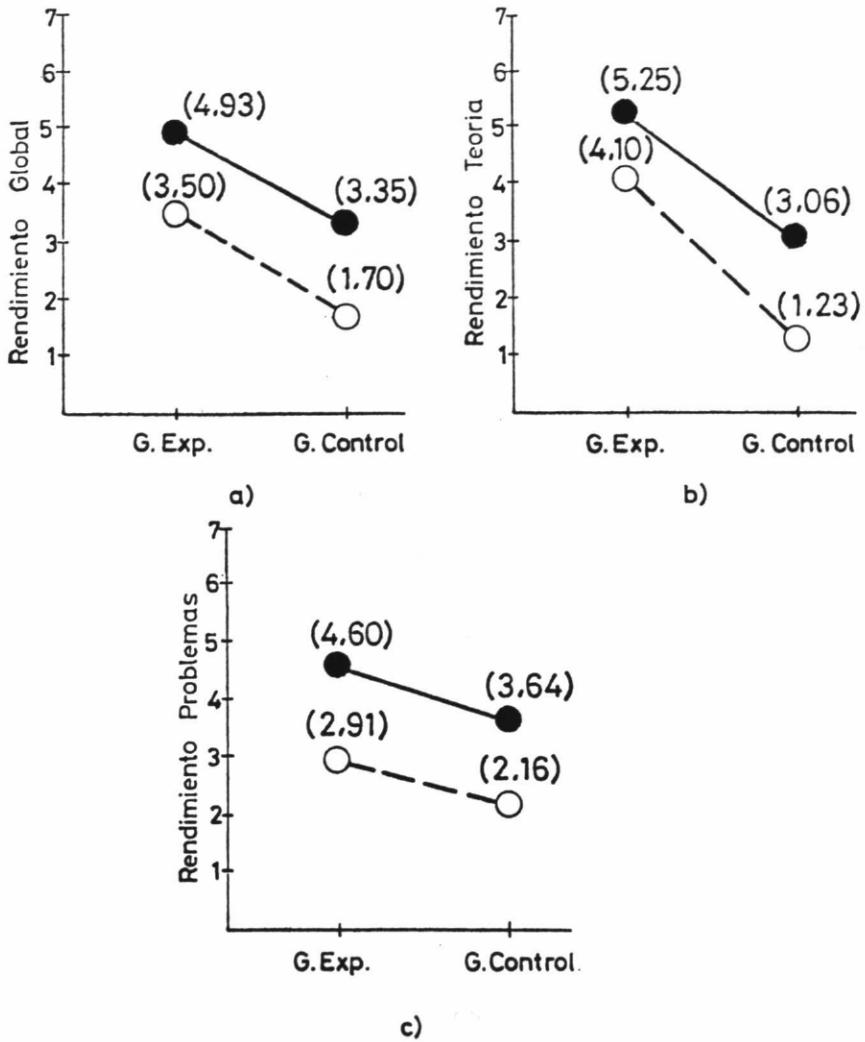


Fig. 1. Experimento SANBRU. Resultados promedio en el examen criterio; a) Global, b) Teoría y c) Problemas. Los círculos negros y línea continua representan a los alumnos de rendimiento previo alto y los círculos blancos y línea de trazos representan a los alumnos de rendimiento previo bajo.

El hecho de que el efecto positivo en el examen criterio de teoría haya sido algo superior para los alumnos de rendimiento previo bajo, puede tener la explicación en que dichos alumnos son los que menos capacidad de síntesis poseían seguramente y su ganancia relativa con el tratamiento es lógicamente superior que en aquellos que poseían una capacidad de síntesis más elevada.

## B. Experimento JUSREB

Las características de la matriz  $X(x_{ijkl})$  son exactamente igual que en el experimento SANBRU. La única diferencia está en que corresponden a calificaciones del tercer parcial y que en este caso el tratamiento ( $i = 1$ ) consistía en la entrega de problemas resueltos fuera del período lectivo normal, tras finalizar una unidad didáctica (nivel 2 del tratamiento).

Los resultados en los contrastes han sido los mostrados en la tabla y la figura que aparecen a continuación.

La tabla 2 y figura 2 nos muestran resultados de cierto interés y que pueden resumirse de la siguiente manera:

El efecto global del nivel 2 del tratamiento ha sido positivamente significativo al nivel ( $\alpha = 0,05$ ) en su rendimiento en problemas, mientras que tal efecto no se ha producido en teoría.

Para los alumnos de rendimiento previo bajo se observa un cierto efecto positivo tanto en teoría como en problemas, pero que no son significativos considerando cada examen criterio aislado. El efecto se hace significativo para estos alumnos al nivel ( $\alpha = 0,05$ ) si consideramos el examen criterio global (teoría + problemas).

Las interacciones tratamiento  $\times$  rendimiento previo no son estadísticamente significativas, salvo aisladas para un grupo u otro de alumnos como se detalla en los dos puntos anteriores. Sin embargo, se observa cierta tendencia en este sentido en lo relativo a teoría.

Es necesario resaltar los curiosos efectos que pudiéramos llamar de interacción del estímulo con el criterio evaluador que se han producido en este experimento. La causa de que el beneficio mayor en cuanto a rendimiento en problemas se haya producido para los alumnos de rendimiento previo alto, hay que suponerla en que posiblemente han sido este tipo de alumnos los que trabajaron los problemas fuera del aula, centrándose en su resolu-

TABLA 2  
*Contrastes en el experimento JUSREB*

Hipótesis	Efecto	F	(n <sub>1</sub> , n <sub>2</sub> )	Decisión
1. C <sub>i</sub> = (11 — 1 — 1) M' = (11) L = 4,44	Global del tratamiento (TG)	5,62	(1,52)	Rechazada (**)
2. C <sub>i</sub> = (11 — 1 — 1) M' = (10) L = 1,01	Tratamiento en Teoría (TT)	0,52	"	Aceptada
3. C <sub>i</sub> = (11 — 1 — 1) M' = (01) L = 3,43	Tratamiento en problemas (TP)	7,58	"	Rechazada (***)
4. C <sub>i</sub> = (1 — 1 — 11) M' = (11) L = — 1,55	Global de la interacción (tratamiento × rendimiento) (TRG)	0,69	"	Aceptada
5. C <sub>i</sub> = (1 — 1 — 11) M' = (10) L = — 2,12	En teoría de la interacción (tratamiento × rendimiento) (TRT)	2,29	"	Aceptada
6. C <sub>i</sub> = (1 — 1 — 11) M' = (01) L = 0,57	En problemas de la interacción (tratamiento × × rendimiento) (TRP)	0,21	"	Aceptada
7. C <sub>i</sub> = (10 — 10) M' = (10) L = — 0,56	En teoría de la interacción (tratamiento × rendimiento alto) (TR1T)	0,31	"	Aceptada
8. C <sub>i</sub> = (10 — 10) M' = (10) L = 1,57	En teoría de la interacción (tratamiento × rendimiento bajo) (TR2T)	2,5	"	Aceptada
9. C <sub>i</sub> = (010 — 1) M' = (01) L = 2	En problemas de la interacción (tratamiento × × rendimiento alto) (TR1P)	5,16	"	Rechazado (**)
10. C <sub>i</sub> = (010 — 1) M' = (01) L = 1,43	En problemas de la interacción (tratamiento × × rendimiento bajo) (TR2P)	2,63	"	Aceptada
11. C <sub>i</sub> = (010 — 1) M' = (11) L = 2,8	En (Teor. + Probl.) de la interac. (tratamiento × × rendimiento bajo) (TR2G)	4,52	"	Rechazada (**)

(\*\*\*) F<sub>0,01</sub> (1,52) = 7,21.

(\*\*) F<sub>0,05</sub> (1,52) = 4,05.

(\*) F<sub>0,1</sub> (1,52) = 2,81.

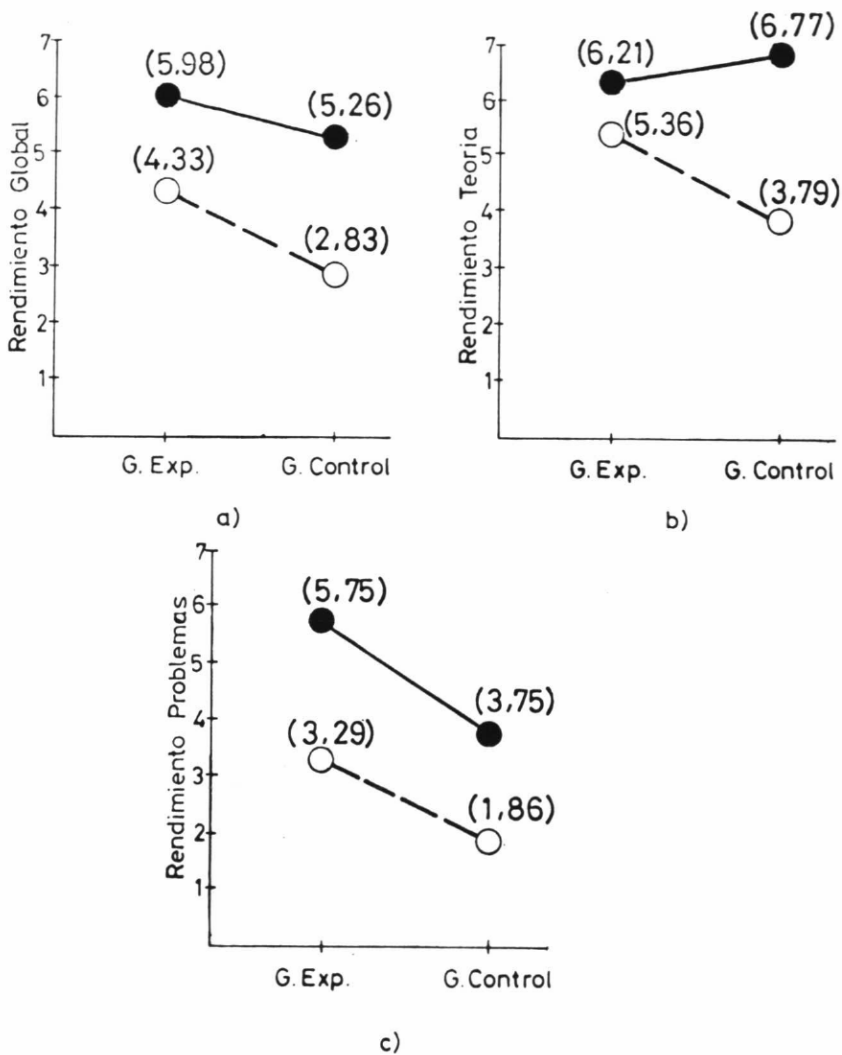


Fig. 2. *Experimento JUSREB*. Resultados promedio en el examen criterio: a) Global, b) Teoría y c) Problemas. Los círculos negros y línea continua representan a los alumnos de rendimiento previo alto y los círculos blancos y línea de trazos representan a los alumnos de rendimiento previo bajo.

ción, esto es, incidiendo en la mera capacidad de aplicación de conceptos y leyes que manejaban y comprendían relativamente bien.

Este tipo de alumnado no ha tenido seguramente que salirse del mero nivel comportamental de *aplicación* para seguir adecuadamente el tratamiento y por ello su beneficio se ha centrado única y exclusivamente en ese aspecto.

Los alumnos de rendimiento previo bajo, que seguramente seguían con mayor esfuerzo y dificultades este tipo de tratamiento, han tenido que usar más recursos para poder seguirlo, tales como ayudas de compañeros, consultas más frecuentes a la teoría etcétera, de forma que aunque no han aumentado tan sensiblemente su capacidad de aplicación, se han beneficiado sustancialmente en aspectos de comprensión y capacidad de síntesis, de forma que en conjunto, su beneficio por el tratamiento ha sido equiparable, y quizá ligeramente superior, al obtenido por los alumnos de rendimiento previo alto. En definitiva, un beneficio más amplio aunque menos intenso.

A nuestro entender, volvemos a encontrar la lógica en los resultados generales, al ver que los efectos se vuelven a producir en aquellos niveles de comportamiento en los que más directamente incide el tratamiento. Sin embargo, no siempre esa incidencia tiene el mismo signo para los distintos tipos de alumnos. El tratamiento JUSREB es una muestra clara de que no es así.

### C) *Experimento CORBIZU*

Volvemos a tratar con un diseño ( $2 \times 2$ ) idéntico a los anteriores en el que los valores de  $X(x_{ijkl})$  representan calificaciones del examen-criterio (3.<sup>er</sup> parcial), con el subíndice  $k$  igual a tres valores posibles  $k = 1$  (teoría), 2 (problemas), 3 (cuestiones cortas). Por lo tanto,  $X$  es una matriz ( $N \times 3$ ) en vez de ( $N \times 2$ ) como en los dos experimentos anteriores. Lo que se refleja en la forma de las matrices  $M$  de los contrastes.

Los resultados de las hipótesis contrastadas en este experimento en donde el tratamiento experimental ( $i = 1$ ) consistía en la aplicación periódica de cuestiones objetivas sobre el contenido de unidades didácticas explicadas previamente (nivel 1 del tratamiento) quedan reflejadas en la tabla 3 y figura 3 que aparecen a continuación.

TABLA 3

*Contrastes en el experimento CORBIZU*

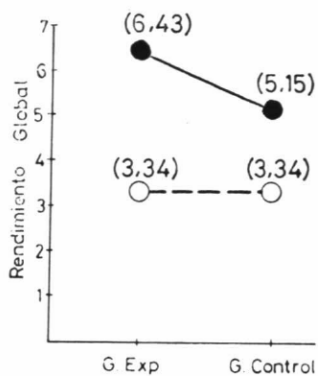
Hipótesis	Efecto	F	(n <sub>1</sub> , n <sub>2</sub> )	Decisión
1. C <sub>i</sub> = (11 — 1 — 1) M' = (111) L = 3,89	Global del tratamiento (TG)	1,90	(1,46)	Aceptada
2. C <sub>i</sub> = (11 — 1 — 1) M' = (100) L = — 0,24	Tratamiento en teoría (TT)	0,04	"	Aceptada
3. C <sub>i</sub> = (11 — 1 — 1) M' = (010) L = 1,93	Tratamiento en problemas (TP)	1,57	"	Aceptada
4. C <sub>i</sub> = (11 — 1 — 1) M' = (001) L = 2,20	Tratamiento en cuestiones (TC)	2,39	"	Aceptada
5. C <sub>i</sub> = (11 — 1 — 1) M' = (111) L = 3,81	Global de la interacción (tratamiento × rendimiento) (TRG)	1,83	"	Aceptada
6. C <sub>i</sub> = (1 — 1 — 11) M' = (100) L = 0,99	En teoría de la interacción (tratamiento × rendimiento) (TRT)	0,7	"	Aceptada
7. C <sub>i</sub> = (1 — 1 — 11) M' = (010) L = 0,24	En problemas de la interacción (tratamiento × × rendimiento) (TRP)	0,02	"	Aceptada
8. C <sub>i</sub> = (1 — 1 — 11) M' = (001) L = 2,59	En cuestiones de la interacción (tratamiento × × rendimiento) (TRC)	3,29	"	Aceptada
9. C <sub>i</sub> = (10 — 10) M' = (100) L = 0,37	En teoría de la interacción (tratamiento × rendimiento alto) (TR1T)	0,19	"	Aceptada
10. C <sub>i</sub> = (010 — 1) M' = (100) L = — 0,61	En teoría de la interacción (tratamiento × rendimiento bajo) (TR2T)	0,56	"	Aceptada

Hipótesis	Efecto	F	(n <sub>1</sub> , n <sub>2</sub> )	Decisión
11. C <sub>i</sub> = (10 — 10) M' = (010) L = 1,08	En problemas de la interacción (tratamiento × × rend. alto) (TR1P)	0,95	"	Aceptada
12. C <sub>i</sub> = (010 — 1) M' = (010) L = 0,84	En problemas de la interacción (tratamiento × × rend. bajo) (TR2P)	0,63	"	Aceptada
13. C <sub>i</sub> = (10 — 10) M' = (001) L = 2,4	En cuestiones de la interacción (tratamiento × × rend. alto) (TR1C)	5,42	"	Rechazada (**)
14. C <sub>i</sub> = (010 — 1) M' = (001) L = 0,19	En cuestiones de la interacción (tratamiento × × rend. bajo) (TR2C)	0,04	"	Rechazada (*)
15. C <sub>i</sub> = (11 — 1 — 1) M' = (011) L = 4,13	Tratamiento en problemas y cuestiones (TPC)	3,39	"	Rechazada (*)
16. C <sub>i</sub> = (1 — 1 — 11) M' = (011) L = 2,83	En problemas y cuestiones de la interacción (trat. × rend.) (TRPC)	1,58	"	Aceptada
17. C <sub>i</sub> = (10 — 10) M' = (011) L = 3,48	En problemas y cuestiones de la interacción (trat. × rend. alto) (TR1PC)	4,62	"	Rechazada (**)
18. C <sub>i</sub> = (010 — 1) M' = (011) L = 0,65	En problemas y cuestiones de la interacción (trat. × rend. bajo) (TR2PC)	0,18	"	Aceptada

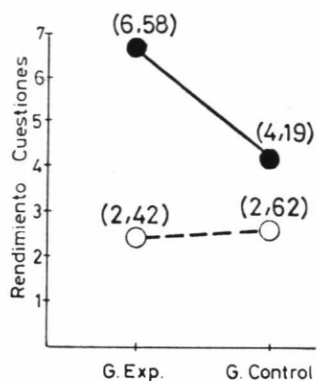
(\*\*) F<sub>0,05</sub> (1,46) = 4,26.

(\*) F<sub>0,1</sub> (1,46) = 2,83.

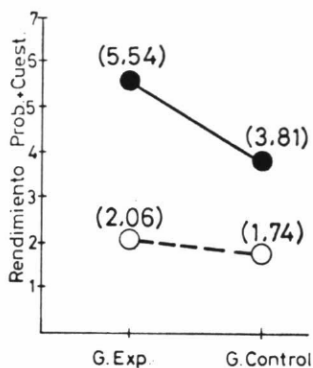




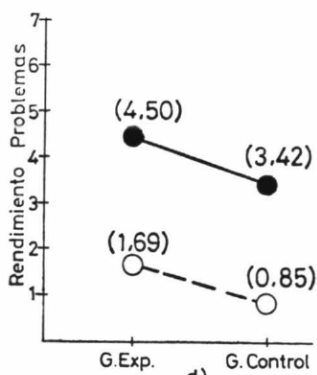
a)



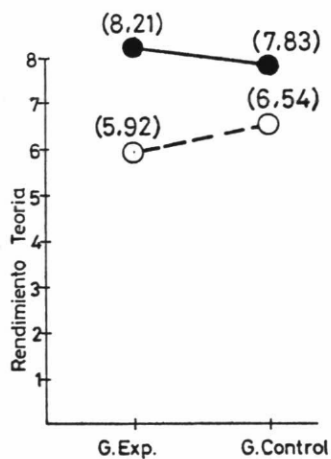
b)



c)



d)



e)

Fig. 3. Experimentos CORBIZU. Resultados promedio en el examen criterio: a) Global, b) Cuestiones, c) Problemas + Cuestiones, d) Problemas y e) Teoría. Los círculos negros y línea continua representan a los alumnos de rendimiento previo alto y los círculos blancos y línea de trazos representan a los alumnos de rendimiento previo bajo.

Observamos que nuestro nivel 1 del tratamiento ha afectado positivamente al rendimiento del alumnado en global, problemas y cuestiones aunque el efecto no sea estadísticamente significativo. Sin embargo, si consideramos las calificaciones de problemas y cuestiones en conjunto el efecto del tratamiento es significativo al nivel ( $\alpha = 0,1$ ).

Similarmente al experimento anterior, este efecto positivo sobre el rendimiento en problemas y cuestiones se centra en los alumnos de rendimiento previo alto con un nivel de significación ( $\alpha = 0,05$ ). Sin embargo, el peso fundamental de este efecto significativo está en el rendimiento en el examen de cuestiones, parte del examen-criterio donde la interacción con el tratamiento ha sido significativa al nivel ( $\alpha = 0,1$ ) para todo el grupo experimental y al nivel ( $\alpha = 0,05$ ) para los alumnos del grupo experimental con rendimiento previo alto.

Independientemente de los niveles de significación, los resultados de este experimento aparecen de nuevo como lógicos, el efecto positivo se ha vuelto a producir en aquel tipo de destreza mental en la que el tratamiento ofrecía un refuerzo. En este caso han sido los alumnos de rendimiento previo alto los únicos beneficiados, muy posiblemente por ser los únicos que han seguido seriamente un tratamiento cuyas características requerían un seguimiento individual inmediato y continuo del desarrollo del curso. Algo que no era requerido en el caso de los resúmenes (SANBRU), ni de forma tan clara, debido a la posibilidad de ayuda externa, compañeros, texto, etc., en el caso de problemas (JUSREB).

#### D. Experimento BLATUR

Ya hemos explicado con anterioridad las razones que nos forzaron a llevar un análisis de covarianza con el rendimiento previo en física como variable concomitante, en vez de dividir a los grupos en dos niveles de rendimiento como en los tres experimentos anteriores.

La matriz  $X(x_{ikl})$  de dimensiones  $(75 \times 2)$  representa los resultados del tercer examen parcial en teoría y problemas, teniendo presente que lo que en este caso denominamos examen-criterio de problemas venía a ser una mezcla de lo que venimos denominando problemas y cuestiones. Razones de índole práctico nos imposibilitaron separar esta calificación global en dos (problemas y cuestiones).

En este caso, el experimento no se reducía a dos grupos (experimental y control), sino que contábamos con cuatro grupos, tres experimentales (nivel 1<sub>1</sub> y 1<sub>2</sub> y nivel 3) y un grupo de control.

Recordemos que diferenciamos entre nivel 1<sub>1</sub> y 1<sub>2</sub> porque había ciertas diferencias en la aplicación del mismo tratamiento (tests objetivos) tal y como ha sido explicado con anterioridad.

El hecho de tratarse de un análisis de covarianza y no un diseño factorial ( $2 \times 2$ ), como en los casos anteriores, influye en la forma de la matriz  $\hat{\beta}$  y, por lo tanto, en la  $C_1$  diseñada para realizar los diferentes contrastes multivariantes.

Los resultados de las hipótesis contrastadas se recogen en la Tabla 4.

En el experimento BLATUR, el tratamiento evaluación periódica no ha tenido efecto estadístico significativo en ninguno de sus niveles estudiados (nivel 1 (1<sub>1</sub>, 1<sub>2</sub>) y el nivel 3).

Por lo que respecta al rendimiento en el examen-criterio en aspectos de aplicación (problemas y cuestiones) el tratamiento puede decirse que prácticamente no ha afectado en nada aunque los valores de los contrastes L son siempre positivos.

En lo relativo al rendimiento en teoría, el efecto de los distintos niveles del tratamiento ha sido sistemáticamente contrario, aunque en ningún caso significativo al nivel ( $\alpha = 0,1$ ). Este hecho no tuvo ninguna repercusión sobre la evaluación final de los alumnos porque los profesores habían decidido a priori dar más peso relativo (80 por 100) a los aspectos de aplicación.

La chocante diferencia entre los resultados de los Experimentos SANBRU, JUSREB y CORBIZU con los del Experimento BLATUR nos introducen en un tema de una gran relevancia y que sin ninguna duda merece el análisis más detallado que le dedicamos en el siguiente apartado.

En palabras llanas podríamos sintenzar los resultados de este experimento diciendo que el esfuerzo dedicado por los alumnos a los tratamientos experimentales ha producido unos frutos demasiado débiles como para compensar, de cara al rendimiento global, el tiempo de trabajo individual empleado en dicho tratamiento y que pudiera haber sido dedicado al estudio de temas.

TABLA 4

*Contrastes en el experimento BLATUR*

Hipótesis	Efecto	F	(n <sub>1</sub> , n <sub>2</sub> )	Decisión
1. C <sub>i</sub> = (11 — 310) M' = (11) L = — 2,87	Global del tratamiento (TG)	1,14	(1,70)	Aceptada
2. C <sub>i</sub> = (11 — 310) M' = (10) L = — 3,35	Tratamiento en teoría (TT)	2,65	"	Aceptada
3. C <sub>i</sub> = (11 — 310) M' = (01) L = 0,48	Tratamiento en problemas (TP)	0,16	"	Aceptada
4. C <sub>i</sub> = (10 — 100) M' = (1,1) L = — 0,81	Global del tratamiento 1 <sub>1</sub> (T1G)	0,54	"	Aceptada
5. C <sub>i</sub> = (10 — 100) M' = (10) L = 0,89	Del tratamiento 1 <sub>1</sub> en teoría (T1T)	1,14	"	Aceptada
6. C <sub>i</sub> = (10 — 100) M' = (01) L = 0,09	Del tratamiento 1 <sub>1</sub> en problemas (T1P)	0,03	"	Aceptada
7. C <sub>i</sub> = (01 — 100) M' = (11) L = — 0,86	Global del tratamiento 3 (T2G)	0,58	"	Aceptada
8. C <sub>i</sub> = (01 — 100) M' = (10) L = — 1,23	Del tratamiento 3 en teoría (T2T)	2,02	"	Aceptada
9. C <sub>i</sub> = (01 — 100) M' = (01) L = 0,37	Del tratamiento 3 en problemas (T2P)	0,25	"	Aceptada
10. C <sub>i</sub> = (00 — 110) M' = (11) L = — 1,20	Global del tratamiento 1 <sub>2</sub> (T4G)	1,18	"	Aceptada
11. C <sub>i</sub> = (00 — 110) M' = (10) L = — 1,23	Del tratamiento 1 <sub>2</sub> en teoría (T4T)	2,11	"	Aceptada
12. C <sub>i</sub> = (00 — 110) M' = (10) L = 0,03	Del tratamiento 1 <sub>2</sub> en problemas (T4P)	0,00	"	Aceptada
13. C <sub>i</sub> = (100 — 10) M' = (11) L = 0,39	Diferencial entre 1 <sub>1</sub> y 1 <sub>2</sub> en rendimiento global	0,13	"	Aceptada

F<sub>0,1</sub> (1,70) = 2,79.

## 6.5. Interacción de la evaluación periódica con el contexto educativo

En los experimentos SANBRU, JUSREB y CORBIZU el efecto de la evaluación periódica sobre el rendimiento de los alumnos aparece como sistemáticamente positivo aunque la significación de este efecto se centra principalmente en aquellas destrezas sobre las que el nivel del tratamiento incide de forma más directa.

En todos los casos en que aparecen efectos estadísticamente significativos lo es a favor del tratamiento, esto es, de acuerdo como están planteadas las hipótesis contrastadas, el valor del contraste  $L$  es siempre positivo.

En los contrastes en los que no aparece un efecto estadísticamente significativo, la tendencia es que  $L$  sea positiva y cuando  $L$  tiene valor negativo, el valor absoluto del contraste es muy pequeño.

En el experimento BLATUR, en donde en definitiva llevamos a cabo tres experimentos aislados tratados de forma conjunta en un mismo diseño, vemos que en todos los grupos experimentales se producen sistemáticamente unos efectos similares:

a) Influencia práctica nula del tratamiento en el rendimiento en aspectos de aplicación (problemas y cuestiones) independientemente de que el nivel del tratamiento incida directamente en dichos aspectos (niveles  $1_1$  y  $1_2$ ) o que lo haga en aspectos de síntesis (nivel 3). En todos los contrastes sobre aspectos de aplicación,  $L$  tiene un valor positivo pero muy pequeño.

b) Influencia contraria sin significación estadística ( $L$  es siempre negativa) en el rendimiento de los alumnos en aspectos de síntesis (examen de teoría), tanto en los niveles del tratamiento que inciden sobre aspectos de aplicación ( $1_1$  y  $1_2$ ) como del nivel 3 que incide sobre aspectos de síntesis.

Esta disparidad de resultados nos llevó a examinar cuidadosamente donde podría estar la causa.

De entre todas las posibles, la única que nos aparece como clara es la de que los experimentos SANBRU, JUSREB y CORBIZU fueron desarrollados en un contexto educativo —la Facultad de Ciencias— completamente diferente de la Academia General Militar en el que se desarrolló el experimento BLATUR.

El hecho de que el control académico de la asignatura física del curso selectivo para el ingreso en la Academia General Militar fuera llevado a cabo por el mismo equipo y departamento

que controlaba dicha asignatura en el curso selectivo de la Facultad de Ciencias, disminuye la probabilidad de que la causa esté en la diferencia de enfoque disciplinario.

Tampoco encontramos plausible la explicación de que el estudiante con vocación de químico tenga una actitud apriorística diferente ante la evaluación periódica que el estudiante con vocación de militar. Lo que sí parece evidente es que a posteriori y evidentemente esto debe ser consecuencia del contexto educativo general del estudiante, el futuro militar parece reaccionar de forma diferente al tratamiento-estímulo que el futuro químico. Sin embargo, esta diferencia factual no es percibida tan claramente por el estudiante, como indican los resultados que debiera existir, ya que al preguntarles sobre qué pensaban de la evaluación periódica como estímulo didáctico (datos recogidos en el capítulo cuarto) los futuros militares venían a responder de forma parecida a los futuros químicos.

Nosotros creemos ver la causa, y esto es algo que evidentemente debiera ser investigado de forma científica con experimentos "ac hoc", en las diferencias en lo que vamos a denominar *rigidez* del contexto educativo.

Nuestros tratamientos estímulos suponen la adición de un cierto grado de control al desarrollo académico normal que antes no existía. Bien es cierto que la mayor parte de este nuevo control se plasma en auto-control por parte del alumno de su propio aprendizaje y muy poco en control de dicho aprendizaje por parte del profesor, pero, en definitiva, se introduce más control en el proceso.

El contexto educativo de una Facultad tradicional podríamos calificarlo de *relajado*, al menos si lo comparamos con el existente en una Academia Militar.

El estudiante de una Facultad es libre de hecho de asistir a clase, presentarse a las evaluaciones, programar su estudio, etcétera, el estudiante de una Academia Militar está dentro de un contexto *rigido* en este sentido. En definitiva, soporta mucho más control externo que el estudiante de una Facultad.

El Gabinete Psico-pedagógico del curso selectivo para el ingreso en la Academia General Militar nos proporcionaba datos recogidos en una encuesta en el mismo curso en que se realizaba el experimento BLATUR, que vienen a reforzar empíricamente los supuestos que acabamos de exponer. Algunos datos de esta encuesta eran los siguientes:

a) El 56 por 100 de los alumnos pensaban que la delimitación y obligatoriedad de las horas dedicadas a estudio era un factor que afectaba negativamente al rendimiento.

b) El 38 por 100 veían efectos negativos en las condiciones ambientales (horarios, lugar, densidad humana).

- c) El 26 por 100 se sentían en una situación de tensión por:
- incompatibilidad vida académica universitaria-vida académica militar;
  - compañerismo negativo (situación permanente de oposición a plazas limitadas);
  - influencia negativa de los repetidores.

Aunque estemos ante un tema con connotaciones mucho más amplias que las puramente académicas y por supuesto fuera del alcance de nuestro interés inmediato y capacidad de análisis, vemos a los alumnos del experimento BLATUR dentro de un contexto lo suficientemente controlado como para que el añadir cierto control, aunque sea en forma de auto-control, no produzca ningún efecto claramente positivo, sino que hasta pudiera ser negativo caso de que el tratamiento hubiera sido más rígido y continuado.

Pensemos que compaginar una disciplina académico-militar, con varias asignaturas clave para el ingreso-oposición, con el curso selectivo de ciencias, desemboca necesariamente en un contexto social que va a tener seguramente una influencia mucho más decisiva que cualquier tratamiento didáctico que pudiera introducirse.

El hecho aparentemente negativo de contrastar que nuestro tratamiento en la Academia Militar no ha tenido los efectos positivos que hubiéramos deseado, nos lleva a descubrir, al menos pasa nosotros, algo que pudiera ser el resultado más importante de todo el trabajo en su conjunto, esto es, el hecho de que el aumento del control académico, sea cual sea la forma en que éste se plasme, pudiera tener una influencia positiva en el rendimiento escolar mientras no se alcance lo que pudiera ser el *punto de saturación*. El aumentar el control por encima de dicho punto podría tener, posiblemente, efectos contrarios.

Está lejos de nuestra intención dogmatizar con esta explicación, ya que nuestro trabajo no muestra más que indicios al respecto, sin embargo, el problema se nos presenta lo suficientemente trascendente como para resaltarlo de cara a futuros trabajos de investigación.

## 6.6. Otros resultados de nuestros experimentos

En nuestro análisis histórico sobre la evaluación periódica había aparecido en algún trabajo el efecto de homogeneización de los rendimientos en los grupos en los que se había aplicado el tratamiento. En nuestros experimentos las medias y desviaciones típicas posteriores al tratamiento experimental que dan reflejadas en las siguientes tablas.

### A. Experimento SANBRU

	Experimental		Control	
Teoría ... ..	(4,66	1,26)	(2,18	2,44)
Problemas ... ..	(3,73	1,86)	(2,92	2,08)
(Teoría + Problemas) ... ..	(4,18	1,43)	(2,61	1,97)

### B. Experimento JUSREB

	Experimental		Control	
Teoría ... ..	(5,79	2,96)	(5,28	2,54)
Problemas ... ..	(4,52	2,44)	(2,80	2,55)
(Teoría + Problemas) ... ..	(5,21	1,94)	(4,05	1,96)

### C. Experimento CORBIZU

	Experimental		Control	
Teoría ... ..	(7,05	2,16)	(7,16	2,26)
Problemas ... ..	(3,04	3,01)	(2,08	2,86)
Cuestiones ... ..	(4,42	2,98)	(3,29	2,86)
Global ... ..	(4,83	2,15)	(4,16	1,93)

### D. Experimento BLATUR

	Nivel 1 <sub>1</sub>		Nivel 3		Nivel 0		Nivel 1 <sub>2</sub>	
Teoría ... ..	(2,11	2,19)	(1,84	2,06)	(3,13	3,59)	(1,51	2,68)
Problemas ... ..	(4,25	1,82)	(4,61	1,62)	(3,43	2,24)	(3,88	1,89)
Global ... ..	(3,27	1,74)	(3,23	1,4)	(3,73	2,52)	(2,84	1,98)

Aunque el contraste de hipótesis analizado con anterioridad nos ofrece un cuadro de resultados más completo y preciso, la inspección de los datos que acabamos de presentar nos muestra cómo en los tres primeros experimentos se observa en las me-



días el efecto positivo introducido por los tratamientos, sobre todo si tenemos en cuenta datos similares sobre el rendimiento previo expuestos en el capítulo cuarto.

En lo relativo al experimento BLATUR se observa, analizando estos datos con los de rendimiento previo, ya que se trata de una covariación, que el efecto experimental no ha sido positivo.

Por lo que respecta a la dispersión de las distribuciones de los datos de rendimiento, y teniendo en cuenta la distribución de rendimientos previos, se observa:

a) En el experimento SANBRU una reducción sensible de la dispersión en lo relativo a rendimiento en teoría y algo menos en el global. Esto es lógico si observamos que el efecto experimental producido se ha acentuado en los alumnos de rendimiento previo bajo, sobre todo en teoría.

b) En los experimentos JUSREB y CORBIZU no se aprecia efecto sobre la dispersión.

c) En el experimento BLATUR se aprecia cierta reducción de la dispersión en los grupos experimentales pero también aparece como lógico debido a esa tendencia a reducir calificaciones de teoría que ha introducido el efecto experimental.

En definitiva, no parece que la homogeneización de los grupos sea algo fácil de conseguir con los tratamientos empleados, salvo en el caso de que el efecto se produzca en los alumnos más desventajados como en el caso SANBRU.

## **6.7. Límites de nuestro trabajo**

Antes de concluir nuestro análisis de resultados queremos, por razones de rigor científico, establecer claramente los límites de nuestro estudio de la evaluación frecuente como estímulo didáctico.

La disciplina (Física), nivel educativo (Universitario) y tipo de alumnos (científicos con especialización principal distinta de la Física) son el marco de referencia para generalizar nuestros resultados. Dentro de una interpretación rigorista habría que perfilar y acotar mucho más estos tres puntos de referencia, pero entendemos que aun sin hacerlo, el generalizar a ese marco no nos induciría a cometer errores significativos. Las variables de matiz que habría que tener en cuenta nos aparecen como muy débiles para modificar sustancialmente el contexto definido por los tres puntos de referencia citados.

Hemos limitado nuestro estudio a los efectos producidos por los niveles del tratamiento en las evaluaciones inmediatas a la interacción didáctica. Por ello, nuestros tratamientos han sido aplicados en trimestres o cuatrimestres, ya que con esa periodicidad se evalúa el *rendimiento inmediato* del alumnado.

Ante este hecho, no hemos podido estudiar efectos relativos al tiempo de aplicación del tratamiento experimental, porque hubiéramos necesitado una capacidad de decisión muy superior en la estructuración temporal del desarrollo de los cursos, muchos más medios y muchos más recursos humanos de alumnado y profesorado.

Algo similar podría decirse respecto al *rendimiento retardado* de los alumnos. Este es otro tema importante que requeriría muchos más medios para su estudio.

Metodológicamente hablando, el estudio de efectos de tiempo de aplicación del tratamiento e influencia en el rendimiento retardado no supone un trabajo más complejo que el que nosotros hemos realizado. Sin embargo, sería necesario intervenir y modificar convenientemente el calendario académico con vistas a proporcionar un contexto acorde con el diseño experimental que se necesita para ese trabajo. Por otra parte, estudiar el rendimiento retardado plantearía en este momento problemas de realización práctica provenientes de dificultades de seguimiento de los alumnos.

Es evidente que éstos y otros muchos aspectos relevantes se escapan de la atención de nuestro trabajo, pero ello es algo que ocurre con todo trabajo de esta índole. Abordar toda la problemática de la evaluación frecuente en un trabajo individual, aunque se contara con muy valiosas colaboraciones, era algo irrealizable.

El punto final de esta discusión pudiera ser la afirmación de que en nuestro trabajo se ha intentado estudiar todos aquellos problemas relativos a la evaluación frecuente que eran factibles de ser estudiados sin modificar el contexto académico natural que ofrecían los centros educativos en los que se han realizado los experimentos.



## 7. CONCLUSIONES

1) De los trabajos previos sobre evaluación periódica como estímulo para el aprendizaje, deducimos que existen efectos generalmente positivos de este tipo de tratamiento, bien elevando rendimientos u homogeneizando grupos.

El conocimiento inmediato de los resultados de las pruebas-estímulo es un factor decisivo para que el tratamiento tenga efectos positivos. Sin embargo, la frecuencia de aplicación no parece ser un factor crítico, debiendo ser dictada por la estructura de las unidades didácticas.

Asimismo, se infiere la complementariedad de los distintos tipos de pruebas de evaluación.

2) La actitud hacia la física misma no aparece como el factor determinante para explicar los bajos rendimientos que se observan en estudiantes con otros campos de especialización, caso de los químicos y militares, que siguen los cursos con regularidad. Entendemos que los factores más determinantes pueden ser la propia estructura de la física, que conjuga aspectos lógico-deductivos e inductivos y las dificultades de coordinación entre los niveles previos del alumnado, posiblemente bajos, y los exigidos en los cursos, así como en dificultades de conjugación entre los procesos de enseñanza y evaluación.

3) El tratamiento de evaluación estímulo SANBRU, basado en la entrega periódica de resúmenes de la materia explicada en clase, y realizado con alumnos de la asignatura Física de 1.º de Químicas, ha tenido un efecto positivo significativo para todos los alumnos al nivel  $\alpha = 0,01$  en el examen criterio global y de teoría, y al nivel  $\alpha = 0,1$  en el de problemas. Aunque las interacciones no son significativas, se observa un efecto superior en teoría para los alumnos de rendimiento previo bajo.

4) El tratamiento de evaluación estímulo JUSREB, basado en la entrega periódica de problemas tipo, y realizado con alumnos de la asignatura Física de 1.º de Químicas, ha tenido un efecto positivo significativo para todos los alumnos al nivel  $\alpha = 0,05$  en el examen criterio global y al nivel  $\alpha = 0,1$  en el examen criterio de problemas.

Para los alumnos de rendimiento previo alto el efecto positivo se ha centrado en el examen criterio de problemas al nivel  $\alpha = 0,05$ , no produciéndose efecto positivo en teoría, mientras que para los alumnos de rendimiento previo bajo el efecto no ha llegado a ser significativo ni en teoría ni en problemas de forma aislada, pero sí lo ha sido en el examen criterio global al nivel  $\alpha = 0,05$ .

5) El tratamiento de evaluación estímulo CORBIZU, basado en la aplicación periódica en clase de pruebas objetivas cortas y realizado con alumnos de la asignatura Electricidad y Óptica de 3.º de Químicas, ha tenido un efecto positivo significativo para todos los alumnos en el examen criterio de (problemas + cuestiones) al nivel  $\alpha = 0,1$ .

En el examen de cuestiones se ha producido una interacción significativa (tratamiento  $\times$  rendimiento previo) al nivel  $\alpha = 0,1$ .

Para los alumnos de rendimiento previo alto se ha producido efecto positivo al nivel  $\alpha = 0,05$  en el examen criterio de cuestiones y en el de (problemas + cuestiones).

6) Los experimentos BLATUR, realizados con alumnos de la asignatura Física del curso selectivo para el ingreso en la Academia General Militar, y basado en tratamientos similares a los SANBRU y CORBIZU, no han tenido efecto significativo alguno.

7) La disparidad de resultados entre los experimentos realizados en la Facultad de Ciencias y el Curso Selectivo para el Ingreso en la Academia General Militar, nos ofrece un claro ejemplo de interacción entre el tratamiento de la evaluación-estímulo y el contexto didáctico. Es evidente que los estímulos empleados incluyen cierto control en forma de auto-control del estudiante, produciendo efectos positivos en el rendimiento siempre que el contexto educativo no sea rígido, esto es, no haya superado un hipotético punto de saturación de control externo.

8) Los resultados de los experimentos nos confirman la validez general de la evaluación periódica para estimular el rendimiento de los alumnos en física. Sin embargo, existe una interacción clara entre el tipo de evaluación estímulo y el tipo de

evaluación criterio. Los efectos inciden más positivamente en aquellos comportamientos a los que van dirigidos, tienen alguna incidencia en comportamientos cercanos y no suelen tener mucho efecto en comportamientos ajenos.

El hecho de que los distintos tipos de tratamiento hayan incidido de forma diferente en los alumnos de rendimiento previo alto o bajo, tiene su causa en las dificultades que plantea cada tratamiento para ser seguido por unos y otros, así como en el tipo de comportamiento en el que se centran. Así, por ejemplo, el tratamiento de resúmenes es de más fácil seguimiento por parte de los alumnos de rendimiento previo bajo que el de pruebas objetivas.

9) Los distintos tratamientos presentan un elevado grado de complementariedad con vistas a su uso como estímulo didáctico, dependiendo del tipo de comportamiento que se desee estimular y del tipo de alumno de que se trate.

El tratamiento SANBRU incide muy positivamente en la capacidad de síntesis para todos los alumnos aunque de forma más decisiva para los de rendimiento previo bajo. Su efecto sobre la capacidad de aplicación también es positivo aunque de menor significación.

El tratamiento JUSREB incide positivamente en la capacidad de síntesis, comprensión y aplicación de los alumnos de rendimiento previo bajo y en la capacidad de aplicación de los alumnos de rendimiento previo alto.

El tratamiento CORBIZU incide fundamentalmente en la capacidad de comprensión y aplicación de los alumnos de rendimiento previo alto.

## BIBLIOGRAFIA

- ALLEY, R. E.: "Physics in undergraduate engineering education. Report of a survey", *American Journal of Physics*, 40 (1972), 1063-1069.
- ANDERSON, T. W.: *An Introduction to Multivariate Statistical Analysis*, New York, Wiley, 1958.
- ANGELL, G. W.: "The effect of immediate knowledge of quiz results on final examinations scores in freshman chemistry", *Journal of Educational Research*, 42 (1949), 391-394.
- ATKIN, J. Myron: "Research styles in science education". *Journal of Research in Science Teaching*, 5 (1967-1968), 338-345.
- AUSUBEL, D. P.: *Educational Psychology: A Cognitive View*, New York, Holt, Rinehart & Winston, 1968.
- BALCH, J.: "The influence of the evaluating instrument on students' learning", *American Educational Research Journal*, 1 (1964), 169-182.
- BARTLETT, M. S.: "The vector representation of a sample", *Proceeding of the Cambridge Philosophical Society*, 30 (1934), 327-340.
- BATES, C. A.: "Physics degree courses", *Physics Education*, 8 (1973), 135-141.
- BAYLEY, J. M.: "Physics and everything: A bibliography", *American Journal of Physics*, 39 (1971), 1347-1352.
- BEECHING, Lord: "Opening address to the 3<sup>rd</sup> international conference on physics education", en *The Education of a Physicist*, Brown, S. C., y Clarke, N. (ed), London, Oliver & Boyd, 1966, 1-11.
- BEESON, R. O.: "Immediate knowledge of results and test performance", *Journal of Educational Research*, 66 (1973), 224-226.
- BOSE, R. C.: "The fundamental theorem of linear estimation", *Proceedings of 31<sup>st</sup> Indian Scientific Congress*, 1944.
- BRIDGHAM, R. G.: "Methods in research in science education", *Journal of Research in Science Teaching*, 11 (1975), 169-174.
- BROWN, J. A. C.: *The Social Psychology of Industry*, Middlesex, England, Penguin Books, 1954.
- BROWN, W. F. y HOLTZMAN, W. H.: "A study-attitudes questionnaire for predicting academic success", *Journal of Educational Psychology*, 46 (1955), 75-84.
- BRUNER, J. S.: *The Process of Education*, Cambridge, Mass., Harvard University Press, 1960.
- CALFEE, R. C.: *Human Experimental Psychology*, New York, Holt, Rinehart and Winston, 1975.
- CAMPBELL, D. T. y STANLEY, J. C.: "Experimental and quasi-experimental design for research on teaching", en *Handbook for Research on Teaching*, Gage, N. L. (ed.), Chicago, Rand McNally, 1963.
- CASAS, J.: *Optica*, Universidad de Zaragoza, Departamento de Física Fundamental, 1974.
- COHEN, E. G.: *A New Approach to Applied Research: Race and Education*, Columbus, Ohio, Merrill, 1970.
- COLVIN, S. S.: *The Learning Process*, New York, The Macmillan Co., 1913.
- COTHERN, C. R.: "Teaching science to nonscience majors. Some attitudes, ideas, and approaches", *American Journal of Physics*, 41 (1973), 8-12.
- COWAN, D. J.: "Environmental topics in an undergraduate physics curriculum", *American Journal of Physics*, 40 (1972), 1748-1756.
- CRAIG, G. S.: "Certain techniques used in developing a course study in science for the Horace Mann elementary school", *Teacher College Contributions of Education*, Columbia University, New York, Teacher College Press, 1927.
- CRONBACH, L. J. y SNOW, R. E.: *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*, New York, Irvington Publishers, 1977.
- CRONBACH, L. J. y SUPPES, P.: *Research for Tomorrow's Schools*, National Academy of Education, London, Mac Millan, 1969.
- CURTIS, F. D.: "Testing as a means of improving instructions", *Science Education*, 28 (1944), 29-31.

- CURTIS, F. D. y WOODS, G. G.: "A study of the relative teaching value of four common practiches in correcting examination papers", *School Review*, 37 (1929), 615-623.
- DAVIS, A.: "Testing and the course of classroom learning", *Journal of Educational Psychology*, 34 (1943), 526-531.
- Decreto 1678/1969, de 24 de julio, sobre creación de los Institutos de Ciencias de la Educación ("B. O. del E." de 15 de agosto de 1969).
- DEPUTY, E. C.: "Knowledge of success as a motivating influence in college work", *Journal of Educational Psychology*, 20 (1928), 327-334.
- DIXON, W. J. (ed.): *BMD-Biomedical Computer Programs*, Los Angeles, University of California Press, 1973.
- EBEL, R. L.: *Measuring Educational Achievement*, Englewood Cliffs, N. J., Prentice Hall, 1965.
- EISNER, E. W.: "Instructional and expressive educational objectives. Their formulation and use in curriculum", en *Instructional Objectives*, Popham, W. J.; Eisner, E. W.; Sullivan, H. J., y Tyler, L. L., AERA Monograph Series on Curriculum Evaluation, n.º 3, Chicago, Rand McNally, 1969.
- ERAUT, M.: "Promoting innovation in teaching and learning: Problems, processes and institutional mechanisms", *Higher Education*, 4 (1975), 13-26.
- ESCUDERO, T.: *Final Project*, Education, 493A, Stanford University, California, 1974.
- ESCUDERO, T. y FERNANDEZ URIA E.: *Proyecto de Instituto de Ciencias de la Educación. Estructura y funciones. II-III Plan Nacional de Investigación del INCIE*, ICE de la Universidad de Zaragoza, 1975.
- ESCUDERO, T., y FERNANDEZ URIA, E.: "Enfoques conductistas y humanistas en la formación práctica del profesorado", *Educadores*, 90 (1976), 641-656.
- ESCUDERO, T. y FERNANDEZ URIA, E.: "La formación del profesorado de ciencias. Un análisis de la problemática", *Bordón*, 214 (1976), 301-315.
- FITCH, M. L., DRUCKER, A. J. y NORTON, J. A. Jr.: "Frequent testing as a motivating factor in large lecture classes", *Journal of Educational Psychology*, 42 (1951), 1-20.
- FRIEDMAN, B.: "Mathematics in the training of a physicist", en *The Education of a Physicist*, Brown, S. C., y Clarke, N. (ed.), London, Oliver & Boyd, 1996, 63-67.
- GAGNE, R. L.: *The Conditions of Learning*, New York, Holt, Rinehart & Winston, 1965.
- GAMOW, G.: *Biography of Physics*, New York, Harper-Torchbooks, 1964.
- GLASS, G. W.: "Design of evaluation studies", en *Educational Evaluation: Theory and Practice*, Worthen, B. R., y Sanders, J. R. (ed.), Worthington, Ohio, Jones, 1973, 225-232.
- GLUCKSMAN, M. C.: "The use of retesting as a teaching device in an elementary algebra course", *School Science and Mathematics*, 73 (1973), 725-729.
- GOLDBERG, H. S.: "An introductory mechanics laboratory at U.I.C.C.", *American Journal of Physics*, 41 (1973), 1319-1327.
- GREEN, A. E. S.: "Report on the conference on tradition and change in physics graduate education", *American Journal of Physics*, 43 (1975), 214-222.
- HECK, D. L.: "Charts of some upper percentage points of the distribution of the largest characteristic root", *Annals of Mathematical Statistics*, 31 (1960), 625-642.
- HERTZBERG, O. E., HEILMAN, J. D. y LEUENBERGER, H. W.: "The value of objective tests as teaching devices in educational psychology classes", *Journal of Educational Psychology*, 23 (1932), 371-380.
- HOTELLING, H.: "Multivariate quality control, illustrated by the air testing of sample bombrights", en *Selected Techniques of Statistical Analysis*, Eisenhart, C. y al. (ed.), New York, McGraw Hill (1947), 11-184.
- INGRAM, D. J. E.: "The changing face of the physics", *Physics Education*, 10 (1975), 135-137.
- JACOBSON, W. J.: "Approaches to science education research: Analysis and criticism", *Journal of Research in Science Teaching*, 7 (1970), 217-225.
- JERSILD, A.: "Examination as an aid to learning", *Journal of Educational Psychology*, 20 (1929), 602-609.



- JONES, H. E.: "Experimental studies of college teaching", *Archives of Psychology*, n.º 68, 1923.
- JOSSEM, E. L.: "More options for the physics major", *American Journal of Physics*, 40 (1972), 1889-1890.
- KEOHANE, K. W.: "Teaching physics in the future", *Physics Education*, 3 (1968), 338-339.
- KEYS, N.: "The influence on learning and retention of weekly as apposed to monthly tests", *Journal of Educational Psychology*, 25 (1934), 427-436.
- KEYS, N.: "The influence of true-false items on specific learning", *Journal of Educational Psychology*, 27 (1936), 511-520.
- KIRK, R. E.: *Experimental design: Procedures for the Behavioral Sciences*, Belmont, California, Brooks/Cole, 1968.
- KIRKPATRICK, J. E.: *The Motivating Effect of a Epecific Type of Testing Program*, Ph. D. Thesis, Iowa University Research Studies, 1933.
- KITCH, L. V.: *An Experiment in Integrating Testing with Learning in High School Biology*, M. A. Thesis, University of Southern California, 1932.
- KRATHWOL, D. R. y PAYNE, D. A.: "Defining and assessing educational objectives", en *Educational Measurement*, Thorndike, R. L. (ed.), Washington, American Council of Education, 1971, 17-45.
- KUHN, T.: "The function of measurement in modern physical science", en *Quantification*, W. Harry (ed.), Indianapolis, Bobbs-Merril, 1961.
- LAFOURCADE, P. D.: *Planeamiento, Conducción y Evaluación en la Enseñanza Superior*, Buenos Aires, Kapelus, 1974.
- LASHIER, W. S. Jr. y WREN, E. L.: "Effect of pretest feedback and mathematics skills overview on IPS (Introductory Physical Science) achievement", *Science Education*, 61 (1977), 513-518.
- LAWLEY, D. M.: "A generalization of Fisher's Z-test", *Biometrika*, 30 (1938), 180-187.
- MACH, G. R. Jr.: *A comparative Study of Students Performance in an Intermediate Calculus Class as a Result of Different Evaluation Programs*. Ph. D. Thesis, Purdue University, 1963.
- MCKENZIE, G. R.: "Some effects of frequent quizzes on inferential thinking", *American Educational Research Journal*, 9 (1972), 231-239.
- MILLER, J. S.: "Matters of importance in the teaching of physics", en *Teaching Physics Today. Some important topics*, Paris, O.C.D.E., 1965, 47-70.
- MONK, J. J., y STALLING, W. M.: "Another look at the relationship between frequency of testing and learning", *Science Education*, 55 (1971), 183-188.
- MOOD, A. M., y GRAYBILL, F. A.: *Introduction to Theory of Statistics*, New York, McGraw Hill, 1963.
- MORRISON, D. F.: *Multivariate Statistical Methods*, New York, McGraw Hill, 1967.
- MOZER, F. S., y NAPELL, S. M.: "Instant replay and the graduate teaching assistant", *American Journal of Physics*, 43 (1975), 242-244.
- NELSON, C. H.: "Evaluation objectives of science teaching", *Science Education*, 43 (1959), 20-27.
- NOLL, U. H.: "The effect of written tests upon achievement in college classes: An experiment and a summary of evidence", *Journal of Educational Research*, 32 (1939), 345-358.
- ONN, D. G.: "A medical physics course based upon hospital field experience", *American Journal of Physics*, 40 (1972), 1147-1152.
- OPPENHEIMER, F.: "Teaching and learning", *American Journal of Physics*, 41 (1973), 1310-1313.
- OVERALL, J. E. y KLETT, C. J.: *Applied Multivariate Analysis*, New York, McGraw Hill, 1972.
- PEASE, G. R.: "Should teachers give warning of tests and examinations", *Journal of Educational Psychology*, 21 (1930), 273-277.
- PETERSON, J.: "The effect of attitude on immediate and delayed retention: A class experiment", *Journal of Educational Psychology*, 7 (1916), 523-532.
- PHYSICS EDUCATION: N.º monográfico "Physics courses in higher education", 10 (1975).
- PIKUNAS, J. y MAZZOTA, J.: "The effect of weekly testing in the teaching of science", *Science Education*, 49 (1965), 373-376.

- PLOWMAN, L. y STROUD, J. B.: "Effect of informing pupils of the correctness of their responses to objective test questions", *Journal of Educational Research*, 36 (1942), 16-20.
- PRESSEY, S. L.: *Psychology and the New Education*, New York, Harper's, 1933.
- PRICE, R. M. y BRANDT, D.: "Walk-in laboratory: A laboratory for introductory physics", *American Journal of Physics*, 42 (1974), 126-130.
- PROGER, B. B.: *The Relationship Between Four Testing Programs and the Resultant Achievement and Test Anxiety Levels of High and Low Previous Achievement Sixth Grade Arithmetic Students*, Ed. D. Thesis, Lehigh University, 1968.
- RAO, C. R.: *Advanced Statistical Methods in Biometric Research*, New York, Wiley, 1952.
- RAO, C. R.: *Linear Statistical Inference and its Applications*, New York, Wiley, 1965.
- REITZ, J. R. y MILFORD, F. J.: *Fundamentos de la Teoría Electromagnética*, México, D. F., UTEHA, 1969.
- REMMERS, H. H. y REMMERS, E. M.: "The negative suggestion effect of true-false examination question", *Journal of Educational Psychology*, 17 (1926), 52.
- ROSENTHAL, R. y JACOBSON, L.: *Pygmalion in the classroom*, New York, Holt, Rinehart and Winston, 1968.
- ROSS, R. T. y PIRIE, M.: "The persistence of errors in successive true-false tests", *Journal of Educational Psychology*, 25 (1934), 422-426.
- ROTHKOPF, F. Z.: "Some conjectures about inspection behavior in learning from written sentences and the response mode problem in programmed self-instruction", *Journal of Programmed Instruction*, 2 (1963), 31-46.
- ROTHKOPF, F. Z.: "Some theoretical and experimental approaches to problems in written instruction", en *Learning and the Educational Process*, Krumboltz, J. D. (ed.), Chicago, Rand McNally, 1965, 193.
- ROTHKOPF, F. Z.: "Learning from written instructive materials: An exploratory of the control of inspection behavior by test-like events", *American Educational Research Journal*, 3 (1966), 241-249.
- ROY, S. N.: "On a heuristic method of test contribution and its uses in multivariate analysis", *Annals of Mathematical Statistics*, 24 (1953), 22-230.
- ROY, S. N.: *Some aspects of Multivariate Analysis*, New York, Wiley, 1957.
- SAVIRON, J. M. y YARZA, J. C.: *Física General, Termodinámica, Electromagnetismo y Ondas*, Zaragoza, 1974.
- SCANDURA, J. M. y ANDERSON, J.: "Educational research and the science educator", *Science Education*, 52 (1968), 353-357.
- SCOTT, I. O.: "Use of examination to stimulate learning", *California Journal of Secondary Education*, 13 (1938), 223-225.
- SCHOENE, H.: "Essentials for a minimum course in physics", en *Teaching Physics To-day. Some important topics*. Paris, O.C.D.E., 1965, 13-25.
- SCHWAB, J. J.: "The teaching of science as enquiry", en *The Teaching of Science*, Schwab, J. J., y Brandwein, P. F., Cambridge, Mass., Harvard University Press, 1962.
- SELAKOVICH, D.: "An experiment attempting to determine the effectiveness of frequent testing as an aid to learning in beginning college courses in american government", *Journal of Educational Research*, 55 (1962), 178-180.
- SHULMAN, L. S. y TAMIR, P.: "Research on teaching in the natural sciences", en *Second Handbook of Research on Teaching*, Travers, A. (ed.), Chicago, Rand McNally, 1973.
- SMITH, H. A.: "Educational research related to science instruction for the elementary and junior high school: A review and commentary", *Journal of Research in Science Teaching*, 1 (1963), 199-222.
- SONES, A. M. y STROUD, J. B.: "Review, with special reference to temporal position", *Journal of Educational Psychology*, 31 (1940), 602-609.
- SPEARS, J. y ZOLLMAN, D.: "Orientation for the new teaching assistant. A laboratory based program", *American Journal of Physics*, 42 (1974), 1062-1066.
- STANDLEE, L. S. y POPHAM, W. J.: "Quizzes contribution to learning", *Journal of Educational Psychology*, 51 (1960), 322-325.

- STONE, G. R.: "The training function of examinations: Retest performance as a function of the amount and kind of critique information", *Research Report No. AFPTRC-TN-55-8*, San Antonio, Texas, USAT Personnel Training Research Center, Lackland Air Force Base, 1955.
- STOVER, D. W.: "Testing and grading: Procedures for improved student motivation", *The Mathematics Teacher*, 70 (1977), 498-503.
- STRASSENBURG, A. A.: "College Physics teacher preparation. How to do it", *American Journal of Physics*, 39 (1971), 1307-1310.
- SUCHMAN, J. R.: *The Elementary School Training Program in Scientific Enquiry*, Project n.º 216, National Defense Education Act, University of Illinois, 1962.
- TAWNEY, D. A.: "The preparation of physics teachers for secondary schools: The role of university physics departments", *Physics Education*, 8 (1973), 377-379.
- TUCKMAN, B. C.: *Conducting Educational Research*, New York, Harcourt, Brace & Jovanovich, 1972.
- TURNER, A. H.: "The effect of frequent short objective tests upon the achievement of college students in educational psychology". *School and Society*, 33 (1931), 760-762.
- TYLER, R. W.: "Resources, models and theory in the improvement of research in science education", *Journal of Research in Science Teaching*, 5 (1967-68), 43-51.
- TYLER, R. W.: "Analysis of strengths and weakness in current research in science education", *Journal of Research in Science Teaching*, 5 (1967-68), 52-63.
- TYLER, R. W.: "Research in science teaching in a larger context", *Journal of Research in Science Teaching*, 11 (1974), 133-139.
- TYLER, F. T., y CHALMERS, T. M.: "The effect on scores of warning junior high school pupils of coming tests", *Journal of Educational Research*, 37 (1943), 290-296.
- VICK, F. A.: "The case for teaching physics", *Physics Education*, 3 (1968), 338.
- VON WEIZSACKER, D. F.: *La importancia de la Ciencia*, Barcelona, Ed. Labor, 1968.
- WEINSTOCK, H. R.: "The concepts of 'model' and educational research", *Journal of Research in Science Teaching*, 4 (1966), 45-51.
- WILKS, S. S.: "Moment-generating operators for determinants of product moments in samples from a normal system", *Annals of Mathematics*, 135 (1935), 312-340.
- WINER, B. J.: *Statistical Principles in Experimental Design*, New York, McGraw Hill, 2.ª ed., 1971.





