



Universidad Nacional Autónoma de México

FACULTAD DE FILOSOFIA Y LETRAS

División de Estudios de Posgrado

EJEMPLAR UNICO

ANALISIS DE LAS RELACIONES EXISTENTES ENTRE LAS
MEDICIONES PERIODICAS Y EL RENDIMIENTO ESCOLAR.

T E S I S



que para optar por el grado de
Maestro en Enseñanza Superior
P r e s e n t a

FACULTAD DE FILOSOFIA Y LETRAS
ESTUDIOS SUPERIORES

JUAN MARTINEZ DEL CERRO

TESIS CON
FALLA DE ORIGEN

1985



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ANALISIS DE LAS RELACIONES EXISTENTES ENTRE LAS
MEDICIONES PERIODICAS Y EL RENDIMIENTO ESCOLAR.

TESIS CON
FALLA DE ORIGEN

A mis padres:

Arq. Juan Martinez del Cerro.

Sra. Paz Escalante de Martinez
del Cerro.

TESO CUI
FALLA DE ORIGEN

Prólogo.

En los prefacios de tesis anteriores he expresado algunos pensamientos que considero aún válidos y vigentes y por ello los reitero ahora.

"Muchas tesis profesionales se inician con una serie de agradecimientos a todas aquellas personas de las que el sustentante se siente, en alguna forma, deudor y me parece justo y loable que así se haga. Sólo que esos sustentantes son normalmente bastante jóvenes, su vida no ha sido todavía larga y quizá por eso el número de personas a las que muestran su agradecimiento sea más o menos reducido, o demasiado genérico: A mis maestros, a mis amigos, etc."

"Pero....si yo quisiera agradecer en forma específica a todas aquellas personas a las que en este momento de mi vida debo algo, y por supuesto que si quisiera hacerlo, la lista sería tan larga que sólo los nombres ocuparían muchas páginas, aún sin mencionar los motivos de mi agradecimiento. ¿ Se deberá eso a que he vivido un buen número de años?, ¿o a que soy muy afortunado y debo mucho a muchas gentes, a muchos amigos que me han ayudado?

"Sea cual fuere la respuesta, quien lo haya hecho lo sabe y en ese caso al leer estas líneas reciba mi más sincero y afectuoso agradecimiento.

No obstante que mi pensamiento sigue siendo en ese sentido el mismo, desde la fecha inicial en la que lo



expresé y la actual he acumulado nuevas experiencias y limitándome solamente a lo que estrictamente concierne a este trabajo, pues de otra manera volvería a estar en la situación descrita de llenar páginas y páginas con mis reconocimientos, no puedo menos que mencionar a dos personas que con su ayuda, paciencia y estímulo lo han hecho posible, ellos son el Maestro Jesús Aguirre Cárdenas y la Maestra Libertad Menéndez y Menéndez

TESIS CON
FALLA DE ORIGEN

I N D I C E

	Página.
Prólogo	I
Introducción	III
I PRIMERA PARTE.	
<u>Algunos aspectos teóricos sobre la medición y</u> <u>evalaución del rendimiento escolar.</u>	
1. La medición en pedagogía	1
1.1. Breve visión retrospectiva	2
1.2. Conceptos generales.	33
1.3. El proceso de medición	38
1.4. Escalas de medición.	42
1.5. Instrumentos de medición	50
1.5.1. Tipos de pruebas	50
1:5.2. Características de los instrumentos de medición. .	53
1.6. Medición y evaluación.	61
2. La evaluación de programas educativos.	64
2.1. Modelo de Astin y Panos.	66
2.2. La teoría de sistemas.	74
2.3. Las fuentes de información	82
2.4. Teoría de las decisiones	87
I SEGUNDA PARTE	
<u>Informe sobre el trabajo experimental realizado.</u>	
3. Fases iniciales de la investigación.	98
3.1. Planteamiento del problema	100
3.2. Antecedentes	105
3.3. Planteamiento de la hipótesis.	152

282



3.4.	Variabes	159
3.5.	Variabes extrañas.	161
3.6.	Control de variables extrañas	170
3.7.	Definición de términos.	180
3.8.	Aparatos e instrumentos	195
4.	Metodología	214
4.1.	Muestreo.	216
4.2.	Selección del diseño.	223
4.3.	Procedimiento del diseño.	230
5.	Presentación y análisis de resultados	239
5.1.	Análisis descriptivo de los datos	239
5.2.	Tratamiento estadístico de los datos.	251
5.3.	Conclusiones de la investigación.	265
	Conclusiones generales.	268
	Referencias bibliográficas.	272
	Referencias hemerográficas.	280

283

TESIS CON
FALLA DE ORIGEN

Introducción.

De las diversas técnicas que se aplican en el proceso enseñanza-aprendizaje (E-A) las que quizá sean motivo de más frecuentes debates y en las que -hasta donde mi experiencia alcanza, entre los profesores de la Facultad de Arquitectura de la UNAM- a un menor consenso se haya llegado, sean las que conciernen a la evaluación del rendimiento escolar (R-E).

Como parte de ellas, el procedimiento a seguir pudiera ser el punto de partida para que las divergencias disminuyeran. En ese sentido, en la asignatura de Administración de Proyectos y Obras I y II hemos recurrido a los exámenes al final del semestre, a la presentación por parte de los alumnos de trabajos representativos de los contenidos de los cursos, a las pruebas para cada una de las unidades programáticas de la materia y a cualesquiera otros procedimientos que los profesores han considerado como adecuados.

El que los docentes no hayamos llegado hasta ahora a un acuerdo sobre cuál de ellos pudiera ser el idóneo, no creo que se deba a una discrepancia de tipo teórico, pues la mayoría concordamos con los autores que señalan a las pruebas, en especial a las que se administran periódicamente durante el ciclo lectivo, como factores que suelen propiciar una mejoría en el aprendizaje. Lo que estimo que suscita el desacuerdo es la falta de evidencia empírica producto de investigaciones debidamente realizadas en nuestro medio.

Por creerlo así, me propuse hacer una investigación de tipo experimental para suplir la deficiencia anotada, para obtener el necesario apoyo empírico que hiciera posible, inicialmente, llegar a algún consenso sobre cuál sería el procedimiento de evaluación adecuado.



Para ello, procuré ante todo, darle al estudio el debido fundamento teórico y a ese aspecto dedico la primera parte de este trabajo.

En ella se examinan, si bien sólo en forma somera, los principios teóricos en los que se fundamentan las técnicas de medición y evaluación del R-E.

Después de hacer una revisión retrospectiva de la medición en pedagogía y de tratar sobre los aspectos teóricos específicos de ella, se intenta establecer en que difiere de la evaluación y cuáles son las relaciones entre una y otra.

A continuación, se hace un breve análisis del modelo para evaluación de Astin y Panos y se hace una tentativa, siguiendo las indicaciones de los autores, de ampliar su sustento tanto en la teoría de sistemas, como en la de las decisiones.

Con ese apoyo teórico general para la medición y evaluación, en la segunda parte se incluyen los aspectos teóricos específicos en los que se fundamenta la experimentación, así como el examen de los principales factores que inciden en su desarrollo.

Enseguida se aborda lo concerniente a la metodología directamente aplicable a los estudios del tipo del que nos ocupa y se hace después una relación sucinta de la manera en la que fue realizado.

Por último se presentan los resultados obtenidos, se analizan a la luz de la estadística para obtener algunas probables inferencias y se proponen ciertas conclusiones posibles.

Por la experiencias que me ha dejado el estudio experimental someramente descrito, estoy ahora en aptitud de apreciar las consecuencias que las limitaciones metodológicas tuvieron en los resultados.

En las fases iniciales de la investigación no me percaté plenamente de los efectos que ellas tendrían y no fue, sino



cuando dispuse de algunos datos concretos y traté de derivar de ellos algunas inferencias, cuando tomé clara conciencia de la medida en la que, como veremos en su oportunidad, nos habían afectado en detrimento de la validez de dichas inferencias.

Aún así, a pesar de la variedad de factores que atentan en su contra, estimo que este trabajo experimental podría servir, cuando menos, como un antecedente para alentar a otros investigadores del área a emprender trabajos similares en los que estuviera presente un mayor control de las variables extrañas.

TESIS CON
FALLA DE ORIGEN

I. PRIMERA PARTE.

ALGUNOS ASPECTOS TEORICOS SOBRE LA MEDICION
Y EVALUACION DEL RENDIMIENTO ESCOLAR.

TESIS CON
FALLA DE ORIGEN

La medición en pedagogía.

En las interacciones que el hombre intencionalmente emprende con el medio ambiente que lo rodea, el conocimiento y evaluación de los efectos que ocasiona son para él de particular importancia, pues en esa información se apoyará para regular sus acciones futuras.

Cuando esa evaluación, ese juicio de valor, se refiere, como es frecuente el caso en materia educativa, al grado de aptitud alcanzado por alguna persona en el desempeño de ciertas tareas específicas, se ha recurrido, desde tiempos bastante remotos, al uso de instrumentos de medición, a la aplicación de pruebas escolares con el fin de fundamentarlo tan objetivamente como sea posible.

La revisión de la historia de la evolución de las técnicas de medición y evaluación, al permitirnos saber cuáles fueron los problemas que al paso de los años surgieron y las soluciones que se intentaron, es probable que nos ofrezca, como dice Robert L. Ebel: "...un marco de referencia esencial para la construcción de una estructura coherente de conocimiento en el campo..." (1) de la medición y evaluación del rendimiento escolar.

TESIS CON
FALLA DE ORIGEN

1. EBEL, ROBERT L. Fundamentos de la medición educacional, 15.

1.1. Breve visión retrospectiva.

Entre los antecedentes más tempranos del uso de exámenes, quizá el de mayor antigüedad, lo podamos encontrar en China durante la dinastía Han (202 A.C.- 221 D.C.), época en la que el nombramiento de los funcionarios del gobierno se hizo depender, ya de manera usual, de sus méritos. No fue sin embargo, sino hasta el advenimiento de la dinastía Tang (618-906 D.C.) cuando los reglamentos respectivos fueron completados.

Esos exámenes eran de tres grados mayores, además de otros muchos preliminares e intermedios, lo que aunado a la necesidad de cumplir numerosas condiciones especiales hacían que fuera bastante difícil aprobarlos.

Pero el esfuerzo valía la pena, pues quien los presentaba y aprobaba podía tener acceso a los puestos en el gobierno y conseguir así, no sólo las prerrogativas propias de la burocracia, sino también algún ascenso en la escala social.

Resulta que de ese modo, en la China Imperial los exámenes eran el medio idóneo para que los individuos, al esforzarse y demostrar su capacidad, pudieran sobreponerse a las desventajas de su origen social y alcanzar los privilegios del estrato al que ascendían.(2)

TESIS CON
FALLA DE ORIGEN

2. WEBER, MAX, Essays in Sociology, 423.

Se puede decir que en esta forma se inician los esfuerzos de los educadores en pos de los medios que habrían de permitirles constatar con la mayor objetividad posible los avances de los estudiantes.

En el camino hacia el fin indicado encontramos algunos logros de especial importancia y entre ellos destacan los siguientes:

En la época del renacimiento todavía predominaban en Europa los exámenes orales en uso desde la Edad Media. Este tipo de pruebas, al paso de los años, había sufrido un considerable deterioro, pues su administración se convirtió en una mera ceremonia en la que, prácticamente, todos los alumnos resultaban aprobados. Esta circunstancia evitó que surgieran objeciones o que se manifestara algún deseo de modificar los procedimientos acostumbrados.

Hacia fines del siglo XVI los Jesuitas, preocupados por la situación imperante y en su afán por dar a la educación algún rigor científico expresaron sus ideas sobre la teoría y la práctica de la misma, incluyendo un conjunto de reglas para los exámenes que ellos proponían que fueran escritos. (3)

No obstante, las pruebas orales, de dudosa eficacia, siguieron siendo las que con mayor frecuencia se impartían y fue solo alrededor de 1760 que los exámenes escritos se empezaron a usar en las

3. EBEL, ROBERT L. Op. cit., 18.



universidades de la Gran Bretaña, mientras que en las de los E.E.U.U. su uso se inició, aparentemente, cerca de 1830.

En los años próximos a 1845, en las escuelas de Boston se encargó a un subcomité la tarea de examinar a los alumnos y éste, debido sobre todo a la gran cantidad que de ellos había, elaboró una serie de pruebas escritas. Estas pruebas, a pesar de no recibir una aceptación generalizada hicieron una buena impresión en algunos educadores, entre ellos en Horace Mann, quien propugnó por su utilización aduciendo las ventajas siguientes:

- a) Se pueden lograr más elementos de juicio sobre el rendimiento de cada alumno.
- b) Se dispone de una constancia escrita de ese rendimiento.
- c) A cada alumno se le formulan las mismas preguntas; de ese modo todos reciben igual trato.
- d) Hay menos probabilidad de favoritismo o animadversión respecto de un alumno o maestro concreto.

Estos argumentos aunados a la fuerte personalidad de su autor hicieron que las pruebas escritas empezaran a predominar sobre las orales.

Hacia falta, sin embargo, establecer criterios objetivos para su calificación y en este sentido se dirigió la

TESIS CON
FALLA DE ORIGEN

proposición del Reverendo George Fisher, quien diseñó y publicó escalas acerca de los niveles de calidad en ortografía, matemáticas y otros temas de estudio, las que no obstante tener una significativa validez de contenido, no fueron apreciadas en esas fechas y no fue sino hasta principios de este siglo cuando Thorndike y otros autores como Ayres, Curtis, Starch y Odell retomaron el enfoque de Fisher para promover el movimiento científico en la educación. (4)

Para este movimiento fueron particularmente importantes las aportaciones de Wilhelm Wundt, James McKeen Cattell, Francis Galton y Alfred Binet.

El primero de ellos, Wundt, estableció en 1879 el primer laboratorio de psicología experimental en Leipzig, Alemania.

En las primeras etapas de su trabajo se interesó, al igual que los fisiólogos, en la medición de las sensaciones: vista, tacto, oído, etc., pero poco a poco fueron los aspectos claramente psicológicos, como la medición del campo perceptivo, el ritmo de aprendizaje y otros similares, los que atraieron su atención.

Estudiando con Wundt, en Alemania, James McKeen Cattell adquirió sólidos conocimientos en materia de psicología cuantitativa y experimental. En esa misma época conoció los trabajos de Galton y se interesó en ellos, especialmente en sus estudios sobre diferencias



4. EBEL, ROBERT L. Op. cit., 20

individuales y método estadístico.(5)

Sir Francis Galton, el célebre primo de Darwin, fue un dedicado estudioso de la medición y le concedió un alcance tal que llegó a sostener que con suficiente trabajo e ingenio todo podía ser medido.

No obstante su valía y seriedad, su sentido del humor lo condujo a proponer lo siguiente: "...Muchos procesos mentales pueden ser medidos aunque sólo sea en una forma burda. Por ejemplo: el grado en el que la gente se aburre, puede ser obtenido al través del número de los movimientos con los que demuestra su inquietud".(6)

Bromas aparte, Galton hizo importantes estudios sobre la posibilidad de que las características individuales fueran hereditarias, y en 1869 publica sus conclusiones en una obra titulada 'Hereditary Genius'. Postula en ella que las características personales se heredan, no sólo las físicas, sino también las aptitudes y rasgos de la personalidad. Suponía, asimismo, que las tendencias delictivas y los trastornos psicológicos se transmiten de padres a hijos por herencia directa.(7)

Se percató, sin embargo, que era necesario dar a sus postulados el debido apoyo empírico y se dedicó a medir una considerable variedad de atributos humanos.

Entre ellos, el poder de discriminación sensorial, pues pensaba que quien tuviera los sentidos más agudos tendría también una mayor capacidad intelectual.

5. THORNDIKE, ROBERT L. y Elizabeth Hagen. Tests y técnicas de medición en psicología y educación, II, 13.
6. GOULD, STEPHEN J. The Mismeasure of Man, 75.
7. NUNNALLY, JUM C. Introducción a la medición psicológica, 58.



De igual forma, al regresar J. M. Cattell a los E.E. U.U. en 1888, imbuido del respeto que por el método científico le había transmitido Wundt y del gran interés por la medición que había adquirido de Galton, se ocupó también de hacer mediciones acerca de la agudeza sensorial tratando de determinar si tenían una relación directa con la capacidad intelectual.

Este tipo de relación no pudo ser empíricamente probado. Los estudios de Seashore, Wissler, Bagley y otros, no rindieron ninguna prueba positiva en ese sentido. (8)

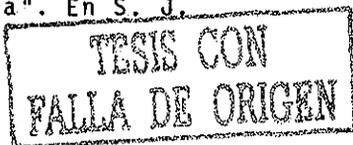
Tanto Cattell, como el cuarto de los autores mencionados, Alfred Binet, ejercieron una gran influencia en la medición en psicología y educación.

El segundo de ellos (1857-1911) se dedicó, siendo director del Laboratorio de psicología de la Sorbona, al estudio de la medición de la inteligencia aplicando para ello el método propuesto por Paul Broca. Se abocó, dicho brevemente, a medir cráneos, sin poner en duda la conclusión básica de la escuela de ese autor: "La relación entre la inteligencia de los sujetos y el volumen de sus cabezas... es muy real y ha sido confirmada por todos los investigadores metódicos, sin excepción...". (9)

En el curso de los tres años siguientes al de la aparición del 'Año Psicológico', en 1895, Binet

8. EBEL, ROBERT L. Op. cit., 21.

9. BINET, ALFRED. "Historia de las investigaciones sobre las relaciones entre la inteligencia y el tamaño y forma de la cabeza". En S. J. Gould, Op. cit., 146.



continuó haciendo estudios sobre craneometría y al cabo de ellos su seguridad en los planteamientos de Broca se había desvanecido. Su fe original fue destruida como resultado del examen de cinco cabezas de niño de edad escolar, pues las diferencias en las medidas de los cráneos no pudieron correlacionarse con los niveles de inteligencia mostrados por los sujetos en su desempeño ordinario.

En 1904 fue comisionado por el ministro de educación pública para realizar un estudio con un propósito específico y práctico: desarrollar técnicas para identificar aquellos niños cuya falta de éxito en las clases normales parecía indicar la necesidad de alguna forma de educación especial.

Binet eligió un procedimiento puramente pragmático. Reunió una serie de tareas cortas relacionadas con problemas de la vida diaria (contar monedas, decidir qué cara es más 'bonita', por ejemplo), pero que supuestamente implicaban algún proceso básico de razonamiento, como ordenamiento, comprensión, invención y corrección. El esperaba que la mezcla de varias pruebas, cada una para cierta capacidad, le permitiría expresar el potencial general de un niño en un solo número.

De la escala creada por él, Binet alcanzó antes de su muerte en 1911, a publicar tres versiones, haciendo

TESIS CON
FALLA DE ORIGEN

notar en 1905 que: "La escala, en rigor, no permite medir la inteligencia, porque las capacidades intelectuales no se pueden sobreponer y por lo tanto no puede medirse como si se tratara de una magnitud lineal". (10)

Reconoció siempre la complejidad de la inteligencia y era un teórico demasiado apto para caer en el error lógico que John Stuart Mill identificó y que consiste en creer que todo aquello que recibe un nombre debe ser una entidad o un ser con su propia existencia independiente. (11)

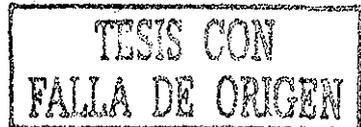
En la misma época en la que Binet desarrollaba su escala para medir la inteligencia Edward L. Thorndike publica el primer libro de texto sobre la medición en educación, pues creía firmemente en su valor. "Todo lo que existe, decía, existe en cierta cantidad". Conocerlo exhaustivamente implica conocer su cantidad lo mismo que su calidad". (12) Fue autor, junto con sus alumnos, de las primeras escalas para medir el rendimiento educativo.

Con esa finalidad, propuso pruebas objetivas que fueran motivo de fuertes objeciones; se dijo que eran fragmentarias y superficiales y que tal vez fueran adecuadas para comprobar la memorización de hechos, pero no los procesos mentales de razonamiento lógico, evaluación crítica o síntesis creativa. Estas

10. GOULD, STEPHEN J. Op. cit., 151.

11. Ibidem, 151.

12. EBEL, ROBERT L. Op. cit., 23.



capacidades, según los impugnadores de las pruebas objetivas, sólo podrían apreciarse por medio de las de ensayo.

Daniel Starch replicó a esas críticas objetando las pruebas de composición. En colaboración con E. C. Elliott realizó una serie de estudios en los que puso de manifiesto que las puntuaciones que los profesores daban a un alumno en una prueba de ese tipo, eran notoriamente distintas.

Al respecto, si bien las pruebas de composición pueden ser calificadas de manera confiable cuando la tarea se realiza en forma cuidadosa y bajo la debida supervisión, no por ello las medidas de rendimiento que producen son, en términos generales, subjetivas y poco confiables.

Starch y Elliott pudieron, en efecto, comprobar lo arriba expresado acerca de las pruebas de ensayo, pero no lograron refutar el cargo principal contra las pruebas objetivas, ni demostrar que sirvieran para medir todo aquello que con una prueba de composición es posible.

Las múltiples aportaciones que sobre el tema surgieron en ese entonces se traslapan cronológicamente, intentaré sin embargo, hacer mención de cuando menos las principales de ellas, las que sin duda contribuyeron a enriquecer el conocimiento de nuestro tema.

TESIS CON
FALLA DE ORIGEN

En 1904, Charles Spearman publicó un escrito titulado 'La inteligencia general objetivamente medida y determinada' en el que exponía las conclusiones a las que había llegado al observar que si dos tests mentales se administran a un gran número de personas, el coeficiente de correlación entre los resultados de uno y otro es casi siempre positivo.

Esa correlación positiva, pensó, era indicio de que cada test no medía un atributo independiente de la función mental y, en consecuencia, debía haber alguna estructura común subyacente.

Esa estructura podía ser de tipo 'bifactorial'. La resolución de cada test requería, de acuerdo con su punto de vista, la posesión de alguna información específica, a la que llamó 's', pero también reflejaba la operación de otro factor, al que denominó 'g' o inteligencia general. (13)

Como una posible prueba para su hipótesis, Spearman propuso el coeficiente para el rango de correlación que lleva su nombre y que está definido por la ecuación siguiente:

$$r_s = 1 - \frac{\sum_{i=1}^n D_i^2}{n(n^2-1)}$$

13. GOULD, STEPHEN J. Op. cit., 256, 257.

Donde:

n = número de pares de observaciones (x_i, y_i) .

D_1 = rango (x_i) - rango (y_i)

La ecuación hace posible establecer la correlación entre dos conjuntos de variables ordenadas por rango, pero aun en el caso de que fuera positiva y cercana a uno, "...los factores (s y g) pueden considerarse como influencias causales hipotéticas subyacentes que determinan las relaciones observadas entre los conjuntos de variables. Solamente cuando se les considera desde este punto de vista tienen interés y significado en psicología". (14)

En 1916, Lewis M. Terman hizo una revisión de la escala de Binet publicada en 1911 y al extender su rango de aplicabilidad a los adultos amplió el número de tareas de cincuenta y cuatro, propuestas por Binet, a noventa.

Terman era en ese entonces profesor de la Universidad de Stanford y de ahí el nombre de la prueba que propuso: *Stanford-Binet*. Su objetivo principal no fue tanto afinar y ampliar la original, sino hacer que fuera posible aplicarla, prácticamente, a todos los niños, pues tenía la esperanza de establecer una graduación de habilidades innatas para que cada niño encontrara su lugar adecuado en la vida.

14. EYSENK, H. J. "The Logical Basis of Factor Analysis", 113, en S. J. Gould, Op. cit., 268.

Haciendo uso de los métodos estadísticos adecuados, Terman consiguió establecer la escala, de manera que el niño 'promedio' obtuviera 100 puntos en cada edad, o sea que su edad, cronológica fuera igual a la mental.

Estableció también una desviación estándar de 16 puntos para cada edad cronológica.

La validez de este tipo de pruebas para medir la inteligencia sigue siendo motivo de dudas, pues no se ha logrado establecer de manera inequívoca que sea esa la característica que miden y no, como parece más probable, que se limitan a establecer una correlación con normas aceptadas.

No obstante, la prueba que con mayor frecuencia se usa en la actualidad para medir la inteligencia es la resultante de las diversas revisiones realizadas a la que desde 1916 se conoce como la Stanford-Binet. Después de la revisión de Terman en ese año, fue sometida a otras: en 1937 en la misma Universidad de Stanford, en 1960 por el propio Terman y en 1972 cuando se publican una serie de normas para interpretar los resultados.

La importancia de la prueba radica, tanto en el gran número de personas a quienes se administra, en 1975 Thorndike informó que el promedio anual era de 800,000 aproximadamente, como en que para muchos niños el resultado que hayan obtenido se convierte en una

etiqueta que los marca, cuando menos, durante toda su estadía en las escuelas.

Un poco después de la aparición de la versión de 1916 de la prueba Stanford Binet, al inicio de la participación de los E.E. U.U. en la Primera Guerra Mundial, en 1917, Robert M. Yerkes y un grupo de colaboradores como Terman, Goddard y otros basándose en los trabajos de Arthur Otis, elaboraron los tests conocidos como Alfa y Beta con el fin de medir las aptitudes de los reclutas del ejército y de acuerdo con los resultados asignarles las tareas adecuadas.

El primero 'Army Alpha Test', es breve y contiene preguntas sobre aritmética, razonamiento e información general, el segundo el Test Beta, se usó para examinar a personas, que o bien tenían poco dominio del inglés, o habían fracasado en el Alfa. Se trata de la resolución de rompecabezas geométricos y del análisis de ilustraciones.

Los resultados, inicialmente obtenidos, después de haberlo administrado a cerca de un millón setecientos mil reclutas, mayores de 18 años, fueron sorprendentes, pues la edad mental promedio de los de origen caucásico apenas sobrepasaba a los trece años y la de los negros los diez.

El análisis de los contenidos del test dió apoyo a una posible explicación: lo que se estaba midiendo no era

TESIS CON
FALLA DE ORIGEN

la inteligencia, aun suponiendo que exista como una entidad identificable, sino la influencia que el medio sociocultural había tenido en la persona.

Debido a eso, los negros y también los blancos provenientes de los medios socio-económicos menos favorecidos obtenían puntuaciones inferiores que quienes pertenecían a estratos menos bajos.

Los años comprendidos entre las dos guerras mundiales, 1918 a 1939, fueron una época de rápido desarrollo de las técnicas de medición educativa.

En este sentido se pueden citar las aportaciones de William A. McCall en 1920, pues fue el primero en exhortar a los maestros para que emplearan pruebas objetivas; la de Edward K. Strong, quien desarrolló una técnica para medir el interés de los estudiantes hacia las diversas actividades de aprendizaje que emprendían. Merece especial mención la de Carl Brigham, uno de los discípulos de Yerkes, pues fue creador de una prueba para el desempeño escolar (scholar aptitude), de tipo objetivo para medir aptitudes verbales de carácter general y cuantitativo, con el fin de que sirviera de complemento a las de composición que se usaban para evaluar distintos aspectos del rendimiento de los alumnos. Desde 1926, las diferentes versiones de la prueba de Brigham, han servido para que aquellos que quieren ingresar al 'college' demuestren su grado de aptitud.

Unos años antes de esa última fecha, en 1923, un comité integrado por Terman, Thorndike y Yerkes en especial, diseñó una prueba llamada 'National Intelligence Test' (Prueba nacional de inteligencia), en la que retomando la proposición anterior de Terman la modificaban e incluían dos escalas. La A que consistía en reactivos sobre: razonamiento aritmético, complementación de oraciones, selección lógica, sinónimo-antónimo y relaciones de símbolos-dígitos; y la B que comprendía los concernientes a: información, complementación, vocabulario, analogías y comparaciones.

Su principal objetivo era adaptar las pruebas que se habían administrado al ejército a las necesidades de las escuelas. Los autores sostenían que eran fáciles de aplicar, confiables y útiles para clasificar a los niños entre los grados tercero y octavo, en función de sus capacidades intelectuales.

En estas pruebas, como en las anteriormente reseñadas, se intentaba que la medida de la inteligencia se expresara con un solo número, lo cual daba origen, de nueva cuenta a la reificación de la inteligencia, a su conceptualización como una entidad identificable y no como la conjunción de diversas habilidades para realizar diferentes tareas.

En 1928, patrocinado por una beca de la Carnegie

Foundation para el progreso de la enseñanza, Ben D. Wood y William S. Learned, emplearon pruebas objetivas en el primer estudio de amplio alcance sobre el rendimiento escolar. Los resultados mostraron considerables variaciones entre las puntuaciones obtenidas por los estudiantes, no sólo cuando pertenecían a distintas escuelas, sino aun cuando asistieran a la misma institución.

Fue, asimismo notorio, que en las pruebas administradas por ciertos maestros, los alumnos conseguían, en general, muy bajas puntuaciones. A estas pruebas se las criticó por contener una excesiva proporción de preguntas relativas a hechos aislados, pero ni siquiera en las respuestas a esos reactivos que fueron considerados como 'simples', pudieron los alumnos obtener buenas calificaciones. (15)

En la década de los treinta, surgieron de nueva cuenta críticas en el sentido de que las pruebas objetivas sólo eran útiles para medir el conocimiento de hechos aislados. Estas objeciones fueron tomadas en cuenta por algunos de los estudiosos de la medición y en respuesta a ellas Ralph Tyler publicó en 1931 en el 'Educational Research Bulletin' un artículo titulado 'A Generalized Technique for Constructing Achievement Tests' (Una técnica generalizada para construir pruebas de aprovechamiento), que tenía como fundamento la definición de los objetivos en términos de las conductas

15. EBEL, ROBERT L. Op. cit., 28.

que el alumno debía mostrar. Se insistía, sin embargo, más en el desarrollo de las capacidades que en la adquisición de conocimientos.

Los estudios que hizo Tyler al respecto, tratando de encontrar si existía alguna correlación entre las puntuaciones obtenidas en las pruebas de capacidad y en las de conocimiento, no mostraron que la hubiera en forma significativa, pero sí que los estudiantes tendían a conservar mejor las capacidades que los conocimientos.

Aun así, muchos de los maestros manifestaron que encontraban difícil definir en términos conductuales los resultados de las principales enseñanzas que intentaban impartir. (16)

El problema subsiste y sólo puede, hasta cierto punto, subsanarse incluyendo en los objetivos alguna manifestación observable y en consecuencia medible, del aprendizaje que pretendemos hayan obtenido los alumnos.

Hubo en la misma década, otros avances importantes, tales como la formación de una junta examinadora en la Universidad de Chicago que inició sus trabajos en 1932 con L. L. Thurstone como examinador general y un distinguido grupo de especialistas en medición que incluía a John M. Stalnaker y a Ralph Tyler. Su labor consistió en diseñar exámenes y en solucionar diversos problemas en el área de su especialidad.

16. Ibíd., 30.

Por la misma época, hubo otras aportaciones dignas de citarse como lo fueron: la máquina para la puntuación de pruebas, que a instancias de Ben Wood, produjo en 1935 la IBM; las pruebas de aptitud desarrolladas en 1936 por las universidades de Harvard, Yale, Columbia y Princeton, que con el fin de predecir el desempeño de los alumnos en posgrado se aplicaba a quienes habían egresado del pregrado. Su validez predictiva resultó por demás limitada, salvo en aquellos casos en los que las escuelas de posgrado reconocían y recompensaban a los estudiantes que obtenían altas puntuaciones en esas pruebas.

Antes de proseguir con esta breve visión retrospectiva es conveniente recordar, aunque sólo sea en forma somera, la aportación que hizo Spearman pues está directamente relacionada con la de L.L. Thurstone que es a la que nos corresponde referirnos en orden cronológico.

Como vimos, en 1904 Spearman propone una posible estructura para la inteligencia en la que se podrían distinguir dos factores: uno de ellos 's', relativo a la información específica ; y otro, 'g', a la inteligencia general.

Entre los años de 1935 y 1950, L. L. Thurstone, profesor de Psicología en la Universidad de Chicago desde 1925 hasta su muerte en 1955, publica libros y

artículos en revistas especializadas proponiendo un procedimiento para el análisis factorial diferente del creado por Spearman.

La principal diferencia entre uno y otro estriba en la desaparición del factor 'g', quedando incluidas únicamente las 'habilidades mentales primarias' (HMP).

La consecuencia más importante del cambio radica en que al incluirse sólo las HMP y no la 'g', al no haber una habilidad general que medir, el niño no será clasificado linealmente según el grado en el que supuestamente posea una, hasta cierto punto hipotética, inteligencia innata. Se abre así el camino hacia la posibilidad de reconocer que la posesión de ciertas habilidades específicas no es, por sí misma, ni mejor ni peor que la de otras. (17)

Por la misma época, en 1938, al observar Oscar K. Buros que el gran número de pruebas publicadas daba lugar a que muchas de ellas no tuvieran la debida calidad, inició una publicación anual en la que se hacía un examen crítico de las que se ponían en uso con el fin de guiar a los usuarios en la selección que hicieran e inducir a los editores a mejorar la calidad de las que publicarán en el futuro. Estos anuarios, si bien no han escapado a las críticas y no están libres de defectos, han llegado a ser reconocidos como guías indispensables para la selección de las pruebas en los E.E. U. U. (18)

17. GOULD, STEPHEN J. Op. cit., 296-304.

18. EBEL, ROBERT L. Op. cit., 32.

La segunda guerra mundial, al igual que la primera, hizo necesario que las pruebas usadas para clasificar a los reclutas se adaptaran a las condiciones imperantes: mayor número de soldados y tareas más diversificadas.

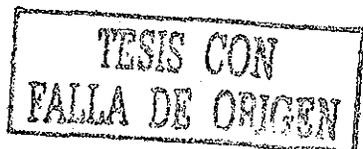
Las innovaciones, sin embargo, no alcanzaron la importancia de las implantadas en la primera gran guerra con la aplicación de las pruebas Alfa y Beta. Consistieron de manera primordial en el uso de pruebas para aptitudes múltiples y en el aumento de los tipos disponibles.

En el otro aspecto, el gran número de reclutas que debían ser clasificados y asignados a labores específicas, hizo necesario el uso de procedimientos más expeditos, tanto para obtener los datos, como para procesarlos. Con ese objeto se diseñó la maquinaria apropiada para acelerar la clasificación e interpretación de los resultados de las pruebas.

No obstante, quizá el mayor avance productó de la segunda guerra lo haya sido el logrado en cuanto a la interpretación y uso de los resultados de las pruebas.

En los años de la posguerra, uno de los logros principales es el que se refiere a la toma de conciencia con respecto al papel que juega la administración de pruebas en educación. Han dejado de ser un fin en sí mismas para convertirse en un elemento más en el proceso E-A. (19)

19. SMITH, FRED M. y Sam Adams. Educational Measurment for the Classroom Teacher, 23.



Por esos mismos años, en 1948, como resultado de la invitación que Ralph W. Tyler extendió a otros examinadores para cambiar impresiones sobre los problemas a los que se enfrentaban, surgió la idea de clasificar y organizar los objetivos de la educación superior. Benjamin S. Bloom y un destacado grupo de investigadores, entre quienes podemos citar a L. J. Cronbach, Robert L. Ebel, Max Engelhart, N. L. Gage y en especial a David R. Krathwohl, se abocaron a dicha tarea y en 1951, los resultados hasta entonces obtenidos, fueron discutidos en un simposio organizado con motivo del Congreso Anual de la American Psychological Association que tuvo lugar en la ciudad de Chicago, Ill. En 1956 se publica el libro de Bloom y sus colaboradores titulado: 'Taxonomy of Educational Objectives. Handbook I: Cognitive Domain'. Pronto surgen las críticas, referidas en especial a los siguientes aspectos: a) La dificultad para encontrar ejercicios específicos para cada categoría taxonómica, b) Las categorías no están claramente definidas y no son, por lo tanto, mutuamente excluyentes, c) La separación entre los tres dominios es más bien artificial y, d) Los nombres de las categorías no parecen poseer gran significado operacional.

Con miras a corregir esas y otras debilidades de la taxonomía de Bloom para el dominio cognoscitivo, G. Madaus, E. Wood y R. Nuttall hicieron un estudio de



La validez de la jerarquía y del posible efecto de la introducción del factor 'g' (inteligencia general) para explicar la estructura taxonómica.

Como resultado del análisis causal que los autores mencionados realizaron, obtuvieron las siguientes conclusiones: En los tres primeros niveles: conocimiento, comprensión y aplicación; la jerarquía, en principio, se conserva correctamente.

En el cuarto nivel, análisis, disminuyen los vínculos entre niveles contiguos y empiezan a aparecer otros entre los que no lo son. Este fenómeno, sin embargo, puede ser explicado al introducir el factor 'g' como complementario.

En los niveles superiores, síntesis y evaluación, casi desaparece la vinculación directa y el factor 'g' se convierte en el principal elemento para explicar los resultados observados. (20)

Tomando en cuenta lo anterior, Robert L. Ebel propone que las medidas clásicas de rendimiento se limiten a los primeros cuatro niveles: conocimiento, comprensión, aplicación y análisis. Los otros dos: síntesis y evaluación serían tratados por separado.

A partir de la taxonomía de Bloom han surgido otras muchas proposiciones, algunas con objeto de modificarla, otras con el de completarla o refinarla. Acerca de las más difundidas de ellas trataré a continuación.

20. DE LANDSHEERE, VIVIAN Y GILBERT. Objetivos de la educación, 96.



Algunos autores, como R. Horn por una parte y N. Metfessel, W. Michael y D. Kirsner por la otra, han conservado en lo fundamental la taxonomía de Bloom y las proposiciones que han hecho tienen por objeto operacionalizarla. Con ese fin, a la denominación del nivel o subnivel se agrega la actividad conducente para alcanzar el objetivo.

Entre quienes la modifican, de manera más o menos sustancial, podemos citar el caso de J. P. Guilford con su proposición de un modelo tridimensional que consiste en situar en el eje de las 'x' los tipos contenidos: figurativos, simbólicos, semánticos y conductuales; en el de las 'y' los productos: unidades, clases, relaciones, sistemas, transformaciones e implicaciones; y en el de las 'z' las operaciones, producción divergente, convergente, memoria y cognición. (21)

Incluyo a continuación el modelo de Gagné-Merrill, porque si bien no se limita al dominio cognoscitivo, pues comprende también el afectivo y el psicomotor, comparte con el de Guilford la crítica de que ninguno de los dos apunta en forma directa hacia los objetivos de la educación. Se podría decir que se trata en realidad de una clasificación de los procesos necesarios para que el aprendizaje indicado en los objetivos sea alcanzado.

21. Ibidem, 110-112.



A partir de los ocho tipos de aprendizaje que Gagné distingue y ordena de lo simple a lo complejo, él y Merrill elaboran una taxonomía que puede resumirse de la manera siguiente:

Dominio	Niveles	Categorías
Afectivo.	1. Emocional (respuesta involuntaria).	1. Emocional (aprendizaje de señales).
Psicomotor.	2. Psicomotor (nivel poco elevado de conciencia).	2.1. Comportamiento topográfico. 2.2. Cadenas. 2.3. Comportamiento hábil.
	3. Memoria. (comportamiento conciente).	3.1. Denominación. 3.2. Memoria serial (asociación verbal). 3.3. Memoria discreta (asociación múltiple).
Cognoscitivo.	4. Cognoscitivo complejo (conciencia intensa del comportamiento, transferencia).	4.1. Clasificación (aprendizaje conceptual). 4.2. Análisis (aprendizaje de principios). 4.3. Síntesis (resolución de problemas).



Además de las citadas, son dignas de mención en cuanto al dominio cognoscitivo, las de V. Gerlach y A. Sullivan y la de De Block.

En el dominio afectivo, los problemas para definir y jerarquizar los objetivos educativos son aún mayores que para el cognoscitivo. En los dos casos es difícil clasificarlos como pertenecientes a un dominio en forma exclusiva, pues en la conducta que se pretende inculcar no es posible discernir el efecto de lo intelectual del de lo afectivo o psicomotor, como sería indispensable hacerlo, de manera precisa por cierto, para especificar los objetivos de cada dominio por separado.

Además de esta dificultad, compartida por los tres dominios, en el afectivo se presenta otra, como lo es la de establecer algún consenso sobre el significado de los términos que se emplean para describir las conductas. v.g. apreciar el arte, valorar la integridad, etc.

Para este dominio, Krathwohl publicó en 1964 una taxonomía que comprende cinco niveles básicos:

- 1.00 Recepción.
- 2.00 Respuesta.
- 3.00 Valorización.
- 4.00 Organización.
- 5.00 Caracterización por un valor o sistema de valores.

TESIS CON
FALLA DE ORIGEN

Varios autores han hecho proposiciones modificando el enfoque de Krathwohl y en vez de establecer la jerarquía de acuerdo con el nivel de interiorización lo hacen en relación al que correspondería a la actividad.

En ese sentido Landsheere propone la siguiente taxonomía:

- I. El individuo responde a un estímulo exterior.
 1. Es simplemente receptivo.
 2. Recibe y reacciona.
 3. Recibe y reacciona aceptando o rechazando.
- II. El individuo toma la iniciativa.
 4. Trata espontáneamente de comprender, de juzgar, de experimentar.
 5. Actúa según sus opciones.

Con un enfoque distinto, W. French publicó una taxonomía que puede considerarse como perteneciente al dominio afectivo porque se centra en el alumno y no en la materia. No obstante, al examinarla es posible percatarse de que en ella no se distingue con claridad lo afectivo de lo cognoscitivo. Su enfoque es claramente normativo y se puede decir que se refiere a los requisitos necesarios para ingresar al sistema social existente y no a las actitudes conducentes para su superación o impugnación. (22)

22. Ibidem, 160-162.



Comprende tres rubros principales:

1. Realizarse.
2. Volverse capaz de mantener relaciones interpersonales positivas.
3. Volverse capaz de mantener relaciones impuestas por la vida en las grandes organizaciones.

Nuestra reseña de las taxonomías es necesariamente somera, pues existe una gran cantidad de ellas de manera que resulta imposible incluirlas en los reducidos límites de este trabajo. Pasaré en consecuencia, al dominio psicomotor.

Para este dominio Anita Harrow fue, si no la primera, pues su libro A Taxonomy of the Psychomotor Domain apareció hasta 1972, si quien elaboró la taxonomía que es generalmente considerada como la mejor desarrollada.

Comprende cinco niveles básicos:

- 1.00 Movimientos reflejos.
- 2.00 Movimientos naturales o fundamentales.
- 3.00 Aptitudes perceptivas.
- 4.00 Aptitudes físicas.
- 5.00 Habilidades motrices.

Como antecedentes podemos mencionar las de C. E. Ragsdale en 1950, R. Dave en 1968, Elizabeth J. Simpson asesorada por B. Bloom en 1966 y la de R. Kibler, L. Barker y D. Miles en 1970, además de la de J. P. Guilford en 1958.

TESIS CON
FALLA DE ORIGEN

Se han hecho, asimismo, intentos para integrar los tres dominios en una sola taxonomía, pero a pesar de los esfuerzos de M. Scriven Y B. M. Tuckman en particular, no se ha producido hasta la fecha una que sea lo suficientemente manejable para que sea posible aplicarla en la práctica educativa.

De las taxonomías, brevemente reseñadas, las de B. Bloom, D. Krathwohl y A. Harrow son las que con mayor frecuencia se usan y de todas las existentes se puede afirmar, en términos generales, que las del dominio cognoscitivo son las más elaboradas, quizá porque la enseñanza escolar se interesa principalmente en los aprendizajes de ese tipo, los que por otra parte se prestan bien a la evaluación.

En comparación, la del dominio afectivo parece poco trabajada, probablemente debido a que las manifestaciones emotivas, por la dificultad que implica el hacerlo, pocas veces son motivo de una medición rigurosa.

En el dominio psicomotor, tal vez por la poca atención, que en general, se concede en las escuelas a la educación física y artística, la taxonomía respectiva raramente se aplica. (23)

No obstante, esas taxonomías al ser estudiadas y analizadas y al hacerse el intento por aplicarlas, suscitan reflexiones y reconsideraciones que son benéficas para el progreso de la pedagogía.

23. Ibidem, 211-212.



Las taxonomías y las pruebas se relacionan estrechamente desde el momento en el que la finalidad de éstas es medir el grado en el que se alcanzan los objetivos ordenados de acuerdo con las primeras.

Paralelamente a los estudios acerca de la jerarquización de los objetivos, en el periodo entre guerras y posteriormente hasta llegar al presente, se han seguido haciendo esfuerzos para que las pruebas sean cada vez mejores.

Con ese fin se han elaborado diversos tipos de normas que al través de un proceso de depuración han llegado a ser las que podemos considerar como vigentes.

Retrocediendo un poco en el tiempo, encontramos que hacia el fin del siglo pasado, en 1895, la American Psychological Association (APA), nombró un comité que se encargaría de investigar sobre la posibilidad de normalizar las mediciones físicas y mentales.

Unos años después, a principios de siglo, varios comités de la APA se propusieron normalizar las pruebas psicológicas y si bien en los siguientes cincuenta años surgieron algunas aportaciones personales, como la de Truman Kelley en 1924, no fue sino hasta la década de los cincuentas cuando la propia APA y otras asociaciones como la American Educational Research Association (AERA) y el National Council on Measurements Used in Education (NCMUE), hicieron algunas recomendaciones en relación a dichas pruebas.

TESIS CON
FALLA DE ORIGEN

En 1953 se publican las primeras Normas Éticas para Psicólogos y en él se concede especial atención a la salvaguarda de las pruebas considerándolas como 'equipo profesional' tratando así de evitar, tanto su uso indebido, como su invalidación.

Una segunda versión de las normas éticas, menos extensa que la primera se publicó en 1963 con un enfoque similar centrado en asegurar el correcto uso de las pruebas y en evitar su invalidación como resultado de una divulgación incontrolada.

La tercera versión, aún más reducida aparece en 1977. En ella, los 19 principios relativos a las pruebas contenidos en la de 1953, se reducen a uno solo, que en su parte medular establece que: "Las personas examinadas tienen el derecho de conocer los resultados, su interpretación y cuando sea apropiado, los datos en los que se hayan basado los juicios emitidos". (24)

Debido a que prácticamente todas las normas de ética se aplican a las actividades de los psicólogos y únicamente la parte citada pudiera concernir a los educadores, sin abundar sobre este aspecto me referiré a continuación a las normas técnicas.

De ellas existen también tres versiones, la última de 1974 es una adaptación de las Normas Técnicas para Pruebas que data de los primeros años de la década de los cincuentas y que aún habiendo sido revisada en 1966,

24. HANEY, WALT. Standards for Tests and Test Use, 12.

resultaba ya obsoleta alrededor de 1970.

Después de tres años de trabajo, las normas revisadas fueron aprobadas por la APA y demás organizaciones afines.

Estas normas constan de tres partes principales:

I. Normas para los manuales de pruebas y reportes.

A. Divulgación de la información.

B. Guías para la interpretación.

C. Normas y escalas.

II. Normas para los informes de investigación sobre confiabilidad y validez.

E. Validez.

F. Confiabilidad y error de medida.

III. Normas para el uso de las pruebas.

G. Preparación profesional, especialidad de los usuarios.

H. Selección o procedimiento para la prueba o método.

I. Administración y calificación.

J. Interpretación de las calificaciones.

De los aspectos de estas normas que mayor interés tienen para nosotros y que son aquéllas que conciernen no sólo a las pruebas normalizadas, sino también a las elaboradas por el profesor iré tratando en cada uno de los incisos correspondientes a: normas y escalas, validez, confiabilidad, administración y calificación, e interpretación de las calificaciones.

1.2. Conceptos generales.

Antes de examinar las normas y principios aplicables a las pruebas es necesario hacer algunas consideraciones sobre la finalidad que se persigue al aplicarlas.

En un sentido amplio, una prueba, según la definición del Webster's New Collegiate Dictionary es: "Cualquier serie de preguntas, ejercicios u otros medios para medir las habilidades, conocimientos, inteligencia, capacidades o aptitudes de un individuo o grupo de ellos".

Son en consecuencia, el medio apropiado para recabar información sobre las conductas de las personas, que será a su vez fundamento para emitir el respectivo juicio acerca de la conveniencia o inconveniencia de esos comportamientos.

En esa información deben incluirse y en ello concuerdan la mayoría de los autores consultados: Robert L. Ebel, Jum C. Nunnally, Norman E. Gronlund Y Georgia Sachs Adams entre otros, tanto los datos cualitativos como los cuantitativos.

Discrepan, sin embargo, en cuanto al significado y alcances del término medición.

Algunos de ellos, entre quienes los más destacados pudieran ser: N. R. Campbell, el distinguido historiador de la ciencia; Herbert Dingle Y Norman E. Gronlund, se

inclinan hacia un concepto restringido de la medida, según el cual sólo lo cuantitativo puede ser medido.

Así, para Gronlund: "La medición está limitada a descripciones cuantitativas del comportamiento del alumno. No incluye descripciones cualitativas...". (25)

Considerando que fue en 1920 cuando N. R. Campbell propuso las siguientes definiciones: "...las cantidades son propiedades para cada una de las cuales existe una operación empírica similar a la suma en aritmética".

En tanto que: "...las cualidades se caracterizan por la ausencia de esa operación aditiva". (26) Creo poder afirmar que en esas definiciones se apoyan, cuando menos tácitamente, los autores que sustentan un concepto restringido de la medida.

Según la concepción de Campbell, sólo los fenómenos físicos y ni siquiera todos ellos, serían susceptibles de ser medidos. Así por ejemplo: la suma aritmética del peso de dos sustancias diferentes es el de la mezcla que pueden formar, pero la suma de las densidades de las dos sustancias no es el de la mezcla y no por eso la densidad escapa a la medición.

Para otros autores, la medición puede ser entendida de una manera más amplia, veamos algunas de sus opiniones:

Robert L. Ebel, afirma que: "...en principio todos los resultados importantes de la educación son mensurables ..." (27) Georgia S. Adams, nos dice que: "...las

25. GRONLUND, NORMAN E. Op. cit., 9.

26. CAMPBELL, N. R. En Suppes & Zinnes, Basic Measurement Theory, 15.

27. EBEL, ROBERT L. Op. cit., 55.

mediciones no son sino observaciones cuidadosas de resultados alcanzados bajo condiciones tipo...". (28)

Jum C. Nunnally, expresa que: "...la medición consiste en un conjunto de normas para asignar números a los objetos de modo tal que esos números representen cantidades de atributos...". (29)

Es, sin embargo, S. S. Stevens, quien al replicar a Campbell, nos ofrece una definición formal de la medida en un sentido lato, pues sostiene que: "...la medición en el más amplio sentido, se define como la asignación de números a objetos o sucesos según ciertas reglas. El hecho de que los números puedan asignarse según diferentes reglas conduce a distintos tipos de escala y clases de medida. El problema es entonces hacer explícitas: a) las diferentes reglas para la asignación de números, b) las propiedades matemáticas de las escalas resultantes, y c) las operaciones estadísticas aplicables a las mediciones hechas con cada tipo de escala". (30)

Esta definición es, cuando menos incompleta, pues nos daría la posibilidad de establecer reglas para asignar a los objetos, sucesos o propiedades números aleatorios, como lo hacemos en el muestreo al azar, sin que por eso estemos realizando algún proceso de medición.

Para que ésta exista es necesario que entre los símbolos, sean o no numéricos, haya la misma relación

28. ADAMS, S. GEORGIA. Medición y evaluación, 20.

29. NUNNALLY, JUM C. Op. cit., 23.

30. ACKOFF, RUSSELL L. Scientific Method, 197.

que entre los objetos, sucesos o propiedades que pretendemos medir.

Los números son los símbolos más frecuentemente usados en la medición porque sus interrelaciones han sido exhaustivamente estudiadas por la matemática y porque comparten con las observaciones algunas de las relaciones que entre ellas pueden establecerse. No obstante, los números tienen también propiedades de las que no participan las observaciones y por eso es importante hacer, en cada caso, explícitas cuáles propiedades son compartidas y cuáles no lo son.

Si tratáramos de especificar cuál es el conjunto de operaciones que definen la medición, el esfuerzo, como la historia lo demuestra, sería vano, pues con el tiempo cambian y mejoran al igual que los demás procedimientos científicos.

Es posible, en cambio, definir la medición en términos de su función: "...como una forma de obtener símbolos para representar las propiedades de los objetos, sucesos o estados, de manera que entre esos símbolos haya la misma relación relevante que la que exista entre lo que representan". (31)

Se puede decir, en consecuencia, que esta concepción amplia de la medida permite que las pruebas cumplan con mayor amplitud el objetivo que les hemos asignado de medir diversos atributos, pues en éstos nos es dable

31. Ibídem, 178, 179.

... incluir, tanto los cuantitativos, como los cualitativos.

Al respecto, veremos en el siguiente inciso como se realiza la medición.

1.3. El proceso de medición.

Para que sea posible llevar a buen término la medición, para que podamos obtener símbolos cuyo conocimiento sea equivalente al de las propiedades que representan, es necesario seguir un proceso, que en principio, consta de tres pasos, a saber:

- a) Señalar e identificar la cualidad o atributo que se habrá de medir.
- b) Determinar un conjunto de operaciones en virtud de las cuales el atributo pueda manifestarse, hacerse perceptible.
- c) Establecer un conjunto de procedimientos o de definiciones para traducir las observaciones a enunciados cuantitativos de grado o cantidad. (32)

Las tres operaciones, que constituyen una secuencia, están interrelacionadas. De la identificación clara y correcta del atributo dependen, tanto que las operaciones que harán posible su manifestación sean las adecuadas, como que la escala que se elija para medirlo también lo sea.

La realización correcta de esta secuencia, aun en las ciencias físicas, puede no ser tan sencilla como a primera vista parece, En algunos casos es prácticamente evidente: la longitud de un objeto, su peso, se identifican con facilidad, se manifiestan al mirarlo o tratar de levantarlo y en un longímetro o en una

32. THORNDIKE, ROBERT L. y Elizabeth Hagen. Op. cit., 17.

báscula podemos leer directamente la cantidad de unidades de cada una de esas medidas. Sin embargo, no en todos los casos es así. Supongamos, por ejemplo, que se tratara de medir la dureza de un cuerpo. Para identificar ese atributo es necesario saber cuál es el significado que se confiere al término dureza en física, pues sólo tomando en cuenta que se trata de la resistencia de un cuerpo a ser indentado o rayado por otro, nos damos cuenta de cuál es la operación necesaria, el instrumento adecuado, para que se manifieste y cómo podemos traducir la observación a enunciados de grado o cantidad.

En las ciencias sociales, específicamente en pedagogía, el proceso es aún más difícil. La naturaleza misma de los atributos que pretendemos medir es la causa principal de que así sea, pues ya no se trata de características físicas, sino de las cognoscitivas, afectivas o psicomotoras de los individuos.

La identificación de ese atributo, su definición conceptual, nos enfrenta en muchas ocasiones con problemas nada fáciles de resolver.

Al respecto, podemos citar el caso, ya clásico, de los intentos que se han hecho para definir la inteligencia sin que, hasta ahora, se haya llegado a algún consenso general al respecto.

Esta falta de una definición clara del atributo afecta al segundo paso del proceso, pues de ella depende cuáles son las operaciones, los instrumentos, para que se ponga de manifiesto.

Si consideramos como Spearman, que la estructura de la inteligencia es de tipo bifactorial, con un componente de índole general 'g' (inciso 1.1.), o si nos adherimos a lo propuesto por Thurstone y Guilford, en el sentido de que ese factor 'g' no existe y se trata en realidad de un conjunto de habilidades mentales primarias, serán distintas las operaciones, las pruebas, que elijamos para que este atributo se haga perceptible y medible.

En el caso específico de la medición del rendimiento escolar habrá que observar, en cada uno de los pasos, ciertos requisitos que hagan posible que los resultados obtenidos sirvan de fundamento para una evaluación acertada.

- a) Al señalar el atributo, el que se elija ha de ser pertinente en relación con los objetivos del programa y del grado en el que hayamos logrado definir éstos con cierta precisión va a depender que la identificación de ese atributo pueda hacerse clara y correctamente.
- b) Para que pueda manifestarse será necesario escoger, a menudo diseñar, el instrumento adecuado, de modo que sea el atributo de nuestro interés y no otro el que se haga perceptible.

c) La traducción de las observaciones a términos cuantitativos implica, tanto aplicar la escala adecuada, como hacer una estimación cuidadosa del grado de dificultad de las diferentes preguntas o problemas, para asignar a cada una de ellas la debida puntuación. (33)

Veremos a continuación cuáles son las principales escalas de las que podemos disponer, así como sus características básicas.

33. MENENDEZ, LIBERTAD y Laura Elena Rojo. Introducción a la medición del rendimiento escolar, 6, 7.



1.4. Escalas de medición.

Como acabamos de ver en el inciso anterior, una vez que el atributo que deseamos medir ha sido identificado y estamos en aptitud de observar su manifestación, el paso siguiente consiste en traducir lo que hayamos observado a símbolos que expresen el grado o cantidad que corresponda a esa manifestación específica.

Por la diversidad de aprendizajes que son motivo de estudio para la pedagogía, la medición de ellos, cualquiera que sea su tipo, tiene una especial importancia. La tiene por igual si se trata de los que se muestran en forma cualitativa que de las que lo hacen cuantitativamente. En ambos casos, la traducción de las observaciones a símbolos que admitan un ordenamiento jerárquico hace posible un mejor análisis e interpretación de los fenómenos educativos.

Para realizar ese ordenamiento se han diseñado cuatro escalas de medición aplicables, cada una de ellas a datos de distinta índole.

Esas escalas son las siguientes:

- a) Nominal.
- b) Ordinal.
- c) De intervalo.
- d) De razón o proporción.

De cada una de ellas anotaremos sus propiedades formales, cómo se aplican y el tratamiento estadístico que admiten.



a) Escala nominal.

i) Propiedades formales:

Es reflexiva: $x=x$, para todos los valores de x .

Es simétrica: $x=y$, luego $y=x$.

Es transitiva: si $x=y$, $y=z$, entonces $x=z$.

ii) Aplicación, ejemplo:

Su aplicación consiste en partir de una clase dada y formar subclases que se excluyan mutuamente. La única relación implicada es la equivalencia, los miembros de cualquier subclase deben ser equivalentes en la propiedad medida.

Se simboliza con el signo = (igual a) (34)

Como ejemplo podríamos ofrecer el que sigue:

<u>Población</u>	<u>Propiedad</u>	<u>Categoría</u>
Personas vivas.	1. Sexo.	1a. Hombre. 1b. Mujer.
	2. Grado educativo.	2a. No terminó la primaria.
		2b. Terminó la primaria, no así la secundaria.
		2c. Terminó la secundaria.

34. SIEGEL, SIDNEY. Estadística no paramétrica, 43.



En esta clasificación las categorías para cada propiedad son exclusivas y exhaustivas para la población considerada. Cada persona tiene que pertenecer a una de las categorías de cada propiedad. (35)

iii) Tratamiento estadístico.

Debido a que los símbolos que designan a las diferentes categorías pueden intercambiarse sin alterar la información esencial de la escala, las únicas operaciones estadísticas admisibles son aquellas en las que los datos no sufren alteración en el proceso: el modo y la frecuencia. La medida de asociación más común para datos nominales es el coeficiente de contingencia 'C'.

b) Escala ordinal o de rango.

i) Propiedades formales:

Incorpora además del signo = (igual a) el de $>$ (mayor que).

Es irreflexiva, pues no es verdad que para cualquier x , x sea $> x$.

Es asimétrica, ya que si $x > y$, y no es $>$ a x ($y < x$).

Es transitiva, pues si $x > y$, $y > z$, luego $x > z$.

Puede haber dos tipos de esta escala: parcialmente ordenada, cuando la relación



mayor que, se sostiene sólo entre algunos pares de clases, pero no en todos y la propiamente dicha ordinal, cuando la relación mayor que, se sostiene en todos los pares de clases de manera que surja el rango completo.

ii) Aplicación, ejemplo:

Podemos citar como ejemplo de la escala ordinal, la aplicación que de ella se hace en la UNAM en las calificaciones que oficialmente reciben los alumnos.

Las notas que se les otorgan se pueden ordenar, de menor a mayor, desde N. A. (no acredita), S. (suficiente), B. (Bien) hasta M. B. (muy bien), sin que en ello esté incluida alguna información sobre la diferencia entre una y otra de las calificaciones.

iii) Tratamiento estadístico.

➤ Para describir la tendencia central de los datos de esta escala puede usarse la mediana.

Las pruebas estadísticas admisibles son no paramétricas, por ejemplo, los coeficientes de correlación basados en rangos, el r_s de Spearman o el r de Kendall, son adecuados. (36)

Ahora bien, la medición en sentido estricto, implica el uso de una unidad constante como es el caso en las escalas de intervalo y de razón.

36. SIEGEL, SIDNEY. Op. cit. ., 43.



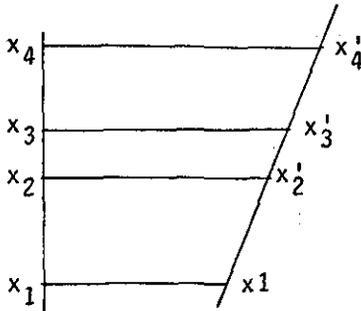
c) Escala de intervalo.

i) Propiedades formales:

Se caracteriza por ser única hasta la transformación lineal, según la siguiente ecuación:

$$x' = ax + b, \text{ siendo } x > 0$$

Esta propiedad se puede expresar gráficamente así:



La escala de intervalo no tiene un cero natural, se establece de manera arbitraria, por conveniencia y se acepta por convención.

No se puede afirmar, en rigor, que algún valor de una escala de intervalo sea el múltiplo de otro. La ecuación que la representa nos permite comprobarlo aun en el caso de que $x_2 = 2x_1$, pues tendríamos:

TESIS CON
FALLA DE ORIGEN

$$ax_2 + b \neq 2(ax_1 + b)$$

ya que: $a(2x_1 + b) \neq 2(ax_1 + b)$

pues: $2ax_1 + b \neq 2ax_1 + 2b$, siempre si $b \neq 0$
pero si $b = 0$, entonces:

$$2ax_1 = ax_2, \text{ pues substituyendo } x_2 = 2x_1$$

obtenemos:

$$2ax_1 = 2ax_1$$

Podemos, eso sí, expresar las diferencias (intervalos) entre los valores de las escalas en cuestión, como múltiplos el uno del otro.

Esto se comprueba comparando los intervalos entre tres puntos: x_1 , x_2 y x_3 , ya que:

$$\frac{(ax_1+b)-(ax_2+b)}{(ax_2+b)-(ax_3+b)} = \frac{ax_1-ax_2}{ax_2-ax_3} = \frac{x_1-x_2}{x_2-x_3} \quad (37)$$

comprobándose así lo propuesto.



ii) Aplicación, ejemplo.

En las pruebas escolares es usual asignar a la respuesta correcta a una pregunta un cierto valor en puntos y la calificación de un estudiante será la suma de los puntos que haya obtenido en sus contestaciones correctas, sin que ello implique que cero puntos correspondan, en realidad, a la ausencia total de conocimiento, ni que el alumno que obtenga 90 puntos sepa el doble del que sacó 45, o el triple del que logró 30, pues para ello sería necesario que la escala tuviera un punto cero real.

iii) Tratamiento estadístico.

Si bien los números correspondientes a los diversos valores de la escala de intervalos no pueden sujetarse a algunas operaciones aritméticas, la suma o la multiplicación, por ejemplo, sin una explicación especial, la mayoría de esas operaciones pueden realizarse con las diferencias (intervalos) entre todos los pares de dichos números. (38)

Las operaciones estadísticas admisibles son todas las paramétricas (media, desviación estándar, etc.) y las pruebas estadísticas aplicables son la 't' de Student, la 'F', etc. (39)

38. *Ibidem*, 191.

39. SIEGEL, SIDNEY. Op. cit., 49.



d) Escala de razones.

i) Propiedades formales.

Se caracteriza por ser única hasta la multiplicación por una constante positiva de acuerdo con la ecuación siguiente:

$$x' = c.x, \text{ siendo } c > 0$$

Es isomórfica a la estructura de la aritmética y tiene un punto cero real.

Se logra una escala de razones cuando es posible obtener: a) relación de equivalencia, b) relación de mayor a menor, c) proporción conocida entre dos intervalos y d) proporción conocida entre dos valores de la escala. (40)

ii) Aplicación, ejemplo:

Se usa para medir longitudes, pesos, temperaturas cuando éstas se expresan en grados Kelvin, que parten de un cero absoluto, equivalente a -273 grados centígrados.

iii) Tratamiento estadístico.

Son permisibles todas las operaciones y pruebas estadísticas.

40. Ibidem, 50.



1.5. Instrumentos de medición.

Los instrumentos que con mayor frecuencia se usan en la medición del rendimiento escolar son las pruebas, definidas según lo expresado en el inciso 1.2.

Pueden clasificarse, de acuerdo con el punto de vista que se adopte, en dos formas distintas:

Si tomamos en cuenta quién las elabora, podemos clasificarlas en estandarizadas cuando su formato y reactivos son uniformes para alguna asignatura o ciclo escolar y en las que cada profesor diseña para su caso específico.

Si atendemos a su estructura, las podemos dividir en varios tipos:

1.5.1. Tipos de pruebas.

- a) Pruebas de base no estructurada.
- b) Pruebas de base semi-estructurada.
- c) Pruebas de base estructurada.

- a) Al primer tipo pertenecen las pruebas de ensayo y entre ellas incluimos, por conveniencia, las de resolución de problemas numéricos.

Las pruebas de base no estructurada, las de ensayo o composición constan: "...de un número relativamente reducido de preguntas de carácter general, que exigen respuestas extensas. Implican



que el alumno planifique su propia contestación y la exprese en sus palabras...". (41)

Para resolverlas es necesario que recuerde los principios pertinentes, además de los hechos concretos, los relacione entre sí y organice sus respuestas en forma lógica y coherente. Hacen posible, por ende, que el estudiante ponga en juego y manifieste procesos mentales de alto nivel. (42)

Si bien tienen estas cualidades, las pruebas de ensayo no carecen de ciertas desventajas, como lo son que sus grados de validez y confiabilidad son reducidos en comparación con las de otro tipo. Su calificación resulta casi siempre laboriosa y en ocasiones difícil para el profesor.

Por otra parte, las de resolución de problemas llegan a tener un alto grado de pertinencia intrínseca, pues en ellas es posible plantear situaciones similares a aquellas que se le presentarán a los alumnos en el ejercicio de su profesión.

Estas pruebas, al definir con mayor precisión, que las propiamente dichas de ensayo, las tareas que el alumno debe cumplir, poseen un mayor grado de objetividad. (43)

- b) Las pruebas de base semi-estructurada, son aquellas en las que las respuestas están limitadas y guiadas por el tipo y forma de la pregunta. Hacen posible,

41. EBEL, ROBERT L. Op. cit., 161-163.

42. MENENDEZ, LIBERTAD y Laura Elena Rojo. Op. cit., 12.

43. EBEL ROBERT L. Op. cit., 133.

en la medida en la que una contestación restringida lo permite, la apreciación del grado de desarrollo alcanzado por los alumnos en los procesos mentales complejos. Su validez y confiabilidad es menos reducida que en las de ensayo.

- c) Las de base estructurada, llamadas también objetivas, las resuelve el alumno eligiendo entre diversas opciones ya formuladas. (44)

Pueden presentarse en distintos formatos según sea el tipo de preguntas que incluyan, siendo los siguientes los principales:

- a) verdadero-falso, b) complementación, c) selección múltiple, d) ordenamiento, e) correlación, etc.

Su principal ventaja consiste en que por medio de ellas, si están bien diseñadas, es posible conseguir buenos niveles de objetividad, validez y confiabilidad.

Por lo que hace a las desventajas, la más notoria se refiere a que no proporcionan indicadores sobre los procesos mentales complejos que el alumno haya podido desarrollar. (45)

No me extenderé sobre las pruebas de base estructurada porque en la investigación motivo de este trabajo las que se aplicaron fueron no estructuradas, en la mayoría de los casos del tipo de resolución de

44. *Ibídem*, 161.

45. MENÉNDEZ, LIBERTAD y Laura Elena Rojo. *Op. cit.*, 12.

problemas.

Las pruebas, de cualquier tipo que sean, para constituirse en instrumentos adecuados para una correcta medición, deben poseer las siguientes características.

1.5.2. Características de los instrumentos de medición.

Son cuatro los principales criterios a satisfacer:

- a) Objetividad.
- b) Validez.
- c) Confiabilidad.
- d) Utilidad.

a) Objetividad.

La objetividad de una prueba se refiere al grado en el que individuos igualmente competentes para calificar otorgan las mismas puntuaciones.

Esta característica, el que la calificación no dependa del juicio subjetivo de quien la da, además de su importancia en pro de la imparcialidad, tiene también la que adquiere al hacer posible que los alumnos queden convencidos de que las puntuaciones que hayan obtenido son el producto de sus aciertos o errores y se esfuercen en corregir los últimos.

b) Validez.

La validez consiste en que la prueba mida en realidad

el atributo que se pretende medir. "...la validez es cuestión de grado. Un test rara vez es completamente válido y casi nunca completamente inválido...". (46)

La mayor precisión con la que un test cumple su cometido de medir cierta cualidad o atributo, es por consiguiente su grado de validez.

Con la intención de unificar los distintos criterios sobre los tipos de validez, la American Psychological Association, APA, la American Educational Research Association, AERA y el National Council on Measurements Used in Education, han publicado en 1954, 1966 y 1974 diferentes versiones de las Normas Técnicas para las Pruebas y su Uso.

En las versiones de 1954 y 1966 se incluyeron cuatro tipos de validez, a saber: de contenido, predictiva, concurrente y de construcción, no obstante en 1966, la validez predictiva y la concurrente fueron consideradas como distintas formas de validez relacionada con un criterio. (47)

En 1974, las Normas Técnicas incluyeron los siguientes tres tipos de validez:

- i) de contenido.
- ii) relacionada con un criterio.
- iii) de construcción.

46. KARMEL, LOUIS J. Medición y evaluación escolar, 14.

47. HANEY, WALT. Op. cit., 22.

i) Validez de contenido.

Se refiere al grado en que una prueba de rendimiento representa el contenido total de una unidad o secuencia didáctica determinada. (48)

Los reactivos de la prueba, en consecuencia, deben ser una muestra representativa de los contenidos del curso para que haya validez de contenido.

ii) Validez relacionada con un criterio.

Este tipo de validez comprende, según quedó anotado, las que hasta 1966 se consideraron como predictiva y concurrente.

Se pudo hacer la fusión debido a que tanto en la una como en la otra se refleja el grado de concordancia entre las puntuaciones obtenidas en una prueba y otras medidas externas independientes (criterios).

La diferencia consiste en que en el primer caso, predictiva, los datos criterio se recaban en fechas posteriores a la prueba y en el segundo, concurrente, aproximadamente al mismo tiempo.

La validez relacionada con un criterio es la que permite generalizar o predecir el éxito que podrá tener una persona en una tarea diferente a aquella en la que su capacidad haya sido medida. (49)

48. LINDEMAN, RICHARD H. Tratado de medición educacional, 53.
49. MEHRENS, WILLIAM A. e Irvin J. Lehmann. Medición y evaluación. En la educación y en la psicología, 119.

Debido a ello, en las Normas Técnicas de 1974, se concede a este tipo de validez cierta primacía sobre los demás, pues tomando en cuenta que en pruebas que supuestamente la tienen, se basa la admisión, tanto de los estudiantes a las escuelas, como la de los empleados a diversas empresas, se hacía necesario establecer los requisitos para que las pruebas tuvieran en realidad esa pretendida validez predictiva. (50)

iii) Validez de construcción.

Se refiere a la medida en la que la prueba nos proporciona información acerca de una característica significativa de un individuo, al grado en el que mide su rendimiento real con respecto a los objetivos didácticos señalados. (51)

c) Confiabilidad.

"La confiabilidad puede definirse como el grado de precisión y consistencia que guarda una prueba para medir lo que se propone". (52)

Puede determinarse si una prueba es o no confiable administrándola dos o más veces al mismo grupo, utilizando otra que sea equivalente o subdividiéndola en dos partes que también lo sean.

La correlación entre los resultados de la prueba inicial y los de su repetición, sustitución o subdivisión será indicativa de su grado de confiabilidad.

50. HANEY, WALT. Op. cit., 22.

51. LINDEMAN, RICHARD H. Op. cit., 52.

52. MENENDEZ, LIBERTAD y Laura Elena Rojo. Op. cit.,

Cuando la misma prueba se administra dos o más veces al mismo grupo, los resultados pueden ser afectados por las modificaciones que el primer examen haya producido en las conductas de los alumnos, siendo ésta una de las razones por las que la Asociación Americana de Psicología expresa que: "... el reexamen no es un método teóricamente deseable para determinar coeficientes de confiabilidad, cuando, como es lo corriente, los reactivos que constituyen un test son solamente uno de lo muchos conjuntos (reales o hipotéticos) que igualmente podrían haberse usado para medir una capacidad o rasgo particular". (53)

Si se usa una prueba equivalente que contenga el mismo número de reactivos e igual formato, referidos aquéllos a contenidos, también, iguales, la confiabilidad se establece de acuerdo con el grado de correlación entre los resultados obtenidos por el sujeto en ambas pruebas. No elimina del todo, pero si disminuye el efecto de la práctica y de la equivalencia real entre las dos pruebas, va a depender la validez de las conclusiones que se obtengan.

La subdivisión de un test en partes equivalentes permite probar la consistencia del contenido, pero no las diferencias entre las respuestas individuales que pudieran darse en los casos en los que se

53. Asociación Americana de Psicología. En Karmel Louis J. Op. cit., 130.

administren dos o más pruebas iguales o equivalentes.

Para la confiabilidad de la prueba completa se usa la ecuación de Spearman-Brown, que es la siguiente:

$$\text{Confiabilidad del test completo.} = \frac{\text{Confiabilidad de medio test} \times 2}{\text{Confiabilidad de medio test} + 1}$$

d) Utilidad.

Se llama utilidad a la cualidad que tiene una prueba si su elaboración y administración son económicas; si es aceptable la relación que los economistas llaman insumo-producto, ya se trate de erogaciones, ya de consumo de tiempo incluyendo el necesario para su calificación.

Con el mismo fin, determinar la calidad de una prueba escolar, Robert L. Ebel propone que se examinen los siguientes aspectos:

1. Pertinencia.

Se refiere a que la prueba contenga las preguntas adecuadas para evaluar los rendimientos deseados.

2. Equilibrio.

Consiste en que la proporción de las preguntas relativas a cada tema concuerde con la intención que el autor de la prueba haya expresado en las especificaciones respectivas.

3. Eficacia.

Una prueba la posee, cuando para su solución y calificación son suficientes los tiempos limitados de los que disponen, tanto el alumno como el profesor.

4. Objetividad.

Es la cualidad que, por ser las preguntas claras y permitir respuestas definidas, hace posible que los expertos coincidan sobre cuáles son las contestaciones correctas y por ende se uniformen las calificaciones.

5. Especificidad.

Consiste en que las preguntas para ser correctamente contestadas exijan el conocimiento del que precisamente se trate, de manera que quienes no lo posean sólo puedan esperar las calificaciones que correspondan a respuestas dadas al azar.

6. Dificultad.

Se refiere a que las preguntas tengan el grado de dificultad adecuado al nivel de los estudiantes a quienes se dirige la prueba y les permita mostrar sus conocimientos.

7. Discriminación.

Se consigue cuando la respuesta a cada una de las preguntas refleja el rendimiento de cada uno de los examinados y también a éste corresponda la puntuación de la prueba completa.

8. Confiabilidad.

Se logra cuando las puntuaciones que se obtienen en la prueba concuerdan con las correspondientes a otras pruebas en las que se midan las mismas capacidades.

9. Ecuanimidad.

Se obtiene si la prueba está construida y administrada de manera tal que todos los alumnos tengan igual oportunidad de demostrar sus conocimientos reales en el campo que cubre dicha prueba.

10. Tiempo para resolver la prueba.

Se considera como adecuado si la extensión de la prueba, en relación al tiempo que se concede para resolverla, no favorece indebidamente a los alumnos más veloces. (54)

Hemos visto, hasta ahora, que la finalidad principal de las pruebas es la de servir como instrumento para la medición de las habilidades, capacidades, conocimientos, atributos, etc. de los individuos.

Vimos, asimismo, como se realiza el proceso de medición, cuáles son las escalas de las que podemos disponer, cuáles son los tipos de prueba más usados y los requisitos que ellas deben satisfacer para cumplir cabalmente con el fin que les asignamos.

Nos corresponde, por consiguiente, relacionar la medición con el objetivo que se persigue al hacerla: la evaluación, en nuestro caso, del rendimiento escolar.

54. EBEL, ROBERT L. Op. cit., 445-471.

1.6. Medición y evaluación.

El proceso de medición nos condujo a obtener una serie de datos cuantitativos, de grado o de cantidad que describen hasta qué punto se posee determinada característica.

La evaluación es, sin embargo, algo más que la descripción de algún atributo, pues como su nombre lo indica, lleva aparejado un juicio de valor.

En la educación, entendida como un proceso que se realiza con la intención de que la conducta de los alumnos se modifique, la evaluación es el procedimiento para determinar en qué consisten esas modificaciones y en estimarlas en relación con los valores representados en los objetivos y así detectar hasta dónde se logran éstos. (55)

Se trata, en tal caso, de una evaluación 'referida a un criterio' y este se establece de acuerdo con el grado de dominio que el alumno alcanza en el desempeño de tareas específicas. Es necesario, por ende, definir el aprendizaje en función de dichas tareas, incluyendo la valoración diferencial de las mismas, por ser esos los elementos en los que se habrá de basar el criterio que sustentemos.

La valoración se establece al margen del desempeño del grupo y tiene una relación directa con el contenido

55. TABA, HILDA. Elaboración del currículo, 410.



educativo. (56)

Hay además otro tipo de evaluación, el que se hace 'de acuerdo con la distribución normal' y que consiste, en principio, en comparar la calificación de un estudiante con los promedios que haya obtenido el grupo del que forma parte, el nivel medio de ejecución de ese grupo se convierte en el del desempeño esperado, según el cual se juzgará, para cada alumno, la calidad que haya alcanzado. (57)

El primer procedimiento, la evaluación con referencia a un criterio, hace posible que el profesor constate el grado en el que sus alumnos hayan logrado el dominio de los contenidos del curso, pues en los objetivos respectivos habrá especificado las metas a las que pretende que los estudiantes lleguen en el proceso E-A. Podrá en consecuencia hacer una evaluación objetiva del grado de aprendizaje conseguido.

En la evaluación, de acuerdo con la distribución normal sólo podrá saber cuál es el aprendizaje obtenido por un alumno en relación con el que haya logrado el grupo en promedio, pero no el grado en el que cada estudiante haya dominado los contenidos del curso.

Esta manera de calificar, si bien cada vez menos usada, hace posible que en un grupo de bajo rendimiento haya alumnos que aunque alcancen una calificación un poco superior al promedio y resulten por eso aprobados, no

56. HOUSDEN, JACK L. y Lannie Le Gear. "La evaluación referida a un criterio". En Sistematización de la enseñanza. 705.

57. Ibidem, 704.



hayan obtenido el mínimo de conocimientos que se ha pretendido impartirles.

Una evaluación así realizada carece de utilidad para cumplir, como veremos en los incisos siguientes, con su principal propósito que es el de servir de fundamento para la toma de decisiones educativas.



2. La evaluación de programas educativos.

Sobre el papel que la evaluación juega en el proceso E-A hay una coincidencia generalizada entre los autores en el sentido de considerar que su principal objetivo es el de servir como medio para recabar información y en ella fundamentar las decisiones educativas que se adopten.

Así por ejemplo: "La medición es el ama de llaves de la enseñanza. Sin medición no puede haber evaluación. Sin evaluación es imposible la retroalimentación. Sin retroalimentación no puede tenerse una idea precisa de los resultados alcanzados. Sin conocer estos resultados, tampoco es posible una mejoría sistemática en el aprendizaje". (58)

O sea que los cambios necesarios para mejorar el aprendizaje, las decisiones que al respecto se tomen, habrán de basarse en el conocimiento de los resultados alcanzados y en su evaluación.

Georgia S. Adams, si bien no con las mismas palabras, nos dice que las decisiones se toman mejor cuando para ello se dispone de una buena información recabada por conducto de la medición y la evaluación. (59)

Para N. Gronlund la evaluación es también la fuente de la información en la que se fundamentan las decisiones que los maestros adoptan. (60)

58. PARNELL, D. "Elementary and Secondary Education Amendments, 2698, en MEHRENS, W. A. e I. J. Lehman. Op. cit., 19.

59. ADAMS S. GEORGIA. Op. cit., 17, 18.

60. GRONLUND, NORMAN E. Op. cit., 3.



En concordancia con las opiniones mencionadas, están las de Astin y Panos, quienes con mayor claridad y amplitud se refieren a la evaluación como origen de la información que servirá de base para la toma de decisiones educativas.

Al modelo que ellos proponen para la evaluación de programas educativos me referiré a continuación.

2.1. Modelo de Astin y Panos.

Antes de prodecer al examen del modelo, haré algunas precisiones sobre el significado que otorgaremos a ciertos términos que vamos a manejar.

Así, por evaluación entenderemos, en el ámbito de la educación, aquel proceso sistemático que se emprende con el fin de determinar hasta qué punto los alumnos alcanzan los objetivos propuestos. Incluye tanto las descripciones cualitativas y cuantitativas del comportamiento de los alumnos, como los juicios valorativos que se refieren a la conveniencia de ese comportamiento. (61)

Un programa educativo, para los autores del modelo, puede ser cualquier actividad pedagógica en proceso que haya sido diseñada para producir cambios específicos en la conducta de los alumnos expuestos a ella. Puede tratarse de una sola lección, del curso semestral o anual de alguna asignatura, de un plan de estudios o de un sistema escolar completo.

"La evaluación implica la recolección de información concerniente al efecto de un programa educativo. Dado que existen muchos usos posibles de tales informes, se asume que el propósito fundamental de la evaluación es producir información que pueda ser usada en la toma de decisiones educativas...". (62)

61. Ibidem, 8, 9.

62. ASTIN, A. W. y R. J. Panos. La evaluación de programas educativos, 6.



Hechas las precisiones anteriores estamos en condiciones de proceder al examen del modelo de Astin y Panos.

Podemos, en primer lugar, decir que está integrado por tres componentes conceptualmente distintos, a saber: salidas, entradas y operaciones.

Las salidas son el resultado de la aplicación de un programa educativo: las metas, logros o variables dependientes, susceptibles de ser expresadas en diferentes niveles de abstracción, desde uno alto, como podría serlo, en el caso de un arquitecto, la adquisición del sentido de la forma y proporción, hasta otros relativamente concretos que pueden ser observados y medidos en forma menos indirecta, tales como su capacidad para calcular los elementos de una estructura, los costos de un edificio, etc.

"Un análisis de las salidas educativas relevantes debe incluir no solamente aquellos resultados para cuya consecución haya sido específicamente diseñado el programa, sino también los 'efectos secundarios' de éste...". (63) Si en un estudio de evaluación no se toman en cuenta esos resultados inicialmente no previstos se puede llegar a conclusiones carentes de validez.

Las salidas, como después veremos, serían equiparables a los productos de un sistema concebido desde el punto de vista insumo-producto.

63. Ibidem, 12.



Con ese mismo enfoque, las entradas equivaldrían a los insumos, en principio a todos los recursos necesarios para la obtención del producto.

Dichos recursos son de índole diversa: humanos, materiales, económicos, tecnológicos, de tiempo, etc., de manera que en el caso de un sistema escolar los siguientes serían los más importantes:

Los alumnos, que pueden ser considerados como la materia prima en la que actúa la educación, con todas sus características y variables: talentos, habilidades, actitudes, aspiraciones, etc., tanto si su origen se concibe como predominantemente genético, como si se piensa que los factores socioculturales son su principal causa.

Los profesores, autoridades, empleados y demás recursos humanos necesarios para la buena marcha de las actividades escolares.

Los recursos materiales y económicos. Los primeros consistentes en los edificios y espacios abiertos que formen parte de la escuela, así como los materiales de docencia respectivos; los segundos formados por el patrimonio y los medios financieros de los que disponga la institución.

Los recursos tecnológicos, como su nombre lo indica, son todas aquellas técnicas que se apliquen en el proceso E-A.



Y si bien, cada uno de esos recursos influye, en alguna forma, en la educación, Astin y Panos nos hacen ver que: "Las características de entrada del estudiante son una consideración importante en la evaluación educativa por dos razones al menos.

En primer lugar, el nivel final de actuación del escolar, en términos de los criterios relevantes de salida, puede a menudo estar condicionado por ciertas de sus características de entrada sin importar cuál pueda ser la naturaleza del programa educativo particular... En segundo lugar, en la medida que el ambiente educativo refleja las características mismas de los discípulos, puede ser influido dramáticamente por la naturaleza de la entrada estudiantil...". (64)

La interacción de las entradas y las operaciones ha merecido especial atención por parte de autores tan destacados como J. Piaget, J. Bruner y otros, quienes nos indican que ningún programa por creativo e imaginativo que sea podrá ser suficiente para que los alumnos que no hayan alcanzado el necesario nivel psicológico lleguen a comprender la enseñanza que se les pretende impartir. Las características de entrada de los alumnos son en cierto sentido, como decíamos, la materia prima en la que actúa la educación, aquello que, como desde tiempo inmemorial han postulado los filósofos orientales, puede ser moldeado, afinado, pulido, sin que por ello su esencia cambie. De un trozo de madera

64. Ibidem, 14, 15.



podemos hacer la más bella escultura de ese material, que de ninguna manera se convertirá en otro.

El segundo aspecto, la influencia que las entradas tienen en ambiente que priva en la escuela, sólo puede ser detectado cuando se visitan distintos centros escolares. Al compararlos nos podremos dar cuenta de que las diferencias son notorias y de que casi siempre van de acuerdo con los antecedentes socioculturales de los alumnos.

El tercer componente, las operaciones o procesos son '...aquellas características del programa educativo capaces de afectar las salidas relevantes de los estudiantes...'. (65)

Comprenden tanto las variables de 'tratamiento' como las 'ambientales'.

Entre las primeras se inscriben los planes y programas de estudios, los objetivos, los contenidos, las técnicas de enseñanza-aprendizaje, las actividades de aprendizaje y la medición misma, así como cualquier otra variable susceptible de ser considerada como de tratamiento, en tanto tuviera algún efecto en los resultados.

Las variables ambientales, ya correspondan al medio humano, ya al físico surten efectos en las salidas y de ellas el que quizá requiera ser atendido con mayor cuidado es el producido por los factores socioculturales.

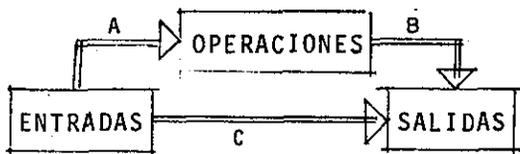


Este efecto, al igual que el de las entradas en el ambiente escolar, puede observarse mejor cuando los estudios de evaluación se realizan simultáneamente en varios escenarios sociales diferentes.

"El principal uso de las medidas de las operaciones educativas en la investigación en evaluación es proveer un marco de referencia analítico dentro del cual se interprete el efecto observado de un programa educativo sobre las salidas relevantes del estudiante en términos de variables ambientales o de 'tratamiento' específicas ...". (66)

Al respecto, podemos ya desde ahora obtener alguna consecuencia que nos servirá para normar la investigación motivo principal de este trabajo y que consiste en la necesidad de ejercer el debido control sobre las variables, pues sólo así es posible interpretar correctamente el efecto de un programa educativo.

Los tres componentes, brevemente delineados, interactúan en la forma que el siguiente esquema muestra con claridad.



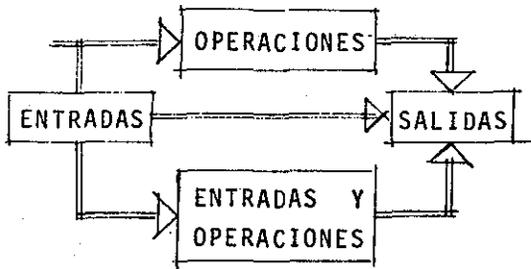
66. Ibídem, 16.

De acuerdo con el esquema, podríamos inclinarnos a pensar que la relación causal más directa es la que se establece entre las operaciones y las salidas cuando lo que tratamos es de tomar una decisión educativa encaminada a elegir la operación más adecuada para que el desempeño de los alumnos sea el deseado (relación B del esquema).

De ser así, la evaluación tendría como propósito principal el de informar del efecto de las operaciones en el desempeño de los alumnos.

El esquema, sin embargo, indica que además del efecto de las operaciones en las salidas, éstas se ven afectadas por las entradas en forma directa (relación C) e indirecta (relación A) al través de las operaciones. Resulta, por lo tanto, que es necesario el conocimiento de las relaciones A y C, antes de que sea posible interpretar correctamente una relación B observada, pues las diferencias en las salidas pudieran deberse en mayor medida a las desigualdades en las entradas que al efecto de las distintas operaciones. (67)

Para expresar de manera más clara el efecto que la combinación de las entradas con las operaciones produce en las salidas, quizá fuera conveniente, como lo sugieren los autores, trazar el siguiente esquema:



Ahora bien, esta primera parte del examen del modelo de Astin y Panosnos proporciona indicios, que estimo ciertos, acerca de la posibilidad de encontrarle el debido sustento teórico en tres disciplinas principalmente.

En la teoría general de sistemas intentaremos apoyar el aspecto estructural del modelo y en el análisis de la validez de la información -según sea la fuente de la cual provenga- y en la teoría de la decisión, trataremos de hallar la fundamentación teórica de sus aspectos funcionales.

A la luz de esas teorías continuaré, en los incisos sucesivos, con el examen del modelo en cuestión, relacionándolo con ellas en la medida en la que, según mi criterio, resulte pertinente.

2.2. La teoría de sistemas.

Antes de proceder al intento de relacionar el modelo de Astin y Panos con la teoría de sistemas, haré una breve exposición acerca de las principales características de dicha teoría.

Se puede afirmar que la teoría de sistemas, en su concepción actual, surge como una reacción al enfoque microscópico de tipo analítico que a partir del siglo XVIII y merced a los trabajos de Anton Van Leeuwenhoek, destacado microscopista danés, había sido el dominante en la labor de los científicos.

La visión microscópica, al concentrar el interés científico en el estudio de las partes, si bien ha propiciado muchos de los avances que en la ciencia se han conseguido, al ser aplicado a fenómenos de tipo social, económico, ecológico, etc., no ha rendido los frutos que inicialmente de él se esperaban.

Sólo en fechas recientes se ha empezado a complementar la visión microscópica, en la que se estudian las partes para tratar de establecer las relaciones causa-efecto, con la macroscópica, propia del enfoque de sistemas, en la que se emplea el conocimiento que se tiene de los elementos para estudiar el comportamiento de todo un conjunto de partes o subsistemas que interactúan. (68)

68. GEREZ, VICTOR y Manuel Grijalva. El enfoque de sistemas, 19.

Sin embargo, el concepto de sistema tiene ya una larga historia, aunque el término no se haya usado. "Como 'filosofía natural' podríamos remontarlo a Leibniz, a Nicolás de Cusa con su coincidencia de los opuestos; a la medicina mística de Paracelso; a la visión de la historia de Vico e Ibn-Kaldun, como sucesión de entidades o 'sistemas' culturales; a la dialéctica de Marx y Hegel- por mencionar unos cuantos nombres de una rica panoplia de pensadores". (69)

Aun así, la necesidad y factibilidad de un enfoque de sistemas no fue evidente sino hasta hace poco. Resultó por necesidad del hecho de que el esquema mecanicista fue insuficiente para enfrentarse a problemas teóricos, especialmente en las ciencias biosociales. (70)

La teoría de sistemas ha mostrado ser una herramienta rica en posibilidades para el desarrollo de la ciencia, los ejemplos podrían ser numerosos y para citar sólo uno elegí el de Piaget, quien según Hahn: "...vinculó expresamente sus conceptos a la teoría general de sistemas de Bertalanffy...".

Al término sistema pueden otorgársele distintos significados dependiendo de cuál de las cuatro características siguientes se considere como definitoria:

- a) La interrelación como aspecto esencial del sistema.
- b) La función teleológica como aspecto esencial de todo sistema.

69. VON BERTALANFFY, L. Teoría general de sistemas, 9.

70. Ibídem, 10.

- c) El sistema como instrumento conceptual.
- d) El sistema como explicación del devenir histórico.

Al incluir las cuatro características en la definición de un sistema, Antoni J. Colom, nos proporciona la siguiente: "...la concepción de la realidad, en cuanto se concibiera dicha realidad como una totalidad contruida por un conjunto de elementos (así como sus atributos, características y propiedades) en interacción dinámica y procesual respecto a un estadio posterior que se toma como objetivo o finalidad del propio sistema...". (71)

De los sistemas, así definidos, podremos alcanzar una mejor comprensión si en su análisis distinguimos tres dimensiones:

- a) el tiempo.
- b) la secuencia lógica.
- c) la estructura formal de la disciplina.

Todo sistema, durante su vida, atraviesa por distintas fases: planeación, desarrollo, implementación, operación y retiro en términos generales y esas fases van teniendo lugar en el tiempo: primera dimensión del sistema.

La solución de un problema sistemático depende también de que sea posible disponer de conocimientos específicos en la disciplina involucrada, siendo asimismo necesario que estén organizados según una estructura formal: tercera dimensión del sistema. (72)

71. COLOM, ANTONI J. Sociología de la educación y teoría general de sistemas, 15.

72. GEREZ, VICTOR y Manuel Grijalva. El enfoque de sistemas, 22.

Si entendemos como estructura, "...un conjunto de elementos entre los cuales existen relaciones tales, que la modificación de un elemento o de una relación entraña una modificación de los otros elementos o relaciones..." (73), nos damos cuenta de que tanto las áreas de conocimiento como los sistemas pueden ser considerados desde el punto de vista de su estructura.

"Desde el momento que una estructura está formada por un conjunto de elementos en interrelación, todo sistema tendrá una estructura". (74)

Podemos decir también, que una estructura o un sistema se caracterizan por tener propiedades que no se encuentran en ninguna de sus partes constitutivas, ni en el mero agregado de ellas. Dependen de sus partes, de sus miembros, pero no equivalen a la sola yuxtaposición de ellos. En muchos casos la relación entre los miembros es más importante que su naturaleza intrínseca. (75)

La importancia de las estructuras en educación ha sido señalada por Jerome S. Bruner, al expresar que:

"...Captar la estructura de una materia es comprenderla en forma tal que nos permita relacionarla significativamente con muchos otros temas....

Aprender la estructura es, dicho brevemente, aprender cómo las cosas se relacionan entre sí.... Hay dos modos en los que el aprendizaje sirve al futuro. Uno es al

73. COLOM, ANTONI J. Op. cit., 19.

74. Ibidem, 19.

75. FRONDIZI, RISIERI. ¿Qué son los valores?, 208, 209.

través de su aplicación específica a tareas muy similares a las originalmente aprendidas. El otro, es por medio de la transmisión de principios o actitudes La continuidad del aprendizaje que se produce con el segundo tipo de transmisión depende del dominio de la estructura de la materia en cuestión". (76)

En este sentido, la teoría de sistemas podría proporcionarnos un marco conceptual adecuado para estructurar nuestras observaciones en materia educativa, para que de manera similar a aquella en la que la estructura de las leyes de la física ha dado ímpetu a la tecnología, la estructura sistémica de la educación también se la diera. (77)

Ahora bien, para razonar sobre un sistema es conveniente considerar los aspectos siguientes:

1. Los objetivos del sistema considerado como un todo y más específicamente las medidas de actuación del sistema completo.
2. El medio ambiente del sistema: las restricciones fijas.
3. Los recursos del sistema.
4. Los componentes del sistema, sus actividades, metas y medidas de actuación.
5. La administración del sistema.

El objetivo, la meta por alcanzar, no sólo debe quedar bien definida, sin que pueda existir confusión al

76. BRUNER, JEROME S. En J. W. Forrester Principles of Systems, 1-4.

77. FORRESTER, JAY W. Principles of Systems, 1-4.

respecto, sino que también es necesario que en la medida de actuación del sistema se tomen en cuenta las consecuencias relevantes que puedan producir las actividades del propio sistema.

El ambiente, por definición, es aquello que está fuera del sistema y que si bien lo afecta, no puede ser controlado.

Los recursos materiales, humanos, económicos y de otra índole son parte del sistema y pueden ser manejados optimizando su uso en relación con el objetivo deseado.

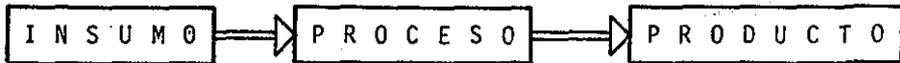
Los componentes son las partes en las que puede dividirse el sistema en atención a las tareas que cada una de ellas debe cumplir.

La administración se ocupa de la planeación del sistema como un todo, así como de su control y resolución de los problemas que pudieran presentarse en relación con los aspectos indicados: objetivos, medio ambiente, recursos, componentes. (78)

El razonamiento que hagamos acerca de un sistema no será, sin embargo, necesaria y estrictamente secuencial, muchas veces habrá que volver sobre un aspecto ya considerado y examinarlo de nueva cuenta, pues quizá sólo al pasar a los siguientes nos percatemos de la existencia de ciertas interacciones que no habíamos detectado en su verdadera dimensión en una primera instancia.

78. CHURCHMAN, C. WEST. El enfoque de sistemas, 47.

Una manera más en la que un sistema puede ser concebido es "...como una clase de entidad identificable a la cual se incorporan diversos insumos y se obtiene algún tipo de productos o servicios. (79)



Si volvemos ahora al modelo de Astin Y Panos, estamos en aptitud de constatar que su estructura es similar a la arriba indicada.

A los insumos corresponden las entradas, al proceso las operaciones y al producto las salidas. Figura 1.

Al tener el modelo de los autores citados una estructura que podemos llamar sistémica y al poseer las principales características que definen a un sistema, como lo son: a) estar formado por componentes en constante interrelación e interacción dinámicas, b) tener una función teleológica (la evaluación de los programas educativos), c) poder constituirse como marco conceptual para el análisis del proceso educativo y de evaluación y, d) al permitirnos recabar información, que en un momento dado, puede sernos útil para conocer el devenir histórico, me parece que contamos ya con un buen apoyo para afirmar que el modelo de Astin y Panos puede considerarse como fundamentado en la teoría de sistemas.

79. Ibidem, 81.

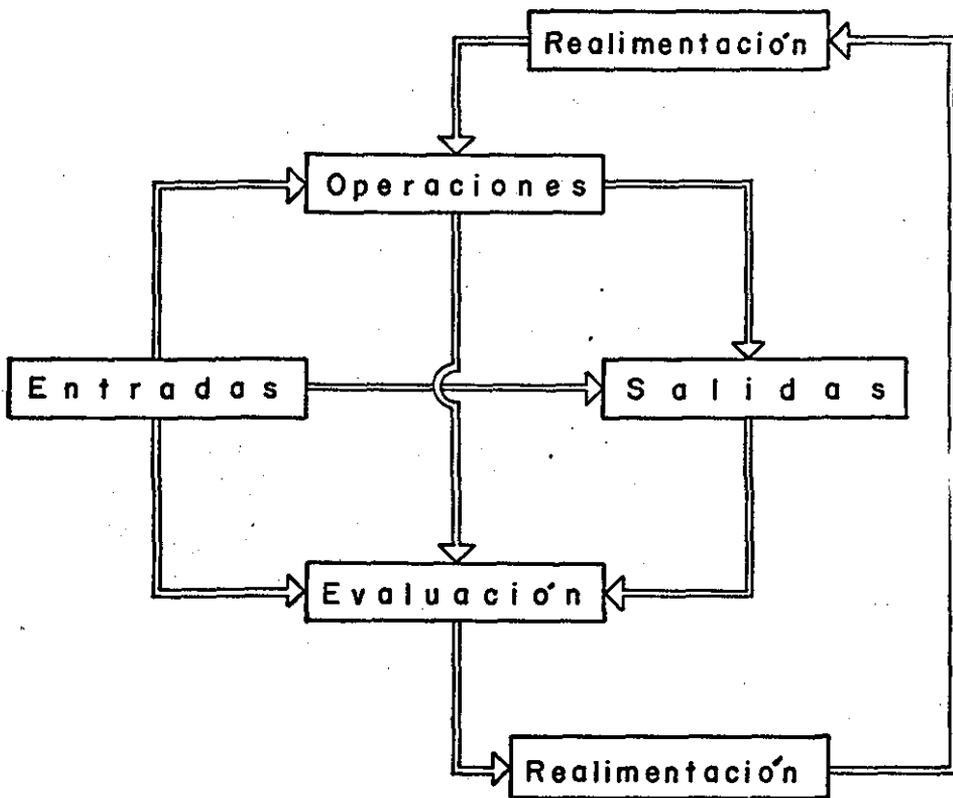


FIGURA 1

TESIS CON
FALLA DE ORIGEN

2.3. Las fuentes de información.

El examen del modelo de Astin y Panos nos condujo a hacer en el inciso anterior, algunas consideraciones acerca de su estructura y nos corresponde ahora referirnos a su funcionamiento.

Por ser según los autores, la principal función de la evaluación proveer, a quien toma las decisiones, de información pertinente en torno a las entradas, operaciones y salidas del programa a considerar, resulta que de la calidad de la información, base para la toma de decisiones, depende la de estas últimas. (80)

La información, en su forma más general, es solamente una colección de proposiciones cuyo valor de verdad es de 0 ó de 1, falso o verdadero. La incertidumbre de los valores de verdad en las proposiciones empíricamente basadas se expresa en términos de la probabilidad asociada a cada una de ellas, como veremos en el inciso siguiente.

Hay, en principio, dos niveles importantes de información que podríamos denominar de tipo 'estado' y de tipo 'relacional'. La primera, como su nombre lo indica, describe el estado de cosas en algún contexto y puede referirse a las condiciones actuales, a las pasadas, o la predicción de las futuras.

La observación de la evolución o permanencia de algún estado a lo largo de cierto tiempo puede servirnos, en

80. ASTIN, A. W. y R. J. Panos. Op. cit., 21.



la evaluación educativa, para pronosticar el posible estado futuro.

La información de tipo relacional especifica algunas conexiones entre estados. En esas relaciones puede, o no, implicarse causalidad. En cualquier caso, la información relacional nos permite reducir un conjunto de proposiciones a partir de otro, mientras que la información tipo 'estado' no contiene ningún elemento deductivo y es puramente un medio de identificación. (81)

No obstante ser la información relacional la que hace posible la deducción de proposiciones a partir de otras y de que en esa forma se constituye como base para la toma de decisiones, su calidad y por ende su confiabilidad pueden variar, de manera que mientras mejor sea la calidad de la información disponible serán también mejores las probabilidades de que se adopte una decisión acertada. (82)

La calidad de la información que se obtiene como resultado de la evaluación depende, según Astin y Panos, de la fuente o forma en la que se recaba.

Llaman anecdótica a la que se obtiene al hacer observaciones esporádicas y asistemáticas del comportamiento de un alumno o grupo de ellas. Puede ser tanto de estado, como relacional, pues a menudo contiene una suposición causal, una relación entre las entradas y las salidas, pero sin que la hipótesis reciba el debido apoyo empírico. Ni siquiera se somete

81. WHITE, D. J. Teoría de la decisión, 177-179.

82. MEHRENS, WILLIAM A. e Irvin J. Lehman. Op. cit., 16.



al consenso, cada profesor tiene su relato o anécdota favorito y probablemente en él basa sus decisiones. El valor de la información anecdótica como fundamento para tomar decisiones es realmente nulo, pues es el resultado de hacer el intento de generalizar una situación particular por medio de analogías improcedentes. (83)

El folklore es otra fuente de información, comparte con la anecdótica las características de ser obtenida de manera asistemática, de carecer de apoyo empírico y de ser, en ocasiones de estado y en otras relacional por referirse a una suposición acerca de alguna relación causal entre un programa y una salida educativos. No obstante, a diferencia de la anecdótica ha pasado por el filtro del consenso, el que, en cierto sentido, le confiere el carácter de tradición. Su confiabilidad para apoyar en ella decisiones educativas es también baja y no porque los conceptos tradicionales sean siempre falsos, sino porque al no haberse sujetado a la contrastación empírica no contamos con las necesarias evidencias sobre su veracidad. (84)

La información de tipo descriptivo, se limita precisamente a éso, a describir un programa educativo y puede ser enteramente cualitativa o a incluir rasgos cuantitativos. A diferencia de las anecdóticas y folklóricas, no contiene afirmaciones explícitas sobre relaciones causales, es una información de 'estado',

83. ASTIN, A. W. y R. J. Panos. Op. cit., 22.

84. Ibíd., 22.



de manera que quien trate de apoyarse en ella para tomar alguna decisión educativa, se verá precisado a establecer, ya sea explícita, o como más a menudo sucede, implícitamente ese indispensable vínculo causal, ausente en la información.

El cuarto tipo, que los autores llaman información de investigación, es de tipo relacional, incluye una afirmación acerca de un vínculo causal entre una operación y una salida educativas.

La información de investigación, por ser producto de este tipo de actividad, comparte con ella lo mismo sus ventajas que la posibilidad de que su validez pueda verse afectada por ciertos factores. Entre ellos, a reserva de dedicarme con mayor extensión al tema en el inciso dedicado a la 'selección del diseño', citaré por ahora sólo dos:

Por una parte el control de las variables. Cuando no se ejerce o es insuficiente, la correcta interpretación de los resultados es virtualmente imposible. Nunca sabremos, a ciencia cierta si son producto de los efectos de las variables que fueron controladas o de aquéllas que no lo fueron.

Por la otra, la selección del diseño que puede tener una estrecha relación con el punto anterior, pues en algunos de ellos no se toman en cuenta ciertas variables y por ende no se controlan.



Tal es el caso que citan Astin y Panos relativo a la aplicación de un diseño en el que sólo se observa el efecto de las operaciones en las salidas y no los que pudieran ejercer las entradas.

En esta situación parecía que las operaciones habían tenido los efectos supuestos en las salidas, sin embargo al ser el caso reexaminado se pudo detectar que los ejercidos por las entradas habían sido aún mayores.

En una investigación debidamente realizada debe, en consecuencia, intentarse el control de todas las variables y estar siempre conciente de que el apoyo empírico los puede recibir tanto la hipótesis alterna, como la de nulidad.

La posibilidad de que en el transcurso de un estudio sean halladas relaciones causales distintas a las inicialmente supuestas, es una de las características que distinguen y otorgan un grado más alto de validez a la información producto de una investigación en comparación con la de tipo folklórico.

TESE CON
FALLA DE ORIGEN

2.4. Teoría de las decisiones.

Merecen también nuestra atención las teorías en las que las decisiones se pueden apoyar, pues como vimos y de acuerdo con Astin y Panos, el objetivo principal de la evaluación es el de recabar información que sea útil para adoptar la decisión, en cada caso, adecuada.

La teoría de las decisiones se ha desarrollado con el fin de proporcionar al decisor las normas conducentes para hallar la solución óptima a los problemas que se le presenten.

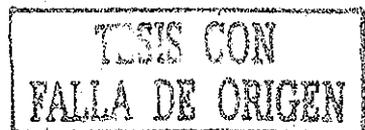
Así pues, veamos en primer lugar, cuáles son las condiciones mínimas y a la vez suficientes para que se pueda considerar que existe un problema.

1. Un individuo que tenga un problema. El decisor.
2. Un resultado que el decisor desea. (p. ej. un objetivo).

Sin el deseo de obtener un resultado todavía no alcanzado no puede haber problema. Un objetivo es un resultado que tiene valor positivo para el decisor.

Cualquier situación problemática debe involucrar cuando menos dos posibles resultados (0 y no 0), pero sólo uno de ellos puede tener valor positivo.

3. Cuando menos dos procedimientos con diferentes grados de eficiencia y cada uno con alguna probabilidad de conducir al objetivo deseado.



4. Un estado de duda por parte del decisor en cuanto a cuál es la selección 'óptima'.
5. Un ambiente o contexto del problema. El ambiente consiste de todos los factores que pueden afectar al resultado y que no están bajo el control del decisor.

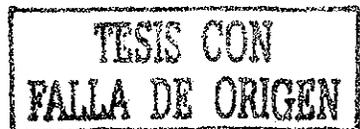
Los problemas pueden, por supuesto, ser más complejos que el mínimo descrito. Las complicaciones pueden resultar de las siguientes condiciones:

- a. Los decisores son varios y no uno solo.
- b. El decisor o decisores toman la decisión pero son otros los que la ejecutan.
- c. La eficiencia de la decisión pudiera verse afectada por la reacción de otras personas en las que hubiera surtido algún efecto.
- d. Varios objetivos pudieran estar involucrados en un problema y no ser consistentes entre sí.
- e. Los procedimientos pudieran ser muy numerosos, tal vez hasta incontables. (85)

Una vez establecida la existencia de un problema, sencillo o complejo, el siguiente paso consiste en elegir el procedimiento que nos conduzca a su 'óptima' solución.

Para ello, debemos primero comprender cuál es para un problema la mejor solución, la 'óptima'.

Podemos decir al respecto que, cuando menos hasta ahora y en el contexto indicado, no se ha llegado a



establecer cuál es el significado definitivo del término.

Aún así, intentaremos hacer un breve análisis acerca del criterio a seguir, en cada una de las tres situaciones que Luce y Raiffa consideran que es necesario tomar en cuenta para hallar cuál podría ser la mejor solución a un problema.

Estas situaciones son las que siguen:

- a) Certidumbre. Cuando se sabe que cada acción conducirá indefectiblemente a un resultado específico.
- b) Azar (risk). Cuando cada acción conducirá a un conjunto de posibles resultados, cada uno de ellos con una probabilidad de ocurrencia conocida. Se asume que esas probabilidades son conocidas por el decisor.
- c) Incertidumbre. Cuando una acción, o varias de ellas tienen como consecuencia un conjunto de posibles resultados específicos cuyas probabilidades de ocurrencia, o no las conocemos, o ni siquiera son significativas. (86)

De estas tres situaciones, nos ocuparemos con alguna extensión de la segunda por las siguientes razones: En la primera, que casi nunca se nos presenta en pedagogía experimental, al conocer el decisor los efectos específicos de cada procedimiento y en caso de que esté en aptitud de determinar el valor de cada resultado,

86. Ibidem, 32.



aspecto al que más adelante me referiré, no tiene sino que elegir el procedimiento que produzca el máximo valor.

La segunda situación, azar, será en la que nos encontremos en la mayoría de las ocasiones, sobre todo si se trata de la evaluación de programas educativos, pues podremos disponer de información sobre la probabilidad de obtener cierto resultado.

Sabremos, controlando el efecto de las variables de entrada el que las operaciones, también controladas, producen en las salidas y podremos calcular la probabilidad de que determinado resultado ocurra.

La tercera situación, la incertidumbre, nos induciría, antes de tomar alguna decisión, a tratar de indagar, ya buscando antecedentes de casos similares, ya realizando un experimento, la probabilidad de que la operación que nos proponemos implantar produzca los resultados que nuestra hipótesis supone.

En esta forma convertiríamos la situación de incertidumbre en una de azar.

En este caso, cuando un problema surge en una situación de azar nuestro análisis revelará que la decisión óptima puede ser definida como aquella que 'maximiza la expectativa matemática del valor o la utilidad'. El significado de este criterio depende, sin embargo, de que los objetivos sean definidos cualitativa o cuantitativamente.

TESIS CON
FALLA DE ORIGEN

Entenderemos por objetivo cualitativo el que se obtiene o no se obtiene y por cuantitativo el que se obtiene, o no, en varios grados.

Para mejor aclarar los conceptos anteriores, veamos dos ejemplos sencillos.

Los dos se refieren a la manera en la que se puede ponderar la eficiencia de un procedimiento.

En el primero, suponemos una situación de azar y un objetivo cualitativo.

Incluimos dos posibles procedimientos C_1 y C_2 , y dos objetivos probables O_1 y O_2 y asignándoles las probabilidades abajo indicadas, tendremos:

	O_1	O_2
C_1	$P_{11} = 0.6$	$P_{12} = 0.4$
C_2	$P_{21} = 0.2$	$P_{22} = 0.8$

Y para ponderar la eficiencia de cada procedimiento usamos la ecuación:

$$\sum_{j=1}^n P_{ij} - V_j$$

Para resolver la ecuación será necesario que a cada objetivo le asignemos un valor relativo V , y si este fuera de 0.7 para O_1 y de 0.3 para O_2 , entonces:

$$\text{Para } C_1 \quad \sum_{j=1}^2 = 0.6(0.7) + 0.4(0.3) = 0.54$$

$$\text{Para } C_2 \quad \sum_{j=1}^2 = 0.2(0.7) + 0.8(0.3) = 0.38 \quad (87)$$

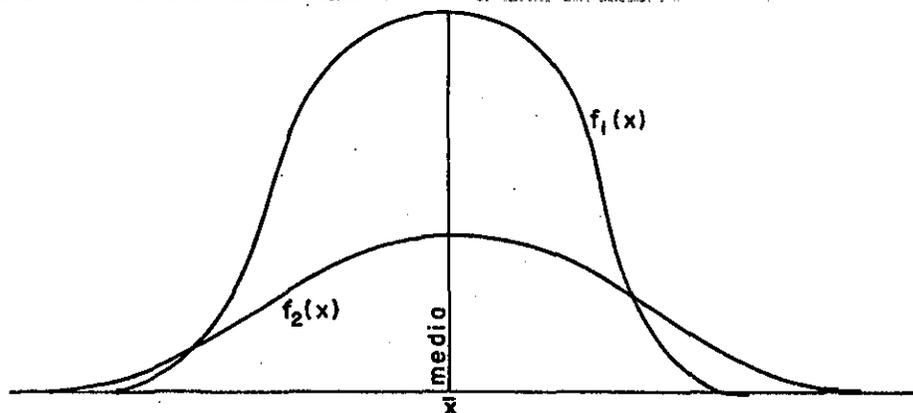
El cálculo es sencillo y los resultados claros: en el caso del ejemplo el procedimiento C_1 tiene una mayor medida de eficiencia (0.54) que el C_2 (0.38).

El segundo caso, al igual que el primero, se nos presenta con frecuencia en la pedagogía experimental. Se refiere a una situación al azar con objetivos

87. Ibidem, 37, 38.

cuantitativos, pues se trata de que los alumnos alcancen las mayores puntuaciones posibles.

Supongamos que con dos procedimientos distintos: C_1 y C_2 , las curvas de distribución tienen la misma media, pero diferente grado de dispersión.



Si sólo nos guiáramos por la media, los dos resultados nos parecerían iguales, pero si observamos la dispersión, nuestra selección dependerá de cuál de los resultados estimemos como preferible, $f_1(x)$ ó $f_2(x)$, pues $f_1(x)$ implica mayor uniformidad y $f_2(x)$ una disparidad más acusada, alumnos con calificaciones más altas y otros con unas realmente bajas, quizá reprobatorias.

En cuanto a los valores relativos, V en el primer ejemplo los elegimos a priori y si bien en el segundo asumimos que una calificación más alta tiene mayor valor que una baja, en ninguno de los casos fundamentamos

nuestra selección en algún criterio objetivo.

La necesidad de proporcionar al decisor una técnica para hacer esa selección de valor, ha dado lugar, en fechas más o menos recientes, a una nueva área en la investigación científica, la teoría del valor.

Los esfuerzos para medir el valor relativo que un individuo confiere a los objetos, sucesos o situaciones se remontan a los años finales del siglo XVIII, cuando en 1798, Jeremy Bentham hizo una proposición que él llamó 'cálculo de valores'. Desde entonces los economistas han especulado constantemente sobre la medida de 'utilidad' (ellos usan el término valor).

En las últimas cuatro décadas los psicólogos han dirigido su atención a la medición de actitudes, un concepto en estrecha relación con el de valor. Pero ha sido sólo a partir de los años cincuenta que se han hecho los mayores esfuerzos con carácter científico para desarrollar y aplicar las medidas de valor.

Una medida de valor relativo de tipo general, aplicable a todas las situaciones, no ha sido aún desarrollada no obstante el empeño y significativos logros de algunos estudiosos del tema. Podemos señalar en este sentido los trabajos de von Neumann y Morgenstern acerca de la teoría de juegos y los de E. A. Singer Jr. en el desarrollo de la teoría del 'experimentalismo'.

Siguiendo los pasos de von Neumann-Morgenstern, en fechas posteriores a la de la publicación de su libro "Theory of Games and Economic Behavior" en 1947, otros autores han tratado de encontrar mejores medidas de valor relativo.

Entre ellos podemos mencionar a Davidson-Siegel-Suppe, quienes han elaborado uno de los procedimientos más refinados al respecto, así como a Churchman-Ackoff con una proposición que por no estar basada, como las otras, en la teoría de juegos, elimina la necesidad de hacer ciertas suposiciones en relación con la probabilidad subjetiva de que el resultado esperado sea el máximo.

En todos los casos, sin embargo, las técnicas disponibles no suprimen el requisito de ordenar previamente los objetivos de acuerdo con el valor a priori que les conceda el decisor.

Los científicos contemporáneos se muestran reacios a involucrarse en cuestiones que conciernen a los valores. Tanto la mayoría de ellos, como la de los filósofos aduce que los problemas de valores no pueden reducirse o traducirse a cuestiones científicas. (88)

Resulta así, que para la selección y jerarquización de los objetivos, en nuestro caso educativos, el decisor tendrá que apoyarse en su preferencia personal o en alguna de las teorías de los valores que han postulado los filósofos y sólo a partir de las premisas que al

88. Ibidem, 81, 82.

respecto adopte, aplicar la técnica adecuada para maximizar las probabilidades de llegar a la meta deseada.

En coincidencia con lo expresado, Astin y Panos nos dicen que: "Cualquier decisión educativa implica una elección entre los medios alternativos disponibles a través de los cuales pueden alcanzarse los objetivos deseados. En el ámbito de un sistema escolar, por ejemplo, tales medios podrían incluir la organización deliberada de ciertas experiencias de aprendizaje (plan de estudios, métodos), la estructuración propositiva del medio ambiente físico (por ejemplo: el diseño y ubicación de las aulas, edificios, campos deportivos), o el establecimiento de ciertas normas y reglamentos...". (89)

En todos esos casos, por tratarse de un sistema escolar, la consideración primaria para tomar una decisión sería de índole educativa, se adoptaría en razón al efecto que la medida pudiera producir en las salidas.

No obstante, hay ocasiones en las que prevalecen los motivos económicos (p. ej. las modificaciones o ampliaciones de los espacios de la escuela, mientras que en otras (la implantación de un plan de estudios o de un programa), son los criterios directamente educativos los que se privilegian.

89. ASTIN, A. W. y R. J. Panos. Op. cit., 7.

En todo caso, en la adopción de un criterio, sea económico, educativo o de otra clase, se involucra un complejo sistema de ponderación subjetiva y juicios de valor. (90)

Al respecto y si bien es cierto que el valor que se asigne a cada objetivo dependerá de la escala que en ese sentido asuma el decisor, también lo es que para alcanzar el que haya estimado como valioso, tendrá que elegir el procedimiento idóneo. Así pues, cuando el decisor haya jerarquizado sus objetivos, podrá disponer de alguna de las técnicas arriba indicadas como guía para ponderar las diversas opciones.

Por el breve examen que hemos hecho en los incisos anteriores de los fundamentos teóricos de la evaluación educativa en general y del modelo de Astin y Panos en particular, estimo que contamos ya con una aceptable base para proceder a la exposición del trabajo experimental que he realizado y que es el motivo central de esta tesis.

90. Ibidem, 8.

II. SEGUNDA PARTE

INFORME DEL TRABAJO EXPERIMENTAL
REALIZADO.

3. Fases iniciales de la investigación.

Por la experiencia que he podido acumular durante treinta años, de 1954 a la fecha, de ser profesor de la que hasta hace poco se llamaba Escuela Nacional y ahora es ya Facultad de Arquitectura, me he podido dar cuenta de algunos de los problemas a los que nos enfrentamos quienes pretendemos impartir enseñanza en ella.

Los he constatado en las dos áreas en la que he dado clase: en la de tecnología y en la de diseño.

En la primera y desde hace ya más de veinte años me ocupo del curso de Administración de Proyectos y Obras II (A.P.O.II), el que si bien en ese lapso ha cambiado varias veces de nombre, ha conservado, en lo esencial, el mismo programa.

Al respecto, en mi tesis para obtener la Maestría en Arquitectura, en el área de tecnología, incluí dos proposiciones: la de un modelo ecléctico para la elaboración de programas y la aplicación del mismo al caso específico de la materia en cuestión.

El programa que propuse lo he podido aplicar sin mayores dificultades, debido tanto, a que sus contenidos y los del que oficialmente sigue en uso son, en lo sustancial, iguales, como a que en el vigente se deja a los profesores, de hecho, en libertad de decidir sobre la manera más conveniente de conducir los demás aspectos del proceso E-A.

Lo anterior es particularmente cierto en cuanto a la evaluación, pues en este rubro no se especifica ningún procedimiento a seguir, a pesar de ser en el que se presentan las mayores dificultades.

3.1. Planteamiento del problema.

Tratando de subsanar esa deficiencia, los profesores de los cursos de A.P.O. II, hemos intentado diversos procedimientos para medir y evaluar el rendimiento escolar de los alumnos.

Hemos recurrido a exámenes anuales cuando era esa la periodicidad de los calendarios lectivos y a semestrales en la actualidad.

Algunos de nosotros, en ocasiones, hemos optado por la administración de varias pruebas en un semestre y aunque, a primera vista, nos ha parecido que este procedimiento es más efectivo para mejorar el rendimiento académico que el que consiste en impartir las pruebas sólo al final de cada período escolar, nuestras observaciones han sido, hasta ahora, esporádicas y asistemáticas y por ende con un reducido grado de validez.

Ante esa situación y tomando en cuenta que, como veremos en el siguiente inciso, las investigaciones realizadas con el fin de establecer los efectos de las pruebas periódicas en el desempeño académico de los alumnos no han rendido resultados ciertos y consistentes, me ha parecido que sería útil y pertinente emprender una investigación en ese sentido.

Esa investigación la he estado realizando en la Unidad Académica de Talleres de la Facultad, específicamente

en el Taller E de la Facultad de Arquitectura de la UNAM.

En ese Taller, como en los demás de la unidad académica a la que está adscrito, los alumnos se inscriben en forma voluntaria al llegar al tercer semestre y deben permanecer en el mismo hasta el sexto, o sea durante toda la etapa de la carrera a la que se denomina 'formativa'. En la etapa 'integral' séptimo y octavo semestres pueden, si así lo desean, cambiarse a otro taller.

El curso de A.P.O. II, se imparte en el quinto semestre y la única materia seriada como antecedente es la de Administración de Proyectos y Obras I, en el cuarto semestre.

Ello da lugar a que alumnos de semestres anteriores al quinto, con el sólo requisito de haber aprobado la asignatura antecedente, se inscriban en el curso que nos ocupa, sin haber adquirido otros conocimientos que son indispensables para su debida comprensión, como pueden serlo los que se imparten en materias relacionadas con los procedimientos de construcción en especial.

Debido a que la clase corresponde a un semestre impar, quinto, en los que según el calendario lectivo son también impares, predominan los alumnos regulares, los que en el quinto semestre a partir de su inscripción inicial cursan ese mismo período escolar; mientras que

TESIS CON
FALLA DE ORIGEN

en los semestres pares es mucho mayor el porcentaje de irregulares o repetidores.

Las sesiones tienen lugar los martes y jueves de las diez y media a las doce horas, durante quince semanas por regla general.

Ya antes, desde las siete de la mañana, los alumnos han asistido a otras clases y algunas de ellas les exigen bastante esfuerzo y atención concentradas, como es el caso de aquéllas en las que se pretende enseñarles a calcular la estructura de un edificio.

Otra circunstancia que he observado, es que en las clases matutinas, como la nuestra, la mayoría de los alumnos dedican todo su tiempo al estudio y sólo de manera eventual al trabajo remunerado.

Los cursos de A.P.O. II se imparten en los locales que le han sido asignados a la Unidad Académica de los Talleres de la Facultad. Se agrupan en varios edificios separados entre sí por áreas jardinadas y comunicados por pasillos cubiertos. En cada uno de ellos se alojan dos talleres: uno matutino y uno vespertino por regla general.

El salón del que podemos disponer se encuentra en la segunda planta de uno de esos edificios. Su espacio es insuficiente para que todos los alumnos puedan sentarse ante alguna mesa o escritorio. El resultado es que cuando los grupos son numerosos y la asistencia nutrida,

TESIS CON
FALLA DE ORIGEN

la incomodidad de los alumnos es notoria y motivo de frecuentes distracciones y dificultades para que se conserve el orden necesario.

Se dispone de un buen pizarrón, cuyo uso no obstante, resulta en ocasiones inconveniente por la inadecuada iluminación tanto natural, como artificial del salón.

Su acústica es asimismo deficiente, por un lado, se oyen todos los ruidos que producen los automóviles y autobuses que pasan cerca y también los que hacen los alumnos que circulan o hablan en los pasillos interiores; por el otro, cuando los alumnos no son capaces de guardar un silencio prácticamente total, cosa que rara vez sucede, los de las últimas filas no pueden oír lo que el profesor está diciendo.

El trabajo experimental en cuestión tuvo lugar en los dos semestres de los años de 1983 y 1984.

Debido a que como más adelante veremos, los resultados no han sido hasta ahora concluyentes, me propongo continuarlo en los semestres venideros, aunque de ser posible en otra área, quizá la de diseño.

El estudio se hizo con los grupos de alumnos que, como ya vimos, se inscriben voluntariamente en el Taller E y cuando les corresponde lo hacen de la misma manera en la clase de A.P.O. II. Constan de alrededor de 50 alumnos, con un máximo de 70 y un mínimo de 40, de ellos un 25% aproximadamente son mujeres y el 75% restante hombres.

TESIS CON
FALLA DE ORIGEN

En las condiciones brevemente descritas se desarrolló la investigación con el fin primordial de obtener datos, lo más confiables que fuera posible, para dar respuesta a la siguiente interrogante: ¿producen algún efecto en el R-E de los alumnos las mediciones parciales que del mismo se hagan en el transcurso de un semestre escolar?

En torno a esta pregunta procederemos en los siguientes incisos al examen de las demás etapas del estudio.

TESIS CON
FALLA DE ORIGEN

3.2. Antecedentes.

Una vez planteado el problema del que nos ocuparemos y antes de entrar de lleno a las particularidades del estudio experimental motivo de este trabajo, estimo que sería conveniente hacer un examen, por breve que sea, de los procedimientos seguidos y de los resultados obtenidos en otras investigaciones que versen sobre el mismo tema que la nuestra.

Se trataría, en términos generales, de aquéllas en las que se haya intentado indagar acerca de las relaciones que pudieran existir entre la medición periódica y el rendimiento escolar, pues el conocimiento de otras investigaciones afines nos es necesario para no caer en inútiles repeticiones.

Con el fin de recabar la mayor cantidad posible de información sobre dichas investigaciones acudí al Centro de Información Científica y Humanística (CICH) de la UNAM.

Una vez que expliqué al asesor en pedagogía el objeto de la consulta que pretendía hacer, nos dedicamos a examinar los contenidos generales de diversos bancos de datos que de acuerdo con el índice general parecían tener alguna relación con el tema de nuestro interés.

Después de eliminar los que no tenían ninguna, llegamos a la conclusión de que los denominados ERIC y PSYCINFO eran los que debíamos revisar de manera



detallada pues eran los más prometedores en cuanto al propósito que perseguíamos.

El primero de ellos abarca el período que va de 1966 a agosto de 1982; y el segundo, el que queda comprendido entre 1967 y junio de 1982.

Entre los dos, los datos generales y resúmenes (abstracts) se refieren a más de trescientas cincuenta investigaciones.

Después de examinar con todo cuidado ese cúmulo de material, me di cuenta de que en realidad era muy diversa la índole de los temas tratados.

Supongo que eso se debe a que las palabras clave que sirvieron como guía en la búsqueda tienen, en pedagogía, tan amplias implicaciones que dan lugar a que sean varios los temas afines incluidos.

Esas palabras fueron en especial, prueba (test), examen (examination), periódico (periodic), recurrencia (recurrence), final, parcial (partial) y desempeño académico (academic achievement).

Por otra parte, la computadora responde a cada término por separado y no a las combinaciones, frases, que con ellos podrían formarse y por eso las investigaciones detectadas no fueron, todas ellas, específicas en cuanto al tema de nuestro estudio. No solamente se obtuvo información sobre mediciones o pruebas periódicas, sino

TESIS CON
FALLA DE ORIGEN

también acerca de las de cualquier tipo, evaluación en distintos aspectos, etcétera. Como ejemplos puedo citar dos estudios: 'Instrucción bilingüe personalizada. Evaluación final' y 'Una prueba parcial sobre las teorías de la cultura de la pobreza, de Oscar Lewis, en las áreas rurales'.

Por las circunstancias someramente reseñadas, de los 350 resúmenes (abstracts) que fueron rescatados de los bancos de información, únicamente ocho se refieren a investigaciones directamente relacionadas con el tema que nos ocupa.

De cada una de ellas obtuve el informe correspondiente y de la parte medular de los mismos haré a continuación una breve exposición.

a) La primera de dichas investigaciones fue realizada por H. Allen Murphy, James S. McMichael y Richard P. Cariello, en el Departamento de Psicología del C. W. Post Center de la Universidad de Long Island y publicada en el Journal of Personalized Instruction en marzo de 1977, Vol. 2, No. 1.

Con esta investigación se pretendía encontrar respuesta a la pregunta concerniente a la probable relación entre el número de unidades de prueba a las que un alumno se somete y su calificación en el examen final.

TESIS CON
FALLA DE ORIGEN

Una característica importante consistió en que cada alumno, como parte del Personalized System of Instruction (Sistema de instrucción personalizada) podía tomar el número de pruebas que juzgara adecuado para alcanzar el pleno dominio de los temas de estudio.

Un curso de iniciación a la estadística en psicología se impartió en veintitres estudiantes aplicando el sistema mencionado.

Consistió en 14 unidades temáticas divididas en 30 partes. De esas 30 partes, 14 eran sobre los contenidos de cada tema y las 16 restantes sobre 'procedimiento'. En estas últimas los estudiantes debían resolver problemas que implicaban la aplicación de los procedimientos estadísticos que se les habían enseñado. Las preguntas sobre contenido eran de respuesta breve y, con menor frecuencia, de selección múltiple.

Las pruebas sobre procedimiento consistían en problemas que el alumno debía resolver con la ayuda de una calculadora. Debido a que se impartían durante las horas de clase, su carácter era más el de tareas que el de exámenes. Se les permitía corregir errores aritméticos y, a menudo, pequeñas fallas en el procedimiento. Solamente en los casos en los que el estudiante se mostraba realmente

TESIS CON
FALLA DE ORIGEN

desorientado se consideraba que había fallado y se le administraba un nuevo examen.

Cada uno de los setenta reactivos de la prueba de contenido en el examen final, fue calificada en forma independiente por el instructor y por el autor principal de dicha prueba, teniendo cada uno la opción de dar crédito parcial a las respuestas que estimara cercanas a las correctas.

La correlación entre las calificaciones otorgadas por cada una de las personas encargadas de asignarlas, mostró que el sistema empleado era altamente confiable ($r=0.97$, $n=23$).

La calificación de la parte relativa al procedimiento fue, en el examen final, más complicada. Cada uno de los diez problemas fue analizado para detectar los posibles errores críticos y por cada uno de ellos que era cometido por el alumno se le descontaba un punto. Para comprobar la confiabilidad, ocho exámenes fueron seleccionados aleatoriamente y un problema de cada uno de ellos fue calificado por el tercer autor de la prueba. El porcentaje de concordancia de las calificaciones fue del orden del 95 al 100%, con una media de 98.75%, probándose así que el método para calificar era confiable.

Los estudiantes estimaron que el promedio adecuado de pruebas para alcanzar el dominio pleno de las

TESIS CON
FALLA DE ORIGEN

catorce unidades temáticas era de 20.7, o sea alrededor de una prueba y media por cada unidad. La media de las calificaciones en el examen final fue de 139.4 puntos sobre 170 posibles. La correlación entre el número de pruebas de contenido y la calificación en el examen final fue: $r = -0.29$ ($n=23$, $p=0.192$).

En consecuencia, los datos obtenidos condujeron a la conclusión de que no existía ninguna correlación significativa entre el número de pruebas tomadas (y falladas) y la calificación en el examen final del curso en cuestión.

- b) La segunda fue realizada por B. C. Goldwater y L. E. Acker, en la Universidad de Victoria en Canadá y sus resultados publicados en la revista 'Teaching of Psychology' de diciembre de 1975, Vol. 2, No. 4.

A los alumnos que se inscribían en el curso de Introducción a la Psicología en la Universidad de Victoria, se les ofrecía una opción entre dos secciones de clase de tipo tradicional y una experimental impartida por los autores.

En la mesa de inscripciones se les proporcionaba un instructivo sobre los procedimientos a seguir en cada caso. Alrededor de la tercera parte, 234, de los alumnos inscritos en las tres secciones del



curso, eligieron el de tipo experimental. La mitad de ellos quedaron adscritos a un grupo de dominio pleno y los restantes al de control.

La asignación se hizo seleccionando, de la lista en orden alfabético, un alumno sí y otro no al grupo de dominio pleno. En el siguiente período de clases se invirtieron los grupos, de manera que todos los alumnos quedaran expuestos a las mismas condiciones, pues quienes habían formado el grupo de control en el primer período pasaron en el segundo al de dominio pleno y viceversa. En esta forma, afirman los autores de la investigación, se resolvió el problema ético que se hubiera suscitado al dar a un grupo de estudiantes desigual trato que al otro. En este sentido y por ser el caso de trato desigual en el que me encuentro, pues durante todo el semestre la mitad de los alumnos forman el grupo de control y el resto el experimental, en cada ocasión procedo a explicarles que para que hubiera en ello una actitud éticamente objetable, sería necesario que de antemano supiéramos cuál va a ser el efecto de la variable experimental. En ese caso, de manera obvia, el experimento carecería de objeto y sería inútil realizarlo. No siendo así me parece que no hay ningún problema ético al aplicar a la mitad del grupo pruebas periódicas y a todos la final.

Volviendo al informe que estoy reseñando, los autores se refieren al problema que se presentó al surgir la posibilidad de que en el segundo período, por los efectos de la transferencia, la comparación del grupo de dominio pleno con el de control no rindiera resultados confiables, pues éste había recibido el curso de dominio pleno en el período anterior.

Por esta razón la comparación entre el grupo de dominio pleno y el de control se hizo analizando, únicamente, los resultados del primer período, no contaminado por el posible efecto de la transferencia.

El libro 'Basic Psychology' de Kendler, en su edición de 1968, fue usado como texto. Se dividió en 22 unidades de 20 a 30 páginas cada una. Cada período abarcó 11 de dichas unidades. Los alumnos, tanto los del grupo de dominio pleno, como los del de control, asistían a una lección común durante la primera hora de clase de cada semana. En la segunda y tercera horas, a los alumnos del grupo de dominio pleno se les administraba una prueba de 10 reactivos (ocho de opción múltiple y dos de complementación), basada en la unidad del texto asignada a esa semana. Los reactivos de cada prueba se elegían con objeto de formar una muestra representativa de la unidad de cada semana.

Los alumnos eran informados en la clase siguiente a la de la prueba acerca de cuáles eran las respuestas correctas y como disponían de una copia de las suyas, podían confrontarlas y saber si habían o no obtenido el mínimo de 8 de 10 puntos, en alguna de las dos pruebas de cada semana, que era lo necesario para recibir el crédito respectivo.

Si no aprobaban la primera de esas pruebas podían tomar la segunda sin penalización, pero no se les permitía hacerlo para mejorar la calificación, cuando hubiera sido aprobatoria.

Se les daba también la oportunidad de recurrir a la tutoría de los profesores asistentes y a la del instructor si la solicitaban expresamente.

Al grupo de control no se le autorizaba a tomar las pruebas, pero cada dos semanas, más o menos, se le aconsejaba en forma enfática que no descuidara sus lecturas para que pudiera avanzar al parejo del otro grupo.

Asimismo, en la segunda hora de cada semana y en grupos pequeños (15 a 25 alumnos), guiados por el profesor asistente, se dedicaban a discusiones orientadas, tanto al texto de Kendler, como a los principios del método científico que Whaley y Sorrat ponen en relieve en su libro 'Attitudes of Science' (1967). La concordancia de los dos autores en cuanto

a dichos principios hacia posible alentar a los alumnos a detectarlos al leer el libro de texto, el de Kendler.

Al término del primer período se administró al total de los alumnos un examen de 100 reactivos de selección múltiple en condiciones formales.

Este cuestionario se acompañó con otro cuyo objeto era que los alumnos evaluaran el curso,

Para los alumnos del grupo de dominio pleno la calificación del período quedó determinada por los resultados del examen final (40%) y el de las pruebas semanarias (60%).

Para el grupo de control se tomaron en cuenta: el examen final (40%), el que se les hizo a la mitad del curso, consistente en 50 reactivos de opción múltiple (30%) y las pruebas correspondientes a la sección de discusión (30%). El grupo de dominio pleno tuvo oportunidad de aprobar cada unidad en dos intentos y recibía de inmediato la información sobre los resultados que iba obteniendo cada semana.

Resultados: Un total de 198 estudiantes tomó el examen final del primer período (101 de dominio pleno, 97 de control) y 190 el del segundo.

Los datos del examen final del primer período revelaron una ventaja sustancial del grupo de

alumnos de dominio pleno sobre los del de control. La mediana fue de 66.5 y 55.5 respectivamente, con $p < 0.005$.

Debido a que las distribuciones de los resultados del examen final mostraron sesgo, se usó la mediana y como medida de significatividad la prueba de dicha mediana.

El éxito del programa de dominio pleno se hace más patente si se toma en cuenta que el 55% del grupo obtuvo calificaciones de A o B (un mínimo de 65% de respuestas correctas) en el examen final del primer período, por sólo 30% del grupo de control.

En el examen del segundo período, la diferencia entre los grupos de dominio pleno y de control fue poco significativa, 70.8 y 67.8 fueron las medianas respectivas. Esta convergencia pudo deberse a alguna de las siguientes causas;

Primera: el contenido de las pruebas del segundo período fue menos difícil que el del primero que incluía: estadística, sensación, percepción y fundamentos biológicos.

Segunda: debido a que, como quedó ya anotado, el grupo de control del segundo período había sido el de dominio pleno en el primero, los resultados del examen de ese segundo período se contaminaron por los efectos de la transferencia.

Fue un hecho que algunos alumnos informaran a los instructores que al pasar del grupo de dominio pleno al de control habían conservado los buenos hábitos de estudio que en el primero habían adquirido.

Los datos del presente estudio sugieren que el procedimiento de pruebas semanarias, aunado al criterio de dominio pleno, es suficiente para generar beneficios en el desempeño académico. Al parecer el motivo principal de ese efecto es el tiempo adicional que los alumnos del grupo de dominio pleno dedicaron al estudio, el doble, según ellos mismos informaron, que los del grupo de control, lo cual no hizo que su interés y gusto por el curso disminuyera, sino al contrario, que aumentara.

- c) La tercera investigación fue realizada por O. Glenn Geitger y Darrel E. Bostow en la Universidad de South Florida, E. E. U. U. y publicada en la revista 'Psychological Reports' de 1976, No. 39, Págs. 707-710.

Keller (1968) describió un método en el que adapta algunos principios del conductismo a la enseñanza escolar.

Las variantes de este método han sido aplicadas con éxito según informan: Ferster, 1968; Malott &

Svinicki, 1969; McMichael & Corey, 1969; y otros.

El método de Keller comparte con sus variantes ciertas características, por ejemplo: pequeñas unidades del material del texto con pruebas para cada una, pruebas sucesivas cuando el alumno no alcanza el desempeño académico especificado; lectura, por parte de los alumnos, de los objetivos del curso; relación cercana con los profesores asistentes y la oportunidad de determinar su propio ritmo de estudio a fin de cumplir los requerimientos del curso.

No obstante que en algunos estudios se ha investigado la contribución de esas variables específicas al rendimiento escolar, los efectos relativos de cada variable o posible combinación de ellas no ha sido, hasta ahora, puesto en claro. Este estudio demuestra los efectos de las pruebas semanarias por unidad de contenido en los exámenes de tipo objetivo.

Cuarenta y nueve estudiantes de un curso de introducción a la psicología de la educación, fueron aleatoriamente asignados a dos grupos, sacando números de un sombrero en la primera clase. Cinco estudiantes se rehusaron a someterse voluntariamente a eses proceso. Se les permitió, sin embargo, elegir el grupo en el que se les administrarían las pruebas,

pero sus calificaciones se omitieron en el análisis de los resultados.

En el trimestre académico, el curso se impartió en clases de cuatro horas los jueves por la noche. En la primera clase los alumnos fueron informados acerca de la participación que tendrían en un programa de evaluación procedimiento. Al principio del curso se dió a los estudiantes un sumario de los temas de la materia que fueron tomados de cuatro textos sobre psicología de la educación.

El contenido se dividió en nueve unidades, una para cada semana. El instructor explicó que al grupo al que no se le administrarían pruebas semanales, recibiría a la mitad y al fin del curso exámenes amplios, los dos a base de reactivos de opción múltiple. Cada uno de esos exámenes contaría como máximo por el 50% de la calificación final.

El grupo con pruebas semanales recibiría esos mismos exámenes, a la mitad y al fin del curso también, pero su valor relativo en la calificación final sería de 25 y 50% respectivamente. Este grupo, de manera adicional, tomaría una prueba de opción múltiple cada semana que cubriría sólo el material de ese lapso. Las calificaciones combinadas de todas las pruebas semanales valdrían un 25% de la calificación final.

Cada semana, durante la primera media hora de clase el grupo de pruebas semanaarias las resolvió sin ninguna clase o discusión de material previos. Inmediatamente después, la clase entera se reunía durante tres horas con un descanso alrededor del mediodía. La primera hora y media se dedicaba a la discusión del material del texto y la segunda, a películas, conferencias cortas y discusiones sobre temas relacionados con la materia.

Ambos grupos se reunían para el examen a mitad del curso. Las pruebas semanaarias constaban de 20 reactivos de opción múltiple cada una, las de medio curso de 50 y la final de 57 del mismo tipo para las tres. Con la excepción de las pruebas semanaarias, los dos grupos estaban completamente integrados uno al otro.

Resultados:

	n	Prueba a mitad del curso.		Prueba final	
		\bar{x}	s	\bar{x}	s
Con pruebas semanaarias	25	45.16	3.57	42.28	7.40
Sin pruebas semanaarias	24	39.38	8.29	37.38	8.82

Para los exámenes a la mitad del curso, la media de las calificaciones del grupo de pruebas semanarias y para el que no las tomó, fue de 45.16 y 39.38 respectivamente.

Estos datos indican un desempeño significativamente mejor por parte del grupo que recibió pruebas semanarias ($t_{47} = 3.09$, $p < 0.01$). Las medias del examen final fueron, en el mismo orden, de 42.28 y 37.08 respectivamente, indicando también un desempeño significativamente mejor para el grupo que tomó las pruebas semanarias ($t_{47} = 2.18$, $p < 0.02$).

Este estudio muestra que la simple adición de pruebas semanarias puede producir diferencias significativas en el desempeño escolar de los alumnos en los exámenes a mitad y fin de los cursos. Los datos apoyan otros estudios previos y sugieren una importante relación entre la frecuencia del estudio y un mejor rendimiento en las pruebas.

- d) La cuarta investigación fue hecha por V. T. Mawhinney y otros investigadores de la 'Southern Illinois University' y publicado en el 'Journal of Applied Behavior Analysis' en el invierno de 1971, No.4 Págs. 257-264.

En fechas recientes se ha encontrado que la aplicación de programas complejos de enseñanza de

tipo experimental, similares a los desarrollados por Keller (1968) y usados por Ferster (1968), han rendido mejores resultados en los exámenes finales en comparación con los conseguidos mediante el uso de los métodos tradicionales de instrucción (McMichael & Corey, 1969; Sheppard & McDermot, 1970).

En los programas educativos es frecuente el uso de dos tipos generales de calendarios de exámenes: el primero consiste en la administración de pruebas a intervalos predecibles que pueden ser fijos o variables, exámenes a la mitad y al final del curso o tan frecuentes como lo son las pruebas diarias; en el segundo los intervalos de interés son variables y las pruebas se imparten en fechas no predecibles.

El uso continuado de esos calendarios de exámenes se basa en las preferencias personales y en las observaciones casuales de los instructores. Muchos investigadores han analizado los hábitos de estudio de los alumnos usando cuestionarios e informes de los mismos estudiantes. En esos trabajos, sin embargo, no se han empleado métodos directos de observación, ni se ha investigado la cantidad o distribución del tiempo dedicado al estudio en función con los calendarios de exámenes.

En este trabajo se investiga el tiempo que los alumnos dedican al estudio de los materiales del

TESIS CON
FALLA DE ORIGEN

curso en función con dos calendarios de exámenes.

Primer experimento.

Se realizó por medio de la observación directa de los hábitos de estudio y haciendo la comparación de la duración y distribución del tiempo dedicado a esa tarea, según fueran los calendarios de pruebas, diarios o semanarios.

Método.

Seis mujeres entre los 18 y 51 años de edad y dos hombres entre los 21 y los 23 que habían aprobado un curso de prerrequisitos en psicología general y eran estudiantes graduados y subgraduados de la Southern Illinois University, se ofrecieron como voluntarios para participar en una sección especial de lectura con miras a obtener los créditos correspondientes al curso. Participando en esta sección los estudiantes satisfacían todos los requisitos del curso regular. Para la selección de los sujetos se aplicaron los siguientes criterios: 1) Que ninguno tuviera experiencia previa en psicología conductista, 2) Que su promedio de calificaciones fuera 'C' como mínimo y, 3) Que no tuvieran ninguna otra actividad académica durante las horas en las que el cuarto para estudio estuviera disponible.

TESIS CON
FALLA DE ORIGEN

Materiales.

Los materiales para estudio consistieron en el texto completo de C. B. Ferster y M. C. Perrott, titulado: 'Behavior Principles', así como de capítulos selectos de la 'Technology of Teaching' de B. F. Skinner.

Hasta donde los textos lo permitieron, se trató de igualar el volumen diario de la lectura asignada a cada alumno. Las lecturas diarias comprendían, aproximadamente, quince páginas. El libro 'Behavioral Principles' está parcialmente programado, después de cada tema de contenido amplio se incluyen los objetivos conductuales correspondientes a esa sección, que informan a los estudiantes sobre las habilidades más importantes que han de adquirir al través de su lectura. En el caso de la 'Technology of Teaching' se prepararon objetivos conductuales que se insertaron en los materiales del texto. Se presentó primero el libro de Ferster y Perrott, 'Behavioral Principles', seguido por el de Skinner, 'Technology of Teaching'.

Lugar de estudio.

Un local para seminario, de aproximadamente cinco por siete y medio metros, bien iluminado y con aire acondicionado sirvió como área de estudio. Los sujetos se sentaban en sillas para escritorio

TESIS CON
FALLA DE ORIGEN

alrededor de una mesa grande en uno de cuyos extremos se hallaban los materiales, numerados, para el estudio. Se prohibía la entrada de personas ajenas a la sesión, indicándolo por medio de carteles colocados en las puertas respectivas.

Definición de las respuestas.

El tiempo dedicado al estudio se definió como aquel período en el que el sujeto permanecía sentado con el material frente a él y se daba por terminado cuando se levantaba de la silla. El observador anotaba el tiempo, el número de minutos, en los que cada estudiante durante las sesiones de estudio asumía la actitud indicada. Se consideraba como tiempo de estudio el que empleaban en pasear la mirada por el cuarto, contar las páginas, encender un cigarrillo, etc. Si un sujeto no acudía al cuarto de estudio se le anotaban cero minutos de labor. Los tiempos se redondeaban al medio minuto más cercano.

Observadores y área de observación.

Los observadores eran cuatro, todos hombres; dos estudiantes y dos miembros del cuerpo de profesores, los primeros ya graduados. Un solo observador estaba al cargo de cada sesión de tres horas.

El observador se situaba en un local adyacente de más o menos 3.30 por 2.50 metros que en uno de sus muros tenía una ventana de alrededor de un metro por

TESIS CON
FALLA DE ORIGEN

noventa cms, provista de un vidrio unidireccional. Un micrófono adosado a la pared del cuarto de estudio permitía oír el ruido que ahí se hacía.

Dos observadores sentados en el cuarto oscurecido anotaban en forma independiente, viendo al mismo reloj, los tiempos de estudio. Se situaban de manera que uno no viera lo que el otro escribía. De la confrontación que durante la décimaprimer sesión se hizo en cuanto a las distintas mediciones del tiempo que se habían hecho, resultó una concordancia del orden de un 92 a un 100%.

Todos los participantes fueron informados acerca de las condiciones en las que el curso se desarrollaría: horario para el uso del área de estudio (3 a 6 p. m. de lunes a jueves), el cambio de impresiones entre los asistentes debía ser mínimo y no traerían acompañantes. Los sujetos debían estudiar sólo los materiales del curso que ahí mismo se les proporcionaban y ni esos materiales ni las notas relativas a ellos podían llevarse fuera.

Se les informó también sobre la relación de los objetivos conductuales con el material del texto, así como acerca de la que habría entre los puntos acumulados en las pruebas y la calificación final.

TESIS CON
FALLA DE ORIGEN

Pruebas.

Las pruebas diarias se impartieron en las semanas primera, segunda, tercera, sexta y novena y las semanarias en las cuarta, quinta, séptima y octava. Fueron de tres tipos: verdadero-falso, de complementación y de opción múltiple. Las diarias valían cinco puntos y veinte las semanarias.

Resultados.

Los tiempos dedicados al estudio variaron en forma considerable, desde un promedio de 31 min. diarios hasta uno de 108. A partir de la décimasegunda sesión en la que se hizo la primera prueba semanal, modificándose así la periodicidad anterior, según la cual las pruebas eran diarias, se pudo notar, entre una y otra sesión, el aumento en la variación del tiempo de estudio. Las ausencias se hicieron más frecuentes en los primeros días de la semana, en la que las pruebas se impartían el último día de la misma. Cuando las pruebas volvieron a ser diarias la asistencia, de nueva cuenta, se uniformó. Resumiendo: El promedio de tiempo total de estudio fue ligeramente mayor en los periodos en los que las pruebas eran diarias que cuando se impartían cada semana.

TESIS CON
FALLA DE ORIGEN

Segundo experimento.

En este caso las pruebas se impartían cada tres semanas para estudiar los efectos de intervalos de interés más prolongados en los hábitos de estudio de los alumnos. Se hizo una comparación entre los resultados del primer experimento, pruebas diarias, y los del segundo, pruebas cada tres semanas, para detectar las posibles variaciones en los tiempos dedicados a la lectura.

Método.

Seis hombres y seis mujeres, cuyas edades fluctuaban entre los 19 y los 26 años, después de tomar un curso para subgraduados en psicología de la educación, se ofrecieron como voluntarios. El criterio para la selección y los demás requisitos fueron los mismos que para el primer experimento. Las pruebas se elaboraron antes del inicio de los trabajos y sólo se les hicieron, en el transcurso del mismo, algunos cambios menores.

Procedimiento.

Se pusieron a disposición de los alumnos los materiales de lectura correspondientes a un mínimo de dos semanas consecutivas. El material de las semanas anteriores no se retiró mientras duró el curso.



Durante las dos primeras semanas las pruebas fueron diarias y cada una tuvo un valor de cinco puntos. Después de tres semanas, en la sexta, se administró una más con valor de sesenta puntos. En la sexta y en la séptima semanas se volvió a las pruebas diarias. Tres semanas más tarde, al término del curso, se impartió la última.

Resultados.

Mientras las pruebas fueron diarias, el tiempo dedicado al estudio y la regularidad en la asistencia de los alumnos fueron similares a los observados en el primer experimento.

Al cambiar la periodicidad, al ser las pruebas administradas cada tres semanas, los tiempos empleados en la lectura variaron en forma considerable entre una y otra sesión. En casi todos los sujetos se observó, que al acercarse las fechas de los exámenes, en uno o en los dos períodos de las pruebas espaciadas cada tres semanas, un notorio aumento en el tiempo que dedicaban al estudio.

Resultados conjuntos de los dos experimentos.

Los dos experimentos demuestran una relación funcional entre la periodicidad de las pruebas y la regularidad en asistencia y tiempos dedicados al estudio.

TESIS CON
FALLA DE ORIGEN

Un intervalo corto entre las pruebas da motivo a una mayor regularidad en el estudio, así como a que se le dedique más tiempo.

Estos resultados conducen a un serio cuestionamiento sobre la efectividad de las pruebas administradas a intervalos largos, más de tres semanas a menudo, como medio idóneo para producir, sólidos y adecuados hábitos de estudio.

- e) La quinta investigación se debe a Jessica Gaynor y a Jim Milham de la Universidad de Houston. E. E. U. U., publicada en el Journal of Educational Psychology en 1976, Vol. 68, No. 3, Págs. 312-317.

Ha habido, relativamente, pocos estudios que demuestren la existencia de diferencias estadísticamente significativas entre los efectos de los distintos métodos de enseñanza que se aplican en las escuelas.

Por otra parte, sin embargo, han surgido hallazgos constantes en cuanto al efecto de los calendarios de medición en el rendimiento escolar -a una mayor frecuencia en las mediciones corresponde un mejor desempeño académico- (Dustin, 1971; Keller, 1968; Martin & Srikameswaran, 1974).

En esa situación, Dubin y Taveggia (1968) sugieren que las investigaciones en las que se comparan los



efectos de los distintos métodos de enseñanza sean abandonadas, afirmando que: "...no hay la menor evidencia que sirva de apoyo para preferir un método de enseñanza a otro en función de sus efectos en el desempeño de los estudiantes en los exámenes".

Si bien es cierto que las dimensiones particulares, observadas en el pasado, no han sido significativas en relación con el desempeño académico, eso no necesariamente quiere decir que no existan.

Desafortunadamente, la afirmación de la hipótesis de nulidad, con respecto a las variaciones tradicionales en la estructura de los cursos, seguirá siendo poco convincente en tanto no se encaren las deficiencias metodológicas del diseño experimental.

Entre las limitaciones metodológicas, las siguientes han sido primordiales: a) Inadecuado control de los efectos ocasionados por el instructor, b) Escaso control durante el proceso de interacción del instructor con el método de enseñanza, c) Falta de una medida consistente de la cantidad de información aprendida por los estudiantes, y d) Falta de evaluaciones por parte de los alumnos, tanto sobre el curso, como acerca de las medidas de rendimiento.

En este estudio se intenta subsanar esas deficiencias:

a) Utilizando el mismo instructor para todas las variantes del método de enseñanza, b) Administrando

TESIS CON
FALLA DE ORIGEN

a los alumnos de cada una de las diferentes condiciones de enseñanza los mismos reactivos de examen, tanto en número, como en contenido y, c) Estableciendo, por parte de los alumnos, una evaluación posterior al curso con el fin de determinar las relaciones entre las preferencias e intereses de los estudiantes y los procedimientos para la enseñanza y la evaluación a los que quedan expuestos.

Método.

Muestra: Participaron 400 estudiantes subgraduados en un curso denominado 'Introductory Psychology' en la Universidad de Houston. La mayoría de ellos estaba en el primero o segundo años de la Universidad. Todos los alumnos del curso participaron en el estudio.

Enseñanza. El objetivo principal del curso consistía en que los alumnos se familiarizaran con el método científico y la manera como se aplica al estudio de la psicología. Se hizo énfasis en la enseñanza de los siguientes aspectos: a) Comprensión de los problemas básicos de la psicología, b) Las habilidades necesarias para observar los fenómenos psicológicos, c) Comprensión de la naturaleza cuantitativa de los datos psicológicos y, d) Reconocimiento de la causación múltiple.

El interés del curso se centró más en el procedimiento que en los contenidos; ésto es, el estudiante no era

simplemente informado sobre las numerosas áreas de estudio en el campo de la psicología, se le pedía, más bien, que cuestionara y evaluara críticamente los métodos usuales, por medio de los cuales los psicólogos hacen frente e investigan los principales problemas de la psicología.

El libro de texto de Dember y Jenkins (1971) y el cuaderno de trabajo de Kammann (1971), al igual que las lecciones y métodos específicos de enseñanza, fueron los medios elegidos para facilitar el cumplimiento de los objetivos.

Condiciones de enseñanza. En cada una de las catorce semanas del semestre hubo dos reuniones. En primer lugar, todos los estudiantes recibían una clase de una hora y media por semana. Las clases eran impartidas por un instructor y una instructora. Para controlar el efecto del sexo del instructor, cada uno daba clase durante el mismo tiempo, rotándose semanariamente. Adicionalmente, todos los estudiantes eran asignados a un segundo grupo, más pequeño, que también se reunía cada semana. Tres de esos grupos participaron en un Laboratorio en el que hacían experimentos en psicología (técnica lección-laboratorio), dos recibían una segunda clase semanal (técnica lección-lección) y los dos restantes asistían a las discusiones de profesores y estudiantes graduados sobre investigaciones en proceso

(técnica lección-investigación-discusión) acerca de temas relevantes en relación con los temas tratados esa semana.

Pruebas. Los exámenes del curso estaban formados con reactivos de opción múltiple y cubrían todo el material presentado en las lecciones semanarias a las que asistían todos los alumnos, así como el contenido del libro de texto.

Los instructores se esforzaron para lograr congruencia entre las cuatro metas que se habían fijado y los reactivos de la prueba.

Dos calendarios de exámenes fueron asignados aleatoriamente a los dos grupos de cada uno de los métodos de enseñanza. Uno consistía en pruebas semanarias que cubrían el material presentado en cada capítulo del libro de texto. Cada semana, la mitad de los estudiantes de los seis grupos recibió una prueba con quince reactivos de opción múltiple.

El otro, aplicado a la mitad restante de los alumnos, consistía en la administración de pruebas con noventa reactivos de opción múltiple a la mitad y al fin del curso.

Los reactivos fueron los mismos para todos los alumnos independientemente de la periodicidad de las pruebas.

Se elaboró también un instrumento de 16 reactivos para la evaluación del curso, en él se pedía a los estudiantes que expresaran sus puntos de vista sobre lo que habían aprendido en el curso y si les había gustado o no.

Resultados.

Para procesar estadísticamente los datos se usó el análisis de varianza y los resultados de este estudio pueden resumirse en la siguiente forma:

- a) Los estudiantes a quienes se aplicó la técnica lección-laboratorio obtuvieron puntuaciones significativamente más altas en los exámenes del curso que aquellos en los que se empleó la de lección-investigación-discusión.
- b) Con la técnica lección-lección se lograron mejores resultados que con la de lección-investigación-discusión.
- c) Los estudiantes a quienes les fueron administradas pruebas semanarias obtuvieron puntuaciones significativamente mejores en los exámenes que aquéllos que sólo las recibieron a mitad y fin del curso.
- d) Los estudiantes que recibieron las pruebas semanarias mostraron mayor aprecio por su calendario de exámenes que aquéllos a quienes se les administraron a la mitad y al fin del curso.
- e) Los alumnos que tuvieron exámenes semanarios

manifestaron haber aprendido más que quienes los recibieron a mitad y fin del curso.

De todo lo anterior parece desprenderse que el rendimiento escolar es significativamente más alto, en función: ya de escuchar una ulterior elaboración y clarificación de materiales directamente relacionados con los anteriormente presentados (lección-lección); ya de una participación activa en la aplicación de los principios de la psicología (lección-laboratorio); pero no es significativamente superior en función de, simplemente, escuchar las descripciones de la aplicación de los principios de la psicología (lección-investigación-discusión).

El hallazgo de que la mayor frecuencia en las pruebas da lugar a puntuaciones significativamente superiores, apoya la hipótesis de que una práctica mejor distribuida (tests semanarios) conduce a resultados más productivos y permite asimismo, suponer que actúa como un fuerte incentivo.

No obstante, es necesario actuar con cautela al evaluar las técnicas de enseñanza y evaluación para no confundir el aprecio que los estudiantes hayan manifestado por el curso, con el grado real de aprendizaje que hayan alcanzado.

f) La sexta investigación fue realizada por Robert E. Salvin de la Universidad Johns Hopkins de

Baltimore, E. E. U. U., publicada en el reporte No. 288 de noviembre de 1979.

Tomando en cuenta que la asignación de calificaciones a los alumnos es uno de los problemas más persistentes e inquietantes a los que todo sistema educativo se enfrenta, el autor se propone probar una técnica que permita eliminar los que considera los principales inconvenientes de la que actualmente se acostumbra aplicar.

Esos inconvenientes surgen al no aplicarse ciertos principios de la psicología motivacional, como lo son: que la realimentación y el refuerzo deben seguir al comportamiento en un tiempo próximo y que las conductas que serán recompensadas deben quedar claramente establecidas y especificadas para que los alumnos sepan lo que se espera de ellos.

Al no hacerse así, al dejar que entre la fecha del examen y aquella en la que los alumnos conocen sus resultados transcurra un tiempo considerable y al asignarse las calificaciones, más que con fundamento en el aprendizaje efectivo, en criterios basados en asistencia, puntualidad, limpieza, docilidad, etc., se contravienen los principios indicados.

Una solución al problema puede consistir en diseñar un sistema en el que los alumnos reciban, regularmente, realimentación y recompensas basadas en la comparación de su rendimiento semanal actual, con el que hubieran tenido con antelación.

De ser eso posible, pues es necesario considerar que las clases se impartirán a un grupo y no individualmente y que el curso deberá desarrollarse a un ritmo normal, podríamos contar con una forma simple, pero efectiva para que todos los alumnos tuvieran la misma oportunidad de alcanzar el éxito académico sin importar cuáles hubieran sido sus habilidades de entrada.

Con ese fin, se aplicó una técnica llamada 'Individual Learning Expectations', ILE, o sea Expectativas individuales de aprendizaje. Consiste, en principio, en que a partir de una calificación base -que para cada alumno se establece por medio de un pretest de 50 reactivos- de haber mejoría en las pruebas subsecuentes, se le recompensa otorgándole puntos extra.

Sujetos.

Fueron 409 alumnos del sexto, séptimo y octavo grados de cuatro escuelas sub-urbanas del estado de Maryland. En tres de ellas todos los estudiantes eran blancos y en la cuarta, la

tercera parte eran negros y las dos restantes blancos.

La muestra total se formó con un 92% de blancos, con hombres y mujeres por mitad. Las clases eran de inglés, en número de 17 y con 11 profesores.

Diseño.

Para el experimento se usó un diseño de dos grupos: el experimental y el de control. El primero tenía 196 alumnos de ocho clases y el segundo, el de control, 213 de nueve clases.

En los casos en los que fue posible, un profesor enseñó, tanto a un grupo experimental, como a uno de control y seleccionó aleatoriamente cuál recibiría uno u otro tratamiento. Cuatro profesores dieron clase a uno o dos grupos experimentales y a uno de control. Los otros siete se encargaron de un solo grupo y fueron asignados al azar a situaciones experimentales o de control. Resultó así que el efecto producido por los profesores fue sólo parcialmente controlado.

Todos los participantes eran alumnos de un curso de nueve semanas para el estudio de la lengua inglesa. En cada semana se impartían dos unidades sobre los siguientes temas: gramática, ortografía, sintaxis, etc.

Las clases a los dos grupos, el experimental y el de control se regían por el mismo calendario y sus materiales de estudio eran iguales.

La semana se dividía en dos ciclos de dos días cada uno, en el primero se impartía la clase y en el segundo se hacían ejercicios sobre los temas tratados. En el segundo día del segundo ciclo, se administraban a los alumnos pruebas de 25 reactivos. El quinto día, casi siempre el miércoles, se destinaba a otro tipo de actividades.

Tratamiento.

La única diferencia en el tratamiento que se dió a uno y otro grupo, el experimental y el de control, radicó en la forma en la que se manejaron las calificaciones de los respectivos alumnos. A los de los grupos experimentales se les decía que su tarea consistía en mejorar las 'calificaciones base' que habían obtenido y se les recordaba cuáles eran; se les informaba, asimismo, tanto sobre la mejoría lograda, como acerca del número de puntos extra que se les otorgarían.

Para los alumnos de los grupos de control se usaron los mismos calendarios y materiales de

estudio, pero no tuvieron calificación base, ni puntos adicionales; sólo se les asignó el número neto de puntos que obtuvieron en las pruebas.

Todos, sin embargo, recibieron cada seis semanas, las calificaciones netas correspondientes a las pruebas, sin puntos adicionales, de manera que el sistema de recompensas usado para el grupo experimental era complementario y no sustitutivo del tradicional.

Instrumentos de medición.

A todos los estudiantes se les administraron formas paralelas del 'Hoyum-Sanders High School English Test', tanto para el pretest como para las otras pruebas.

El Hoyum-Sanders es un instrumento estandarizado para medir el conocimiento del lenguaje y abarca temas acerca de: gramática, ortografía y sintaxis.

Resultados y discusión.

Los resultados muestran que los grupos experimentales obtuvieron un rendimiento significativamente mejor que los de control ($F=6.08$, $g.l.=1,406$, $p < 0.02$).

Los grupos experimentales tuvieron en el pretest una media de puntuación inferior a la de los de control, 69.24 contra 70.90, pero al final del curso los superaron, 77.15 por 75.79. Sin embargo, ni el tratamiento para determinar la interacción con el

pretest, ni la prueba para una tendencia cuadrática ($F < 1$, g. l. = 1,404) se acercaron a tener alguna significatividad estadística, indicando que la mejoría era independiente del rendimiento anterior.

El efecto significativo sobre el rendimiento escolar, indica que el simple cambio en la forma en la que las calificaciones se asignan y el que los alumnos reciban información clara y oportuna, pueden dar cuenta de los logros observados.

Por otra parte, el efecto Hawthorne difícilmente explica los resultados, ya que tanto los grupos experimentales como los de control estuvieron sujetos a nuevos calendarios y programas. Es posible que la información sobre puntuaciones base y de recompensa, recibida únicamente por el grupo experimental, haya producido los resultados observados. En nuevas investigaciones se tratará de eliminar o controlar esa variable.

- g) La séptima investigación se debe a Thomas J. Belzer en el Pasadena City College, presentado a la Universidad Nova como requisito parcial para obtener el grado de 'Doctor of Education'. El estudio se publicó en forma de informe en Dic. de 1974, 32 páginas.

Planteamiento del problema.

El problema de este estudio consiste en comparar

TESIS CON
FALLA DE ORIGEN

los efectos de dos métodos de evaluación usados en el curso de Biología General II en el Pasadena City College y determinar sus relaciones con las calificaciones finales y con la deserción escolar.

Hipótesis.

La hipótesis de nulidad se basó en el supuesto de que las diferentes técnicas de evaluación aplicadas en las clases de biología general no tenían efectos ni en las calificaciones finales ni en la deserción.

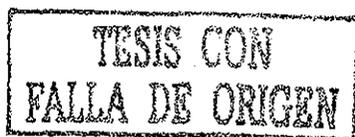
Antecedentes.

Tanto el bajo rendimiento de los alumnos, medido por las calificaciones finales, como las altas tasas de deserción, han sido asuntos de especial interés, en años recientes, para la institución.

Las evaluaciones se realizaban por medio de cuatro exámenes departamentales distribuidos en el semestre y uno al final del mismo.

La relación entre las puntuaciones y las calificaciones en letras era el siguiente:

90 a 100%	A
80 a 89%	B
66 a 79%	C
55 a 65%	D
menor a 55%	E



Este sistema continúa en uso y se aplica al nuevo procedimiento que modifica al anterior en cuanto al tipo y periodicidad de las pruebas que ahora son once, cada una con diez reactivos, y que son administradas al finalizar cada unidad de estudio. A la mitad y al fin del semestre se hacen también pruebas.

Sobre este cambio y sus efectos aparentes en las calificaciones y deserción versa el estudio en cuestión.

El curso de biología general se dividió en once temas principales: historia de la biología, química de la vida, la célula, nutrición, transporte e intercambio de gases, regulación, reproducción, genética, desarrollo, evolución y ecología.

Tres o cuatro lecciones de una hora se impartieron sobre cada tema y para cada unidad se administraron pruebas de diez reactivos a todos los alumnos. La prueba tenía un valor de diez puntos. Los exámenes a la mitad y al final del curso también se administraron.

Restricciones del estudio.

Se usarán únicamente las calificaciones finales expresadas en letras para indicar el rendimiento de los alumnos.

TESIS CON
FALLA DE ORIGEN

Durante el período de estudio del curso no se hizo ningún intento para determinar las razones de la deserción. Tampoco se hizo intento alguno de comparar el rendimiento de los estudiantes con el de los alumnos de otros grupos a cargo de instructores distintos.

No se analizaron los reactivos de las pruebas durante el período de estudio.

Supuestos básicos.

Los tres grupos estudiados eran razonablemente homogéneos. A todos los estudiantes inscritos en los tres grupos se les brindaron iguales oportunidades para aprender, pues el profesor era el mismo.

En cada uno de los tres grupos estudiados la inteligencia estuvo distribuida en forma normal.

Los reactivos seleccionados para las pruebas eran razonablemente comparables en su grado de dificultad.

Datos de tratamiento estadístico.

Se estudiaron tres grupos del mismo número de semestres previos y se establecieron comparaciones entre la distribución de las calificaciones y las tasas de deserción. Se tomaron en cuenta sólo las notas A, B, C y D, pues no se otorgó ninguna F y únicamente dos E fueron dadas. (Se otorga una E cuando un alumno no cumple satisfactoriamente con su trabajo por razones fuera de su control). Para la

TESIS CON
FALLA DE ORIGEN

comparación de las calificaciones finales con los grados de deserción se usó, como prueba estadística, la 'ji cuadrada'.

Significatividad de los datos.

La 'ji cuadrada' para la comparación de la deserción entre los semestres del otoño de 1972 y 1973 y el semestre de primavera de 1974 fue igual a 5.35, con dos grados de libertad y cerca de un 5% como nivel de significatividad ($\chi^2-5\%=5.99$). Como el resultado obtenido estuvo tan cerca de la significatividad, se promediaron los semestres de 1972 y 1973 y se hizo otro análisis de la 'ji cuadrada', que en esta ocasión fue igual a 5.56, con un grado de libertad, lo cual es significativo a un nivel del 2%.

Por lo tanto se rechaza la hipótesis de nulidad, consistente en que no había diferencias significativas en las tasas de deserción en el curso de biología al modificarse las técnicas de evaluación, hacerlas con mayor frecuencia y ser las pruebas menos extensas.

El porcentaje de deserción bajó de un 26% como promedio en los semestres de 1972 y 1973 a aproximadamente un 12% en el de la primavera de 1974.

La 'ji cuadrada' para la comparación de las calificaciones finales, expresada en letras, fue



igual a 15.57, que para seis grados de libertad es significativo a un nivel del 2%; de manera que también se rechaza la hipótesis de nulidad en lo que toca a las calificaciones.

Los resultados indican que la deserción de los estudiantes puede ser significativamente reducida cuando las evaluaciones se realizan sobre pequeños segmentos del material del curso y que ésto también afecta en forma significativa a las calificaciones, expresadas en letras.

Se recomienda ampliar la investigación a más grupos de la misma materia, así como a otras para contar con mayor información y poder determinar las razones de la deserción escolar.

- h) La octava investigación fue hecha por James A. Wiggins y Hallowell Pope de la Universidad de North Carolina y por Donald Bushell Jr. de la Universidad de Kansas. Fue publicada en forma de reporte final en junio de 1968.

El propósito del estudio fue el de explorar los efectos de varios procedimientos prácticos de instrucción en el rendimiento académico.

A seis secciones de un curso de Introducción a la Sociología se les asignaron varias situaciones experimentales que comprendieron pruebas e incentivos de diversos tipos. Los exámenes fueron

TESIS CON
FALLA DE ORIGEN

administrados a intervalos durante el curso y uno al final. Los incentivos incluyeron calificaciones, doble crédito a los mejores, seminarios de honor, exención de asistencia a clases y del examen final.

Los mayores problemas metodológicos que se confrontaron fueron: variaciones en las conductas de los instructores, variación intragrupal en el desempeño de los alumnos, exámenes comparativos, recolección ineficiente de los datos sobre los tiempos de estudio, la brevedad de los períodos para manipular eficientemente las condiciones experimentales.

En términos generales los resultados fueron los siguientes:

- 1) Dos de los cuatro incentivos fueron efectivos para modificar sistemáticamente el desempeño en los exámenes y el tiempo dedicado al estudio,
- 2) La mayor parte de la información adquirida en el curso fue olvidada quince semanas después de su terminación, independientemente de las condiciones experimentales,
- 3) Ni la frecuencia, ni el calendario de pruebas tuvo efecto sistemático en el rendimiento académico y,
- 4) Una relación baja, pero significativa pudo establecerse entre el tiempo de estudio y el rendimiento de los exámenes.

TESIS CON
FALLA DE ORIGEN

Breve síntesis de las investigaciones reseñadas.

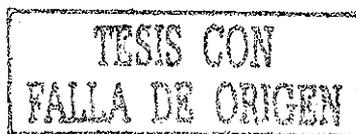
1. En la primera, los autores tratan de establecer alguna correlación entre el número de pruebas que un alumno considera necesario presentar para alcanzar el dominio pleno de la materia y la calificación que obtiene en el examen final.

La correlación fue negativa: $r=-0.29$ ($n=23$, $p=0.192$), de manera que la hipótesis de nulidad es la que recibe el apoyo empírico. No existe correlación significativa entre el número de pruebas presentadas y la calificación final.

2. Para la segunda, se manejaron en forma conjunta dos variables: el dominio pleno, que consistió en la necesidad de obtener un mínimo de ocho puntos sobre 10 para que el resultado fuera considerado aprobatorio y la frecuencia con la que las pruebas se administraban.

La mediana de las puntuaciones para el grupo experimental, dominio pleno y pruebas semanales, fue superior a la del grupo de control. No se hizo, sin embargo, algún intento para determinar el efecto de cada una de las variables.

3. En el tercero, la única variable experimental fue la periodicidad de las pruebas y los resultados apoyaron la hipótesis de que una mayor frecuencia en la administración de las pruebas redundaba en un mejor rendimiento escolar.



4. El cuarto estudio consta de dos experimentos por medio de los cuales se trata de establecer una correlación entre la periodicidad de las pruebas y el tiempo que los alumnos dedican al estudio. Los resultados mostraron que a una mayor frecuencia en la administración de los instrumentos de medición, corresponde una mayor regularidad y tiempo empleado en la lectura del material del curso. No se trató, empero, de determinar la correlación entre los calendarios para la administración de las pruebas y el rendimiento académico.

5. En la quinta investigación, se intenta correlacionar, tanto los efectos de las diferentes técnicas de enseñanza, como los de una mayor frecuencia en la administración de las pruebas, en el rendimiento escolar.

Los resultados dan apoyo a la suposición de que ciertas técnicas de enseñanza-aprendizaje afectan de manera positiva el rendimiento académico en mayor medida que otras. Y lo dan también a la hipótesis de que a una mayor frecuencia en la administración de las pruebas corresponde un mejor rendimiento académico.

6. En el sexto estudio, a todos los alumnos se les administraron pruebas con la misma periodicidad.

TESIS CON
FALLA DE ORIGEN

La diferencia entre el tratamiento que se dió al grupo experimental y al de control, consistió en que al primero se le fijó una calificación 'base', de acuerdo con el resultado que había obtenido en el pretest y se le otorgaban puntos de recompensa si en las pruebas subsiguientes mejoraba su desempeño inicial; se le informaba, asimismo, periódicamente acerca del grado en el que iba consiguiendo esa mejoría. Al segundo, al de control, no se le fijó calificación 'base', ni se le dieron puntos adicionales, sólo la calificación neta de cada prueba.

El grupo experimental obtuvo un rendimiento académico significativamente mejor que el de control, debido muy probablemente al efecto de las recompensas y al del conocimiento de que su desempeño iba o no mejorando, en relación al que inicialmente había tenido en el pretest.

7. En este trabajo, el séptimo en el orden de reseña, se plantea como hipótesis de nulidad que no existe relación alguna entre las diferentes técnicas de evaluación y el rendimiento escolar o tasas de deserción.

Los resultados no apoyaron la hipótesis nula, por el contrario, pudo establecerse que al hacerse las evaluaciones con mayor frecuencia y



sobre menores cantidades de material estudiado, mejoran las calificaciones y decrecen las tasas de deserción, ambas de manera significativa.

8. En esta investigación, la octava, las variables manejadas incluyeron, tanto incentivos que supuestamente afectarían en forma positiva al rendimiento, como distintos calendarios para las pruebas.

Los resultados fueron, en términos generales, los que siguen:

- a. De los cuatro incentivos usados, dos fueron efectivos para mejorar sistemática y significativamente el rendimiento académico.
- b. Ni la frecuencia, ni el calendario de pruebas tuvieron algún efecto en ese sentido.

Los resultados de las investigaciones, brevemente reseñadas, no pueden considerarse como concluyentes, pues no todos dan apoyo a la hipótesis de que a una mayor frecuencia en la pruebas corresponde un mejor rendimiento académico, ni a la de nulidad respectiva. Así, mientras la tercera y la séptima parecen apoyar la hipótesis central, la octava más bien la contradice. Creo, en consecuencia, que vale la pena tratar de encontrar un mayor apoyo empírico a alguna de esas hipótesis que es lo que me propongo con el experimento al que enseguida me referiré.

3.3. Planteamiento de la hipótesis.

Una vez que se ha observado algún fenómeno que suscite preguntas o implique problemas y se inicie la búsqueda de constataciones y/o soluciones, el método científico ha dado muestras de ser el más efectivo para hallarlas.

Ante el problema o la pregunta debidamente planteados, esto es en forma precisa y no vaga, el primer paso consiste en proceder, como lo hemos ya hecho, a examinar los antecedentes pertinentes para saber si se han encontrado respuestas o soluciones satisfactorias. Si ese examen no rinde el fruto deseado, si no hallamos la explicación del fenómeno observado, formulamos alguna hipótesis.

Entendemos como hipótesis una proposición general, que puede ser una teoría tentativa o una suposición provisionalmente adoptada para explicar los hechos detectados.

Por su grado de generalidad, las hipótesis pueden ser universales o existenciales. Las primeras son comprobables en cualquier tiempo y lugar, en tanto que las segundas sólo lo son en un tiempo y lugar específicos.

Las hipótesis que con mayor frecuencia se establecen en las ciencias sociales son las existenciales y pueden

ser descriptivas o causales, Como su nombre lo indica, las primeras se limitan a la descripción del fenómeno, sin tratar de encontrar aquello que lo pudiera originar; las causales en cambio, contienen una posible explicación de los hechos observados, su probable causa.

Las hipótesis causales no son, sin embargo, una mera inferencia inductiva de los hechos, pues a menudo, para formularlas es necesario poner en juego una considerable dosis de inventiva, imaginación e inclusive intuición.

La historia de la ciencia nos ofrece múltiples ejemplos de cómo una serie de hechos, por muchos observados, permanece sin explicación o ésta es deficiente y quien por fin halla la verdadera lo consigue por hacer uso de las cualidades aludidas.

Por otra parte y si bien es cierto que las hipótesis que formulemos no deben contradecir a las teorías probadas y, por supuesto tampoco a los principios básicos de la lógica, ello no implica que en ciertos casos y circunstancias no podamos poner, a las primeras, en tela de juicio cuando los resultados del proceso experimental debidamente realizado no se conformen con ellas.

Quizá uno de los mejores ejemplos al respecto lo sea la explicación que dió Albert Einstein a la contradicción entre la teoría clásica y la observación resultante del

experimento de Michelson-Morley, en el que la velocidad de la luz no fue mayor en el sentido del movimiento del cuerpo emisor que en el transversal.

Se contravenía así un principio de la mecánica clásica: el de la suma de las velocidades. Ese principio lo podemos comprobar en la vida diaria cuando desde un vehículo en marcha, digamos a 30 m/seg = a 108 Km/hora arrojamos un objeto hacia delante. A la velocidad que seamos capaces de imprimirle aumentaremos la del propio vehículo para saber a cuál saldrá despedido.

Al no verificarse ese principio, Einstein no desechó el experimento sino que aplicó una de las ecuaciones conocidas como 'Transformaciones de Lorentz', que explican lo observado, ya que:

$$V = \frac{v_1 + v_2}{1 + \frac{v_1 + v_2}{c^2}}$$

En la que:

V = velocidad total

v_1 y v_2 = velocidades parciales

c = velocidad de la luz

Debido a que en los fenómenos que podemos observar en la vida diaria las velocidades son mínimas en comparación con la de la luz, el valor del segundo término del denominador es, prácticamente nulo y se puede así comprobar el principio de la suma de velocidades.

No obstante, cuando las velocidades son iguales o cercanas a la de la luz, el valor de ese segundo término del denominador se aproxima a la unidad y el total del denominador será casi de dos, de manera que si en el numerador tuviéramos el doble de la velocidad de la luz, al dividirlo entre dos, tendríamos de nueva cuenta el valor inicial de la velocidad de la luz, 300,000 Km/seg, en números redondos.

Y fue así como pudo explicarse la aparente paradoja resultante del experimento de Michelson-Morley, lográndose también dar apoyo empírico a la concepción einsteniana de que la velocidad de la luz es una constante que no aumenta ni disminuye cualquiera que sea la del cuerpo emisor.

No obstante que las teorías de las ciencias factuales pueden ponerse en tela de juicio, pues son producto de la inducción, los postulados de las ciencias formales, deductivas, son incontrovertibles.

Cualquier hipótesis se invalida si se violan los principios de la lógica y resulta también totalmente

vano formular y tratar de probar alguna cuya imposibilidad haya sido matemáticamente demostrada.

En el marco de referencia de la geometría euclídeana, tales son los casos de la cuadratura del círculo y la trisección del ángulo usando únicamente los instrumentos que esa geometría permite: regla no graduada y compás.

La hipótesis, formulada con toda la posible libertad, imaginación, inventiva y si se quiere intuición, se constituye como guía indispensable, como hilo conductor en la realización de las sucesivas etapas que prescribe el método científico.

Sin ella, explícitamente formulada o sólo implícita en la búsqueda de fenómenos observables, el investigador carece de un criterio que le permita discriminar entre los que pudieran tener alguna atinencia y los que no la posean en relación con lo que se haya propuesto indagar.

En consecuencia, con lo hasta ahora expresado y con motivo del caso específico que nos ocupa, estamos ya en aptitud de formular una hipótesis, que como arriba indiqué, será la que nos guíe en las siguientes etapas de la investigación.

Acatando los postulados del método científico, el primer paso que dimos consistió en la observación del fenómeno, en la constatación de que, cuando menos en apariencia, las mediciones del R-E realizadas para cada unidad

programática del curso: especificaciones, costos unitarios, cantidades de obra, estimados de costo y calendarios de obra, surten algún efecto en dicho R-E al medirlo al final del ciclo lectivo.

A continuación procedimos a examinar los resultados de otras investigaciones realizadas con un propósito similar al que perseguimos: conocer el efecto de las pruebas periódicas en el R-E. Al respecto y como quedó ya indicado (inciso 3.2.) son sólo ocho de los 350 informes rescatados de los bancos de datos los que versan sobre estudios, más o menos, directamente relacionados con el tema específico de nuestro interés.

Esos resultados no pueden considerarse, de ninguna manera, concluyentes y por ese motivo juzgamos pertinente emprender una nueva investigación sobre el mismo asunto.

El paso siguiente se refiere a la formulación de alguna hipótesis, para nosotros causal, sobre la posible explicación del fenómeno, que sería la que sigue: si aplicamos mediciones periódicas correspondientes a cada unidad programática del curso de A.P.O. II durante el proceso E-A; entonces obtendremos mayor rendimiento escolar en los alumnos de dicha asignatura.

De esta hipótesis, la principal consecuencia

contrastable que podemos derivar es que el R-E, en el examen final, de los alumnos a quienes se administren pruebas periódicas será mejor que el de aquéllos a quienes no se les impartan.

Para comprobar la veracidad de lo expresado y con el fin de que las inferencias a las que lleguemos posean un aceptable grado de validez, es necesario que para la realización del experimento se adopte un diseño que nos otorgue un cierto grado de garantía en cuanto a la posibilidad de controlar los factores que pudieran atentar contra dicha característica (inciso 4.2.).

Para ello deberá pensarse en una opción metodológica que nos permita comparar dos grupos. A uno de ellos se le impartirían mediciones periódicas y al otro sólo la final.

Antes de proceder a la selección del diseño es necesario que hagamos algunas precisiones acerca de cuáles son las variables que nos pueden afectar y de la manera en la que sus efectos podrían sujetarse a control.

3.4. Variables.

Las variables, los rasgos, características y atributos que distinguen a los objetos o sujetos miembros de una comunidad, pueden clasificarse, si se toma en cuenta el papel que juegan en un estudio experimental, en: independientes, intervinientes y dependientes.

Las primeras, las independientes, son aquellas que en una relación causa y efecto se pueden considerar como las causas.

Las segundas, las intervinientes, si bien en sentido estricto son causas como condiciones necesarias, debido a que pueden convertirse en suficientes para producir el fenómeno observado, es conveniente distinguir su acción de la que ejercen las propiamente independientes por sí solas con el propósito de dilucidar los efectos de cada una de ellas en los resultados obtenidos.

Las últimas, las dependientes, son los posibles efectos en una relación causal.

En el caso que nos ocupa, la variable independiente está constituida por las mediciones periódicas del rendimiento escolar. Para nosotros en una variable activa y dicotómica con valores de 1 y 0, según se administren o no las pruebas periódicas a los grupos experimentales y de control respectivamente.

De las variables intervinientes nos ocuparemos en el inciso siguiente.

La variable dependiente será, en este estudio, el rendimiento escolar. Lo entenderemos como el grado de dominio, medido con una escala de intervalos que va de 0 a 10, que un alumno alcanza para la totalidad de las unidades programáticas del curso de A.P.O. II. Para hacer la medición indicada se usará como instrumento la prueba final.

3.5. VARIABLES EXTRAÑAS.

En la investigación a la que me estoy refiriendo, hemos considerado como variables extrañas o exógenas a todas las intervinientes que de alguna manera pudieran afectar los valores que asuma la variable dependiente.

Las concebimos, según propuse en el inciso anterior, como intervinientes porque ejercen su efecto en forma conjunta con las variables propiamente dichas independientes.

Para efectos de nuestro estudio y con el fin de examinarlas con mayor cuidado, cabe la posibilidad de agruparlas en la forma siguiente:

- a) Variables del medio físico.
- b) Variables biológicas.
- c) Variables psicológicas.
- d) Variables socioculturales.
- e) Variables pedagógicas.

a) Variables del medio físico.

Son todas aquéllas, que como su nombre lo indica, conforman el medio físico en el que se produce el fenómeno.

Entre las que mayor efecto pueden surtir se encuentran: el espacio arquitectónico, abierto o confinado. En este caso las dimensiones, forma y materiales del aula respectiva. De estos últimos, su

TESIS CON
FALLA DE ORIGEN

textura, color y aspecto en general.

El clima y grado de confort que experimentan las personas, la iluminación, natural o artificial, las condiciones acústicas y grado de visibilidad hacia los puntos de interés; el mobiliario y su adecuación para las actividades a las que esté destinado, así como las demás características del espacio arquitectónico, son variables que pueden afectar a la variable dependiente.

Del medio físico particular en el que se llevó al cabo el experimento, someramente descrito en el apartado dedicado al planteamiento del problema, sólo cuento con los datos que pude observar a simple vista, pues no conté con las facilidades necesarias para hacer las mediciones que hubieran sido las pertinentes.

b) Variables biológicas.

De las variables biológicas, de los atributos y funciones somáticas de los seres humanos sujetos al experimento en cuestión, sólo citaré las que en mayor medida y grado de probabilidad podrían afectar a nuestra variable dependiente, el rendimiento escolar.

La edad y el sexo tendrían que ser tomados en cuenta. El nivel general de salud, que entre otros, puede surtir algún efecto en los grados de atención

TESIS CON
FALLA DE ORIGEN

y concentración de alumnos y profesores, así como las capacidades visual y auditiva de los miembros del grupo, deben también ser considerados.

No me extiendo más en el examen de estas variables, ni incluyo otras posibles de tipo biológico, porque acerca de ellas tampoco estuve en aptitud de recabar datos ciertos, de modo que mi conocimiento de ellas se reduce a lo que pude observar durante las clases.

c) Variables psicológicas.

Un tercer aspecto a considerar sería el de las variables de tipo psicológico y de ellas, la que quizá tenga una mayor relevancia para nuestro estudio, es la concerniente a las actitudes que asuman los participantes de los actos educativos.

Entendemos por actitud "...un estado mental y neurofisiológico constituido por la experiencia que ejerce una influencia dinámica en el individuo, preparándolo para reaccionar de una manera especial a un cierto número de objetos y situaciones...". (91) Otros autores consideran que una actitud consiste de tres componentes: "...a) cognoscitivo, b) emocional y c) tendencia a la acción". (92)

Por la importancia de esta variable, tanto por sí misma, como porque al interactuar con el medio físico conforma el medio ambiente que prevalece en

91. ALLPORT, G. Personality a Psychological Interpretation, en M. Grawitz, Métodos y técnicas en las ciencias sociales, 15

92. SUMMERS, GENE. Medición de actitudes, 14



el salón de clase, hubiera sido, a todas luces, conveniente conocer sus valores, medirla aplicando alguno de los procedimientos que indican L. L. Thrustone, S. Cook y C. Sellitz o D. T. Campbell y D. W. Fiske en especial.

Desafortunadamente no pude contar con el auxilio de la persona capaz de hacer dicha medición, de manera que mi conocimiento de las actitudes de los alumnos se reduce a lo que observé durante las sesiones: un aceptable grado de atención e interés en los temas que se exponían.

d) Variables socioculturales.

La formación social y cultural que hayan tenido las personas a lo largo de sus vidas, además de la influencia que para nuestro caso, pudieran tener en la actitud de los estudiantes y profesores, afectará también a lo que R. Gagné llama capacidades intelectuales y estrategias cognoscitivas. No obstante la importancia primordial que el medio familiar y social tiene en el proceso de adquisición y valoración de los conocimientos y en el desarrollo de actitudes hacia las ciencias y las artes, así como en la formación de una escala de valores de todo tipo en los individuos, los años de escolaridad necesarios para llegar al quinto semestre de alguna licenciatura en la UNAM tienden a mitigar esos efectos iniciales. No quiero decir con esto que

TESIS CON
FALLA DE ORIGEN

desaparezcan, ni que no hubiera sido conveniente conocer los antecedentes socioculturales de los sujetos de nuestro experimento, sino solamente que en el nivel académico del mismo, sin soslayar la importancia y efecto diferencial de esas variables, ni olvidarnos de su control, probablemente hayan cedido en jerarquía ante las que a continuación veremos.

e) Variables pedagógicas.

Entre las variables que en este estudio llamamos pedagógicas hemos incluido algunas de las que Astin y Panos consideran como de entrada. Así por ejemplo: los hábitos de estudio de los alumnos, sus habilidades intelectuales, estrategias cognoscitivas, antecedentes académicos, etcétera.

Hemos también incluido algunas de las variables propias de las operaciones, como pueden serlo: el plan de estudios, los programas que de él se derivan, los objetivos, los contenidos, las técnicas de enseñanza aprendizaje, las actividades de aprendizaje y los medios auxiliares para la enseñanza.

Las variables de entrada anotadas son extrañas a nuestro experimento y las relativas a las operaciones lo son en la medida en la que, pudiendo hacerlo, decidimos no incorporarlas como independientes para medir sus efectos.

TESIS CON
FALLA DE ORIGEN

De las primeras no fue posible investigar sus valores y a las segundas les asignamos, como quedó indicado, el carácter de extrañas con valores desconocidos.

Además de las variables someramente examinadas, otros factores, que Campbell y Stanley nos señalan, pueden afectar la validez de las inferencias que hagamos a partir de los resultados del experimento.

La validez interna nos permite afirmar con el debido fundamento que el tratamiento experimental fue la causa de las diferencias observadas al aplicarlo.

Puede verse disminuida por los efectos de:

1. Historia. Los acontecimientos específicos ocurridos entre la primera y la segunda medición, aunados a los de la variable experimental.
2. Maduración. Los procesos internos de los participantes que operan como resultado del paso del tiempo y que incluyen el aumento de edad, el hambre, el cansancio, etc.
3. Administración de tests. El influjo que la administración de un test ejerce sobre los resultados de otro posterior.
4. Instrumentación. Los cambios en los instrumentos de medición o en los observadores o calificadores participantes que pueden producir variaciones en las mediciones que se obtengan.

TESIS CON
FALLA DE ORIGEN

5. Regresión estadística. Opera ahí donde se han seleccionado los grupos sobre la base de puntuaciones extremas.
6. Sesgos resultantes de una selección diferencial de participantes para los grupos de comparación.
7. Mortalidad experimental. Diferencia en la pérdida de participantes para los grupos de comparación.
8. Interacción entre la selección y maduración, etc. (93)

Por lo que hace a la validez externa, entendida como la medida en la que los resultados del experimento pueden generalizarse a diferentes sujetos, contextos, experimentadores y tal vez pruebas, las principales amenazas en contra de ella son las que siguen:

I. Validez de población.

- A. Población experimentalmente accesible en contra de la población a la que se pretende generalizar los resultados. Los resultados del experimento pueden ser aplicables solamente a aquella clase especial de personas entre quienes fueron seleccionados los sujetos experimentales y no a una población más amplia.
- B. Interacción de las variables personales y los efectos del tratamiento. Si la superioridad de un tratamiento experimental sobre otro pudiera invertirse cuando se aplicara a sujetos en los que los valores de las variables personales fueran distintos, ahí existe la interacción indicada.

93. CAMPBELL, DONALD y Julian Stanley. Diseños experimentales y cuasiexperimentales en la investigación social, 17.



II. Validez ecológica.

- A. Descripción explícita de la variable independiente. La generalización y repetición de los resultados de un experimento, presupone un conocimiento completo de todos los aspectos del contexto y tratamiento experimentales.
- B. Inferencia por tratamiento múltiple. Cuando dos o más tratamientos se aplican consecutivamente a las mismas personas en el mismo o en distintos estudios, es difícil y a veces imposible asegurar la causa de los resultados experimentales o generalizarlos a contextos en los que se aplicó un solo tratamiento.
- C. Efecto Hawthorne. El comportamiento de un sujeto puede ser influido al darse cuenta, tanto del experimento, como de la manera en la que debe responder al estímulo experimental.
- D. Efectos de la novedad y el desconcierto. Los resultados del experimento pueden verse parcialmente afectados por el entusiasmo o desconcierto generados por la novedad del tratamiento.
- E. Efectos del experimentador. El comportamiento de los sujetos puede ser inintencionalmente influido por ciertas características o conductas del experimentador.



- F. Sensibilización por pretest. Cuando un pretest ha sido administrado, los resultados experimentales pueden deberse, en parte a la sensibilización al tratamiento.
- G. Sensibilización por post-test. Los efectos del tratamiento pueden permanecer latentes o incompletos y aparecer solamente cuando se administra un test post-experimental.
- H. Interacción de la historia y los efectos del tratamiento. Los resultados pueden ser únicos por la ocurrencia de sucesos 'extraños' durante el experimento'
- I. Medición de la variable dependiente. La generalización de los resultados depende de la identificación de las variables dependientes y de la selección de los instrumentos para medirlas.
- J. Interacción de la fecha de la medición y los efectos del tratamiento experimental. La medición de la variable dependiente en dos fechas distintas, puede producir distintos resultados. (94)

Resulta pues, de importancia capital para conservar la validez de un experimento el control de las variables extrañas o no.

En el siguiente inciso veremos cómo hemos intentado, a veces sin conseguirlo ejercer dicho control.

94. BRACHT H. GLENN y Gene V. Glass. The External Validity of Experiments, 437-439.



3.6. Control de variables extrañas.

La validez de las inferencias que hagamos acerca de los resultados de un trabajo de tipo experimental dependerá, primordialmente, del grado de control que seamos capaces de ejercer sobre las variables.

Nuestro propósito consiste en conocer el efecto de la variable contenida en la hipótesis como supuesta causa del fenómeno y sólo en la medida en la que podamos aislar dicho efecto de los que pudieran ejercer las variables extrañas estaremos en aptitud de afirmar que la variable independiente fue la causa del resultado y no las variables exógenas.

Para controlar, aislar, los efectos de las variables extrañas son varias las técnicas que tenemos a nuestro alcance.

Cuando las variables tienen en los sujetos de la muestra los mismos valores que en los de la población, su efecto será uniforme y no podrán considerarse como la causa de las diferencias en los resultados.

No siendo ese el caso, las variables con valores distintos podrán ser incorporadas al experimento como independientes para medir sus efectos en los resultados.

De no ser esto posible porque no conozcamos sus valores y la distribución de ellos, trataremos de convertir la

situación en la que de manera natural existe en el primer caso, procuraremos que esas variables se distribuyan de igual forma entre los miembros de la muestra y los del universo al que pertenece.

Si de la población seleccionamos aleatoriamente a los miembros de la muestra, tendremos una alta probabilidad de conseguir lo arriba indicado.

La selección al azar, al dar a todos los individuos la misma probabilidad de formar parte de la muestra nos da una cierta garantía, probabilística, de que la distribución de las variables será igual en la muestra que en el universo.

En las investigaciones de tipo experimental, además de elegir la muestra al azar, usaremos el mismo procedimiento para seleccionar a los miembros que integrarán, tanto el grupo de control, como el sujeto a los efectos de la o las variables que pretendemos manipular. y también aleatoriamente asignaremos a cada grupo su carácter, ya de control, ya experimental.

En esta forma intentamos lograr que las diferencias, en caso de haberlas, entre las salidas de los grupos de control y experimental no sean atribuibles sino a la variable manipulada, pues la distribución uniforme de las de entrada nos permite suponer, de manera válida, que no es a ellas a quienes se deben las diferencias en los resultados.

En el caso que nos ocupa, las variables extrañas se controlaron, o cuando menos se intentó hacerlo, de la siguiente manera: Distinguiendo, en primer lugar, entre las variables independientes propiamente dichas y las que con ellas interactúan, las intervinientes.

Las primeras, las mediciones periódicas, son para nosotros la variable manipulada, aquella cuyos efectos queremos medir. Las segundas, las intervinientes son las que consideramos como extrañas.

De acuerdo con la clasificación que propuse en el inciso precedente me referiré en primer lugar a las del medio físico.

Las variables que lo conforman las controlamos impartiendo todas las clases, al grupo entero, en el mismo salón y con igual horario. Con ello, el efecto de estas variables en el grupo de control fue equivalente al que surtieron en el experimental.

En segundo lugar, las biológicas, psicológicas, socioculturales, etc., propias de los alumnos, no fueron en una primera instancia debidamente controladas, ya que la muestra inicial no fue elegida al azar, pues como ya vimos quedó formada por quienes se inscribieron voluntariamente al curso de A.P.O. II en el Taller E.

En una segunda instancia el control se ejerció dividiendo, aleatoriamente al grupo en dos con igual

número de miembros y con el mismo procedimiento se asignó a uno de ellos el carácter de grupo de control y el de experimental al otro.

Las del mismo tipo, relativas al profesor, fueron controladas por haber sido uno sólo quien impartió todas las clases en los tres semestres.

En tercer lugar, las que hemos llamado pedagógicas: plan de estudios, objetivos, contenidos, técnicas de E-A, actividades de aprendizaje y medios auxiliares para la enseñanza, se controlaron aplicando a todos los alumnos el mismo programa, incluyendo todos los aspectos anotados, en los cuatro ciclos lectivos considerados en este estudio.

Los antecedentes académicos, habilidades intelectuales, estrategias cognoscitivas, etc., forman parte de las variables que sólo fueron controladas en una segunda instancia.

De los factores que de acuerdo con Campbell y Stanley atentan contra la validez interna del experimento, algunos fueron debidamente controlados y otros no, como veremos enseguida:

- 1, 2. La historia y la maduración al haber tenido los mismos efectos en los grupos experimentales y de control no pudieron ser las causas de las posibles diferencias en los resultados cuando la comparación se hace para cada semestre por separado, pero sí

cuando la hacemos entre uno y otro.

3. En cuanto al tercer factor, los efectos que las mediciones producen en las subsecuentes, será uno de los problemas que tendremos que afrontar, pues nuestro interés radica en determinar cómo afectan las pruebas periódicas al R-E y no el efecto que podría surtir en el examen final la adquisición de una mayor habilidad para resolver el instrumento. Volveré a este tema al analizar los resultados obtenidos.
4. La instrumentación no pudo ser la causa de las probables diferencias que en cada semestre se hubieran observado entre los resultados del grupo experimental y el de control porque a los dos se les aplicó la misma prueba y el experimentador y calificador fue el mismo para todos los alumnos. En la comparación que estableceremos entre los resultados de diferentes semestres la situación es distinta.

Por una parte, para evitar el efecto de la instrumentación, la estructura de las pruebas fue la misma en los tres semestres y solamente se modificaron los datos.

Por la otra, esa circunstancia pudo ejercer un cierto efecto en el caso de que los alumnos hubieran obtenido alguna información acerca de

cuáles habían sido las pruebas administradas en los semestres anteriores.

Volveré a tratar este punto al examinar los resultados obtenidos.

5. Regresión estadística. Si consideramos que opera en los casos en los que los grupos han sido seleccionados con base en las puntuaciones extremas, no estamos en él.

La selección de los grupos experimentales y de control se hizo por selección simple al azar y no basada en alguna puntuación previa. Tampoco se estableció comparación alguna entre los resultados obtenidos por los alumnos de los grupos experimentales en las pruebas parciales y la final.

6. Sesgos resultantes de una selección diferencial de participantes para los grupos en comparación. No la hubo cuando circunscribimos a un semestre la comparación de los resultados, pues como ya dijimos, la selección de los grupos experimental y de control se hizo al azar.

Al comparar los resultados de los tres semestres no contamos con ninguna garantía de la equivalencia de los grupos cuya formación fue natural. Si acaso podríamos suponer cierta uniformidad entre los alumnos que se inscriben

TESIS CON
FALLA DE ORIGEN

en el Taller E, pues ellos conocen el nivel de exigencia que en él prevalece.

7. Mortalidad experimental. Como veremos al examinar los resultados, sólo en un semestre, el primero, hubo una diferencia apreciable entre el número de alumnos de los grupos en comparación y ésta fue del 17%.
8. Interacción entre la selección y maduración, etc. Este efecto no ocurrió si sólo comparamos los resultados en cada uno de los semestres, pero si pudo haber tenido lugar al equiparar los de dos o más períodos escolares.

Al examinar los resultados intentaré ponderar el efecto de las variables extrañas no suficientemente controladas sobre el posible grado de validez interna de nuestro estudio.

En cuanto a la validez externa, creo que estamos en una situación menos favorable pues casi ninguno de los factores que atentan contra ella, señalados por Bracht y Glass, fueron debidamente controlados.

I. Validez de población.

- A. Debido a que la población accesible fue solamente la del Taller E, no podremos generalizar nuestros resultados a la de toda la Facultad de Arquitectura, ni mucho menos a otras facultades.

- B. En cuanto a la interacción de las variables personales con los efectos del tratamiento, pudiera darse el caso de que al aplicarlo a alumnos de otros talleres o facultades los resultados se invirtieran o fueran, simplemente distintos.

II. Validez ecológica.

- A. Descripción explícita de la variable dependiente. Creo que cumplimos con este requisito, manejando una aceptable definición de ella.
- B. Interferencia por tratamiento múltiple. Al referirnos a la validez interna, este factor quedó incluido y será tomado en cuenta en el análisis de los resultados
- C. Efecto Hawthorne. Cabe la posibilidad de que los miembros del grupo experimental y también del de control lo hayan sufrido.
- D. Los efectos de la novedad en el tratamiento pudieron también afectarnos.
- E. Efectos del experimentador. Si bien en los semestres en los que realicé el experimento mi interés pudo haber sido mayor que en aquéllos en los que no lo intenté, no creo que mi actitud haya diferido notoriamente entre uno y otro de los cuatro en los que se hizo el estudio. No podría,

sin embargo, probarlo.

- F. No pudo haber sensibilización por pretest pues no se aplicó.
- G. Por la misma circunstancia no pudo haber efecto por post-test.
- H. La interacción de la historia y los efectos del tratamiento no pudieron confundirse al comparar los resultados de cada uno de los semestres por separado pues el primer factor fue el mismo para ambos grupos, el experimental y el de control. Fue éste, no obstante, uno de los factores no controlados al comparar los resultados de semestres sucesivos.
- I. Medición de la variable dependiente. Esta variable, para nosotros el R-E, creo que se identificó de manera aceptable y se procuró que los instrumentos para medirla tuvieran la necesaria validez de contenido y de construcción.
- J. Interacción de la fecha de medición y los efectos del tratamiento experimental. Los valores de la variable dependiente que se compararon, las calificaciones obtenidas por los alumnos en la prueba final, corresponden a la misma fecha.

Por lo expresado y a reserva de examinar con un poco más de amplitud los resultados, creo haber conseguido un

regular grado para la validez interna y uno casi nulo para la externa.

3.7. Definición de términos.

El conferir a cada término que se use en una investigación un significado lo más claro y preciso que sea posible, permitirá a otros investigadores interesados en el tema comprenderla y repetirla.

Por esa razón incluyo enseguida las definiciones de los vocablos que con algún sentido particular se utilizan en este informe.

Actitud. Disposición de ánimo de algún modo manifestada. (95)

Como quedó indicado se trata de un estado mental y neurofisiológico que al ejercer sobre el individuo una influencia dinámica lo prepara para reaccionar de una manera especial a ciertos estímulos. Nos referimos a ella en relación con el grado de atención e interés que los alumnos mostraron en las clases.

Aprendizaje memorístico. Es una memorización de una secuencia de palabras y otros símbolos, mediante la repetición verbal o la observación. (96) Durante todo el curso se desalentó a los alumnos a incurrir en este tipo de aprendizaje. Con ese fin se les informó con la debida antelación que en los exámenes podrían usar todo tipo de libros y apuntes.

Aprendizaje significativo. Comprende la adquisición de nuevos significados y, a la inversa, éstos son producto

95. Diccionario de la Real Academia de la Lengua Española.

96. EBEL, ROBERT L. Op. cit., 669-693.

del aprendizaje significativo. (97) Fue este el tipo de aprendizaje que se pretendió que los alumnos adquirieran en todas las sesiones de clase y al estudiar por cuenta propia.

Aptitud. Combinación de capacidades y otras características, innatas o adquiridas, que se suponen indicativas de la capacidad individual para adquirir conocimientos en un campo cognoscitivo determinado. (98)

Para controlar los efectos de esta variable cuyos valores en los alumnos nos fueron desconocidos recurrimos a la técnica de selección al azar.

Aptitud académica, aptitud escolar. Combinación de capacidades, tanto innatas como adquiridas, necesarias para las tareas escolares.

Capacidades mentales. Se refiere, tanto a procedimientos de rutina para resolver problemas, como a tipos más generales de pensamiento teleológico para el reconocimiento, el recuerdo, la comprensión, interpretación, aplicación, análisis, síntesis y evaluación. (99) Los instrumentos de medición fueron diseñados con el propósito de que para resolver adecuadamente los problemas que incluyen fuera necesario que los alumnos hubieran llegado, cuando menos, al nivel de aplicación.

Coefficiente de correlación. Es una medida del grado de relación existente entre dos conjuntos de medidas

97. AUSEBEL, DAVID P. Psicología educativa, 5.

98. ADAMS SACHS, GEORGIA. Op. cit., 791-806.

99. EBEL, ROBERT L. Op. cit., 669-693.



relativos al mismo grupo de individuos. (100)
Empleamos este tipo de coeficiente en el apartado relativo a la confiabilidad de los instrumentos de medición.

Conocimiento. Es una estructura de conceptos y relaciones construida mediante el pensamiento reflexivo a partir de la información recibida. (101) Lo consideramos como uno de los componentes del estado mental y neurofisiológico de los alumnos e importante, tanto por sí mismo, como por ser uno de los factores que conforman sus actitudes.

Correlación. Relación entre dos puntuaciones o medidas. (102) Usando diversos tratamientos estadísticos hemos intentado correlacionar las puntuaciones obtenidas por los alumnos de los grupos experimentales y de control.

Criterio. Es una norma para formular juicios de valor (103) Al hablar de las evaluaciones referidas a un criterio, la norma se hizo consistir en el grado en el que los alumnos habían alcanzado los objetivos señalados.

Creatividad. Es el nombre que se aplica a una capacidad hipotética que los individuos usan para producir: ideas, teorías, poemas, obras teatrales, novela, música, pintura, escultura, arquitectura, etc. Su carácter se considera hipotético pues es muy difícil

- 100. ADAMS SACHS, GEORGIA. Op. cit., 791-806.
- 101. EBEL, ROBERT L. Op. cit., 669-693.
- 102. ADAMS SACHS, GEORGIA. Op. cit., 791-806.
- 103. EBEL, ROBERT L. Op. cit., 669-693.



demostrar su presencia en una determinada persona antes de que haya manifestado poseerla por medio de alguna obra.(104) Nos referimos a ella como una cualidad del profesor que no debe quedar coartada por la rigidez de los programas.

Cuestionario. Es una lista de preguntas escritas relacionadas con un tema particular.(105) Se usó en la prueba sobre la primera unidad temática, especificaciones, sujeta a medición.

Desempeño académico, rendimiento escolar. Conocimientos adquiridos o habilidades desarrolladas por los alumnos en las escuelas.(106) Fue lo que intentamos medir al aplicar las diversas pruebas a los alumnos del curso.

Desviación. En relación con la puntuación de una prueba es la diferencia entre esa puntuación y algún punto de referencia.(107)

Desviación estándar. Es la medida de la dispersión de las puntuaciones en relación con la media de las mismas. La usamos en el análisis descriptivo de los resultados del experimento.

Diferencia significativa. Es una diferencia suficientemente grande entre dos medidas estadísticas comparables calculadas para muestras separadas, de manera que la probabilidad de que esa diferencia se deba al azar esté por debajo de un cierto límite

104. EBEL, ROBERT L. Op. cit., 669-693.

105. Ibíd.

106. Dictionary of Education.

107. EBEL, ROBERT L. Op. cit., 669-693.



definido. Para tratar de probar si existía alguna diferencia de este tipo entre las puntuaciones logradas en los exámenes finales por los grupos de control y experimentales, aplicamos a este estudio diversos tratamientos estadísticos.

Distribución normal. Es una distribución de frecuencia ideal definida por una ecuación matemática. (108)

Nos referimos a ella al tratar sobre la evaluación en la que esa curva es el criterio para determinar si una puntuación es o no aprobatoria.

Distribución sesgada. Es una distribución asimétrica en la cual la mayoría de las puntuaciones está más cerca de un extremo de la curva que del otro. (109)

Error de medición. Es la diferencia entre la puntuación obtenida y la verdadera y puede ser negativa o positiva. (110) En el análisis de varianza aplicamos una ecuación para calcular la suma de los cuadrados de los errores.

Error de muestreo. Es la diferencia entre el valor de una variable calculado para la muestra y el correspondiente a la totalidad de la población. (111) Hemos considerado como muestra, no seleccionada al azar, a los alumnos del Taller E de la Facultad de Arquitectura y como población a la totalidad de los estudiantes de dicha Facultad. Al no haberse elegido la

108. Ibídem.

109. Ibídem.

110. Ibídem.

111. Ibídem.



muestra aleatoriamente, es muy probable que hayamos incurrido en el error de muestreo.

Error estándar de medición. Es una estimación de la desviación estándar de los errores de medición.(112)

Escala. Es una secuencia de números cuyo uso está definido y limitado de manera que tengan una significación especial en cuanto a indicar distintos grados del mismo rasgo o característica.(113) En este estudio hemos conferido al término un significado más amplio al considerar que la secuencia puede ser no sólo de números, sino de cualesquiera otros símbolos.

Evaluación. Es un juicio del mérito del examinado.(114) No obstante ser, en esencia, ese el significado que estamos otorgando al término, en nuestro afán por hacer objetivo el juicio, hemos incluido a la medición como la base necesaria para formularlo.

Evaluación subjetiva. Tiene lugar cuando la puntuación que recibe un estudiante no se basa en alguna de antemano establecida, en la que a cada pregunta se le asigna un determinado número de puntos; sino en el juicio personal del evaluador.(115) Para evitarla, como podemos ver en el apartado dedicado a los aparatos e instrumentos, cada respuesta correcta está tasada con un cierto número de puntos.

Examen. Es aquel proceso de comprobación que tiene por finalidad poner a prueba la capacidad o rendimiento de

112. Ibidem.

113. Ibidem.

114. Ibidem.

115. Ibidem.



los estudiantes en cualquier sector de conocimientos.(116) Tanto en los exámenes finales, como en los parciales, ha sido sólo el rendimiento de los alumnos lo que hemos tratado de determinar.

Experimento. Consiste en que en una situación dada, alguna o algunas de las variables son objeto de manipulación y se observa su efecto sobre las otras.(117) En este estudio, la variable manipulada, la activa, ha quedado constituida por las mediciones periódicas hechas en el transcurso del semestre. Su efecto ha sido observado sobre el rendimiento escolar.

Examen de ensayo. El que está formado por preguntas o indicaciones que requieren una respuesta, más o menos, extensa.(118) Por conveniencia incluimos en este tipo al de resolución de problemas.

Formas equivalentes de una prueba. Se refiere al caso en el que las preguntas de dos pruebas son del mismo tipo, cubren contenidos iguales y lo son también los grados de dificultad y la forma en la que estos se distribuyen en el test. Rinden puntuaciones con valores muy cercanos para la media y tienen la misma variabilidad y confiabilidad. (119) Nos referimos a ellas al tratar sobre la confiabilidad de las pruebas.

Grupo de control. Es el grupo al que en un experimento no se le aplica ese tipo de tratamiento, pero que en todo lo demás se asemeja, en la medida de lo

116. Ibidem.

117. CAMPBELL, DONALD y Jualian Stanley. Op. cit., 15.

118. EBEL, ROBERT L. Op. cit., 669-693.

119. Ibidem.



posible, al grupo o grupos experimentales.(120) Se forma, como lo hicimos, con el fin de aislar los efectos de la variable activa, del que pudieran ejercer las demás.

Grupo experimental. El que se somete a ese tipo de tratamiento.

Instrumentos de medición. En pedagogía son, de manera primordial. las pruebas que se usan para medir el R-E de los alumnos, sus capacidades, interés y otros rasgos. En nuestro caso las empleamos para medir el R-E.

Medida estadística. Es un número empleado para describir o caracterizar algún aspecto de la muestra.(121) En este estudio empleamos la media y la desviación estándar.

Pruebas periódicas. Denominamos así a aquéllas que se hacen durante el ciclo escolar y no sólo al principio o final del mismo. La periodicidad, en nuestro estudio, fue variable pues se realizaron al término de cada unidad programática.

Muestra al azar. Es una muestra seleccionada de tal manera que garantiza igual probabilidad de elección a todas las muestras posibles de ese tamaño que podrían formarse con los miembros del universo del que se trate.((122) En el estudio que nos ocupa, no fue ese el procedimiento empleado para elegir a los grupos motivo del experimento. Trabajamos con grupos naturalmente formados.

120. *Ibidem.*

121. *Ibidem.*

122. *Ibidem.*



Muestra representativa. Es aquélla en la que los valores y la distribución de las variables son iguales que los del universo del que forma parte.

Norma. Cuando este término se aplica a las puntuaciones de las pruebas, se refiere al promedio para los miembros de un mismo grupo. (123) Así la usamos al hablar sobre los dos principales procedimientos de evaluación.

Objetividad. Es una característica de los enunciados que puede ser verificada por un observador o juez independiente. (124) Al hablar de las pruebas en particular y de la evaluación en general, la consideramos como una de las cualidades que deben poseer

Probabilidad. La probabilidad de un suceso es una fracción decimal que expresa la razón matemática entre los acontecimientos que realmente han tenido lugar y las oportunidades que hayan tenido de suceder. (125) La manejamos al hacer el análisis estadístico de los datos obtenidos.

Prueba. test. Es un término de carácter general empleado para designar cualquier recurso o procedimiento para medir la capacidad, rendimiento, interés y otros rasgos. (126) En el texto quedó intercalada la definición que da el Webster's New Collegiate Dictionary.

123. *Ibíd.*

124. *Ibíd.*

125. *Ibíd.*

126. *Ibíd.*



Prueba, confiabilidad de una. Es el grado de precisión y consistencia que guarda el instrumento para medir lo que se propone.(127) La consideramos como una de las cualidades que las pruebas deben poseer.

Pruebas de aptitud. Se administra para determinar las posibilidades de desarrollo que un sujeto tiene en una dirección específica.(128)

Prueba de rendimiento. Es aquella que está concebida para medir hasta qué punto un estudiante domina un cuerpo de conocimientos o posee ciertas habilidades.(129) Fue este el tipo de prueba que se aplicó, tanto en las mediciones parciales, como en la que se hizo al final del curso.

Prueba diagnóstica. Está diseñada para poner de manifiesto los puntos débiles o fallas en algún tema de estudio.(130)

Prueba estandarizada. Es la que se ha construido con especificaciones detalladas, cuyos reactivos se han seleccionado después de haberlos sometido a un ensayo sobre su adecuación en cuanto a dificultad y poder de discriminación.(131)

Prueba final. Es la que se administra al final del curso y cubre todas las unidades programáticas del que consta el mismo.

127. MENENDEZ, LIBERTAD y Laura Elena Rojo. Op. cit.,

15.

128. EBEL, ROBERT L. Op. cit., 669-693.

129. Ibíd.

130. Ibíd.

131. Ibíd.



Prueba periódica. Es la que se administra en el transcurso de un ciclo escolar y no sólo al principio o final del mismo. La periodicidad temporal puede ser fija: semanaria, mensual, etc. o como en nuestro caso variable pues se impartieron al término de cada unidad programática del curso.

Puntuación. Es la suma de los puntos correspondientes a las respuestas correctas, asignando a las preguntas o problemas un cierto número de ellos en relación con su grado de dificultad, que en todo caso puede ser igual para todos. (132) En el inciso dedicado a los aparatos e instrumentos se consignan los valores en puntos que tuvieron las diferentes preguntas y problemas incluidos en las pruebas que se administraron. La suma obtenida se transformó a base 10.

Puntuación aprobatoria. En una prueba es la puntuación mínima que satisface un requisito particular. (133) La consideramos de seis puntos sobre diez posibles.

Rasgo. Es cualquier atributo de una persona, poseído en distintos grados por los miembros de un grupo o clase. (134) Nos referimos a ellos como una de las principales variables a tomar en cuenta y a controlar para conservar la validez del experimento.

Reactivo de opción múltiple. Cuenta con dos partes: el tronco, formado por una pregunta directa o un enunciado

132. *Ibidem.*

133. *Ibidem.*

134. *Ibidem.*



incompleto y dos o más opciones formadas por las respuestas a las preguntas, o por partes de una oración que completan el enunciado del tronco. (135) Los incluimos entre las pruebas objetivas.

Reactivo verdadero-falso. Es el que consiste en enunciados que el examinado debe juzgar como verdaderos o falsos.(136) Se consideran, como los anteriores, propios de las pruebas objetivas.

Reactivos de complemento. El sujeto tiene que añadir el elemento que falta en una oración gramatical, serie o figura.(137) Son también adecuados en las pruebas objetivas.

Rendimiento escolar. Considerando como el grado de dominio que un alumno alcanza en cada una o en la totalidad de las unidades programáticas del curso. Se expresa con una calificación en una escala de 0 a 10. Por ser, aunque no en un sentido estricto una escala de intervalos pues no podemos asegurar que sean iguales, para el tratamiento estadístico de los datos cabe el uso de métodos paramétricos y no paramétricos.

Rendimiento final. Considerando como el grado de dominio alcanzado por un alumno en la totalidad de las unidades programáticas del curso. Se expresa igual que el anterior y se le puede dar el mismo tratamiento estadístico.

135. Ibidem.

136. Ibidem.

137. Ibidem.

Rendimiento parcial. Considerado como el grado de dominio alcanzado por un alumno en cada una de las unidades programáticas del curso. Se expresa en igual forma que los dos anteriores y el tratamiento estadístico adecuado es también el mismo.

Selección al azar, aleatoria. Es la que se realiza dando a todos los miembros de la población la misma probabilidad de quedar incluidos en la muestra. Se obtiene por medio del uso de números aleatorios.

Tabla de números aleatorios. En la secuencia de los dígitos que la forman no existe sistema u orden perceptible.(138) Cuando los números se obtienen por medio de una calculadora programable, una H.P. 67 en el caso de este estudio, se trata en rigor de una pseudoaleatorización. No obstante al ser posible obtener hasta 500,000 datos antes de que se repita la secuencia, contamos con amplio margen de seguridad si sólo requerimos 50 números.

Taxonomía. Es una clasificación ordenada, por lo general de plantas o animales, organizada de acuerdo con sus relaciones naturales o supuestas.(139) El concepto se puede extrapolar, como lo hemos hecho, a cualquier otra clasificación. En nuestro caso a los objetivos educativos.

Unidad programática. Usamos el término para referirnos al agrupamiento de los contenidos de cada uno de los

138. *Ibidem.*
139. *Ibidem.*

temas principales del curso de A.P.O. II: Claves y datos que deben contener los planos, especificaciones, costos unitarios, cantidades de obra, estimados de costo y calendarios de obra.

Validez. Se la suele definir, hablando de una prueba, como el grado en el que mide lo que se pretende medir. (140) En el inciso 1.5.2. examinamos tres tipos de validez: de contenido, relacionada con un criterio y de construcción.

En otros de los apartados de este trabajo hemos otorgado al término un significado más amplio, p.e. al referirnos a la validez de las inferencias que los resultados del experimento nos permiten obtener. Lo que tratamos de determinar es el grado en el que dichas inferencias concordarían con la realidad.

Variable. Es el rasgo que distingue a los miembros de una comunidad.

Variable activa. Es aquélla que se puede manipular.

Variable atingente. Se puede considerar como causa del fenómeno en estudio.

Variable dependiente. Es la posible causa de una relación causal. En nuestro caso se trata del rendimiento escolar.

Variable dicotómica. Para efectos de nuestra investigación es aquélla que sólo puede asumir dos valores.

140. *Ibidem*.

Variable discreta. Es aquélla que sólo aumenta o disminuye en una o varias unidades previamente establecidas. Serán discretas: el número de alumnos y para nosotros las puntuaciones de las pruebas. En contraste con la discreta la variable continua puede asumir cualquier valor.

Variable extraña o exógena. Se considera a aquélla que no está sujeta a un control directo.

Variable independiente. Es la posible causa en una relación causal. En este estudio es la aplicación de mediciones periódicas.

Variable interviniente. Es aquélla que interviene entre una variable independiente y una dependiente en lo que respecta al efecto. (141) A ellas hacemos referencia en el inciso dedicado a las variables extrañas.

Variable politómica. Es la que puede asumir más de dos valores.

Nota: No incluyo en este breve glosario algunos términos que si bien se emplean en nuestro estudio con algún sentido específico, sus definiciones van intercaladas en el texto.

141. HAYMAN, JOHN L. Investigación y educación, 162-181.

8. Aparatos e instrumentos.

Por la índole de la hipótesis que se planteó, las mediciones que nos interesan son las relativas al R-E.

Para realizarlas se aplicaron cinco pruebas parciales a los alumnos del grupo experimental y una final a todos ellos.

Cada prueba parcial abarcó una de las unidades programáticas del curso, a saber: a) especificaciones, b) costos unitarios, c) cantidades de obra, d) estimados de costo y e) calendario de obra. Todas las pruebas fueron escritas.

Las pruebas fueron calificadas con una escala de 1 a 10, dando a cada pregunta o problema un distinto valor en puntos según la ponderación que se hizo en cuanto a su grado de dificultad.

La primera prueba parcial, la relativa a especificaciones fue de respuesta corta para algunos reactivos y un poco más extensa para otros, pero en todos los casos se dio al alumno alguna guía para que normara su criterio.

Las otras cuatro pruebas parciales fueron de resolución de problemas y en todos los casos los alumnos recibieron amplia información sobre los aspectos importantes de cada uno de ellos.

Se les recomendó que antes de dar respuesta a la pregunta o solución al problema, procuraran comprender

su significado y tipo de respuesta que implicaba en la forma más clara que les fuera posible.

Para cada prueba parcial el tiempo concedido fue de hora y media y de dos y media para la final.

a). Primera prueba parcial. Especificaciones.

1. ¿Cuáles son los tres usos principales que tienen las especificaciones?
2. ¿Cuáles son los cuatro capítulos básicos de los que consta una especificación?
3. Explique en forma breve en qué consiste cada uno de ellos y ¿cuáles son los datos mínimos que debe contener?
4. Si se tratara de elaborar una especificación para un muro de carga, ¿qué datos sería necesario investigar?

Para la cuarta pregunta se informó a los alumnos que los datos se debían referir a: materiales, procedimiento de construcción y uso de equipo.

Al primer reactivo se le asignó un punto como valor, al segundo y al tercero dos a cada uno y cuatro al último.

Para obtener un calificación de diez puntos se tomó en cuenta la limpieza y claridad del escrito.

b) Segunda prueba. Costos unitarios.

Se procuró como en todos los casos que no faltara ningún dato y se advirtió a los alumnos que si aparentemente eso sucedía era porque al resolver la parte anterior del problema el resultado obtenido era ese dato que le hacía falta para proseguir.

El problema consistió en el cálculo del costo unitario de una trabe de concreto armado con diseño que se indica en la página 200.

La guía para los alumnos radicó en indicarles cuáles eran los materiales que debían calcular y recordarles que el costo de mano de obra debía ser también incluido. Para ese fin se escribió en el pizarrón la tabla siguiente:

II. Materiales.

Madera de cimbra.....	p.t.
Alambre.....	0.100 kg.
Clavo.....	0.050 kg.
Cimbrafest.....	0.250 lt.
Varilla.....	kg.
Alambrón \emptyset 1/4".....	kg.
Cemento.....	kg.
Arena.....	m ³
Grava.....	m ³

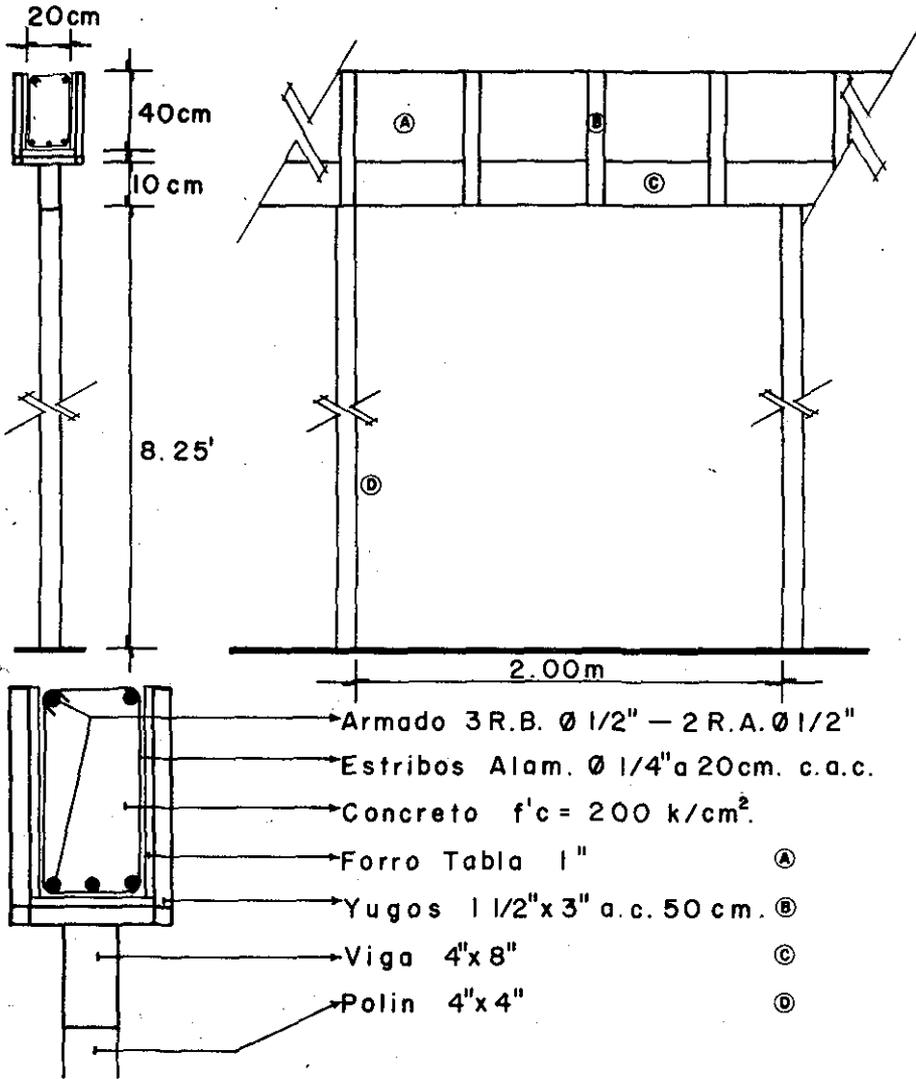
III. Mano de obra.

Albañil y peón..... jo.

Para la mano de obra se les proporcionó el rendimiento medio de un albañil y un peón para el trabajo indicado. Se les indicaron asimismo los precios de cada uno de los materiales y el salario real de la cuadrilla respectiva.

El cálculo del costo de la madera tuvo un valor de dos puntos, el alambre, clavo y cimbrafest uno sólo pues se les dio el dato de la cantidad de cada uno de ellos y no tenían sino que hacer las multiplicaciones adecuadas para obtener el resultado pedido. A los demás materiales: varilla, alambrón, cemento, arena y grava se les asignó un punto a cada uno al igual que a la mano de obra.

La calificación de diez puntos se completaba concediendo el faltante a la buena presentación del trabajo.



ADMINISTRACION DE PROYECTOS Y OBRAS II.
 Segunda Unidad Programática
 Costos Unitarios

TESIS CON
 FALLA DE ORIGEN

c) Tercera prueba. Cantidades de obra.

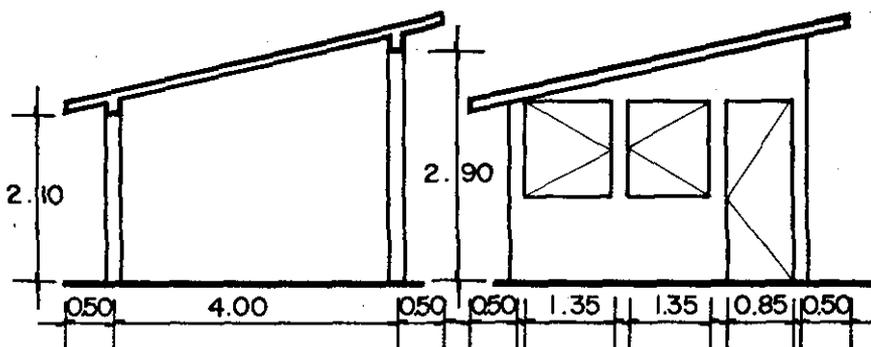
Con objeto que los alumnos calcularan las cantidades de obra de los siguientes conceptos:

1. Excavación.
2. Plantilla.
3. Cadena de liga.
4. Muros de tabique.
5. Cerramientos.
6. Cimbra para techo.
7. Acero de refuerzo en techumbre.
8. Concreto en losa techo.
9. Impermeabilización y enladrillado techo.
10. Relleno en piso.
11. Firme de concreto en piso.

Se les proporcionó el croquis que se reproduce en la página siguiente y que corresponde al local de una bodega en obra.

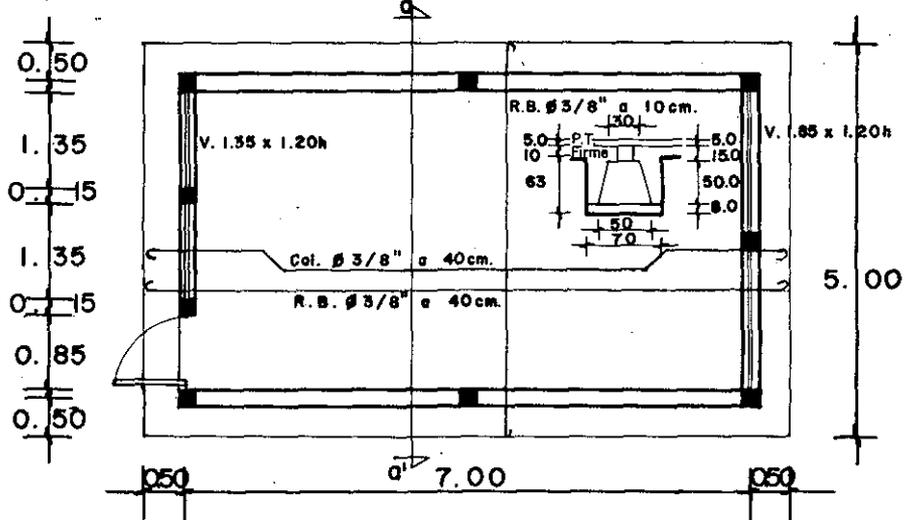
La respuesta correcta a las preguntas 3, 5, 6 y 9 tuvo un valor de 1/2 punto y de uno para las demás. Sumando el que corresponde a la presentación el total es de diez puntos.





Corte a-a'

Fachada



ADMINISTRACION DE PROYECTOS Y OBRAS II.
 Tercera Unidad Programática
 Cantidad de Obra

TESIS CON
 FALLA DE ORIGEN

d) Cuarta prueba. Estimados de costo.

El problema consistió en calcular el costo de una barda para un terreno rectangular de 10 por 20 m. La altura del muro se fijó en 2.50 m.

Como guía a los alumnos, se les proporcionó la lista de los conceptos de obra que debían tomar en cuenta.

1. Excavación.
2. Mampostería de piedra en cimentación.
3. Cadena en cimentación.
4. Castillos de concreto.
5. Muro de block de concreto.
6. Cadena de cerramiento.
7. Puerta para peatones.
8. Puerta para automóviles.

De los ocho conceptos se les pedía el cálculo de los costos unitarios de la mampostería de piedra, de las cadenas de liga (cimentación y cerramiento) y castillos de concreto (los tres conceptos tienen el mismo costo) y el del muro de block, los otros se les proporcionaron, así como los precios de los diversos materiales, salarios reales y rendimientos de las cuadrillas de trabajadores.

A los cálculos de costos unitarios se asignaron: dos puntos a la mampostería de piedra, cuatro a los castillos y cadenas y dos al muro de block; a la

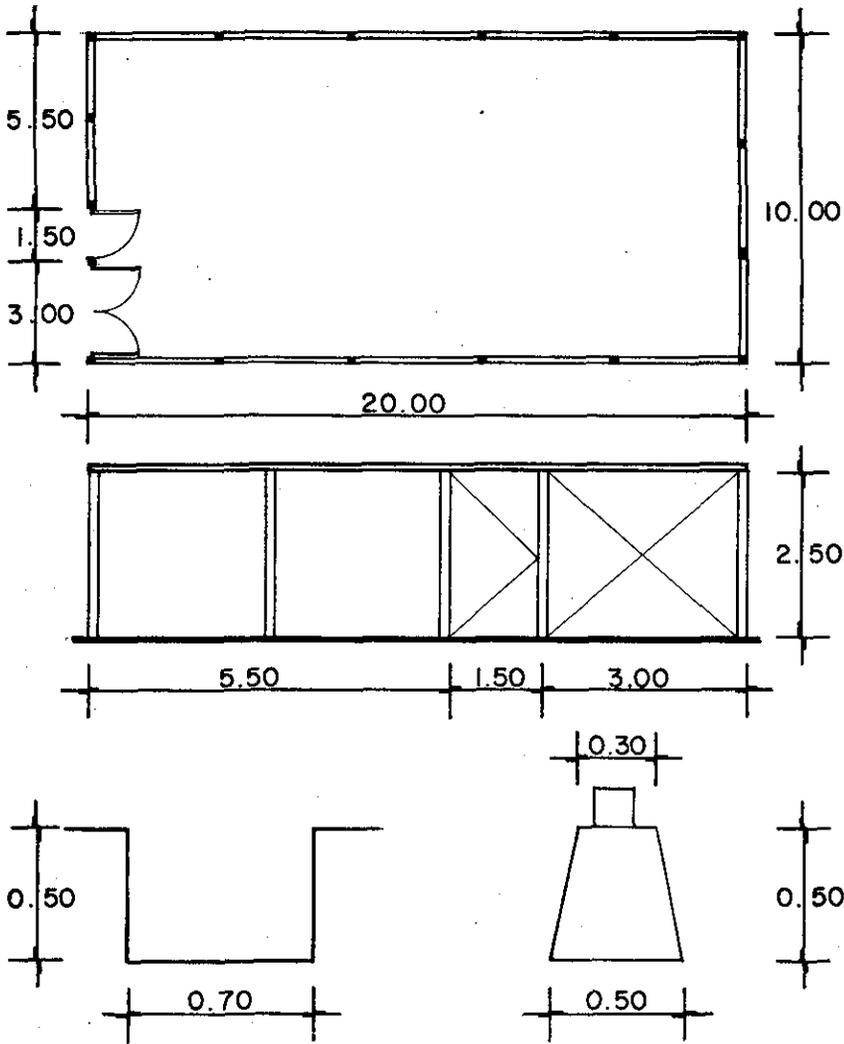


cubicación de las cantidades de obra dos para cada uno de los tres conceptos siguientes: excavación, mampostería y muros de block y uno para cada uno de los cinco restantes.

Sumando estos puntos al que se otorgó en caso de buena presentación, el total es de veinte. El total de puntos obtenidos se dividió entre dos para calcular la calificación.

Croquis en la página siguiente.





ADMINISTRACION DE PROYECTOS Y OBRAS II.
 Cuarta Unidad Programática
 Estimado de Costo

TESIS CON
 FALLA DE ORIGEN

e) Quinta prueba parcial. Calendario de obra.

En esta prueba se pedía al alumno que elaborara el calendario de obra según el método de la ruta crítica (CPM) para una barda de 20 m. de longitud por dos de altura.

El cimiento era una zapata corrida de concreto armado, prefabricada, las piezas para el muro eran también de concreto prefabricado y sólo se construirían en obra los castillos de concreto armado.

Se trataba de resolver los siguientes problemas:

1. Calcular los tiempos de las actividades que a continuación se indican:
 - a. Excavación.
 - b. Colocación de zapatas.
 - c. Castillos de concreto: Cimbra, armado y colado. Descimbrado.
 - d. Colocación de las piezas del muro.
 - e. Elaboración en fábrica de las piezas para el muro.
2. Elaborar la matriz de precedencia y sucesión.
3. Trazar la red o malla.
4. Calcular las iniciaciones más próximas de cada una de las actividades.
5. Calcular las terminaciones más alejadas de cada una de las actividades.

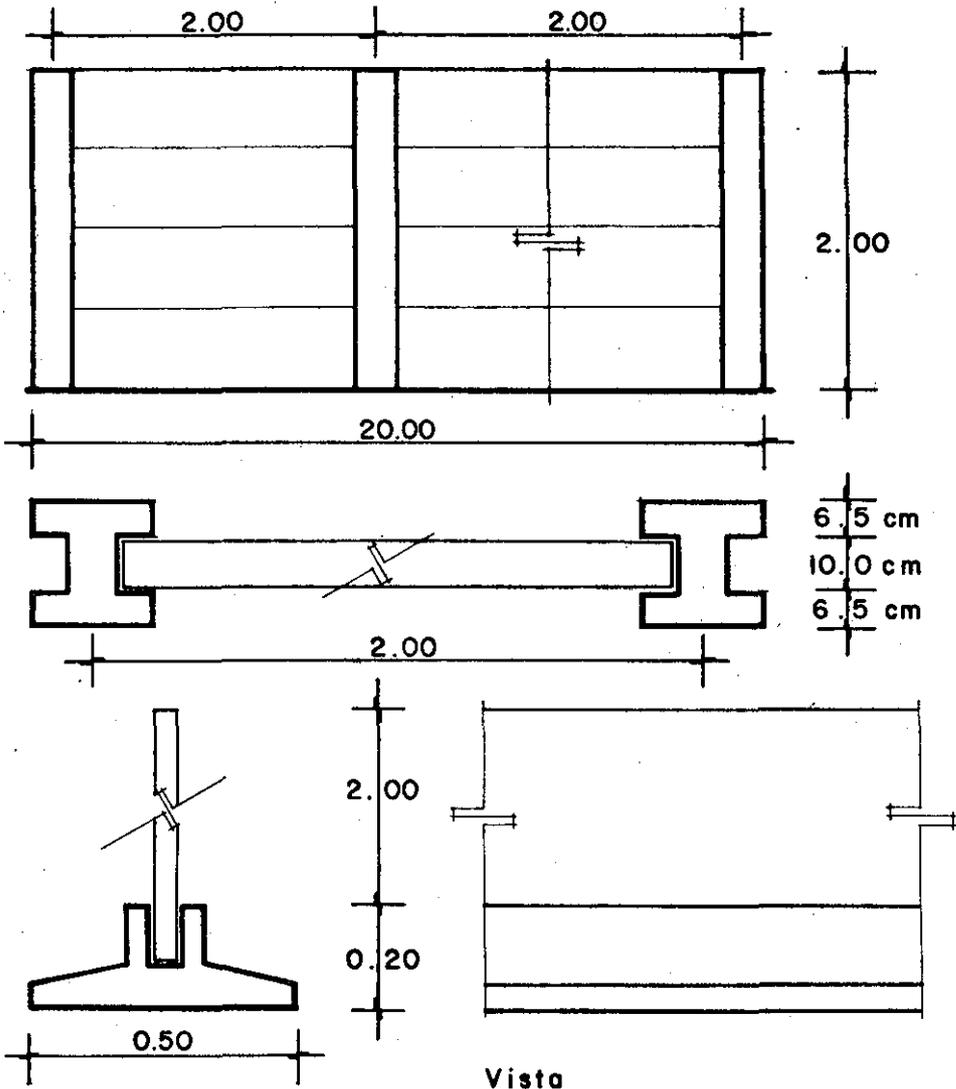


Para calcular los tiempos se indicaron a los alumnos la composición y rendimiento de las cuadrillas encargadas de cada una de las actividades, así como el tiempo que había que esperar, después de colar los castillos, para poder descimbrarlos.

Las preguntas en total fueron nueve, cada una con valor de un punto, los que sumados al correspondiente a la buena presentación, dio un total de diez puntos.

Croquis en la página siguiente.





ADMINISTRACION DE PROYECTOS Y OBRAS II.
 Quinta Unidad Programática
 Calendario de Obra.— Ruta Crítica.

TESIS CON
 FALLA DE ORIGEN

Prueba final.

En esta prueba se incluyeron los principales contenidos del curso, agrupados en cinco unidades programáticas: especificaciones, costos unitarios, cantidades de obra, estimado de costos y calendario de obra.

Las preguntas y problemas a resolver fueron los que siguen:

- a) Con el fin de elaborar una especificación para un muro de block hueco de concreto, para carga, ¿ Cuáles son los datos que trataría de investigar ?.
- b) Calcular el costo de un local para bodega en obra de acuerdo con lo indicado en el croquis de las páginas 212 y 213.

Para tal fin calculará usted los costos unitarios de:

1. Concreto a pie de revolvedora.
2. Acero de refuerzo.
3. Muro de tabique de barro cocido hecho a mano.
4. Losas prefabricadas de techo.

El estimado de costo tomando en cuenta los conceptos que enseguida se anotan:

1. Losa de cimentación.
Madera de cimbra.
Acero de refuerzo.
Concreto.
Vaciado y vibrado del concreto.



2. Muros de tabique.
3. Castillos de concreto.
4. Cadena de cerramiento.
5. Losas para techo.
Fabricación
Colocación.
6. Ventanas.
7. Puerta de entrada.

c) Calendario de obra.

De las Actividades anotadas en la siguiente lista
Calcular los tiempos para aquellas en las que no se
proporciona ese dato.

Elaborar la matriz de precedencia y sucesión.

Trazar la red o malla.

Calcular las iniciaciones más próximas.

Calcular las terminaciones más alejadas.

Para la resolución de los problemas los alumnos
dispusieron de todos los datos necesarios, tales como:
consumo de materiales para cada m^3 de concreto,
porcentaje de desperdicio del acero de refuerzo cantidad
de alambre necesario para armar cada tonelada; dimensiones
de los tabiques, espesor de la junta, cantidades de
materiales para cada m^3 de revoltura, dimensiones y
armado de las losas precoladas; así como los salarios
reales y rendimientos de las cuadrillas de trabajadores.
Se les indicó asimismo cuáles eran los precios de los
diferentes materiales y los datos necesarios para

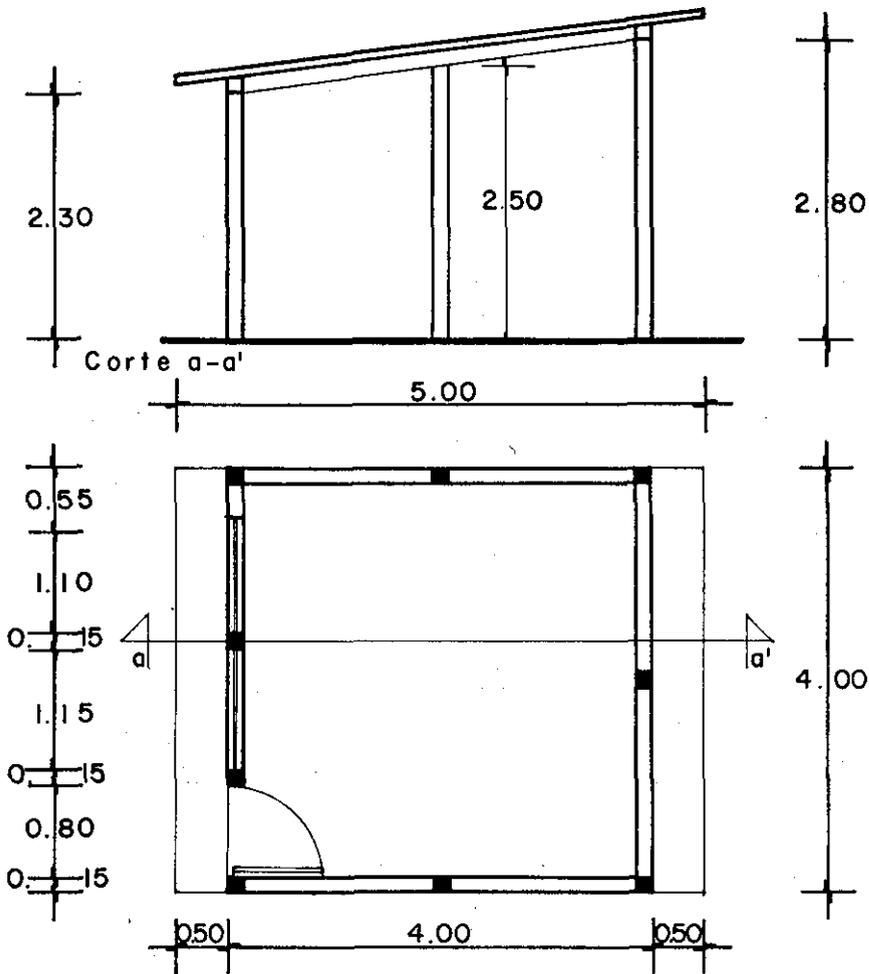
TESIS CON
FALLA DE ORIGEN

calcular el costo correspondiente al uso de equipo. La prueba final, igual que las parciales, fue calificada con una escala de 1 a 10 y las puntuaciones se distribuyeron en la siguiente forma: especificaciones, un punto; costos unitarios del concreto y acero de refuerzo, uno cada uno, dos para el muro de tabique y tres para las losas precoladas. En el estimado de costo se dió un punto a cada uno de los renglones relativos a: madera de cimbra, acero de refuerzo, concreto, vaciado y vibrado, castillos, cadena, fabricación losas techo, colocación de las mismas y uno para las ventanas y puerta en conjunto.

A los muros de tabique correspondieron dos puntos y tres al calendario de obra. La suma de esos puntos, más los dos que se concedieron por buena presentación, dan un total de veinticinco. Para obtener la calificación el total de puntos logrados por cada alumno se multiplicó por 0.4 para transformarlos a la base de 10.

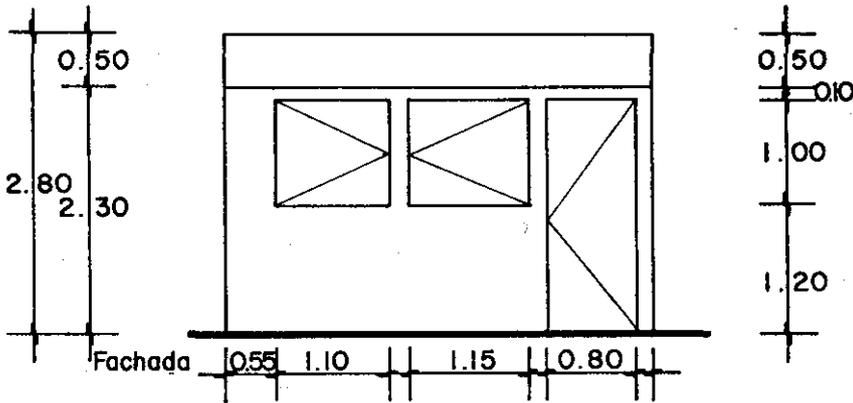
Esta prueba final fue administrada a todos los alumnos en el período A de exámenes y a quienes no la aprobaron o deseaban mejorar su calificación se les impartió en el periodo B otra con la misma estructura, pero con datos diferentes.

Los efectos de la variable respectiva tendrán que ser tomados en cuenta en el análisis de los resultados.



ADMINISTRACION DE PROYECTOS Y OBRAS II.
Examen Final A.

TESIS CON
FALLA DE ORIGEN



Losa Cimentación



Viga madera perimetral
4" x 8"

Transversales y longitudinales

R. A. \emptyset 3/8" a 20 cm.

R. B. \emptyset 3/8" a 20 cm.

Concreto $f'c = 150$ k/cm

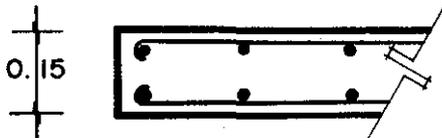
Cimbra 1 1/2"

Refuerzos 2" x 2" a 50cm.

Cuatro varillas \emptyset 5/16"

Estribos alamb. \emptyset 1/4" a 20cm.

Concreto $f'c = 150$ k/cm.



Castillos y Cadenas



ADMINISTRACION DE PROYECTOS Y OBRAS II.

Examen Final A.

TESIS CON
FALLA DE ORIGEN

4. Metodología.

Con el propósito de seleccionar el método adecuado para realizar una investigación, es necesario tomar en cuenta, entre otros factores tales como las variables y su tipo, los fines que se pretende alcanzar.

En nuestro caso, el tema central que abordamos, la evaluación, vista a la luz de Astin y Panos, tiene como fin principal el recabar información que pueda ser de utilidad en la toma de decisiones educativas.

Para ello es indispensable que posea ciertas características que sólo se pueden obtener cuando es producto de una indagación debidamente realizada.

Como método de investigación, el científico ha mostrado ser superior a los demás debido principalmente a que se trata de un proceso controlado y podemos, en consecuencia, dirigirlo eficientemente hacia la consecución de las metas deseadas. (142)

Ese control, sin embargo, se ejerce en distintos grados. El perfecto es un ideal al que la ciencia, en su continuo avance, se acerca cada vez más sin llegar nunca a conseguirlo. En una determinada investigación, algunos aspectos están bajo control y otros no lo están, como hemos visto al referirnos al de las variables extrañas para este estudio.

Aún así, haremos lo posible por ejercer dicho control en las tres etapas principales que podemos distinguir en el proceso científico de investigación.

En la primera, se observan los hechos, para nosotros el posible efecto de las mediciones periódicas en el R-E; en la segunda se plantea alguna hipótesis, como ya lo hicimos y en la tercera se derivan de ella consecuencias que sean susceptibles de ser comprobadas por observación. Si esas consecuencias se verifican la hipótesis se acepta, provisionalmente como verdadera, aunque lo usual es que posteriormente al observarse nuevos hechos, sea necesario modificarla. (143)

Esta tercera etapa, al incluir la contrastación de las consecuencias de la hipótesis con la realidad, es crucial, pues es la que confiere a nuestro conocimiento la cualidad que en especial deseamos obtener, la objetividad, entendida esta como "...La construcción de imágenes de la realidad que sean verdaderas e impersonales..." (144) y por ende transmisibles y verificables.

Con esta finalidad in mente hemos procedido en el desarrollo de este estudio y después de establecer un breve marco teórico de referencia, los pasos siguientes han consistido en plantear el problema, relacionarlo con los antecedentes que pudimos conocer, plantear la hipótesis y derivar sus consecuencias. Nos corresponde, por consiguiente, hacer el intento de contrastarlas con la realidad.

143. RUSSELL, BERTRAND. The Scientific Outlook, 58

144. BUNGE, MARIO. La investigación científica, 21, 22

4.1. Muestreo.

El primer paso que habrá de conducirnos a lo indicado, a la contrastación de la hipótesis con la realidad, consiste en seleccionar a los sujetos u objetos cuyo comportamiento observaremos para verificar si se ajusta o no a nuestras predicciones.

Debido a que sería si no del todo imposible, cuando menos sumamente difícil e impráctico observar a toda la población motivo de nuestro interés, en este caso a la totalidad de los estudiantes, nos tendremos que limitar a hacerlo con una parte de ella, la que nos sea accesible.

Esa parte de la población o ese subconjunto de un conjunto de unidades obtenido con el fin de investigar las propiedades de la población o conjunto del que proceden, se denomina muestra. (145)

Recurrimos a la obtención de una muestra, tanto en el caso señalado, en el que la población es demasiado numerosa, como en aquel en el que por ser lo suficientemente uniforme, desde cierto punto de vista, proporciona una buena representación de toda la población.

En ambos casos, son dos los aspectos que habremos de tomar en cuenta para integrar la muestra.

- a) Su tamaño. El número de sujetos u objetos.
- b) Su diseño. Cuál subconjunto de sujetos u objetos

serán los seleccionados.

Para determinar el tamaño de la muestra es necesario tomar en cuenta la cantidad y dispersión de las variables, así como su tipo.

El conocimiento de la cantidad y dispersión de las variables implica el de la población, que en la mayoría de los casos no poseemos.

Contamos, sin embargo, con el apoyo que nos da la estadística al proporcionarnos una buena probabilidad de acierto cuando consideramos que la distribución de los valores de las variables se ajusta a alguna de las curvas teóricamente determinadas.

Otro criterio que hemos de fijar de antemano es el que se refiere al error máximo que para nosotros será admisible. Ese error, si bien inevitable pues a partir de un sector conocido de la población trataremos de inferir las características de toda ella, es controlable estadísticamente hablando, ya que:

$$V(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}$$

Siendo:

$V(\bar{y})$ = Esperanza de la media muestral.

$\frac{S^2}{n}$ = Variancia de la media muestral.

n = Número de los miembros de la muestra.

N = Número de los miembros de la población.

Ahora bien, la precisión con la que se desea hacer una estimación está indicada por un error 'd', expresado como un porcentaje del valor de la variable para la población y dado que la desviación estándar de la muestra es igual al error estipulado dividido entre el valor de la abscisa 't' en la curva de distribución, o sea:

$$S = \frac{d}{t}, \text{ o } V = \frac{d^2}{t^2}, \text{ o bien } \frac{1}{V} = \frac{t^2}{d^2}$$

Entonces para que la muestra tenga una variancia máxima igual a: $\frac{d^2}{t^2}$ y después de hacer las operaciones matemáticas adecuadas, obtenemos la siguiente ecuación para datos cuantitativos.

$$n = \frac{S^2 \cdot t^2}{d^2}$$

Siendo:

n = Tamaño de la muestra.

S² = Variancia.

t = Valor de la abscisa en la curva de distribución.

Si en vez de datos cuantitativos los que vamos a manejar son cualitativos, sustituimos el valor 't' de la abscisa con el producto del porcentaje esperado y su complemento y así obtenemos:

$$n_0 = \frac{P \cdot Q}{V}$$

y dado que:

$$v = \frac{d^2}{t^2}$$

entonces:

$$n_0 = \frac{P \cdot Q}{\frac{d^2}{t^2}} = \frac{P \cdot Q \cdot t^2}{d^2}$$

Si conocemos el tamaño de la población estamos en posibilidad, si así lo deseamos, de reducir el tamaño de la muestra aplicando la siguiente ecuación:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad (146)$$

Esta muestra reducida de tamaño n , tiene la misma validez desde el punto de vista estadístico que la original n_0 .

Una vez calculado el tamaño de la muestra, podemos proceder a la selección de los sujetos u objetos que la formarán.

El criterio que nos guiará para hacer dicha selección será el relativo a la representatividad de la muestra, porque si las observaciones que de ella hagamos

146. ABAD, ADELA y Luis A. Servín. Introducción al muestreo, 69-76

pretendemos generalizarlas de manera válida a toda la población, entonces la distribución de los valores de las variables deberá ser igual en ambas, en la muestra y en la población.

Con ese fin, existen diversos procedimientos y de ellos sólo me referiré someramente a los que estimo que mejor pudieran haberse adaptado a las circunstancias de nuestro estudio.

El primero, el aleatorio simple, consiste en la selección de 'n' elementos entre los 'N' que constituyen la población, de modo que todas las muestras de tamaño 'n' tengan las mismas probabilidades de ser obtenidas. (147)

Podemos conseguir que ese requisito se satisfaga, haciendo la selección de los sujetos al azar, pues así cada uno de ellos y por consiguiente el grupo que formen, la muestra de tamaño 'n', tendrán las mismas probabilidades de ser elegidos.

Un procedimiento para lograr la aleatoriedad, consiste en asignar a los sujetos un número diferente a cada uno, que sirva sólo para identificación. Se consulta a continuación una tabla de números en los que no haya una secuencia discernible.

Al obtenerse de ella el número de identificación de algún sujeto, éste pasará a formar parte de la muestra. Se repite la operación hasta completar el tamaño, previamente fijado, de dicha muestra.

En nuestro caso, si bien no tuvimos la posibilidad de

elegir la muestra al azar, si aplicamos el procedimiento brevemente descrito para dividir al grupo que recibimos, naturalmente formado, en los respectivos de control y experimental. Sobre este punto volveremos a tratar sobre el procedimiento del diseño.

El segundo método que podríamos haber usado para nuestro estudio, es el de muestreo estratificado aleatorio, en el que la población se divide en un cierto número de subpoblaciones o 'estratos' y de cada uno de ellos se eligen al azar los sujetos de la muestra.

Como criterio para formar los estratos podríamos haber utilizado el rango de calificaciones de los alumnos en el curso antecedente de Administración de Proyectos y Obras I, entre otros.

Un tercer tipo de muestreo al que pudimos haber recurrido, es el que se hace por conglomerados. Su característica distintiva radica en sustituir las unidades físicas, elementales o últimas, a las cuales se refiere el estudio, por unidades de muestreo que comprendan un grupo de aquellas.

En la selección de conglomerados "... interesa que cada uno de ellos represente en lo posible a la población y, por tanto que esté constituido por unidades últimas heterogéneas." (148)

Para aplicar este procedimiento hubiera sido necesario seleccionar aleatoriamente al grupo de estudio tomando en cuenta, cuando menos, a todos los del quinto



semestre de la Facultad de Arquitectura, Esta posibilidad tampoco estuvo a nuestro alcance.

Aún así y a pesar de todas las dificultades que nos impidieron hacer la selección de la muestra al azar, no creo que fuera demasiado aventurado suponer que en un grupo naturalmente formado por alumnos que por diversos motivos y de manera voluntaria se han inscrito en él, la distribución de los valores de las variables se asemeje, en buena medida, a la que tendrían en un grupo aleatoriamente formado.

A esta suposición, empero, no estamos en aptitud de darle el debido fundamento y por ello sólo podemos concederle un peso muy relativo en las inferencias que hagamos a partir de los resultados de este estudio.

TESIS CON
FALDA DE ORIGEN

4.2. Selección del diseño.

Del diseño que se elija para realizar una investigación va a depender, en gran medida, que el control de las variables pueda ejercerse en la debida forma.

En los trabajos de tipo experimental, como el presente, el objetivo primordial del investigador consiste en determinar el efecto de la variable independiente sobre la dependiente, distinguiéndolo del que pudieran ejercer las intervinientes, extrañas o exógenas.

Esos efectos, como vimos al tratar sobre las variables extrañas, pueden ser de diversa índole y por esa razón su control se realiza de distintas maneras.

Por una parte, la muestra se escoge con apego a alguna de las técnicas adecuadas para que podamos tener cierto grado de garantía probabilística acerca de su representatividad.

Por la otra y una vez que contamos con un buen fundamento para suponer que la distribución de las variables de entrada es la misma en la muestra que en la población, nos corresponde hacer el intento de controlar a las que pudieran afectar a la dependiente en el transcurso del experimento.

Para ello podemos, a partir de la muestra, formar dos grupos, el de control y el experimental. Al primero no se le aplica la variable activa en tanto que al segundo si se lo somete a su acción.

Con ese procedimiento, al comparar los valores que al

TESS CON
FALLA DE ORIGEN

término del estudio asuma la variable dependiente en cada uno de dichos grupos, estaremos capacitados para hacer una inferencia válida acerca de los efectos de la variable independiente, distinguiéndolos de aquellos que las demás variables pudieron haber ejercido en todos los sujetos de la muestra.

Hay, sin embargo, otra serie de factores que pueden atentar contra la validez de nuestro estudio y que no han sido controlados por la selección de la muestra, ni con su división en un grupo de control y otro experimental. Entre ellos, los que quizá en mayor medida podrían afectarnos son los que cito a continuación: administración de pruebas con sus diversos efectos, instrumentación, regresión estadística, mortalidad experimental en cuanto a la validez interna. Por lo que hace a la externa: interferencia por tratamiento múltiple, efecto Hawthorne, los efectos de la novedad del tratamiento, efectos del experimentador, sensibilización por prueba previa y/o posterior y medición de la variable dependiente.

Para ejercer el debido control sobre esos factores disponemos de algunos diseños que lo hacen posible y a pesar de que la mayoría de los que proponen los autores, en especial Campbell y Stanley, no permiten que todas las posibles causas de invalidez sean controladas, son el mejor recurso a nuestro alcance para intentar dicho control.



Con el fin de seleccionar el diseño que mejor pudiera convenirnos, hemos tomado en cuenta las variables y su tipo; la indole de la hipótesis que manejamos y en general las circunstancias en las que se realiza el experimento, incluyendo las limitantes.

Una de ellas, nos impidió cumplir con todos los requisitos de uno de los diseños que hubiera sido adecuado a nuestro caso, el que Campbell y Stanley llaman 'Diseño de grupo de control con postest únicamente' y F. N. Kerlinger, 'Diseño de sujetos asignados aleatoriamente al grupo de control y al grupo experimental'. El requisito no cumplido es el que consiste en seleccionar a los grupos de control y experimental, ya de una muestra tomada al azar, ya de utilizar este procedimiento para elegirlos a partir de un grupo lo suficientemente grande como para que se pueda considerar que en él están presentes todos los posibles valores de las variables.

Algunas de las consecuencias de dicha falta de cumplimiento han quedado anotadas en el inciso correspondiente al control de las variables extrañas. En él examinamos los efectos de los factores que Bracht y Glass señalan como atentatorios a la validez externa de los experimentos y nos percatamos que por no haber ejercido el debido control sobre la mayoría de ellos, ese tipo de validez sufriría en nuestro estudio un considerable grado de menoscabo.

TESIS CON
FALLA DE ORIGEN

No obstante y con miras a que su validez interna no se viera igualmente disminuida, adopté un diseño similar a los arriba citados cuyo esquema podría ser el siguiente.

R	X	O ₁
R		O ₂

Siendo:

R = Asignación aleatoria a distintos grupos de tratamiento.

X = Exposición de un grupo a una variable experimental.

O = Proceso particular de observación o medición.

A partir de ese diseño introduje la siguiente modificación:

R	X	O ₁	X	O ₂	X	O ₃	X	O ₄	X	O ₅	O ₆
R											O ₇

Resulta así, que en vez de aplicar la variable experimental y la medición en una única instancia al grupo experimental, lo hacemos repetidas veces y por ello quedamos expuestos al efecto de las mediciones sobre las subsecuentes.

El diseño es similar al que proponen los autores citados, tanto por no incluir prueba previa, como porque los miembros de los grupos de control y experimental son asignados al azar.

No cumple, sin embargo con el requisito señalado en cuanto a la elección de la muestra y por ello el carácter de nuestro estudio será, en el mejor de los casos, cuasiexperimental.

Queda en pié, empero, la posibilidad de que algún otro diseño pudiera haber sido más adecuado para la realización del estudio que nos ocupa.



Al respecto, si consideramos que la limitación que en mayor medida nos ha afectado es la que repercute en el grado de representatividad de nuestro grupo, quizá consiguiéramos que fuera un poco más alto aplicando el diseño de Solomon cuyo esquema sería el que sigue:

R	O ₁	X	O ₂
R	O ₃		O ₄
R		X	O ₅
R			O ₆

Este diseño, si bien no resuelve el problema de la selección de los grupos a partir de uno que sea lo suficientemente numeroso, si nos permite, por incluir una prueba previa, adquirir algún conocimiento de los valores de las variables de entrada.

De ello se pueden derivar ciertas ventajas: por una parte tendríamos un punto de comparación para los resultados finales y por la otra, que aún sin recurrir a la selección de la muestra al azar, si lográramos someter a esa prueba previa a otros grupos de alumnos, con base en los datos que obtuviéramos, podríamos formarnos una imagen más clara en cuanto a la representatividad del nuestro.

Con este diseño, según nos indican tanto Campbell y Stanley, como Kerlinger, se controlarían todas las variables que afectan a la validez interna y la mayoría de las que Bracht y Glass señalan como

TESIS CON
FALLA DE ORIGEN

atentatorias a la externa.

La validez de población se podría, hasta cierto punto, controlar con el conocimiento y comparación de los valores de las variables de entrada de varios grupos. Entre las que disminuyen la validez ecológica no se controlarían el efecto Hawthorne, las de la novedad y el desconcierto y en nuestro caso específico tampoco la interferencia por tratamiento múltiple.

Estaríamos sujetos a su efecto por la modificación que introduciríamos al diseño de Solomon, pues además de las pruebas previa y posterior haríamos cinco mediciones a lo largo del curso.

En estas circunstancias el modelo quedaría como sigue:

R	O ₁	X	O ₂	X	O ₃	X	O ₄	X	O ₅	X	O ₆	O ₇
R	O ₁											O ₇
R		X	O ₂	X	O ₃	X	O ₄	X	O ₅	X	O ₆	O ₇
R												O ₇

Podríamos explicar el esquema diciendo que al grupo inicial, con una cierta equivalencia con otros, lograda por el conocimiento de los valores de las variables de entrada, lo dividiríamos aleatoriamente en cuatro. Al primero le aplicaríamos la prueba previa, las cinco mediciones parciales y la final; al segundo la prueba previa y la final; al tercero de ellos, no le aplicaríamos la prueba previa, pero sí las cinco mediciones parciales y la final; por último al cuarto



grupo, sólo esta última.

Ahora bien, a pesar de las ventajas que pudimos haber obtenido al aplicar este diseño, las limitaciones a las que ya he aludido nos lo impidieron y tuvimos que conformarnos con utilizar el que señalé en la primera parte de este inciso.

Las consecuencias que de lo expresado se derivaron las examinaremos en los apartados siguientes.

TESIS CON
FALLA DE ORIGEN

4.3. Procedimiento del diseño.

Una vez seleccionado el modelo de diseño que regirá la realización del experimento, el procedimiento para hacerlo efectivo fue como sigue:

El grupo de alumnos para la asignatura de Administración de Proyectos y Obras II quedó integrado con los que de manera voluntaria se inscribieron en el Taller E para cursar el tercer semestre. Desde ahí hasta terminar el sexto no pudieron cambiarse a otro Taller, de manera que los que en esta ocasión pasaron al quinto, que es el que corresponde a la materia en cuestión, estaban por tercera vez consecutiva en el mismo Taller. El primer día de clase, conforme los alumnos entraban al salón, anotaban su nombre en una lista, correspondiéndole a cada uno un número que sólo serviría para su identificación. Se les explicó cuál era el plan del curso de Administración de Proyectos y Obras I, que es el único antecedente seriado con el que nos ocupa. A continuación se les informó cuál era, en términos generales, el programa que debían haber cubierto. Si la mayoría manifestaba desconocer algún tema que para el curso actual fuera un antecedente indispensable, se les indicaba la bibliografía que podían consultar y que de ser posible, de haber suficiente tiempo, la repasaríamos brevemente. Se les proporcionó, asimismo, la bibliografía mínima para el semestre. En la segunda clase, recibieron la explicación



necesaria sobre la forma en la que se intentaría realizar un experimento con miras a obtener información lo más objetiva que fuera posible sobre el efecto de las pruebas periódicas en el rendimiento escolar. Se les hizo notar, en especial, que desde el punto de vista ético el trabajo experimental era, en ese caso específico, inobjetable, debido a que, a ciencia cierta, en ninguno de los estudios realizados sobre el tema, hasta donde estábamos enterados, había sido posible llegar a conclusiones que permitieran establecer con certidumbre si ese efecto era positivo, negativo o no se producía ninguno. Se les hizo ver, también, la importancia de su participación y colaboración para que el estudio pudiera llevarse a buen fin.

La reacción de los estudiantes fue, casi unánimemente, positiva y un considerable número de ellos mostró decidido interés en que se llevara al cabo ese tipo de experimentos.

Una vez que el grupo se estabilizó, pues a las primeras clases asisten alumnos que sólo quieren enterarse del programa y se inscriben en la lista sin tener, en realidad, intenciones de continuar, se les explicó que el grupo se dividiría, al azar, en dos, que a uno de ellos se les administrarían pruebas parciales, una por cada unidad programática y a todos las finales.

Por haber recibido los alumnos esa información, por saberse copartícipes de un experimento, nos exponemos

tanto el efecto Hawthorne, como a los de la novedad por el tratamiento.

Esos efectos, si bien quizá no sea imposible evitarlos, son muy difíciles de eludir pues en el transcurso del experimento los alumnos se darán cuenta, casi siempre, de que el curso no sigue por los causes a los que están acostumbrados.

Por otra parte, en el estudio que nos ocupa, nuestro principal interés radica en medir los efectos que las pruebas periódicas pueden ejercer en el R-E, lo cual en cierta forma implica el conocimiento del grado en el que el efecto Hawthorne actúa.

Una vez que los alumnos estuvieron enterados del experimento que se intentaba realizar, se procedió a explicarles en qué consiste la selección al azar, porqué era necesario hacerla y cómo se puede usar para ello uno de los programas de una calculadora Hewlett Packard 67.

Con la ayuda de dos alumnos se corrió el programa y los números aleatorios que se iban calculando se escribían en el pizarrón. Al completarse los correspondientes a la mitad del grupo se suspendió la ejecución.

En seguida, cada uno de los grupos formados, ya por los estudiantes cuyos números de identificación aparecían en el pizarrón, ya por aquellos que no habían salido sorteados, eligió un representante para que 'echara un volado', el 'sol' correspondería al grupo

de control y el 'aguila' al experimental.

Al quedar los alumnos informados acerca de sus respectivos papeles en el estudio y al haberlos, no sin alguna objeción, todos ellos aceptado, no estamos sujetos a la observación que John Jung cita en el sentido de que: "En ausencia de un conocimiento informado por parte del sujeto, estas mediciones representan una invasión a la vida privada." () [4] El tercer día de clases se inició la exposición de los temas del curso y a cada una de las unidades programáticas se les dedicó el siguiente número de sesiones:

Introducción, programa, etc.	2	sesiones
3.1. Planos.	4	" "
3.2. Especificaciones.	2	" "
3.3. Costos unitarios.	6	" "
3.4. Cantidades de obra.	2	" "
3.5. Estimados de costo y pronóstico.	5	" "
3.6. Programación.	2	" "
4. Ejecución de la obra.	2	" "
5. Operación del edificio, etc.	1	" "
	<hr/>	26 sesiones
Pruebas parciales.	5	" "
Total.	<hr/>	31 sesiones

De las treinta y dos posibles clases en un semestre se perdió una sólo por la falta de asistencia de los alumnos debido a que fueron, en otra asignatura, a

hacer una visita a un edificio en construcción. El conocimiento que en ella hubieran podido adquirir sería muy útil para mejorar su comprensión en algunos de los temas del curso que nos ocupa, A.P.O. II. En términos generales las unidades programáticas fueron tratadas en la siguiente forma:

3.1. Planos.

Se les explicó a los alumnos el papel que juegan y la importancia que tienen los planos, bien elaborados, para que la obra se pueda realizar sin tropiezos.

Se les informó acerca de los diversos tipos de planos que son necesarios y de las claves con las que cada uno de ellos se identifica, así como de cuáles son los principales datos que deben contener.

Se les indicó la bibliografía que podían consultar para completar y ampliar lo visto en clase.

3.2. Especificaciones.

Se indicó a los alumnos cuáles son los fines con los que se redactan las especificaciones, cuál es la forma adecuada de hacerlo y se escribieron en el pizarrón las claves y datos básicos que deben contener.

3.3. Costos unitarios.

Por ser este uno de los temas centrales del curso y por existir la posibilidad de impartirlo mediante la técnica de solución de problemas, se dieron a

los alumnos las explicaciones mínimas necesarias y a continuación se les planteó, como ejercicio en clase, un ejemplo que ellos tuvieron que resolver. Después de darles un tiempo prudente para que lo intentaran, se les proporcionó la información correctiva pertinente.

Se les hizo ver que los únicos costos unitarios que era posible calcular de la manera en la que se les había explicado eran los de la estructura y la albañilería.

Que para los demás el proceso se invertía, pues primero se calculaban los totales y luego el unitario.

3.4. Cantidades de obra.

Se siguió el mismo procedimiento que para los costos unitarios. Las explicaciones fueron las mínimas, indispensables y en seguida se les ponían ejercicios a resolver en la clase.

3.5. Estimados de costo y pronóstico.

Se explicó a los estudiantes la forma usual en la que se calculan los costos de las instalaciones y los complementos y cuáles son las partidas presupuestales que integran el estimado de costos, así como los porcentajes que cada una de ellas, en promedio, representa en el total.

Se les indicó, también, la manera en la que los índices de costo se pueden usar y la forma en la

que es posible pronosticarlos con un aceptable grado de aproximación.

3.6. Programación.

En forma por demás somera se informó a los alumnos sobre las distintas formas en las que se puede elaborar un calendario de obra y de erogaciones.

Se les explicó en que consiste el C. P. M. (Método de la ruta crítica) y se desarrolló, con su colaboración, un ejemplo sencillo en el pizarrón.

4. Ejecución de la obra.

Antes de abordar el tema, se les hizo ver a los alumnos que en este momento del trabajo se había ya completado el material necesario para la ejecución de la obra y que con base en él era posible proceder a la siguiente etapa, la construcción del edificio.

Se trataron todos los pasos a seguir, desde los trámites oficiales indispensables, la elaboración de los contratos en relación al procedimiento elegido: por administración o por contrato, la supervisión de los trabajos y la recepción y liquidación de los mismos.

5. Operación.

Se hizo un examen de cuáles eran los aspectos en los que mayor cuidado había que poner una vez que el edificio estuviera funcionando: conservación y mantenimiento y de la importancia de evaluar los resultados obtenidos así como del proceso de realimentación.

Al término de cada unidad programática se citaba al grupo experimental y se le administraba la prueba respectiva. Para ello se usó el mismo salón en el que se daba la clase y se concedió un tiempo igual al de cada una de ellas, una hora y media. Ese lapso resultó suficiente para casi todos los alumnos. Se les autorizó a consultar los libros o apuntes que consideraran conveniente.

Se procuró que los días para prueba fueran los jueves, de manera que en la siguiente sesión, martes, recibieran la información acerca de las calificaciones que habían obtenido. Debido tanto a que en presencia de los alumnos del grupo de control se les proporcionó la información correctiva, como a que, muy probablemente estos obtenían la relacionada con cada una de las pruebas, nos será muy difícil distinguir los efectos que las mediciones periódicas hayan podido tener en los grupos respectivos.

Podemos, ya desde ahora, afirmar que los efectos de la variable experimental no se circunscribieron al grupo al que intencionalmente se le administró, sino que en realidad se extendieron a la totalidad de los alumnos. Para que ese tipo de contaminación no nos hubiera afectado tendríamos que haber evitado el intercambio de información que hubo entre los grupos experimental y de control y esto es imposible lograrlo cuando todos los alumnos pertenecen al mismo Taller,

Al final del curso, a los dos grupos se les administró el examen final que fue el mismo para todos los alumnos. Al análisis de los resultados obtenidos dedicaremos los siguientes apartados.

Presentación y análisis de los resultados.

En este apartado, veremos en primer lugar, cuáles fueron las calificaciones que los alumnos obtuvieron en los exámenes finales de los cuatro semestres en los que, hasta ahora, se ha realizado el experimento.

Los criterios aplicados para la asignación de puntuaciones se especifican en el inciso 3.8., dedicado a los aparatos e instrumentos que fueron administrados. A continuación haremos un análisis descriptivo de los resultados y después los someteremos al tratamiento estadístico adecuado para que las inferencias, que por último trataremos de extraer tengan la necesaria validez.

Los resultados de los exámenes aparecen en la Tablas 1 y 2, así como en los Histogramas o Gráficas 1, 2, 3 y 4.

Análisis descriptivo de los resultados.

Para este análisis aplicaremos los principios de la estadística descriptiva, pues las normas que nos proporciona para ordenar y clasificar los datos hace posible que los comprendamos mejor.

Para ese fin podemos recurrir a los polígonos de

frecuencias, tables e histogramas como los que están a continuación incluidos.

Una vez ordenados y clasificados los datos, procedemos a calcular las medidas de tendencia central que de manera principal son tres.

La primera de ellas es la media aritmética, es decir el promedio de las puntuaciones que se obtiene aplicando la siguiente ecuación:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Siendo:

\bar{x} = Media de la muestra

n = Número de datos.

x_i = Cada uno de los datos.

La segunda es la mediana, que en las distribuciones de frecuencias agrupadas, se define como el dato o el dato potencial de una distribución, por arriba y por debajo del cual caen la mitad de las frecuencias.

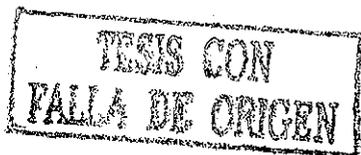
Las medianas para nosotros son las siguientes:

1er. semestre.	Examen A.	G. Experimental	Mdn. = 5
	Examen A.	G. de control	" = 5
	Examen B.	G. Experimental	" = 8
	Examen B.	G. de control	" = 8
2o. semestre	Examen A.	G. Experimental	" = 6
	Examen A.	G. de control	" = 6
	Examen B.	G. Experimental	" = 7

	Examen B.	G. de control	Mdn. = 7
3er. semestre	Examen A.	G. Experimental	" = 9
	Examen A.	G. de control	" = 9
	Examen B.		-
4o. semestre	Examen A.	G. Experim.	" = 7
	Examen A.	G. Control.	" = 8
	Examen B.	G. Experim.	" = 5
	Examen B.	G. Control.	" = 6

La tercera medida de tendencia central es el modo, el número que en un conjunto de datos ocurre con mayor frecuencia y que en nuestro caso tiene los siguientes valores:

1er. Semestre	Examen A.	G. Experim. Mod. = 5
	Examen A.	G. Control. Mod. = 5
	Examen B.	G. Experim. Mod. = 8
	Examen B.	G. Control. Mod. = 8
2o. semestre	Examen A.	G. Experim. Mod. = 5
	Examen A.	G. Control. Mod. = 5 y 6
	Examen B.	G. Experim. Mod. = 6, 7 y 8
	Examen B.	G. Control. Mod. = 7 y 8
3er. semestre	Examen A.	G. Experim. Mod. = 9
	Examen A.	G. Control. Mod. = 9
	Examen B.	- -
4o. semestre.	Examen A.	G. Experim. Mod. = 8
	Examen A.	G. Control. Mod. = 8
	Examen B.	G. Experim. Mod. = 5
	Examen B.	G. Control. - -



Otra característica que nos interesa conocer es la tendencia de los datos hacia una mayor o menor dispersión en relación con la media general, una de cuyas medidas es la desviación estándar que se calcula aplicando la siguiente ecuación:

$$s = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n - 1}}$$

Siendo:

\bar{x} = Media de la muestra.

x_i = Cada uno de los datos.

n = Número de datos.

s = Desviación estándar de la muestra.

La otra medida de dispersión es la variancia y es igual al cuadrado de la desviación estándar.

$$v = s^2$$

Y si en lugar de los datos de la muestra, manejamos los que corresponden a la población, entonces:

= Media de la población

= Desviación estándar de la población.

Para su cálculo usamos las ecuaciones anotadas, sustituyendo $n-1$, por n .

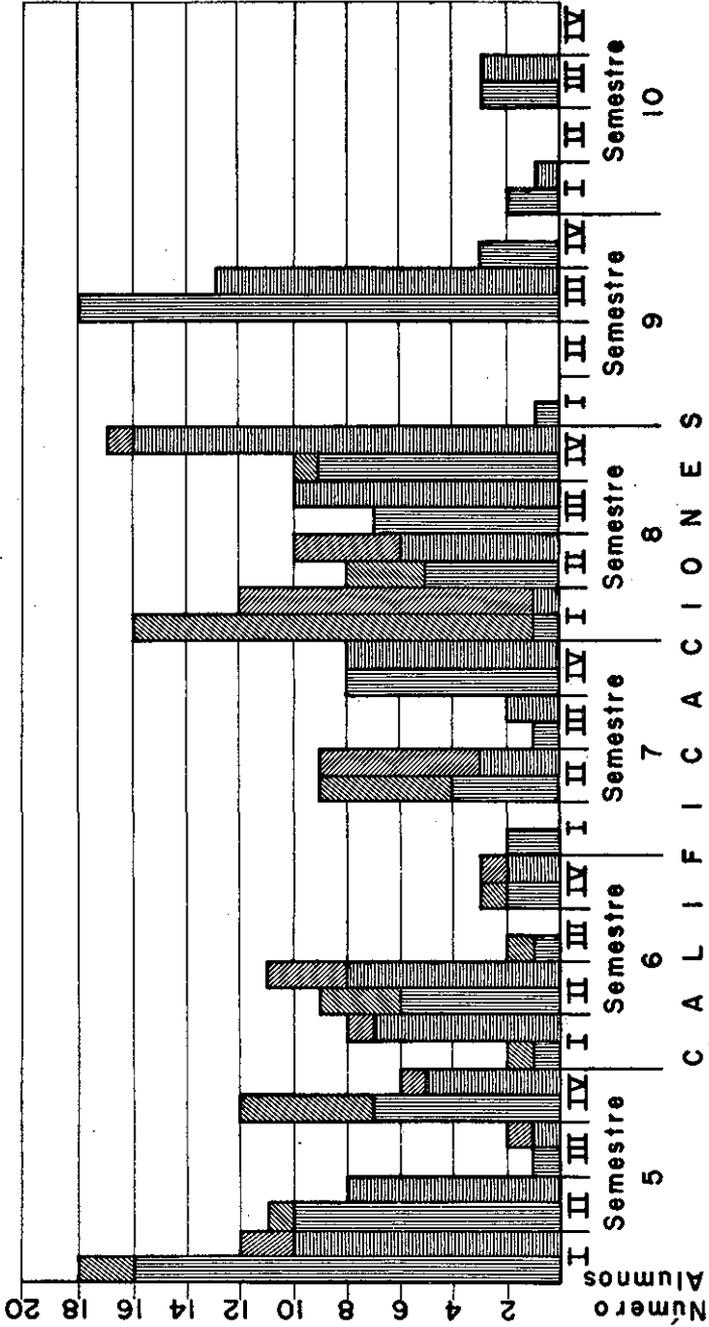
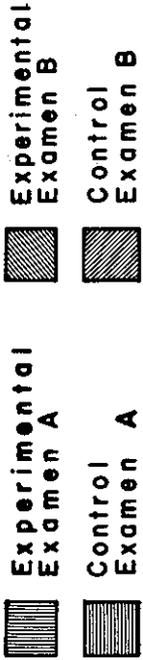
Con las puntuaciones que obtuvimos en nuestro estudio, las medidas de tendencia central y de dispersión se muestran en las tablas y gráficas respectivas.

Semestre	Primero		Segundo		Tercero		Cuarto	
	Número Alumnos	Calificación						
Examen Final A Experimental	16	5	10	5	1	5	7	5
	1	6	6	6	1	6	2	6
	2	7	4	7	1	7	8	7
	1	8	5	8	7	8	9	8
	2	9	-	9	9	18	9	9
	<u>23</u>	<u>25</u>	<u>25</u>	<u>25</u>	<u>31</u>	<u>29</u>	<u>29</u>	<u>10</u>
Suma No. Alumnos Control	10	5	8	5	1	5	5	5
	7	6	8	6	-	6	2	6
	1	7	3	7	2	7	8	7
	1	8	6	8	10	8	16	8
	1	9	-	9	13	9	-	9
	<u>19</u>	<u>25</u>	<u>25</u>	<u>29</u>	<u>31</u>	<u>31</u>	<u>10</u>	<u>10</u>
Suma No. Alumnos Examen Final B Experimental	2	5	1	5	1	5	5	5
	1	6	3	6	1	6	1	6
	15	7	3	7	1	7	1	7
	1	8	3	8	1	8	1	8
	1	9	-	9	1	9	1	9
	<u>19</u>	<u>10</u>	<u>10</u>	<u>10</u>	<u>7</u>	<u>7</u>	<u>10</u>	<u>10</u>
Suma No. Alumnos Control	2	5	1	5	1	5	1	5
	1	6	3	6	1	6	1	6
	12	7	4	7	1	7	1	7
	1	8	4	8	1	8	1	8
	2	9	-	9	1	9	1	9
	<u>17</u>	<u>12</u>	<u>12</u>	<u>12</u>	<u>1</u>	<u>1</u>	<u>10</u>	<u>10</u>

TESIS CON FALLA DE ORIGEN

TABLA DE RESULTADOS

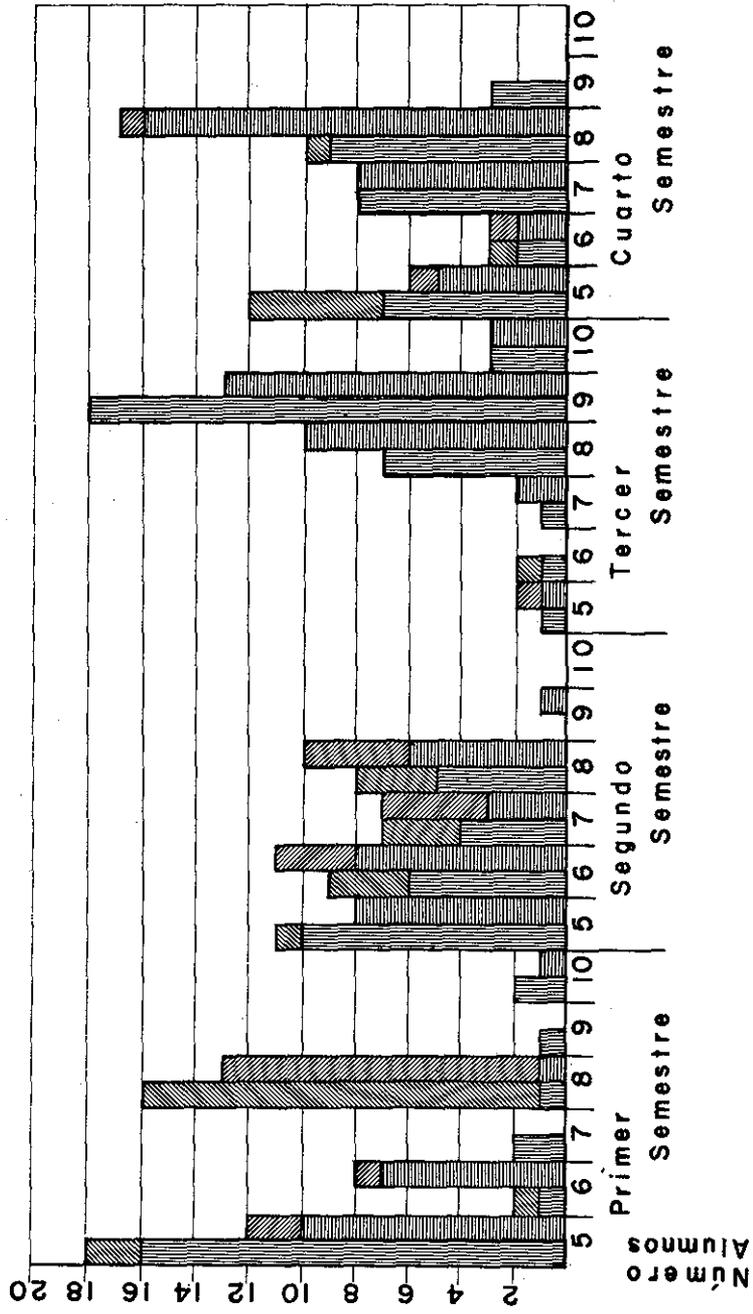
SEGUN CALIFICACION



TESIS CON FALLA DE ORIGEN

Gráfica 2
 TABLA DE RESULTADOS
 POR SEMESTRES

Experimental Examen A
 Control Examen A
 Experimental Examen B
 Control Examen B



TESIS CON
 FALLA DE ORIGEN

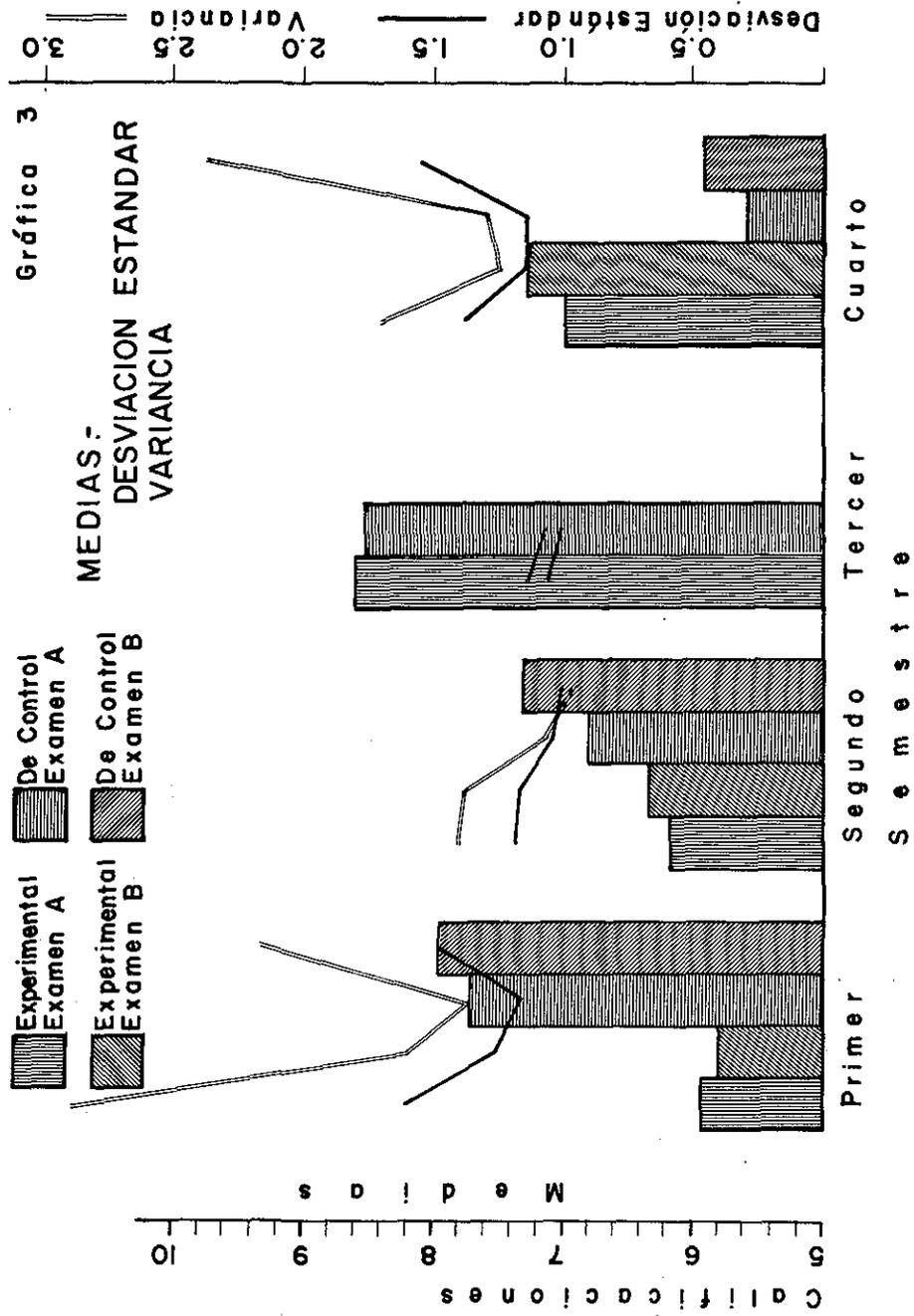
TABLA 2

ANALISIS DESCRIPTIVO DE LOS RESULTADOS

	GRUPO EXPERIMENTAL		GRUPO DE CONTROL	
	NÚMERO ALUMNOS.	MEDIDAS	NÚMERO ALUMNOS.	MEDIDAS
PRIMER SEMESTRE				
EXAMEN FINAL A	23	$\bar{X} = 5.95$ $s = 1.69$	19	$\bar{X} = 5.79$ $s = 1.27$
EXAMEN FINAL B	19	$\bar{X} = 7.68$ $s = 1.16$	17	$\bar{X} = 7.89$ $s = 1.47$
SEGUNDO SEMESTRE				
EXAMEN FINAL A	25	$\bar{X} = 6.16$ $s = 1.18$	25	$\bar{X} = 6.28$ $s = 1.17$
EXAMEN FINAL B	10	$\bar{X} = 6.80$ $s = 1.03$	12	$\bar{X} = 7.25$ $s = 0.97$
TERCER SEMESTRE				
EXAMEN FINAL A	31	$\bar{X} = 8.58$ $s = 1.06$	29	$\bar{X} = 8.48$ $s = 1.02$
EXAMEN FINAL B	0		2	$\bar{X} = 5.50$ $s = 0.71$
CUARTO SEMESTRE				
EXAMEN FINAL A	29	$\bar{X} = 6.97$ $s = 1.35$	31	$\bar{X} = 7.13$ $s = 1.12$
EXAMEN FINAL B	7	$\bar{X} = 5.57$ $s = 1.13$	3	$\bar{X} = 6.33$ $s = 1.53$



TESIS CON FALLA DE ORIGEN



De las tablas y gráficas anteriores se desprende que las diferencias intrasemestrales en las medidas estadísticas incluidas para los grupos experimentales y de control son, en realidad, mínimas.

En los tres primeros semestres las medianas para los grupos experimentales y de control son iguales y para el cuarto las diferencias alcanzan un sólo punto.

El modo asume también los mismos valores para los dos grupos en los cuatro semestres.

En las medias las diferencias máximas son: 0.21 de punto en el primer semestre, 0.45 en el segundo, 0.10 en el tercero y 0.16 en el cuarto.

Todo ello nos faculta a hacer una primera suposición, que podrá ser confirmada o no al aplicar a esos datos el tratamiento que indica la estadística inferencial, en el sentido de que no hay diferencias significativas entre los resultados obtenidos por los grupos experimentales y de control en ninguno de los cuatro semestres considerados.

No trataré de establecer comparaciones entre los resultados de los períodos A y B de exámenes, porque la información que los alumnos recibieron después del primero, me parece que la invalida en cuanto a que la mejoría observada pudiera ser efecto de la aplicación de la variable activa.

Aún así, dicha mejoría fue la misma para los grupos experimentales y de control.

TESIS CON
FALLA DE ORIGEN

Sin embargo, al hacer la comparación intersemestral de los resultados si encontramos diferencias. Los resultados se observan en la siguiente tabla.

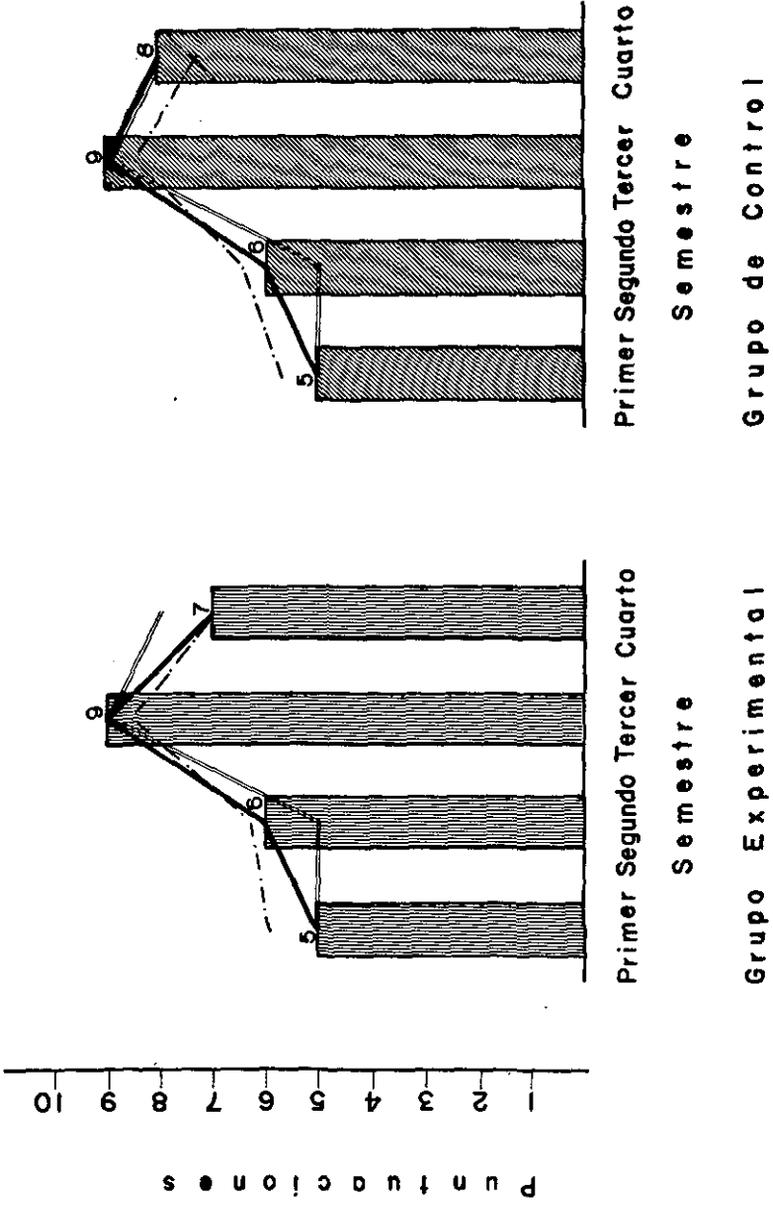
MEDIDAS.	1er. Semes.	2o. Semes.	3er. Semes.	4o. Semes.
<u>Mediana.</u>				
G. Experim.	5	6	9	7
G. Control.	5	6	9	8
<u>Modo.</u>				
G. Experim.	5	5	9	8
G. Control.	5	5	9	8
<u>Media</u>				
G. Experim.	5.95	6.16	8.58	6.97
G. Control.	5.79	6.28	8.48	7.13

La mejoría es notoria del primero al tercero de los semestres y aunque la tendencia se invierte en el cuarto, se conserva en relación con los dos primeros. (ver gráfica) No obstante que las diferencias son apreciables, antes de que pretendamos extraer alguna inferencia es necesario que sometamos los datos al tratamiento que indica la estadística adecuada para ello, como lo haremos en el siguiente inciso.

TESIS CON
FALLA DE ORIGEN

Gráfica 4
 TABLA DE MEDIANAS

Medianas ———
 Modo ———
 Media - - - - -



TESIS CON
 FALLA DE ORIGEN

5.2. Tratamiento estadístico de los datos.

De la misma manera que en el apartado anterior recurrimos al auxilio de la estadística descriptiva, para los análisis que haremos a continuación nos apoyaremos en la rama inferencial de dicha disciplina. Por la índole de los datos que manejamos, pues en estricto rigor no podemos asegurar que los intervalos en nuestra escala de puntuaciones sean iguales, aplicaremos las técnicas no paramétricas al hacer las comparaciones intrasemestrales de los datos obtenidos. En el análisis itersemestral usaremos, de manera primordial, las mismas técnicas, ya que como resultado de no haber seleccionado los sujetos al azar, a la circunstancia arriba indicada hemos de aunar nuestra incapacidad para aseverar con el debido fundamento, que la distribución de los valores de las variables en nuestros grupos se ajuste, ya a la curva normal, ya a la resultante de la ecuación de Student.

a) Análisis intrasemestral.

Haremos en primer lugar, el intento de determinar si las diferencias en los resultados son significativas aplicando la prueba de la 'ji cuadrada'. Esta prueba es útil para encontrar si la hipótesis alterna, consistente en que dos grupos difieren con respecto a cierta característica, se puede aceptar. En nuestro caso hallaríamos la diferencia al comparar

las puntuaciones obtenidas por los grupos experimentales con las logradas por los de control.

Para el cálculo de la 'ji cuadrada' se aplica la siguiente ecuación:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Siendo:

O_{ij} = Número de casos observados en la fila i de la columna j .

E_{ij} = Número de casos esperados conforme H_0 en la fila i de la columna j .

$\sum_{i=1}^r \sum_{j=1}^k$ = Sumas de las filas r y columnas k

gl. = Grados de libertad. $(r-1)(k-1)$

Sustituyendo las literales de la ecuación con los datos que se consignan en la tabla 3, obtenemos los resultados que siguen:

Primer semestre. $\chi^2 = 8.89$ 5 gl. $p(H_0) > 0.16$

Segundo Semestre. $\chi^2 = 0.74$ 3 gl. $p(H_0) > 0.87$

Tercer semestre. $\chi^2 = 2.57$ 5 gl. $p(H_0) > 0.75$

Cuarto semestre. $\chi^2 = 6.55$ 4 gl. $p(H_0) > 0.15$

En consecuencia, debido a que las probabilidades para la hipótesis de nulidad (H_0) que hemos calculado fluctúan entre 0.15 y 0.87, en ningún caso podemos considerar que el criterio para rechazarla y aceptar la alterna se haya cumplido, pues para ello su probabilidad tendría

Tabla 3

ANALISIS INTRASEMESTRAL - PRUEBA χ^2 DATOS c.o.=Casos Observados
c.e.=Casos Esperados

Calificación		5	6	7	8	9	10	Totales
Primer Semestre								
Experimental	c.o.	16	1	2	1	1	2	23
	c.e.	14.24	4.38	1.10	1.10	0.55	1.64	23
Control	c.o.	10	7	0	1	0	1	19
	c.e.	11.76	3.62	0.90	0.90	0.45	1.36	19
Sumas	c.o.	26	8	2	2	1	3	42
	c.e.	26	8	2	2	1	3	42
Segundo Semestre								
Experimental	c.o.	10	6	4	5	-	-	25
	c.e.	9	7	3.5	5.5	-	-	25
Control	c.o.	8	8	3	6	-	-	25
	c.e.	9	7	3.5	5.5	-	-	25
Sumas	c.o.	18	14	7	11	-	-	50
	c.e.	18	14	7	11	-	-	50
Tercer Semestre								
Experimental	c.o.	1	1	1	7	18	3	31
	c.e.	1.03	0.52	1.55	8.78	16.78	3.10	31
Control	c.o.	1	-	2	10	13	3	29
	c.e.	0.97	0.48	1.45	8.22	14.98	2.90	29
Sumas	c.o.	2	1	3	17	31	6	60
	c.e.	2	1	3	17	31	6	60
Cuarto Semestre								
Experimental	c.o.	7	2	8	9	3	-	29
	c.e.	5.80	1.93	7.73	12.08	1.45	-	29
Control	c.o.	5	2	8	16	-	-	31
	c.e.	6.20	2.07	8.27	12.92	1.54	-	31
Sumas	c.o.	12	4	16	25	3	-	60
	c.e.	12	4	16	25	3	-	60

TESIS CON FALLA DE ORIGEN

que ser igual a 0.05 como máximo.

Estamos, no obstante, en posibilidad de hacer un segundo intento para confirmar o modificar la apreciación anterior.

Con ese fin aplicaremos la prueba de Mann-Whitney, adecuada para probar "...si dos grupos independientes han sido tomados de la misma población. Es una de las pruebas más poderosas y constituye la alternativa más útil ante la prueba paramétrica 't' cuando el investigador desea evitar las suposiciones que ésta exige o si la medición en la investigación es más vaga que la escala de intervalos." (150)

Por lo tanto, calcularemos la 'U' de Mann-Wtiney de acuerdo con la siguiente ecuación:

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 \cdot n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Siendo:

n_1 = Número de datos de la muestra 1

n_2 = Número de datos de la muestra 2

R_1 = Suma de los rangos de la muestra 1

R_2 = Suma de los rangos de la muestra 2

Resolviendo las ecuaciones con los datos que aparecen en la tabla 4, obtenemos los siguientes valores:

PRUEBA "U" DE MANN-WHITNEY

Semestre	Primero			Segundo			Tercero			Cuarto			
	Número Casos	Califi- cación	± Rangos Región	Número Casos	Califi- cación	± Rangos Región	Número Casos	Califi- cación	± Rangos Región	Número Casos	Califi- cación	± Rangos Región	
Experimental	16	5	472	10	5	415	1	5	59.5	7	5	381.5	
	1	6	12.5	6	6	153	1	6	58	2	6	93	
	2	7	15	4	7	60	1	7	56	8	7	292	
	1	8	5.5	5	8	30	7	8	322	9	8	144	
	1	9	4	—	9	—	18	9	396	3	9	6	
	2	10	4	—	10	—	3	10	35	—	10	—	
	23		513	25		658	31		926.5	29		916.5	
	Control	10	5	295	8	5	332	1	5	59.5	5	5	272.5
		7	6	87.5	8	6	204	—	6	—	2	6	93
		—	7	—	3	7	45	2	7	112	8	7	292
1		8	5.5	6	8	36	10	8	460	16	8	256	
—		9	—	—	9	—	13	9	286	—	9	—	
1		10	2	0	10	—	3	10	35	—	10	—	
19			390	25		617	29		952.5	31		913.5	

TESIS CON
FALLA DE ORIGEN

Primer semestre.	$U_1 = 200$	$U_2 = 237$
Segundo semestre.	$U_1 = 292$	$U_2 = 333$
Tercer semestre.	$U_1 = 468$	$U_2 = 381$
Cuarto semestre.	$U_1 = 478$	$U_2 = 420$

Tomando en cuenta que en casi todos los casos las muestras exceden de 20 sujetos y en todos la mayor sobrepasa ese límite, podemos suponer, siguiendo a Mann-Whitney, que la distribución de las variables en ellas se acerca rápidamente a la normal y por ende la siguiente ecuación es aplicable.

$$z = \frac{U - \frac{n_1 \cdot n_2}{2}}{\sqrt{\left[\frac{n_1 \cdot n_2}{N(N-1)} \right] \left[\frac{N^3 - N}{12} - \sum T. \right]}}$$

Siendo:

n_1 = Número de datos de la muestra 1

n_2 = Número de datos de la muestra 2

N = Número total de datos.

T = $\sum \frac{t^3 - t}{12}$

t = Número de datos iguales.

z = $\frac{U - \text{Valor de la media}}{\text{Desviación estándar}}$

Sustituyendo las literales con los valores calculados para la 'U' y con los datos que figuran en la tabla 4, los resultados son los siguientes.

Primer semestre. $z_1 = 0.48$ $z_2 = 0.48$ $p(H_0) > 0.31$

Segundo semestre. $z_1 = 0.40$ $z_2 = 0.40$ $p(H_0) > 0.34$

Tercer semestre. $z_1 = 0.29$ $p(H_0) > 0.39$

$z_2 = -1.02$ $p(H_0) > 0.15$

Cuarto semestre. $z_1 = 0.43$ $p(H_0) > 0.33$

$z_2 = -0.44$ $p(H_0) > 0.33$

En los valores que hemos obtenido para la probabilidad de la hipótesis de nulidad (H_0), que van de 0.15 a 0.39, no encontramos fundamento para rechazarla, pues de acuerdo con el criterio que sustentamos sólo cuando son menores a 0.05 podemos aceptar la hipótesis alterna. Los resultados de las dos pruebas, no paramétricas, la 'ji cuadrada' y la 'U' de Mann-Whitney aplicadas al análisis intrasemestral de los resultados, no hacen sino dar apoyo a la apreciación que hicimos con base en el análisis descriptivo de los mismos, en el sentido de que no existe diferencia estadísticamente significativa entre las puntuaciones obtenidas en el examen final por los alumnos de los grupos experimentales y de control en ninguno de los cuatro semestres considerados. Veremos a continuación, si al comparar entre ellos los resultados de los cuatro semestres se confirma también la apreciación inicial, consistente en que en este caso

si hay diferencias significativas.

b) Análisis intersemestral de los datos.

Con la finalidad indicada, determinar si existen o no diferencias estadísticamente significativas entre los resultados correspondientes a los cuatro semestres incluidos en el estudio, emplearemos en primer lugar una prueba no paramétrica, la 'ji cuadrada', aplicable, como antes vimos, a casos en los que no se puede asegurar que los datos correspondan a una escala de intervalos iguales.

Esta prueba, al igual que las otras paramétricas, tampoco requiere que se hagan suposiciones en cuanto a que la distribución de los valores de las variables se ajuste a alguna de las curvas matemáticas correspondientes. Para calcular la 'ji cuadrada' usaremos la ecuación que aparece en la página y los datos que se consignan en la tabla 5.

Haciendo las operaciones necesarias obtenemos:

$$\chi^2 = 151.79 \quad \text{gl. } 15 \quad p(H_0) < 0.001$$

Por ser la probabilidad para que la hipótesis de nulidad se confirme de 0.001, inferior a 0.05 que es el límite que hemos fijado para aceptarla, podemos considerar que cabe su rechazo a afirmar que si existen diferencias estadísticamente significativas entre los resultados de los cuatro semestres incluidos.

Sin embargo, para dar mayor apoyo a la inferencia que de

ANALISIS INTERSEMESTRAL
 PRUEBA χ^2 - DATOS

Tabla 5

Calificaciones	5	6	7	8	9	10	Totales
Períodos							
Primer Semestre							
Casos Observados	26	8	2	2	1	3	42
Casos Esperados	11.49	5.35	5.55	10.90	6.93	1.78	42
Segundo Semestre							
Casos Observados	18	14	7	11	-	-	50
Casos Esperados	13.68	6.37	6.60	12.97	8.26	2.12	50
Tercer Semestre							
Casos Observados	2	1	3	17	31	6	60
Casos Esperados	16.42	7.64	7.92	15.57	9.91	2.54	60
Cuarto Semestre							
Casos Observados	12	4	16	25	3	-	60
Casos Esperados	16.42	7.64	7.92	15.57	9.91	2.54	60
S u m a s							
Casos Observados	58	27	28	55	35	9	212
Casos Esperados	58	27	28	55	35	9	212

acuerdo con la prueba anterior estamos en aptitud de extraer, en el sentido de que las mediciones periódicas tienden a mejorar el R-E, someteremos nuestros datos a otra prueba no paramétrica, el análisis de variancia de una clasificación por rangos de Kruskal-Wallis. Esta prueba es útil para decidir si k muestras independientes son de poblaciones diferentes. La hipótesis de nulidad consiste en suponer que las k muestras proceden de la misma población o de poblaciones idénticas con respecto a los promedios. P. 10

En nuestro caso, por ser la muestra de un tamaño mediano y por contener datos ligados, una misma puntuación para varios sujetos, usamos la siguiente ecuación.

$$H = \frac{\frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)}{1 - \frac{\sum T}{N^3 - N}}$$

Siendo:

N = Número total de datos.

n_j = Número de datos de cada grupo

R_j = Suma de rangos para cada grupo.

$\sum T$ = Suma ($t^3 - t$), t = número de datos iguales

Sustituyendo las literales de la ecuación con los valores que aparecen en la página siguiente, obtenemos los siguientes resultados:



ANÁLISIS INTESEMESTRAL
PRUEBA DE KRUSKAL-WALLIS

Tabla 6
DATOS

S e m e s t r e												
P r i m e r o			S e g u n d o			T e r c e r o			C u a r t o			
Número Casos	Califi. cación	Σ Rangos Renglón	Número Casos	Califi. cación	Σ Rangos Renglón	Número Casos	Califi. cación	Σ Rangos Renglón	Número Casos	Califi. cación	Σ Rangos Renglón	
26	5	4,771	18	5	3,303	2	5	367	12	5	2,202	
8	6	1,128	14	6	846	1	6	141	4	6	564	
2	7	227	7	7	794	3	7	340	16	7	1,816	
2	8	144	11	8	792	17	8	1,224	25	8	1,800	
1	9	27	—	9	—	31	9	837	3	9	81	
3	10	15	—	10	—	6	10	30	—	10	—	
42		6,312	50		5,735	60		2,939	60		6,463	
T o t a l e s												

TESIS CON
FALLA DE ORIGEN

$$H - 21.78 \text{ gl } 3 \quad p(H_0) < 0.001$$

De la misma manera que al aplicar la prueba de la 'ji cuadrada', la probabilidad de que se verifique la hipótesis de nulidad es sumamente baja, 0.001 y estamos por ello en aptitud de rechazarla y aceptar que el análisis estadístico de los resultados tiende a dar apoyo a la alterna en la dirección ya indicada, de que entre las puntuaciones de los diversos semestres si existen diferencias significativas. Por último y para dar un mejor fundamento a los resultados de las pruebas anteriores, aplicaremos ahora una de tipo paramétrico.

Lo haremos sin ignorar que los datos que manejamos, al no poder asegurar que correspondan a una escala de intervalos iguales, no son del todo adecuados para esa clase de prueba. Tampoco ignoramos que, al no haber sido seleccionados los sujetos al azar, carecemos de la base teórica necesaria para establecer la suposición de que la distribución de los valores de las variables se ajusta a alguna de las curvas matemáticas respectivas. Aún en esa situación, el rigor matemático del análisis de variancia en dos sentidos, nos será útil y quizá decisivo para normar nuestro criterio en cuanto a que aceptemos o no la hipótesis alterna.

Los resultados de la aplicación de dicha prueba son los que abajo se consignan:



Datos	\bar{x} Exper,	\bar{x} Control	
Primer semestre.	5.96	5.79	Suma $R_1 = 11.75$
Segundo semestre	6.16	6.28	Suma $R_2 = 12.44$
Tercer semestre.	8.58	8.48	Suma $R_3 = 17.06$
Cuarto semestre.	6.97	7.13	Suma $R_4 = 14.10$
Sumas columnas.	<u>26.67</u>	<u>27.86</u>	
Suma cuadrados renglones.		8.38	
Suma cuadrados columnas.		0.000013	gl ₁ 3
Suma cuadrados errores.		0.039	gl ₂ 1
Suma total cuadrados.		8.421	
$F_1 = 212.54$		$p(H_0) < 0.01$	
$F_2 = 0.001$		$p(H_0) < 0.20$	

La aplicación del análisis de variancia en dos sentidos nos permite comprobar, tanto las diferencias en los resultados entre uno y otro de los semestres, como las existentes entre los grupos experimentales y de control en cada uno de los ciclos escolares.

En la primera comparación de los resultados, intersemestral, el valor de F_1 es igual a 212.45 y la probabilidad de la hipótesis de nulidad es de 0.01, inferior a 0.05 y por ello podemos afirmar, de acuerdo con el criterio adoptado, que cabe su rechazo, así como la aceptación de la alterna consistente, como ha quedado ya indicado,



en que las mediciones periódicas si afectan al R-E y tienden a mejorarlo.

En la segunda, el valor de la probabilidad para la hipótesis de nulidad está por encima de 0.20, superior al límite que nos fijamos para rechazarla. En consecuencia hemos de aceptarla en el sentido de que, en cada uno de los cuatro semestres incluidos en el estudio, no existen diferencias estadísticamente significativas entre los resultados alcanzados por los grupos experimentales y de control.

TESIS CON
FALLA DE ORIGEN

5.3. Conclusiones de la investigación.

De los resultados obtenidos en este estudio podemos obtener algunas conclusiones particulares en relación con la hipótesis que sustentamos consistente en que: si administramos pruebas periódicas para medir el R-E en cada una de las unidades programáticas del curso de A. P. O. II; entonces habrá una mejoría en el desempeño académico de los alumnos de dicha asignatura.

De ella, derivamos como principal consecuencia contrastable que las puntuaciones, en el examen final, de los alumnos de los grupos experimentales, sujetos a la variable activa, serían más altas que las respectivas de quienes integran los de control; no afectados por dicha variable. Los datos que obtuvimos en los exámenes finales de los cuatro semestres considerados fueron analizados en diversas formas.

En primer lugar, a la luz de los principios de la estadística descriptiva y nos dimos cuenta de que las



diferencias en las medias, modos y medianas entre los grupos experimentales y de control son mínimas. Nos percatamos asimismo de que entre las puntuaciones correspondientes a los diversos semestres si parece haber diferencias y que éstas podrían ser significativas. Para comprobar la veracidad de esas apreciaciones, para establecer si recibían el apoyo de la estadística inferencial; sometimos los datos a varias pruebas no paramétricas: la 'ji cuadrada' y la 'U' de Mann-Whitney, al comparar los resultados que corresponden a los grupos experimentales con los que pertenecen a los de control en cada uno de los cuatro semestres. Empleamos asimismo, dos pruebas no paramétricas y una paramétrica para cerciorarnos de la significatividad de las diferencias entre las puntuaciones observadas en los sucesivos ciclos escolares. Estas pruebas fueron: la 'ji cuadrada' y el análisis de variancia por rangos de Kruskal-Wallis -no paramétricas- y el análisis de variancia en dos sentidos -paramétrica-, cuyos resultados examinaremos enseguida.

La 'ji cuadrada' y la 'U' de Mann-Whitney, rindieron como probabilidad para la hipótesis de nulidad valores muy por encima de los que en las ciencias sociales se consideran como el máximo para rechazarla y, en consecuencia, hemos de aceptarla.

Al no haber diferencias estadísticamente significativas entre las puntuaciones de los grupos de control y



experimentales, estamos capacitados para derivar algunas inferencias,

a) La primera de ellas: que las mediciones periódicas no tuvieron ningún efecto en el R-E.

b) la segunda: Que si lo tuvieron, pero que fue el mismo en los alumnos de ambos grupos.

De ellas, me inclino por la segunda por las siguientes razones:

Por una parte, la separación de los alumnos de los grupos experimentales y de control fue, a todas luces insuficiente; no sólo en las clases y en el Taller en general hubo entre ellos contacto permanente, sino que la información correctiva relativa a las pruebas presentadas por los grupos experimentales la recibían de consumo.

Por otra parte y pesar de que los datos con los que cuento acerca de las puntuaciones obtenidas por los alumnos en los años o semestres en los que no se administraban pruebas periódicas carecen de orden y no son producto de observaciones continuas y sistemáticas, tengo la clara impresión de que eran, en general, inferiores a las más recientes, a las que corresponden a los cuatro semestres en los que se realizó el trabajo experimental.

Por lo que hace a la comparación de los resultados de los diversos semestres entre sí, las tres pruebas: 'ji cuadrada', análisis de variancia por rangos de

TESIS CON
FALLA DE ORIGEN

Kruskal-Wallis y análisis de variancia en dos sentidos, rindieron para la hipótesis de nulidad valores inferiores al de 0,05 que consideramos como mínimo para aceptarla y podemos en consecuencia decir que las pruebas estadísticas están dando apoyo a la central, en el sentido de que las mediciones periódicas si afectan y tienden a mejorar el R-E.

De ser así, nos queda aún una pregunta por contestar: ¿ por qué los resultados del primero al tercero de los semestres mejoraron y tuvieron un cierto decaimiento en el cuarto, si en todos ellos se aplicó el mismo procedimiento y fue igual la variable activa ?

Para explicar ese fenómeno, sólo puedo proponer como posibles las siguiente causas:

- a) la mejoría del primero al tercero de los semestres se debió a que el procedimiento empleado, si bien fue en esencia el mismo, comp producto de las observaciones hechas en sus sucesivas aplicaciones, se estuvo puliendo y depurando, de manera que su efectividad se incrementó,
- b) El decaimiento del tercero al cuarto semestre, quizá se haya debido a la disminución en el efecto de la novedad en el tratamiento, pues en ese ciclo se aplicó por cuarta vez consecutiva.
- c) No podemos, por último, descartar la posibilidad de que las diferencias observadas en los resultados hayan sido producidas por una desigual distribución de las variables de entrada en los alumnos integrantes de los cuatro grupos.



Conclusiones Generales.

El estudio experimental que nos ha ocupado a lo largo de este trabajo, me permite en forma breve y a manera de conclusiones, resaltar algunos aspectos que considero dignos de ser motivo de especial revisión en las investigaciones pedagógicas.

- a) En atención a la hipótesis de trabajo, ésta debe constituirse en una guía efectiva para los siguientes pasos del proceso y para ello es necesario que su planteamiento sea el adecuado en cada caso, que se apoye en observaciones, hasta donde sea posible objetivas, del fenómeno de interés e incluya la probable o probables causas del mismo con la suficiente precisión y debida claridad para que también posean esas cualidades las consecuencias contrastables que de ella se deriven.
- b) En cuanto a las variables y a los medios que se prevean para controlarlas, la clasificación y la selección respectivas, se harán tomando en cuenta, tanto la hipótesis, como el modelo de diseño que se elija para la investigación. En este aspecto, a lo largo del estudio, el modelo de Astin y Panos braudo para nosotros una guía de gran utilidad. *ha sido*
- c) La selección del modelo de diseño, por la reciprocidad de las relaciones que guarda con las variables, se

TESIS CON
FALLA DE ORIGEN

hará con base, lo mismo en un análisis de ellas, que en la índole del estudio.

- d) Para proceder con acierto a dicha selección, hemos de reflexionar acerca del grado de probabilidad que podamos tener de cumplir con los diversos requisitos que los distintos tipos de modelo señalan.
- e) Sobre ese particular, si bien sería de desearse que fueran los propiamente experimentales los que aplicáramos por las ventajas que su mayor rigor nos otorga, es para nosotros, investigadores universitarios, sumamente difícil cumplir con todas las condiciones que estipulan, en especial la que se refiere a la selección al azar de los sujetos, de manera que muy pocas veces los podemos usar.
- e) Están a nuestro alcance, sin embargo, los modelos cuasiexperimentales, que no por ser menores las formalidades metodológicas que especifican, dejan de ser de gran utilidad para el investigador, quien al aplicarlos puede lograr muy aceptables grados de validez en sus estudios.
- f) En todo caso, al igual que las variables de entrada, es necesario controlar las que son propias de las operaciones.

Me refiero en particular a ello, porque quizá uno de los factores que en mayor medida afectó a nuestro estudio haya sido la insuficiente separación entre

TESIS CON
FALLA DE ORIGEN

los alumnos de los grupos experimentales y de control.

Por esa razón, entre otras, nos es imposible discriminar los efectos de las mediciones periódicas de los que haya producido la información correctiva que sobre los resultados de ellas se proporcionó a ambos grupos.

- g) No obstante que como hemos visto en varias partes de este trabajo, no es nada fácil controlar los factores que atentan en contra de la validez interna de nuestros estudios, lo es todavía menos, en muchas ocasiones casi imposible, eliminar los que operan en menoscabo de la externa. La posibilidad de derivar a un ámbito más amplio las inferencias que podamos obtener es, por ende, mínima.
- h) No por ello ni por las dificultades que encontramos en nuestro medio quienes consideramos a la investigación como uno de los mejores recursos para acrecentar nuestros conocimientos, se justifica que nos dejemos vencer por el desánimo y permanezcamos inactivos. Por reducida que sea la validez que consigamos a pesar de los esfuerzos que realicemos para lograrla, los resultados que sean fruto de una investigación debidamente conducida siempre serán valiosos y un elemento útil para ir comprendiendo mejor los fenómenos educativos.

TESIS CON
FALLA DE ORIGEN

i) Para concluir, si ninguna de las explicaciones que he propuesto para la mejoría observada en los cuatro semestres estudiados fuera satisfactoria, se me ocurre una última: que el profesor después de veintiocho años de dar clase, por fin a partir del vigésimonono ha aprendido la manera correcta de hacerlo.

Referencias bibliográficas.

- ABAD, ADELA y Luis A. Servín. Introducción al muestreo. México, Limusa, 2a. ed., 1984. 216 p.
- ACKOFF, RUSSELL L. The Scientific Method. New York, John Wiley & Sons, Inc., 1968. 464 p.
- ADAMS SACHS, GEORGIA. Medición y evaluación en educación psicología y "guidance". Trad. de Alfredo Pastor Bodmer. Barcelona, Herder, 2a. ed., 1975. 820 p.
- ADKINS WOOD, DOROTHY. Elaboración de tests. Trad. y pról. de Rogelio Díaz-Guerrero. México, Trillas, 1981. 160 p.
- AEBLI, HANS. Una didáctica fundada en la psicología de Jean Piaget. Trad. de Federico F. Monjardín. Pról. de Jean Piaget. Buenos Aires, Kapelusz, 1973. 208 p.
(Biblioteca de Cultura Pedagógica.)
- ANDERSON, RICHARD y Gerald W. Faust. Psicología educativa, la ciencia de la enseñanza y el aprendizaje. Trad. de Carlos Villegas. México, Trillas, 1977. 569 p.
- ARAUJO E OLIVEIRA, J. B. Tecnología educacional y teorías de instrucción. Trad. de Ana C. Zubasti. Buenos Aires, Paidós, 1976. 177 p. (Biblioteca del Educador Contemporáneo. Serie Didáctica 7)
- ARNAU GRAS, JAIME. Diseños experimentales en psicología y educación. México, Trillas, 1981, 2 Vol. 357 y 348 p.
- ARY, DONALD y otros. Introducción a la investigación pedagógica. Trad. de José M. Salazar y José C. Pecina H. México, Interamericana, 2a. ed. 1983. 410 p.

- ASTIN, ALEXANDER W. y Robert J. Panos. La evaluación de programas educativos. Trad. de Diana Bessoudo Salvo y Enrique Moreno y de los Arcos. México, UNAM, 1983. 52 p.
- AUSUBEL, DAVID P. Psicología educativa. Un punto de vista cognoscitivo. Trad. de Roberto Helier Domínguez, México, Trillas, 1978. 769 p.
- AZORIN POCH, F. Curso de muestreo y aplicaciones. Madrid, Aguilar, 1969. 376 p. (Colección Ciencia y Técnica.)
- BELLERT, STANISLAS y otros. El concepto de información en la ciencia contemporánea. Coloquios de Royamount. Trad. de Florentino M. Torner. México, Siglo XXI, 6a. ed., 1982. 310 p.
- BERTALENFFY, LUDWIG VON. Teoría general de sistemas. Trad. de Juan Almela. México, F. C. E. 1983. 311 p.
- BIGGE, MORRIS L. Teorías de aprendizaje para maestros. Trad. de Agustín Contín. México, Trillas, 1978. 414 p.
- BIRZEA, CESAR. Hacia una didáctica por objetivos. Trad. de Rafael Rodríguez Marín. Madrid, Morata, 1980. 240 p.
- BLOOM, BENJAMIN S. y otros. Taxonomía de los objetivos de la educación. Trad. de Marcelo Pérez Rivas. Pról. de Antonio F. Salonia. Buenos Aires, El Ateneo, 7a. ed., 1979. 355 p.
- BRUNER, JEROME. Hacia una teoría de la instrucción. Trad. de Nuria Parés. México, UTEHA, 1969. 234 p.
- BUNGE, MARIO. La investigación científica. Trad. de Manuel Sacristán. Barcelona-Caracas-México, Ariel, 8a. ed., 1981. 955 p.
- CAMPBELL, DONALD y Julian Stanley. Diseños experimentales y cuasiexperimentales en la investigación social. Trad. de Mauricio Kitaigorodzki. Buenos Aires, Amorrortu, 1982. 158 p.

- COCHRAN, WILLIAM G. y Gertrude M. Cox. Diseños experimentales. Dir. de trad. Basilio A. Rojas. México, Trillas, 1981. 661 p.
- COHEN, MORRIS R. y Ernest Nagel. Introducción a la lógica y al método científico. Trad. de Néstor A. Míguez. Buenos Aires, Amorrortu, 1979. 2 tomos, 268 y 284 p.
- COLOM, ANTONI J. Sociología de la educación y teoría general de sistemas. Barcelona, Oikos-tau, 1979. 196 p. (Colección: Ciencias de la Educación.)
- COPI, IRVING M. Introducción a la lógica. Trad. de Néstor A. Míguez. Buenos Aires, EUDEBA, 25a. ed., 1983. (Manuales EUDEBA)
- CHURCHMAN, C.W. El enfoque de sistemas. Trad. de Alberto García Mendoza. México, Diana, 1978. 270 p.
- EBEL, ROBERT L. Fundamentos de la medición educacional. Trad. de Ramón Alcalde. Buenos Aires, Guadalupe, 1977, 710 p. (Biblioteca Pedagógica.)
- EDWARDS, W. y otros. Toma de decisiones. Trad. de Eduardo L. Suarez. México, F.C.E., 1979. 425 p. (Serie Lecturas 27)
- EISNER, ELLIOT W. Cómo preparar la reforma del currículo. Trad. de Juan Jorge Thomas. Buenos Aires, El Ateneo, 1976. 206 p. (Biblioteca Nuevas Orientaciones de la Educación.)
- EINSTEIN, ALBERT. El significado de la relatividad. Trad. de Carlos E. Prélat. Buenos-México, Espasa-Calpe Argentina, 1948. 165 p.
- FORRESTER, JAY W. Principles of Systems. Cambridge, Mass., Wright-Allen Press, Second Preliminary Edition, 1976. 387 p.

- FRIEDMAN, SOFIA y otros. Sistematización de la enseñanza. Pról. J. Sánchez Azcona. México, UNAM, 1979. 934 p.
- FRONDIZI, RISIERI, ¿Qué son los valores?, Introducción a la axiología. México, F.C.E., 2a. ed. 1981. 236 p. (Breviarios 135.)
- GAGNE, ROBERT M. Las condiciones del aprendizaje. Trad. de José Carmen Pecina. México, Interamericana, 1979. 265 p.
- _____ Principios básicos del aprendizaje para la instrucción. Trad. de Paulina Díaz Cortés. México, Dina, 1979. 199 p.
- GAGNE, ROBERT M. y Leslie J. Briggs. La planificación de la enseñanza. Trad. de Jorge Brash. México, Trillas, 1978. 287 p.
- GAGNE, ROBERT M. y otros. Especificación de objetivos de la educación. Trad. de Federico Patán López. Pról de A. C. Van Dusen. México, Guajardo, 1979. 109 p.
- GARCIA CORTES, FERNANDO. Paquete de autoenseñanza de la evaluación del aprovechamiento escolar. México, UNAM, 1979. 309 p.
- GARCIA HOZ, VICTOR. Principios de la pedagogía sistemática. Madrid, Rialp, 9a. ed., 1981. 694 p. (Biblioteca de Educación y Ciencias Sociales.)
- GARRETT, HENRY E. Estadística en psicología y educación. Trad. de Juan J. Thomas. Buenos Aires, Paidós, 1976. 509 p. (Biblioteca de Psicometría y Psicodiagnóstico. Serie Mayor 22)
- GEREZ, VICTOR y Verónica Czitrom. Análisis de sistemas e investigación de operaciones. México. Representaciones y Servicios de Ingeniería S. A., 1978. 299 p. (Serie Métodos para el Análisis de Sistemas de Ingeniería. Vol. 10.)
- GEREZ, VICTOR y Manuel Grijalva. El enfoque de sistemas. México, Limusa, 1980. 560 p.

- GLAZMAN, RAQUEL y María de Ibarrola. Diseño de planes de estudios. México, UNAM, 1978. 536 p.
- GOULD, STEPHEN JAY. The Mismeasure of Man. New York, London, W. W. Norton & Co., 1981. 352 p.
- GRAWITZ, MADELEINE. Métodos y técnicas en las ciencias sociales. Trad. de Enrique Muñoz Latorre. Pról. de José Balcells Junyet. Barcelona, Hispano Europea, 1975. 2 tomos, 455 y 491 p.
- GRONLUND, NORMAN E. Medición y evaluación en la enseñanza. Trad. de Salvador Sumano. México, Pax-México, 1983. 630 p.
- HANEY, WALT. Standards for Tests and Test Use. Cambridge, Mass. Huron Institute, 1978. 48 p.
- HARROW, ANITA J. Taxonomía del dominio psicomotor. Trad. de Eddy Montaldo. Buenos Aires, El Ateneo, 1978. 119 p. (Biblioteca Nuevas Orientaciones de la Educación)
- HAYMAN, JOHN L. Investigación y educación. Trad. de Eduardo J. Prieto. Barcelona, Paidós. (Biblioteca del Educador Contemporáneo 105)
- HILGARD, ERNEST y Gordon H. Bower. Teorías del aprendizaje. Trad. de F. González Aramburo y otros. México, Trillas, 1978. 718 p. (Biblioteca Técnica de Psicología.)
- HOEL, PAUL G. Introduction to Mathematical Statistics. New York, John Wiley & Sons Inc., 1971. 409 p.
- HYMAN, RONALD T. y otros. Contemporary Thought in Teaching. Englewood Cliffs, N. J., Prentice-Hall, 1971. 292 p.
- JOYCE, B y M. Weill. Models of Teaching. Englewood Cliffs, N. J. Prentice-Hall, 1972. 402 p.
- JUNG, JOHN. El Dilema del experimentador. Trad. de Carlos Villegas, México, Trillas, 1979. 314 p.

- KARMEL, LOUIS J. Medición y evaluación escolar. Trad. de Javier Aguilar V. México, Trillas, 1981. 546 p.
- KAUFMANN, ARNOLD. La ciencia y el hombre de acción. Introducción a la praxeología. Trad. Bartolomé Coll. Madrid, Guadarrama, 1967. 251 p. (Biblioteca para el Hombre Actual.)
- KAUFMAN, ROGER A. Planificación de sistemas educativos. Trad. de Agustín Contin. México, Trillas, 1982. 189 p.
- KERLINGER, FRED N. El enfoque conceptual de la investigación del comportamiento. Trad. de Lucy Ridly, México, Interamericana, 1981. 335 p.
- _____ Investigación del comportamiento. Trad. de José Rafael Blengio y José Carmen Pecina. México, Interamericana, 2a. ed. 1984. 525 p.
- LANDSHEERE, VIVIANE Y GILBERT DE. Objetivos de la educación. Trad. de Alexandre Ferrer. Barcelona, Oikos-tau, 2a. ed. 1981. 315 p. (Colección Ciencias de la Educación.)
- LEVINE, SAMUEL y Freeman F. Elzey. Introducción a la medición en psicología y en educación. Trad. Silvia Vetrano. Buenos Aires, Paidós, 1973. 196 p. (Biblioteca del Educador Contemporáneo, Serie Didáctica 2)
- LINDEMAN, RICHARD H. Tratado de medición educacional. Trad. de Elvira Rissech de Wiñar. Buenos Aires, Paidós, 1971. 218 p. (Biblioteca del Educador Contemporáneo, Serie Mayor 16)
- MAGER, ROBERT F. La medición del intento educativo. Trad. de Elisa Paris. Buenos Aires, Guadalupe, 2a. ed. 1982. 181 p. (Biblioteca Pedagógica.)
- MAGNUSSON, DAVID. Teoría de los tests. Trad. de Javier Aguilar. Profr. R. Díaz-Guerrero. México, Trillas, 1983. 318 p. (Biblioteca Técnica de Psicología.)

- MAYNTZ, RENATE y otros. Introducción a los métodos de la sociología empírica. Trad. de Jaime nicolás Muñiz. Madrid, Alianza Universidad, 3a. ed. 1983. 310 p.
- McGUIGAN, F. J. Psicología experimental. Trad. de María Luisa Avalos de Palmeras. México, Trillas, 3a ed. 1984. 480 p. (Biblioteca Técnica de Psicología.)
- MEHRENS, WILLIAM e Irvin J. Lehmann. Medición y evaluación. En la educación y en la psicología. Trad. de Helene Lavesque D. México, CECSA, 1982. 754 p.
- MENENDEZ, LIBERTAD. Planeación didáctica. México, UNAM. F. de F. y L. Col. de Pedagogía. 1979. 15 p.
- MENENDEZ, LIBERTAD y Laura E. Rojo. Introducción a la medición del rendimiento escolar. México, UNAM. F. de F. y L. Col. de Pedagogía. 1979. 16 p.
- NUNNALLY, JUM. C. Introducción a la medición en psicología. Trad. de Leticia Halperin Donghi. Buenos Aires, Paidós, 1973. 619 p. (Biblioteca de Psicometría y Psicodiagnóstico. Serie Mayor 36.)
- RUSSELL, BERTRAND. The Scientific Outlook. London, George Allen & Unwin Ltd., 1962. 285 p.
- SANCHEZ, ALVARO. Sistemas arquitectónicos urbanos. México, Trillas, 1978. 605 p.
- SELLTIZ, CLAIRE y otros. Métodos de investigación en las relaciones sociales. Trad. de Manuel Rico Vercher. Madrid, Rialp, 8a. ed. 1976. 670 p. (Biblioteca de Educación y Ciencias Sociales.)
- SIEGEL, SIDNEY. Estadística no paramétrica. Trad. de Javier Aguilar Villalobos. México, Trillas, 1982. 344 p. (Biblioteca Técnica de Psicología.)
- SKINNER, B. F. y otros. Aprendizaje escolar y evaluación. Trad. de Manuel Barbera y otros. Buenos Aires, Paidós, 1978. 206 p. (Biblioteca de Psicopedagogía Paidós.3.)

- SMITH, FRED M. y Sam Adams. Educational Measurement for the Classroom Teacher. New York, Harper & Row, 1972. 370 p.
- SUMMERS, GENE F. Medición de actitudes. Trad. de Javier Aguilar V. México, Trillas, 1982. 688 p. (Biblioteca Técnica de Pedagogía.)
- SUPPES, PATRICK y Joseph L. Zinnes. "Basic Measurement Theory" En Handbook of Mathematical Psychology. New York. John Wiley & Sons Inc. 1963. 76 p.
- TABA, HILDA. Elaboración del currículo. Trad. de Rosa Albert. Pról. de W. B. Spalding. Buenos Aires, Troquel, 4a. ed. 1979. 662 p.
- THORNDIKE, ROBERT L. y Elizabeth Hagen. Tests y técnicas de medición en psicología y educación. Trad. de Francisco González Aramburo. México, Trillas, 1978. 733 p. (Biblioteca Técnica de Psicología.)
- TYLER, RALPH W. Principios básicos del currículo. Trad. de Enrique Molina de Vedia. Buenos Aires, Troquel, 2a. ed. 1977. 136 p.
- VAN DALEN, D. B. y W. J. Meyer. Manual de técnica de la investigación educacional. Trad. de César Musiero y Oscar Moyano. Barcelona-Buenos Aires-México, Paidós, 1983. 542 p. (Paidós Educador 4)
- VARGAS, JULIE S. Redacción de objetivos conductuales. Trad. de Federico Patán López. México, Trillas, 1979. 179 p.
- VON CUBE, FELIX. La ciencia de la educación. Trad. de María del Carmen Vicencio. Barcelona, CEAC, 1981. 192 p.
- WEBER, MAX. Essays in Sociology. Trad. y pról. de H. H. Gerth y C. Wright Mills. London, Henley and Boston, Routledge & Kegan Paul, 1977. 490 p.
- WHITE, D. J. Teoría de la decisión. Trad. de José Luis García Molina. Madrida, Alianza Universidad, 1979. 228 p.

Diccionarios.

Diccionario de la Lengua Española. Madrid, 20a. ed. 1984

Dictionary of Education. Carter V. Good. Editor. New York,
McGraw-Hill, 1945.

Webster's New Collegiate Dictionary. Springfield, Mass,
G. & C. Merriman Co. Publishers, 1956.

Referencias hemerográficas.

- BARRAL, MARY E. "The Effects of Learning Partners and Retests on Pretest-posttest Scores, Final Course Grades and Student Attitudes. En: Research in Higher Education. E.E. U.U., 1978. Vol. 8(2). 177-187
- BELZER, THOMAS J. "A Comparison of Two Methods of Evaluation and its Effect on Attention and Final Grades in General Biology". En: Research Report. E.E. U.U. Nova University, 1974. 143 p.
- BRACHT, GLENN H. y Gene V. Glass. "The External Validity of Experiments". En: American Educational Research Journal. E.E. U.U. 1968, Vol. 5(4). 437-474
- CARVER, RONALD P. "A Model for Using the Final Examination as a Measure of the Amount Learned in Classroom Learning. En: Journal of Educational Measurement. E.E. U.U., 1969. Vol 6(2). 59-68
- GAYNOR, JESSICA. "Student Performance and Evaluation Under Variant Teaching and Testing Methods in a Large College Course". En: Journal of Educational Psychology. E.E. U.U., 1976. Vol. 68(3). 312-317
- GEIGER, GLENN y Darrel E. Bostow. "Contingency-managed College Instruction: Effects of Weekly Quizzes on Performance on Examinations". En: Psychological Reports. E.E. U.U., 1976. Vol. 39(34). 707-710
- GOLDWATER, B. C. y L. E. Acker. "Instructor-paced, Mass Testing for Mastery Performance in an Introductory Psychology Course". En: Teaching of Psychology. Canadá, 1974. Vol. 2(4). 152-155.
- HADDAD, NABIL F. y otros. "Programmed Student Achievement: A Hawthorne Effect ? En: Research in Higher Education. E.E. U.U., 1975. Vol. 3(4). 315-322

MAWHINNEY, V. T. y otros. "A Comparison of Students Studying Behavior Produced by Daily, Weekly and Three-week Testing Schedules". En: Journal of Applied Behavior Analysis. E.E. U.U., 1971. No. 4. 257-264.

MURPHY, ALLEN H. y otros. "Correlating Unit-quizz Retakes and Final Examination Performance". En: Journal of Personalized Instruction. E.E. U.U., 1977. Vol. 2(1) 49-51

SLAVIN, ROBERT E. "Effects of Individual Learning Expectations on Student Achievement". En: Report 228. E.E. U.U. Johns Hopkins University, 1979. 20 p.

WIGGINS, JAMES A. "Learning Contingencies in a College Classroom: A Pilot Study". En: Research Report. E.E. U.U.. University of North Carolina, 1968. 30 p.

