



UNIVERSIDAD DE BARCELONA

INSTITUTO DE CIENCIA DE LA EDUCACION

**La Investigación Experimental  
en Educación:  
DISEÑOS DE EXPERIMENTACION**

Jaime ARNAU

**2º Seminario sobre:  
"METODOLOGIA CIENTIFICA  
APLICADA A LA INVES-  
TIGACION EDUCATIVA"**

## 2.- DISEÑOS DE EXPERIMENTACION

2.1. Los objetivos de la investigación experimental.

2.2. Diseño experimental y control.

2.3. Validez interna y validez externa.

2.3.1. Validez interna.

2.3.2. Validez externa.

2.4. El diseño de experimentación.

2.4.1. Diseños experimentales.

2.4.2. Diseños de factor único.

2.4.2.a.- Diseño antes y después.

2.4.2.b.- Diseño solamente después.

2.4.2.c.- Diseño de Solomon de cuatro grupos.

2.4.3. Diseños factoriales.

2.4.4. Diseños casi-experimentales.

2.4.4.a.- Diseño antes y después con grupo control no equivalente.

2.4.4.b.- Diseños contrabalanceados.

2.4.4.c.- Diseño de series cronológicas.

2.5. Valor del diseño de investigación en el estudio educacional.

2.6. Resumen.

## 2.- DISEÑOS DE EXPERIMENTACION.

### 2.1. Los objetivos de la investigación experimental.

En principio se puede afirmar que cada investigación experimental tiene un objetivo propio. Ahora bien; podemos clasificar los objetivos de la investigación en cuatro grandes grupos: 1). El establecimiento de las posibles relaciones entre dos o más variables. Muchas veces los experimentos se conciben como una investigación exploratoria o un estudio diseñado para determinar las condiciones bajo las cuales ciertos fenómenos ocurren.

En efecto, los experimentos son intentos aproximativos para formular problemas con mayor precisión, o bien, para la explicitación de hipótesis mucho más precisas. Así, podemos estudiar si las recompensas tienen algún efecto sobre el aprendizaje, o si los condicionamientos sociales afectan o no a los C.I. de los niños. En todos estos casos se pretende conocer si dos o más variables tienen alguna relación o no.

2). En segundo lugar, los experimentos están pensados para probar con más precisión las características propias de un determinado individuo o grupo. Tampoco, para este caso, son necesarias hipótesis de trabajo previas sobre la forma como van a presentarse dichas características. Nos interesa, pues, conocer las formas comportamentales propias de un determinado individuo, o clase de individuos, evitando, en lo posible las tendencias sistemáticas debidas a la técnica de recogida de datos, a fin de aumentar la fiabilidad de nuestros resultados.

3). Un tercer objetivo suele ser el incremento de la fiabilidad de los resultados hallados en anteriores investigaciones. Se pretende, por tanto, conocer la frecuencia con que un determinado factor se halla asociado con otro. Esto implica, naturalmente, la réplica de experimentos anteriormente realizados, y por tanto, partir ya de hipótesis previas, lo cual es práctica común en la ciencia. En efecto, hemos de considerar la réplica como una función importante de los experimentos, aunque jamás podamos tener una certeza absoluta a pesar de que se hayan tomado todas las precauciones para evitar las desviaciones.

4). Otra de las razones, es la verificación de una teoría o hipótesis causal. Las investigaciones de comprobación de relaciones causales entre variables requieren diseños que no sólo reduzcan las tendencias o errores sistemáticos y aumenten la fiabilidad de los resultados, sino que además, permitan inferir la posible ley existentes entre los diversos aspectos analizados. Esta sería, uno de los objetivos fundamentales de toda ciencia, y por lo mismo, de la investigación educativa. (1)

### 2.2.- Diseño experimental y control.

El concepto de "control" se halla íntimamente relacionado con el de

(1) Sellitz, C. Jahoda, M. Deutsch, H. y Cook, S. W. Métodos de investigación social.

diseño de experimentación. No obstante, en relación al término control, existe comúnmente bastante confusión debido a su múltiple uso. Boring (1954) en su artículo sobre la historia del concepto de control, nos indica que han sido tres los usos que se han hecho del mismo. En primer lugar, el término control ha sido utilizado como sinónimo de "limitación de condiciones". En este sentido se eliminan todas aquellas condiciones que pueden llegar a interferir con los resultados. Un control riguroso se logra en situaciones propias de laboratorio en donde cualquier fuente de distorsión, como ruidos, luces, colores, etc., puede ser, sino eliminada totalmente, si al menos mantenida en condiciones constantes para todos los sujetos.

En segundo lugar, se entiende por control a la "acción directa" que posee el experimentador sobre la variable independiente. En efecto, la manipulación activa de dicha variable es sinónimo, también, de control. Así, por ejemplo, la presentación de programas televisados o no, a fin de estudiar su efecto sobre el rendimiento escolar o nivel científico de los alumnos, es un tipo de variable que es manejada por el investigador, y por tanto se halla bajo su control.

Por último, el investigador suele referirse, en muchos casos, a experimentos con "grupo control". La introducción de este procedimiento es debido a John Stuart Mill (1846) (1) quien propuso dos métodos para el descubrimiento de las relaciones causales: el método de la concordancia y el método de la diferencia. En el método de la Concordancia establece que "si dos o más instancias del fenómeno investigado tienen una sola circunstancia en común, en la cual todas las instancias concuerdan, es la causa (o efecto) del fenómeno estudiado" (2). Es decir si un fenómeno "X" va siempre seguido por un fenómeno "Y", se puede concluir que "presumiblemente" "X" sea la causa de "Y". Ahora bien, según J.S. Mill este método puede ir acompañado por el de la Diferencia. Este establece que "si dos o más momentos en los que se produce un fenómeno tiene un único elemento en común, en tanto que entre dos o más momentos en los cuales el fenómeno no se manifiesta, no existe nada en común salvo la ausencia de esa circunstancia; la circunstancia por la cual difieren los dos momentos del fenómeno es efecto o causa, o bien elemento constitutivo de la causa del fenómeno" (3). Evidentemente, dicho método es una extensión del anterior y supone la introducción del control o "ausencia de la circunstancia X", que constituye la posible causa del fenómeno que se observa.

La utilización del "control" como grupo de contraste es práctica común en la experimentación educativa, y constituye uno de los métodos corrientes con los que suelen diseñarse los experimentos. Ahora bien, la introducción de un grupo control a fin de poder inferirse la eficacia de un determinado tratamiento o condición antecedente, implica la "suposición" previa de que ambos grupos, o bien, los grupos

---

(1) Mill, Stuart John., A system of logic. Harper y Brothers, Nueva York, 1846.

(2) Mill, J.S., op. cit., 1846, p. 224.

(3) Mill, J.S., op. cit., 1846, p. 229



experimentales y de control, sean equivalentes con respecto al fenómeno bajo observación. La presuposición de la equivalencia de los grupos obedece, fundamentalmente, a dos razones básicas. La primera se refiere a la validez del experimento, y por tanto, a la correcta inferencia de la hipótesis experimental. Dada la importancia que tiene este tema para el diseño de experimentación, trataremos este aspecto con más detalle en el siguiente apartado. El segundo propósito a que obedece la equivalencia, es el "aumento de la sensibilidad del experimento". Se entiende por sensibilidad a la precisión o capacidad que nos ofrece el diseño para captar las mínimas diferencias entre los tratamientos experimentales. Esto supone partir de grupos homogéneos con respecto a los factores independientes extraños, y existe para ello una serie de técnicas que van desde el apareo de individuos hasta la formación de grupos con igual media y desviación estándar (1). En la medida que se logran grupos homogéneos con respecto a una serie de características que pueden afectar la variabilidad de los datos experimentales, seremos capaces de obtener las mínimas diferencias entre los tratamientos experimentales. Hemos de tener en cuenta, que este segundo objetivo se cumple para aquellos estadios de investigación bastante avanzados, ya que inicialmente, lo que básicamente interesa es la validez del diseño.

### 2.3.- Validez interna y validez externa.

Uno de los principales aspectos de todo diseño de experimentación es su válido para inferir la hipótesis experimental. Es decir, si se puede llegar a establecer la existencia de una relación o no, entre los factores antecedentes con respecto a los consecuentes. Estamos, por tanto, ante un problema de eficacia de la cual dependerá, fundamentalmente, el logro de los objetivos de la investigación.

Según Campbell y Stanley (1963) (2) pueden distinguirse dos clases de validez: a) la validez interna y b) la validez externa. Vamos, pues, a analizar estos dos tipos de validez, y a estudiar con detalle los factores que comprometen, en cada uno de los dos casos, la validez experimental.

#### 2.3.1.- La validez interna.

Cuando se habla de la validez interna el investigador se pregunta hasta que punto la variación observada en la variable dependiente ha sido realmente provocada por la variable  $X$  independiente. Como afirma Campbell (1957) "la vali-

---

(1) Para una ampliación del tema puede consultarse, Arnau, J., "El diseño experimental: su aplicación a las ciencias de la conducta", en Metodología científica aplicada a la investigación educativa. I.C.E., Universidad de Barcelona, Informe nº 10; diciembre 1974, p. 71-73.

(2) Campbell, D.T., y Stanley, J.C., Experimental and quasi-experimental designs for research. Rand McNally and Comp., Chicago, 1963.



dez interna representa la exigencia mínima; cuando se habla de ella es para preguntar si la estimulación experimental ha determinado la diferencia en la concreta situación de tal experiencia". (1). Antes de responder de una forma afirmativa, es necesario que nos aseguremos que algunas de las variables extrañas al experimento no hayan contaminado los resultados. Vamos a presentar a continuación algunas de estas posibles variables extrañas al experimento, y que deberán ser objeto de control.

- a).- La historia o conjunto de factores diferentes de "X" (o variable independiente) que pueden producirse, dentro o fuera del marco del experimento y, consecuentemente, afectar los resultados del mismo. La presencia de estas variables que suelen aparecer a lo largo del proceso de experimentación pueden llegar a contaminar los efectos de la variable estudiada. Como es lógico estas variables suelen aparecer en experimentos de larga duración como ocurre en la mayoría de las investigaciones educativas.
- b).- La maduración que recoge el conjunto de procesos biológico y psicológicos que se producen en los sujetos en función del paso del tiempo. Son pues, procesos de orden interno, como por ejemplo, el crecimiento, fatiga, el hambre, etc., que pueden interferir la acción de la variable independiente.
- c).- El efecto de la medida que se ha aplicado a los sujetos antes de someterse a la situación experimental. Las medidas tomadas de los sujetos o bien los tests previamente aplicados, pueden influir en los resultados finales independientemente de la acción del tratamiento experimental.
- d).- Los instrumentos de medida pueden, también, sufrir cambios. Estos cambios no sólo pueden afectar a los aparatos de medición, sino sobre todo, tratándose de experimentos en el campo educativo, a las personas encargadas de evaluar o registrar las observaciones.
- e).- La regresión estadística se produce sobre todo, cuando la selección de los sujetos se realiza en función de los puntajes extremos. Supongamos que hemos pasado un test de rendimiento y hemos seleccionado a los sujetos que han obtenido las puntuaciones más altas (por ejemplo, superiores a 85). Cuando, posteriormente, les pasemos otro test de rendimiento, independientemente de la aplicación del tratamiento experimental, la puntuación promedia de este grupo de sujetos, en el segundo test, se desplazará hacia el valor medio de la población original. Es decir, en el segundo test habrá una distribución más amplia en los puntajes obtenidos por los sujetos. Lo mismo hubiera ocurrido si hubiésemos elegido los sujetos con peores puntuaciones. En el segundo test, la media del grupo,

---

(1) Campbell, D.T., "Factors relevant to validity of experiments in social

baja en el primero, se desplazaría hacia la media real de la población. Se da, por tanto, un fenómeno de regresión, debido a la acción del azar, a causa de las imperfecciones en los instrumentos de medición.

- f).- La selección diferencial de los sujetos hace que muchas veces los resultados acusen más bien las diferencias existentes entre los grupos, que la acción del tratamiento experimental. Si queremos probar un método de aprendizaje de lectura y seleccionamos dos clases de primer grado, puede ocurrir que, independientemente, del método, el grupo experimental tuvo se más habilidad para la lectura que el grupo control.
- g).- La mortalidad experimental o pérdida de sujetos puede, también, afectar la validez interna del experimento, sobre todo, si ésta pérdida se produce entre los sujetos que han obtenido las puntuaciones más altas o baja en el "pre-test". Esta posibilidad es muy alta, sobre todo en aquellos casos en que la experimentación tiene una duración bastante larga. Los alumnos pueden enfermar o no asistir a clase en el momento que se pasan las pruebas de control, lo cual, puede afectar notablemente los resultados.
- h).- La interacción entre selección y maduración, selección e historia, y otras condiciones pueden comprometer seriamente la validez de los resultados experimentales. Puede ocurrir, por ejemplo, que si bien tanto el grupo experimental como el control hayan obtenido puntuaciones iguales en el pre-test, dado que el grupo experimental esté más motivado, o sea más rápido en captar las explicaciones, obtenga un rendimiento superior en el re-test que el grupo control, independientemente del tratamiento. Estas posibles interacciones deberán tenerse en cuenta a fin de no comprometer la validez interna del experimento.

### 2.3.2. La validez externa.

Si la validez interna constituía la exigencia mínima para la adecuada inferencia de la hipótesis, y la consecución de los objetivos de la experimentación, la validez externa se refiere al problema de la generalización de los resultados. La validez externa está asociada, pues, a la generalización y representatividad de los logros de la investigación. "Cuando hablamos de ella -añade Campbell (1957)- es para preguntarnos a qué poblaciones, a qué contextos, a qué variables puede extenderse el efecto constatado". Es especialmente importante para las investigaciones educativas el tener asegurada la representatividad de los resultados ya que de ella depende la capacidad de aplicación de los mismos, lo cual constituye uno de los objetivos primordiales. A continuación vamos a presentar aquellos factores que deberán ser controlados a fin de asegurar la validez externa del diseño.





- a).- El efecto reactivo o interactivo del pre-test puede restringir la posibilidad de generalización de los resultados experimentales. La previa aplicación de un test puede incrementar o disminuir la sensibilidad de un grupo de sujetos hacia la variable independiente. De ahí que los resultados obtenidos de un experimento en el que los sujetos han sido sometidos a una previa prueba no serán representativos del efecto de la variable experimental para una población que no ha sido sometida a dicha pre-prueba.
- b).- Los efectos de la interacción entre los errores de selección y la variable experimental pueden comprometer, en gran manera, la generalidad de los resultados. Esto nos lleva a un problema de muestreo. Si el grupo experimental y control no están seleccionados al azar de la población de donde proceden, difícilmente los resultados podrán extenderse a toda la población de referencia. Sobre todo, si se parte de grupos naturales o muestras no representativas, es posible de un determinado tratamiento interactúe con el grupo experimental debido a sus características particulares.
- c).- Los efectos interactivos de los procedimientos experimentales restringen, también, la amplitud de la generalización de los resultados. Esto suele ocurrir en aquellos experimentos en los que existe, por parte de los sujetos, una predisposición en participar en los mismos. Esto suele ocurrir, sobre todo, en aquellos casos en que los experimentos se realizan sistemáticamente con los alumnos de primer año de especialidad o bien, con sujetos voluntarios. Como ha señalado Rosenthal (19-65) (1), "los sujetos voluntarios difieren considerablemente de los sujetos no voluntarios", y consecuentemente, no constituyen una muestra al azar. Aspecto éste que atenta completamente al principio del muestreo al azar. De ahí que los resultados obtenidos en estos casos en los que los sujetos o bien están predispuestos, o bien, tienen previa familiaridad con las técnicas experimentales, no puede ser considerados como representativos.
- d).- La interferencia del tratamiento múltiple viene a ser una extensión de lo dicho en el punto anterior. Para aquellos casos en que el experimento exige una exposición repetitiva a dos o más tratamientos, es posible que la acción de los tratamientos anteriores interactúe con los tratamientos posteriores, a no ser que su acción quede eliminada con el tiempo. Siempre que no interese, precisamente, conocer el efecto de esta exposición sucesiva como ocurre en los estudios sobre el aprendizaje y la transferencia.

---

(1) Rosenthal. R.. "The volunteer subject". Human Relations. 1965. 18. 280-296

Tenemos, por tanto, que en todo diseño se dan dos tipos de validez cuya importancia no se puede poner en duda. Ahora bien, el problema se presenta debido a que muchas veces existe una incompatibilidad entre estos dos tipos. En efecto, los controles que se utilizan para asegurar la validez interna tienden a comprometer la validez externa. A medida que aumentamos las restricciones sobre la situación experimental, desfiguramos totalmente el contexto y limitamos automáticamente la representatividad de la misma. Esto nos lleva sin duda a preguntarnos, en el caso que se hallen comprometidos los dos tipos de validez, cuál de los dos deberá prevalecer. Según Campbell (1957), la respuesta es clara, pues deberemos dar prioridad a la validez interna. Lo ideal sería un diseño que nos permitiera asegurar las dos clases de validez. No obstante, siempre será mejor para la ciencia conseguir probar hipótesis, que sacrificarlas en aras a generalizaciones vagas o imprecisas.

#### 2.4.- El diseño de Experimentación.

Una de las características más importantes del diseño de experimentación consiste en el alto grado de control que permite. Si lo comparamos con otro tipo de investigación, el diseño de experimentación no sólo nos "permite poner en evidencia la influencia de los factores experimentales, sino también, y esto es quizá lo más importante, eliminar la influencia de factores extraños" (1) Estos son, sin duda, los dos principales objetivos que deberán tenerse en cuenta en la construcción o aplicación de un diseño experimental. Tenemos, por tanto, en el diseño experimental un adecuado plan de actuación, que nos indicará las observaciones que es necesario realizar, los análisis estadísticos que son apropiados aplicar y las conclusiones que podemos inferir.

No vamos a extendernos sobre el concepto de diseño experimental, aspecto que ya tratamos en el Seminario anterior, sólo plantearemos el estudio del diseño experimental en función de su validez interna y externa.

Si bien, Campbell y Stanley (1963), nos presentan una sistematización del diseño, agrupándolos en tres grandes categorías (diseños pre-experimentales, diseños experimentales y diseños casi-experimentales), nos vamos a referir fundamentalmente a los diseños experimentales, para referirnos luego, brevemente, a los diseños casi-experimentales.

##### 2.4.1.- Diseños Experimentales.

Como hemos indicado, en el paradigma de la investigación experimental el investigador manipula directamente ciertos aspectos de la situación o variable independiente y pretende determinar los efectos de dicha variable mediante la medición de ciertas características comportamentales de los individuos. La técnica

(1) Natalon, B., "La logique des plans d'experience", en Lamaine, G., y Lamaine J.N. (eds.) *Psychologie sociale et experimentation*. Mouton, París, 1971.

Por tanto, que en todo diseño se han de tener en cuenta  
 cuya importancia no se puede poner en duda. Ahora bien, el problema no reside  
 debido a que muchas veces existe una interrelación entre estos tipos. En  
 efecto, los resultados que se obtienen con estos tipos de diseños se  
 comparan con los resultados de los diseños experimentales. A veces que aumentan los resultados  
 la influencia de los factores, de los factores totales y de los factores  
 relativamente la representatividad de la muestra. Pero nos lleva a una  
 punto, en el caso que se obtienen comparaciones los dos tipos de diseños, en el  
 los dos diseños prevalece. Según Campbell (1957), la respuesta es clara, pues  
 dependen del método e la validez interna. En idealmente se dice que no  
 permite comparar los dos tipos de diseños. No obstante, siempre será mejor  
 para la ciencia conseguir probar hipótesis, que hacerlas en una o general-  
 mente vale o importante.

#### 2.4.- El diseño de Experimentación

Una de las características más importantes del diseño de experimentación  
 está centrado en el alto grado de control que permite. Si la comparación con  
 otro tipo de investigación, el diseño de experimentación no sólo nos permite po-  
 ner en evidencia la influencia de los factores experimentales, sino también, y es  
 lo más importante, eliminar la influencia de los factores extraños. (1)  
 Esto es, en definitiva, los dos principios básicos que deben tenerse en cuenta  
 en la construcción o aplicación de un diseño experimental. Por tanto,  
 en el diseño experimental se debe tener en cuenta, que nos indican las op-  
 ortunidades que se necesitan realizar, los análisis estadísticos que son necesi-  
 dos aplicar y las conclusiones que podemos inferir.

No vamos a extendernos sobre el concepto de diseño experimental, as-  
 pecto que ya tratamos en el capítulo anterior, sólo plantearemos el estudio del  
 diseño experimental en función de su validez interna y externa.

Si bien, Campbell y Stanley (1963), nos presentan una clasificación  
 del diseño, agrupados en tres grandes categorías (diseños pre-experimentales,  
 diseños experimentales y diseños cuasi-experimentales), nos vamos a referir funda-  
 mentalmente a los diseños experimentales, para referirnos luego, brevemente, a  
 los diseños cuasi-experimentales.

#### 2.4.1.- Diseños Experimentales

Como hemos indicado, en el parágrafo de la investigación experimental  
 el investigador manipula directamente ciertos aspectos de la situación o variable  
 independiente y pretende determinar los efectos de dicha variable mediante la ob-  
 servación de ciertas características conductuales de los individuos. La técnica

(1) Véase, E. "La lógica del plan de experimentación", en J. G. y J. G.



experimental no sólo implica la "producción", en un momento dado de las condiciones cuyo efecto se pretende medir, sino también, el control de todos los restantes aspectos que pueden llegar a alterar la acción de la variable independiente. Esta es una de las razones por las que uno de los objetivos primordiales del diseño experimental sea el control o neutralización de todas aquellas variables extrañas que de alguna manera pueden llegar a influir en los resultados y que hemos analizado anteriormente. Hasta hace poco la investigación educativa estaba utilizando diseños de grupo único por lo que el control de los "factores extraños" era completamente impreciso e ineficaz. Para superar las deficiencias propias de este tipo de diseño los investigadores educativos han decidido el empleo del "grupo control" y la utilización de diseños propiamente experimentales. Con la introducción del grupo control el investigador posee una base lógica de contrastación, cumpliendo, así, con el criterio de comparabilidad exigido por la ciencia.

Por otra parte, no es suficiente, para cumplir con los criterios de contrastabilidad exigidos por la ciencia, la utilización de un "grupo de control" o de referencia, sino que además es necesario que los grupos que se comparen sean equivalentes. Es decir, el grupo experimental y el de control deben ser equivalente con respecto a todos los factores que puede influir sobre la variable dependiente, a excepción en la exposición del tratamiento experimental. A fin de que ambos grupos queden compensados con respecto a todas las variables extrañas, el experimentador suele utilizar, con frecuencia, la técnica de la "distribución al azar de los sujetos". Es decir, la aleatorización de los grupos constituye la única garantía de que el "azar" es la única razón que interviene en la distribución de los puntajes de los individuos. Y por tanto, es de esperar de que el azar actúe por igual en ambos grupos y que, como consecuencia, presentarán idénticas distribuciones.

Con la aleatorización, como es lógico, no eliminamos todas aquellas variables que como la edad, el sexo, el C.I., etc., pueden llegar a afectar los resultados, sólo se pretende que afecten a los mismos, tanto en un sentido como en otro. Como señala Fisher (1935) (1) "la simple precaución de la aleatorización será suficiente para garantizar la validez de la prueba de significación, por la cual será juzgado el resultado de la experimentación". Hemos de tener en cuenta, que por el hecho de la simple aleatorización de los grupos no se garantiza de una manera absoluta la equivalencia de ambos grupos, ya que ello depende del número de sujeto. A medida que aumenta el número de sujetos en ambos grupos existe una mayor probabilidad de una distribución de puntajes según la ley normal, y por tanto, de una mayor equivalencia de los mismos. Sólo en función de este criterio podremos aplicar adecuadamente las pruebas estadísticas de significación, que se reducen, en último término, en comparar las diferencias reales entre los grupos, con las diferencias teóricas en el supuesto que sólo hubiese actuado el azar.

(1) Fisher. R.A.. The design of experiments. Hafner. Nueva York. 1925

Con la técnica de la distribución al azar de los sujetos a los grupos, resolvemos un problema de equivalencia, gracias al control que dicha técnica nos permite de los factores extraños, con lo cual garantizamos la validez interna del diseño.

No obstante, el experimentador no debe conformarse con detectar diferencias sino en aumentar el nivel de precisión de las mismas. Debe estar, por tanto, empeñado en aumentar la sensibilidad de su diseño. Con el azar aseguramos la validez interna, pero no reducimos el error experimental. Fijémonos que el error experimental depende de la variabilidad de los individuos dentro de cada grupo, y que, por tanto, si formamos de grupos con características psicológicas muy dispares dicha variabilidad aumentará enormemente y la posibilidad de probar la hipótesis será mínima. Será muy difícil para grupos de gran dispersidad de puntajes probar la eficacia de tratamientos débiles, ya que su posible acción sistemática sobre los resultados quedará absorbida por la dispersión existente. Esta es una de las razones por la que todo el experimentador no sólo se conforme con formar grupo equivalente sino también homogéneos. Será muy diferente probar la eficacia de un determinado método de enseñanza (por ejemplo, los métodos audiovisuales) si los grupos experimental y control presentan una gran diversidad en relación a los C.I. En la medida que podamos conseguir grupos más homogéneos, por ejemplo, con un C.I. superior a 105, aumentaremos la sensibilidad de nuestro experimento.

Greenwood (1945), nos propone dos métodos para lograr la formación de grupos homogéneos: a) el control de precisión, y b) el control de la frecuencia de distribución. El primer caso se trata de la formación de los grupos por pares de sujetos idénticos. Es decir, con la misma característica psicológica, o bien relevantes con respecto a la variable de medida. Este sistema ofrece bastantes dificultades, sobre todo, a medida que aumentamos el número de características en función de las cuales apreamos los individuos. Es difícil encontrar pares de sujetos con dos o más características iguales. Por esta razón, a veces suele emplearse el segundo sistema o control de la frecuencia de distribución. Mediante esta técnica se pretende formar los dos grupos (experimental y control) en términos del total de la distribución de un factor. Así, si se considera por ej., que la edad es un factor significativo para los efectos estudiados, se intentará conseguir que las edades medias que ambos grupos sean iguales. Si hacemos lo mismo con otros factores significativos, conseguiremos, por este procedimiento, formar dos grupos homogéneos. Tampoco en esta situación podemos olvidar la "asignación del azar". Asignación que deberá observarse no sólo en la formación de los grupos, sino también, en la aplicación de tratamientos.

Hemos analizado las principales características que reúnen los diseños experimentales, estableciendo aquellas condiciones que nos aseguren una máxima eficacia en la inferencia de la hipótesis. A continuación vamos a analizar

Con la técnica de la distribución al azar de los sujetos a los grupos, resolvemos un problema de equivalencia, evitamos el control que afecta al grupo experimental, con lo cual permitimos la validez interna del diseño.

No obstante, el experimentador no debe confundirse con el control de los grupos, sino en el nivel de prestación de las mismas. Debe estar, por tanto, pendiente en cualquier momento de su diseño. Con el grupo experimental, en validez interna, pero no reduciendo el error experimental. El diseño que el autor experimental debe de la validez de los individuos dentro de cada grupo, y que, por tanto, al formarse de grupos con características psicológicas muy diferentes, debe tener en cuenta el nivel de aprendizaje y la posibilidad de probar la hipótesis de los mismos. Será muy difícil para grupos de gran heterogeneidad de grupos probar la validez de las representaciones de ellos, ya que en muchos casos no se puede obtener los resultados que se desearían por la heterogeneidad interna. Esta es una de las razones por la que todo el experimentador no debe confundirse con los grupos equivalentes sino también homogéneos. Será muy difícil probar la validez de un determinado efecto de aprendizaje (por ejemplo, los sujetos más hábiles) si los grupos experimentales y control presentan una gran diversidad en relación a los G.I. En la medida que podamos conseguir grupos más homogéneos, por ejemplo, con un G.I. superior a 100, aumentaremos la validez de nuestro experimento.

Greenwood (1945), nos propone dos métodos para lograr la formación de grupos homogéneos: a) el control de pretestes, y b) el control de la técnica de distribución. El primer caso se trata de la formación de los grupos por pares de sujetos idénticos. Es decir, con la misma característica psicológica, o bien relevantes con respecto a la variable de medida. Este sistema ofrece ventajas diferentes, sobre todo, a medida que aumentamos el número de controlaciones en función de las cuales se forman los individuos. En el caso de control por pares de sujetos con las mismas características iguales. Por otra parte, a veces puede aplicarse el segundo sistema a control de la frecuencia de distribución. Mediante esta técnica se pretende formar los dos grupos (experimental y control) en términos del total de la distribución de un factor. Así, si se controla por el, que la edad es un factor significativo para los efectos de aprendizaje, se intentará conseguir que los dos grupos sean iguales. El hecho de estar con otros factores significativos, como el sexo, por ejemplo, puede ser un factor importante. Después de estos factores podemos obtener la "representación del sexo". Asimismo que deben observarse no sólo en la formación de los grupos, sino también, en la aplicación de los tratamientos.

Como resultado de las principales características que tienen los diseños experimentales, estadísticos y psicológicos que nos permiten una mejor comprensión de la información de la hipótesis. A continuación vamos a analizar

los principales diseños experimentales en relación a su validez interna y externa. Para ello vamos a dividir los diseños en clásicos de variación de un solo factor, y los diseños factoriales o de variación de dos o más factores.

#### 2.4.2.- Diseños de factor único.

El paradigma de experimentación del diseño de factor único o clásico consiste en la comparación de las observaciones obtenidas que un grupo que ha sido sometido a un determinado tratamiento experimental con las de un grupo que no ha sido sometido a tal tratamiento. Dos problemas nos podemos plantear a raíz de estos diseños: a) en primer lugar, surge el problema de los niveles o modalidades de variación del factor experimental. Hemos de tener en cuenta, que el grupo control, que se caracteriza por la ausencia del factor experimental, constituye ya de por sí una modalidad de dicho factor. No obstante, según la exigencia de nuestra hipótesis, podemos planificar experiencias en las que sólo nos interese detectar diferencias y en este caso sólo emplearemos dos modalidades de la variable independiente (o tratamientos) como suele ocurrir en los clásicos experimentos de presencia y ausencia del tratamiento experimental, o bien, podemos introducir varios niveles del factor independiente a fin de precisar mejor el tipo de relación existente entre la variable independiente y la variable de medida.

Según, pues, la cantidad de tratamientos que utilicemos, tendremos los diseños de tipo bivalente o multivalente. Lógicamente todo proceso de investigación se interesará inicialmente por los diseños de tipo bivalente, para proseguir, posteriormente, con diseños multivalentes.

b) Un segundo problema a plantearnos se refiere al "momento en el que vamos a tomar medidas de las variables dependientes". Según este criterio, los diseños de factor único, suelen dividirse en diseños "antes y después" y diseños "solamente después". Este va a ser el criterio que vamos a adoptar en el análisis que vamos a hacer de los diseños de factor único en relación a su validez interna y externa.

##### 2.4.2.a.- Diseño antes y después.

El experimentador puede estar, muchas veces, interesado en tomar las medidas de la variable dependiente no sólo después de la aplicación del tratamiento, sino también, antes de la aplicación. Y esto por una serie de razones que vamos a enumerar:

1.- El investigador puede estar interesado en aumentar la sensibilidad de su diseño, para ello las medidas "antes" le pueden ser de gran utilidad para formar dos grupos de sujetos que inicialmente tengan, en relación a la variable dependiente, unos puntajes equivalentes. De esta forma se consigue una mayor homogeneización de los grupos, y consecuentemente, una mayor precisión en detectar los valores más débiles de la variable independiente. Si nos interesa conocer la



eficacia de un nuevo método en el aprendizaje de lectura, podemos pasar a los sujetos, previamente, un test de lectura y formar dos grupos con puntajes similares con respecto a este test. Por otra parte, no hemos de olvidar la asignación al azar de los sujetos a los grupos, para conseguir un total control de otras variables que podrían interferir con el tratamiento (sexo, C.I., interés del alumno, etc.).

2.- Se puede descartar, también, comprobar si una variable experimental es pertinente o no en una situación concreta. Así, p. ej., supongamos que se pretende estudiar la eficacia de una nueva estructuración del sistema escolar. En este caso, es interesante tener un conocimiento del rendimiento habido antes de la introducción del cambio para comprobar la efectividad o no de la nueva estructuración escolar.

3.- En otros casos la hipótesis que se pretende estudiar se halla relacionada con la posición inicial de la variable dependiente. Esta relación puede constituir unas de las condiciones antecedentes previas para la posible efectividad de un tratamiento. La introducción de un sistema de enseñanza más abierto, puede ser diferente según la actitud previa del alumnado con respecto al profesorado sea neutra o bien totalmente contraria. Será por tanto conveniente, en esta situación, conocer cual es la actitud del alumnado con respecto al tipo de enseñanza que se imparte.

4.- Incluso en el supuesto de que dicha relación con respecto a la hipótesis experimental no exista, el investigador puede desear conocer las variaciones que se producen con respecto a la variable dependiente, una vez se haya introducido el tratamiento.

5.- En la situación en que el experimentador no pueda formar los grupos en base a un criterio de azar, es útil tener un conocimiento de las medidas "antes" de la variable dependiente, para poder tener una garantía de la equivalencia de los grupos, y por tanto, una seguridad en el momento de inferir conclusiones con respecto a la hipótesis.

Tenemos, por tanto, que las medidas "antes" nos ofrecen la posibilidad de conocer si los grupos, independientemente del tratamiento experimental ("X"), diferían o no, más de lo que se puede esperar del azar. Esta suposición o condición estadística fundamental para el tratamiento de los datos, queda, también, garantizada en los diseños "solamente después" mediante la asignación de los sujetos al azar, o mediante algún sistema de aparos que asegure su equivalencia.

Los "diseños antes y después" pueden dividirse, teniendo en cuenta los niveles del factor independiente, en diseños bivalente y multivalente. Vamos a presentar el esquema de ambos diseños, para pasar, luego, a discutir su grado de validez interna y externa.

Esquema de diseño antes y después con dos grupos.

| <u>Grupo</u> | <u>Asignación</u> | <u>Observ. antes</u> | <u>Tratamiento</u> | <u>Observ. después</u> |
|--------------|-------------------|----------------------|--------------------|------------------------|
| E            | Azar              | $O_1$                | $X_1$              | $O_2$                  |
| C            | Azar              | $O_3$                | $X_0$              | $O_4$                  |

(En donde, E = grupo experimental; C = grupo control; O = medidas de las observaciones; y, X = niveles de la variable independiente o tratamientos).

Este diseño puede ampliarse al aumentar el número de tratamientos. Con ello se obtiene un mejor conocimiento del efecto diferencial de la variable independiente, y ofrece la posibilidad de determinar el tipo de relación existente entre el factor independiente y la variable de medida. A continuación presentamos el esquema del diseño antes y después con más de dos grupos.

Esquema de diseño antes y después con más de dos grupos.

| <u>Grupo</u> | <u>Asignación</u> | <u>Observ. antes</u> | <u>Tratamiento</u> | <u>Observ. después</u> |
|--------------|-------------------|----------------------|--------------------|------------------------|
| $E_1$        | Azar              | $O_1$                | $X_1$              | $O_2$                  |
| $E_2$        | Azar              | $O_3$                | $X_2$              | $O_4$                  |
| C            | Azar              | $O_5$                | $X_0$              | $O_6$                  |

Este diseño añade al anterior la posibilidad no sólo de comparar el efecto de los diversos tratamientos sobre el grupo control, sino también la comparación de los tratamientos entre sí. Este diseño, como el anterior, han sido fundamentalmente utilizados por la investigación educativa para el estudio de los diversos métodos de instrucción ya sea entre sí, o con el que se considera como "regular" o "tradicional".

Validez interna.

Hemos ya indicado cuáles eran los factores o variables extrañas que pueden llegar a comprometer la validez interna (historia, maduración, efecto de medida, etc.). Los diseños, como los propuestos, es decir, con grupo control, permiten controlar todas las potenciales fuentes de falta de validez interna.

Si consideramos los posibles factores extraños que pueden llegar a interferir la acción del tratamiento "X", podemos concluir que unos están más bien asociados al cambio que pueden darse desde la observación antes hasta la observación después (cambio intersección), y otros, más bien se hallan relacionados con los posibles cambios que pueden darse entre el grupo experimental y el de

control. (cambio intrasección). La presencia, en este tipo de diseño, del grupo control, nos garantiza la validez interna del diseño en relación a los factores asociados al cambio intersección. En efecto, todas las posibles circunstancias extrañas o acontecimientos que ocurran y que pueden afectar a los sujetos, influirán, por igual a ambos grupos. De ahí que cualquier diferencia que hallemos en las medidas u observaciones después las podremos atribuir a la presencia del tratamiento experimental. Por otra parte, la clase de diseño de medidas "antes y después", nos ofrece la posibilidad de medir los posibles cambios que pueden producirse, mediante la comparación de los puntajes "antes y después".

Medida del cambio en el grupo experimental:  $D_o = O_2 - O_1$

Medida del cambio en el grupo control:  $D_o = O_4 - O_3$

Si con la introducción del grupo control el diseño queda protegido de los posibles efectos debidos al cambio intersección, mediante una adecuada técnica de asignación al azar de los sujetos a los respectivos grupos evitaremos la acción distorsionante de los factores asociados al cambio intrasección. En la medida que el investigador haya procurado de asignar al azar, antes de tomar las medidas pertinentes, no sólo los sujetos, sino también, los observadores, los instrumentos mecánicos, las técnicas de medidas, instructores, etc., tiene la garantía de haber controlado todas aquellas posibles variables que podían haber afectado diferencialmente los grupos. Por su parte, con la comparación de las puntuaciones antes, el investigador puede tener una prueba sobre la equivalencia inicial de ambos grupos. Mediante la comparación de las puntuaciones después puede probar la hipótesis experimental.

Prueba de equivalencia:  $E = O_3 - O_1$

Prueba de la hipótesis:  $H = O_4 - O_2$

#### Validez externa

El ideal de todo diseño de investigación es conseguir junto a una validez interna una máxima validez externa, y como consecuencia una mayor representatividad y generalización de sus resultados. Una de las posibles fuentes potenciales que pueden atentar a la validez externa, se refiere, sin duda, a la posible interacción del tratamiento "X", con cualquier factor extraño. En efecto, el tratamiento puede interactuar con la selección de los grupos, con el posible origen de las muestras, con la acción de las medidas tomadas antes, etc. Es bastante difícil llegar a una total prevención contra estas posibles fuentes de distorsión. No obstante, el investigador puede tener en cuenta una serie de condiciones. Así, una de ellas puede ser la utilización de un sistema de muestreo aleatorio que le garantice la representatividad de las unidades muestreadas. Por otra parte, deberá evitar en lo posible la utilización de sujetos voluntarios, o bien, sujetos procedentes de contextos altamente caracterizados (por ejemplo, estudiantes de los cursos





También, con respecto a la aplicación de las posibles medidas antes o tests, éstos deberán ser presentados como formando parte de las tareas propias del sistema de enseñanza. Es decir, eliminando su carácter de prueba psicotécnica o de control. En todo caso, los sujetos deberán desconocer su posible participación en una investigación experimental, para eliminar cualquier posible efecto de predisposición o desgana.

Por último, podemos señalar, que la repetición del experimento en otro lugar y tiempo, puede significar el decisivo argumento que nos aseguren si las relaciones descubiertas se mantienen y, por tanto, existe la posibilidad de generalizar la hipótesis.

#### 2.4.2.b. - Diseño solamente después.

En este tipo de diseño, los grupos, experimental y de control, sólo son observados, en relación a la variable de medida, después de la aplicación del tratamiento al grupo experimental. Como en cualquier experimento los sujetos son asignados al azar a ambos grupos, lo cual constituye una garantía sobre su equivalencia. Sobre la base de la teoría de la probabilidad se puede suponer que las posibles diferencias habidas en los rendimientos en relación a la variable dependiente son previas a la introducción del tratamiento y son efecto de la asignación al azar. Estas diferencias son tenidas en cuenta por la prueba de significación estadística, sobre cuya base se puede inferir la acción del tratamiento experimental. En el caso de que se haya utilizado alguna técnica de apareamiento, se elimina la influencia interferidora de algún factor, con lo que se aumenta la equivalencia de los grupos. Toda técnica de apareo deberá ir acompañada, por una asignación de los sujetos al azar. Esto es evidente, porque con el apareo se controlan algunas variables, pero no todas. Es, pues, incluso en este caso el empleo del procedimiento al azar, para aleatorizar el efecto de posibles factores extraños no controlados.

#### Esquema de diseño solamente después.

| <u>Grupo</u> | <u>Asignación</u> | <u>Observ.antes</u> | <u>Tratamiento</u> | <u>Observ.después</u> |
|--------------|-------------------|---------------------|--------------------|-----------------------|
| E            | Azar              | -                   | $X_1$              | $O_1$                 |
| C            | Azar              | -                   | $X_0$              | $O_2$                 |

#### Validez del diseño "solamente después".

Partiendo del supuesto que los dos grupos, experimental y de control, han sido formados según un criterio al azar y dado que los sujetos no han sido sometidos a ningún tipo de pre-prueba, este diseño permite controlar los principales efectos que pueden comprometer la validez interna. Así, las variables tales como el efecto de la historia, maduración, pre-prueba, etc., quedan totalmente controla-

recogida de medidas se puede efectuar simultáneamente, como ocurre con los experimentos de efecto inmediato.

Por otra parte, puesto que en este diseño no se toman "medidas antes", se anulan las posibles interacciones entre el tratamiento y tales medidas, con lo cual queda neutralizada una de las fuentes de error de la validez externa. A su vez, este diseño implica un tratamiento estadístico mucho más simple lo cual convierte a este diseño en una forma económica y práctica de investigar.

Ahora bien, puesto que la mayoría de los trabajos de investigación educativa tiene, por objetivo, el estudio del rendimiento escolar, este diseño tiene una aplicación bastante limitada. En efecto, mediante el "diseño solamente después", el investigador sólo puede llegar a probar la hipótesis mediante la comparación de los puntajes después.

Prueba de la hipótesis:  $H = O_1 - O_2$

No obstante, en la mayoría de los casos, los rendimientos vienen determinados por los puntajes que presentan los grupos antes de la aplicación de los tratamientos. Este diseño, no permite por tanto, comprobar este extremo, y consecuentemente, no dispone de una prueba adicional para concluir sobre la equivalencia de los grupos. Si bien, como ya hemos indicado, la teoría de la probabilidad permite inferir esta equivalencia a partir de las puntuaciones después.

Dada la rapidez de ejecución, este diseño se presenta como una herramienta muy práctica de investigación. Ahora bien, siempre que el investigador pueda disponer de medidas antes, es aconsejable la utilización del diseño "antes y después", ya que ofrece mucha más información, y permite, la aplicación de análisis, que de otra manera no será posible realizar.

#### 2.4.2.c.- Diseño de Solomon de cuatro grupos.

Solomon (1949) (1), ha propuesto un paradigma de investigación, que puede ser considerado como una combinación de los dos diseños que hemos propuesto. Así, junto al control que un tipo de diseño antes y después, permite de los factores de variación cotóneos y de desarrollo, la introducción de una réplica, del mismo, en forma de diseño "solamente después" permite al investigador un control de la interacción del tratamiento con el efecto de medida. Con este tipo de diseño la cantidad de análisis aumentan sensiblemente, siendo especialmente eficaz para un cálculo estadístico de la interacción (tratamiento con la medida antes).

---

(1) Solomon, R.L., "An extension of control group design", Psychol. Bull., 1949, 46, 137-150.

Esquema del diseño de Solomon con cuatro grupos.

| <u>Grupo</u>   | <u>Asignación</u> | <u>Observ. antes</u> | <u>Tratamiento</u> | <u>Observ. después</u> |
|----------------|-------------------|----------------------|--------------------|------------------------|
| E <sub>1</sub> | Azar              | O <sub>1</sub>       | X <sub>1</sub>     | O <sub>2</sub>         |
| C <sub>2</sub> | Azar              | O <sub>3</sub>       | X <sub>0</sub>     | O <sub>4</sub>         |
| E <sub>3</sub> | Azar              | -                    | X <sub>1</sub>     | O <sub>5</sub>         |
| C <sub>4</sub> | Azar              | -                    | X <sub>0</sub>     | O <sub>6</sub>         |

Validez del diseño de Solomon con cuatro grupos.

Con respecto a la validez interna, podemos afirmar que este diseño puede controlar y medir, el efecto de las posibles variables coetáneas y de desarrollo que pueden interferir con el tratamiento experimental. La acción de estos efectos combinados, puede medirse, al comparar O<sub>6</sub> con O<sub>4</sub>, y O<sub>5</sub> con O<sub>2</sub>. Al mismo tiempo, se puede comprobar la posible acción del "pretest" sobre los resultados experimentales al comparar O<sub>4</sub> con O<sub>3</sub>, y O<sub>6</sub> con O<sub>3</sub>.

Al mismo tiempo, como ya hemos indicado, con este diseño se pueden controlar y medir, el efecto de la interacción "pretest-tratamiento", al comparar el promedio de puntajes en O<sub>2</sub> y O<sub>6</sub>, con el promedio en O<sub>4</sub> y O<sub>5</sub>. Queda, con esta forma, garantizada la neutralización de uno de los factores que pueden llegar a distorsionar la validez externa.

Este diseño, supone la réplica del mismo experimento, lo cual aumenta la posibilidad de obtener unos resultados mucho más fiables que en cualquier otro diseño presentado hasta aquí.

2.4.3.- Diseños factoriales.

Hasta este punto sólo hemos presentado los diseños clásico o de un solo factor. Estos diseños implicaban la variación de una sola variable independiente y el control o neutralización de los restantes factores independientes. La objeción más clara hecha a tales diseños se desprende del hecho que éstos desfiguran totalmente la realidad de los fenómenos comportamentales. En efecto, la mayoría de los hechos sociales y educativos, se presentan con una gran complejidad y es ingenuo pensar que los factores varían independientemente. En general, la alteración o cambio de un factor influye sobre la conducta, concomitantemente, al cambio experimentado por los restantes factores que influyen coetáneamente sobre la misma. El investigador, al producir los cambios en su variable experimental, no puede ignorar que estos cambios pueden alterar el sistema interfuncional de variables presentes, siempre, en toda manifestación conductual. De todo ello se puede concluir que los diseños factoriales, al aislar un solo factor, no logran captar la complejidad de la realidad que se pretende estudiar.

# Diagrama del diseño de Johnson con cuatro factores

| Grupo | Factores | Factores | Factores | Factores |
|-------|----------|----------|----------|----------|
| 1     | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
| 2     | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
| 3     | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
| 4     | Factor 1 | Factor 2 | Factor 3 | Factor 4 |

## Valores del diseño de Johnson con cuatro factores

Con respecto a la validez interna, podemos afirmar que este diseño puede considerarse y mejor, el efecto de las posibles variaciones sistemáticas y de los errores que pueden introducirse en el experimento. La validez de estos efectos combinados puede reducirse al controlarlos con  $0.1$ ,  $0.2$  y  $0.3$ . Al mismo tiempo, se puede controlar la posible acción del "factor" sobre las variaciones experimentales al controlar  $0.1$  con  $0.2$  y  $0.3$ .

Al mismo tiempo, como ya hemos indicado, con este diseño se pueden controlar y medir, el efecto de la interacción "factor-factor", al controlar el promedio de puntos en  $0.1$  y  $0.2$  con el promedio en  $0.1$  y  $0.2$ . Queda, con este factor, la neutralización de uno de los factores que pueden llegar a afectar la validez externa.

Resumiendo, sobre la validez del mismo experimento, la cual asegura la posibilidad de obtener unos resultados sobre los factores que en cualquier otro diseño presentaría hasta aquí.

## 3.4.3 - Diseño factorial

Desde este punto más allá de los presentados los diseños clásicos o de un solo factor. Estos diseños implican la variación de una sola variable independiente y el control o neutralización de los restantes factores independientes. La validez de estos diseños se basa en la validez de los factores que afectan los resultados. En efecto, la validez de los factores sociales y educativos, se presentan con una gran complejidad y es difícil controlarlos. En general, la validez interna de los factores variables independientes, al controlar o cambiar de un factor a otro, se reduce, por lo tanto, la validez externa por los restantes factores que influyen considerablemente sobre el mismo. El investigador, al producir los cambios en su variable experimental, no puede ignorar que estos cambios pueden afectar el sistema experimental de los factores presentes, siempre, en una neutralización completa. De este modo se puede



muy limitada, y en muchos casos, carentes de significación. Es obvio, que si un educador estudia el posible efecto de diferentes programas educativos no deberá limitarse a investigarlo con sujetos que presenten un elevado C.I., o con sujeto procedentes de un determinado nivel socio-económico. Podría ocurrir que un determinado programa fuese muy eficaz para sujetos procedentes de un nivel social alto, y completamente inoperante para sujetos de estratos sociales bajo. En este caso, no podría concluir sobre la bondad de dicho programa.

La sistemática utilización de diseños de factor único, falsea, totalmente, la realidad educativa y restringe, totalmente, la posibilidad de obtener resultados válidos. Por esta razón, cada vez es mayor el interés que ha suscitado entre los investigadores de la educación las técnicas de análisis propuestos por Fisher (1935), y por los diseños factoriales que dicho autor desarrolló.

El diseño factorial, permite al investigador educacional, superar las limitaciones propias de los esquemas clásicos, y una aproximación más real a los fenómenos educativos.

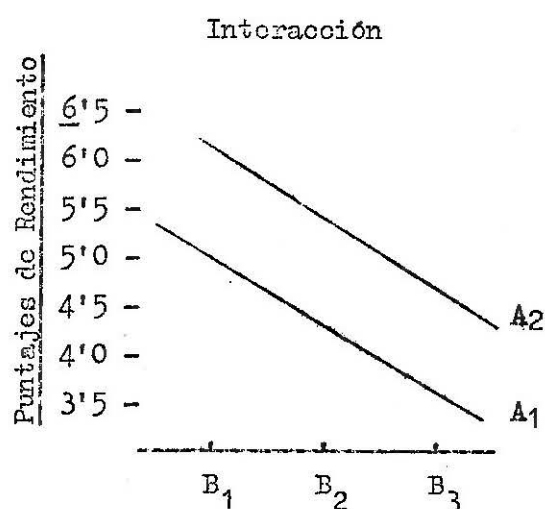
#### Disposición experimental.

Todo diseño experimental supone, la introducción o manipulación simultánea, de dos o más variables independientes, cada una de las cuales pueden presentar dos o más valores. En principio, todo diseño factorial implica la "ortogonalidad", es decir, la combinación de cada uno de los niveles de un factor, con cada uno de los restantes. Las diferentes combinaciones que pueden resultar de dicha ortogonalidad determina no sólo los llamados "tratamientos experimentales", sino también los grupos que constituirán el diseño. Por esta razón, tales diseños suelen simbolizarse de la siguiente forma:  $2 \times 2$ ;  $2 \times 2 \times 2$ ;  $2 \times 2 \times 2 \times 2$ ; (o bien,  $2^2$ ;  $2^3$ ;  $2^4$  etc.) lo cual significa diseños de dos, tres, o cuatro factores con dos valores respectivamente. Aunque este tipo de diseño factorial ha sido el más estudiado por los investigadores, no por ello significa que sean los únicos. Pueden darse, también, diseños factoriales del tipo:  $2 \times 3$ ;  $2 \times 4$ ; etc., todas las posibles combinaciones.

Vamos a presentar un diseño factorial " $2 \times 3$ ", y por tanto, un total de seis posibles combinaciones y seis grupos de sujetos. Supongamos que deseamos estudiar el efecto que, sobre el nivel de aprendizaje de lectura, ejercen las variaciones de los métodos de enseñanza (variable independiente A), y las variaciones en los niveles intelectuales de los niños (variable independiente B). Los niveles de "A", pueden ser: método no-audiovisual ( $A_1$ ) y método audiovisual ( $A_2$ ). Los niveles de "B", pueden ser: C.I. alto ( $B_1$ ); C.I., medio ( $B_2$ ); y C.I. bajo ( $B_3$ ). Asignamos al azar, a cada grupo un número determinado de "n" sujetos, con lo que aseguramos la equivalencia de los grupos experimentales. La tabla de datos hipotéticos para un experimento de este tipo podría representarse de la siguiente forma:

Tabla de datos hipotéticos para un experimento factorial "2 x 3", relativo a la influencia de los "métodos de enseñanza" y "cocientes intelectuales" sobre "el aprendizaje de lectura".

|      |                | Métodos de Enseñanza |                |       |
|------|----------------|----------------------|----------------|-------|
|      |                | A <sub>1</sub>       | A <sub>2</sub> | Media |
| C.I. | B <sub>1</sub> | 5'4                  | 6'4            | 5'9   |
|      | B <sub>2</sub> | 4'5                  | 5'5            | 5'0   |
|      | B <sub>3</sub> | 3'6                  | 4'9            | 4'25  |
|      | Media          | 4'5                  | 5'6            | 5'05  |



La información que nos permite obtener un diseño factorial es sumamente valiosa, porque no sólo podemos analizar, por separado, la eficacia de cada uno de los dos factores sobre el rendimiento final, sino también, su posible acción interactiva. Es decir, nos permite calcular hasta qué punto el efecto de cada uno de los valores que toma, por ejemplo "A", depende o no, de los valores que toma "B". En la medida que la acción de A<sub>1</sub> y A<sub>2</sub> sobre el rendimiento no queda alterado por los diversos valores de "B", y esto queda reflejado en el gráfico mediante líneas paralelas, se puede concluir que no existe interacción entre A y B.

Es muy interesante disponer de un medio para conocer el posible efecto interactivo de las variables independientes, debido a que la investigación educativa trabaja siempre con fenómenos conductuales. Como sabemos éstas son siempre el resultado de la acción de múltiples variables. Es evidente, que es muy diferente el conocimiento que nos pueden proporcionar los diseños tradicionales en relación al comportamiento separado de tales variables, que su real incidencia sobre la conducta. Muchas veces la eficacia de un tratamiento queda mediatizado o alterado por la presencia de otras variables que pueden cambiar, sensiblemente, el sentido de su acción. Por todo ello hemos de concluir que es notoria la importancia que tienen los diseños experimentales para la futura investigación educativa, no sólo por razones de orden práctico y económico, sino, principalmente, porque nos permiten una aproximación real a los fenómenos educativos.

Hemos de tener en cuenta, por otra parte, que los análisis estadísticos que normalmente suelen utilizarse en los diseños factoriales se basan, como propuso Fisher (1935), en el cálculo de las variancias. No hemos presentado aquí, este tipo de cálculo, porque nuestro propósito ha sido poner de manifiesto la lógica interna del diseño y porque dicho cálculo obedece, simplemente, a una técnica de análisis que es fácil encontrar en cualquier manual de estadística (1).

Sólo hemos pretendido destacar la importancia del diseño factorial, dentro del amplio contexto de los diseños de investigación para el estudio de los problemas educativos.

#### 2.4.4.- Diseños casi-experimentales.

Campbell y Stanley (1963) nos hablan, en su interesante exposición sobre el diseño, de los "diseños casi-experimentales". En este tipo de diseños el investigador no posee un control total de todas las posibles variables que puedan afectar los resultados. Esto suele ocurrir sobre todo en trabajos de campo en lo que es muy difícil poseer un control absoluto de las condiciones de producción de los fenómenos. Por esa razón, algunos autores, se refieren a estos diseños, con el título de diseños "con control parcial" (1).

Hemos de añadir, por tanto, que si bien en muchas situaciones, sobre todo educativas, no nos es posible aplicar el ideal de experimentación, no por ello hemos de renunciar a investigar e intentar utilizar diseños que aunque estén faltos de un completo control, nos ofrezcan la posibilidad de obtener resultados científicamente válidos. Dado, pues, este deficiente control que implican los diseños "casi-experimentales", el investigador ha de estar atento a aquellas posibles variables carentes de control, y admitir, siempre, la posibilidad de que sus resultados se deban más a la acción de tales variables que a la del tratamiento.

Vamos a presentar, a continuación, los principales diseños "casi-experimentales" que pueden ser utilizados en la investigación educativa.

##### 2.4.4.a.- Diseño antes y después con grupo control no equivalente.

Este diseño, como señalan Campbell y Stanley (1963) (2), es uno de los más difundidos dentro de la investigación educativa, e implica, como en los diseños antes y después con grupo control, la recogida de datos previamente y después de la aplicación del tratamiento. La única diferencia existente, es que en este caso no se parte de la inicial equivalencia muestral de los grupos (experimental y control). La razón estriba en que se dan investigaciones en las que se parte de grupos formados de una manera natural, como las clases de un Colegio, y es muy difícil asegurar su equivalencia, por más similares que sean. De ahí, que la gran diferencia que encontramos entre los diseños "casi-experimentales" y los vistos anteriormente, es que los sujetos no son asignados a los grupos mediante procedimientos de azar. Sino que se parte de grupos previamente formados, procurando, lógicamente, que muestren características muy similares.

---

(1) Van Dalen, D.B., y Meyer, W.J., op. cit., 1971, p. 301

(2) Campbell, D.T., y Stanley, J.C., op. cit., 1963, p. 47.

Esquema del diseño antes y después con grupo control no equivalente

| <u>Grupo</u> | <u>Asignación</u> | <u>Observ. antes</u> | <u>Tratamiento</u> | <u>Observ. después</u> |
|--------------|-------------------|----------------------|--------------------|------------------------|
| E            | No azar           | $O_1$                | $X_1$              | $O_2$                  |
| C            | No azar           | $O_3$                | $X_0$              | $O_4$                  |

Si bien los grupos no se forman mediante una técnica aleatoria, sino que se consiguen de conjuntos ya formados, como clases completas, se pretende, en la medida de lo posible que ambos presenten características similares. Ello puede comprobarse, mediante la comparación de los puntajes obtenidos de las pruebas pasadas "antes". No obstante, en la aplicación del tratamiento, es conveniente utilizar una técnica al azar.

Validez externa.

En general, se puede afirmar, que este diseño puede controlar los factores principales que comprometen la validez interna. Es decir, gracias a la presencia del grupo control, se puede controlar y medir, los efectos principales de la historia, maduración, medida o instrumentos de medida, y por tanto, se puede esperar que sus efectos no se confundan con el del tratamiento.

Quizá uno de los factores más importantes que pueden llegar a comprometer la validez interna del diseño que consideramos sea la posible interacción que puede producirse entre la selección y la maduración, o entre la selección y la historia, etc. Debido a que los grupos no han sido seleccionados al azar, y a pesar de que el experimentador haya elegido grupos con puntajes medios similares, puede existir dentro de los grupos variables que pueden llegar a interactuar con los factores enumerados, y consiguientemente, confundir los efectos del tratamiento experimental. Supongamos que se haya elegido, para un estudio educacional, una clase de tercer grado para el grupo experimental y una clase de segundo grado para el grupo control. El experimentador puso el máximo cuidado en comprobar que en los puntajes obtenidos en la preprueba ambos grupos dieron unos promedios similares. Ahora bien, la posible ganancia que experimentó el grupo experimental, podría interpretarse no ya como simple efecto del tratamiento que les fue aplicado, sino al hecho que este grupo poseyera ya más destreza y por tanto más facilidad de mejorar el rendimiento que el grupo control. De ello puede deducirse que, independientemente al tratamiento, el grupo control hubiera presentado en el post-test una media de puntajes superior a la del grupo control.

También la regresión estadística puede plantear problemas a la validez interna. Si bien dicha dificultad puede evitarse, no obstante suele incurrirse en ella, sobre todo si los grupos se forman en base a puntajes extremos o medidas correlacionadas. Supongamos que se utilizan dos clases de una escuela



El diseño del estudio debe ser controlado y aleatorio.

| Grupo | Intervención | Observación | Tratamiento    | Observación |
|-------|--------------|-------------|----------------|-------------|
| 1     | No usar      | 0           | X <sub>1</sub> | 0           |
| 2     | No usar      | 0           | X <sub>0</sub> | 0           |

El diseño de grupo no se forma mediante una técnica aleatoria, sino que se consigue de cualquier forma, como grupos completos, no aleatorios, en la medida de lo posible que ambos presenten características similares. El grupo de control, mediante la comparación de los puntos obtenidos de los grupos de control "antes". No obstante, en la aplicación del tratamiento, un conveniente utilizar una técnica de azar.

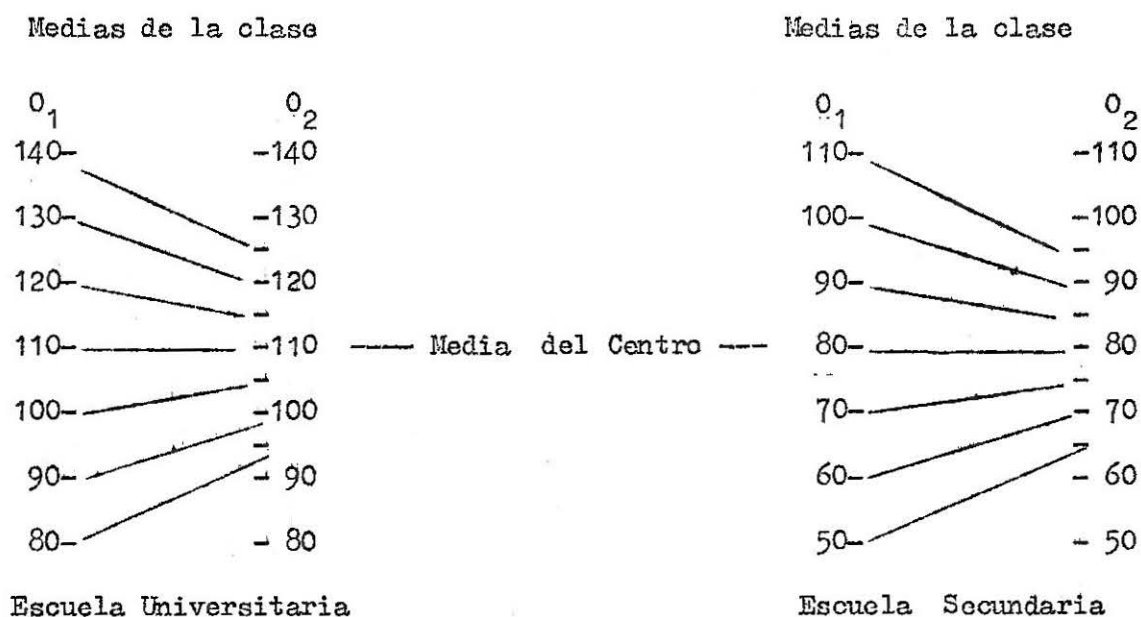
### Validez externa

En general, no puede afirmarse que este diseño pueda controlar los factores extrínsecos que condicionan la validez interna. En efecto, gracias a la presencia del grupo control, se puede controlar y medir, los efectos extrínsecos de la historia, maduración, medida o tratamiento de medida, y por tanto, se puede asegurar que una elección no se confundan con el tratamiento.

Queda uno de los factores más importantes que pueden llegar a comprometer la validez interna del diseño que consideramos sea la posible interacción que puede producirse entre la selección y la maduración, o entre la selección y la historia, etc. Debido a que los grupos no han sido seleccionados al azar, y a pesar de que el experimentador haya elegido grupos con puntajes iniciales similares, puede existir dentro de los grupos variables que pueden llegar a interferir con los factores extrínsecos, y consecuentemente, confundir los efectos del tratamiento experimental. Supongamos que se haya elegido, para un estudio sobre el efecto de tener grupo para el grupo experimental y una clase de no grupo para el grupo control. El experimentador para el estudio controla en su grupo que en los puntos obtenidos en la prueba ambos grupos tienen una puntuación similar. Ahora bien, la posible interacción que experimenta el grupo experimental, podría interferir con los efectos del tratamiento que los grupos, sino el hecho que este grupo pasara ya una historia y por tanto una facilidad de mejorar el rendimiento que el grupo control. De este modo de hecho que, independientemente al tratamiento, el grupo control hubiera que compararlo en el post-test una medida de puntaje superior a la del grupo control.

Finalmente la investigación puede plantear problemas a la validez interna. Si bien dicho diseño puede controlar, no obstante puede tener efectos en ella, sobre todo si los grupos se forman en base a puntajes extremos o medidas correlacionadas. Supongamos que se utilizan dos clases de una muestra

universitaria, en que la clase utilizada, para un programa terapéutico (grupo experimental, tiene un puntaje medio de 80, y que la clase utilizada como grupo control, un puntaje medio de 120 (1). El puntaje medio para dicha escuela es de 110. En este caso, la regresión será de diferente dirección para ambos grupos. Es decir, el grupo experimental regresará, en un sentido ascendente, hacia la media de la población, y el grupo control se desplazará en un sentido descendente. Como es obvio, el efecto del tratamiento podrá confundirse con el de la regresión.



Graf. nº 2. Ejemplo de regresión estadística.

Obviamente, los investigadores no suelen cometer el error de comparar grupos que inicialmente sean disparos. Ahora bien, muchas veces suelen aparear sujetos que tengan los mismos puntajes para formar grupos equivalentes. En este caso hemos de tener en cuenta que si los sujetos pertenecen a diferentes poblaciones, podrá producirse el efecto de "regresión estadística".

En efecto, supongamos que el investigador está interesado en probar un nuevo sistema de enseñanza, y para ello compara su efecto en la escuela universitaria y una escuela secundaria, cuya media general es de 80. Para ello puede seleccionar dos clases, en ambos centros, que tengan un mismo puntaje medio de 90. En este caso, ocurrirá, que la clase perteneciente a la Escuela Universitaria, puesto que la media de su población original es de 110, presentará un desplazamiento ascendente en la segunda prueba. En cambio, la clase de la Escuela Secundaria, puesto que la media de su población original es de 80, mostrará, en el segundo test, un desplazamiento descendente. Estos efectos, debidos sin duda a la regresión estadística, interferirán la acción del tratamiento experimental. Cuando un investigador se halla ante un tipo de diseño como el descrito es aconsejable no practicar el procedimiento de apareo, que como hemos visto puede llegar a confundir la acción

(1) Ejemplo toma de Van Dalen y Meyer, op. cit., 1971. p. 303.

universitaria, en que la clase universitaria, para su programa de estudios (grupos espe-  
cializados), tiene un promedio medio de 80, y que la clase universitaria para grupo control,  
un promedio medio de 100 (1). El promedio medio para dichos grupos es de 100. En  
este caso, la representación de los datos de los grupos es la siguiente:  
El grupo experimental representado en un sentido ascendente, hacia la media de la po-  
blación, y el grupo control en un sentido descendente, hacia la me-  
dia, el efecto del tratamiento sobre las variables con el de la población.



Gráfico No. 2. Ejemplo de regresión estadística.

Ovviamente, los investigadores no pueden cometer el error de comparar  
grupos que inicialmente son diferentes. Ahora bien, muchas veces existen grupos mi-  
nutos que tengan las mismas características pero formen grupos equivalentes. En este caso  
podemos tener en cuenta que si los sujetos pertenecen a diferentes poblaciones, po-  
dría producirse el efecto de "regresión estadística".

En efecto, supongamos que el investigador está interesado en probar un  
nuevo método de enseñanza, y para ello compara un grupo en la escuela universitaria  
con un grupo control, cuyo nivel promedio es de 80. Para ello puede seleccionar  
algunos de los grupos, en ambos centros, que tengan un mismo promedio medio de 90. En  
este caso, ocurrirá que la clase perteneciente a la Escuela Universitaria, puesto  
que la media de un población original es de 100, presentará un comportamiento as-  
cendente en la segunda prueba. En cambio, la clase de la Escuela Secundaria, puesto  
que la media de un población original es de 80, mostrará, en el segundo test,  
un comportamiento descendente. Estos efectos, debidos sin duda a la regresión es-  
tadística, interfieren la acción del tratamiento experimental. Cuando se investi-  
ga el efecto de un tipo de medida como el descrito es recomendable no presentar  
el procedimiento de selección, que como hemos visto puede llegar a confundir la acción

(1) Ejemplo tomado de Van Dusen y Rogers, op. cit., 1971, p. 208.

del tratamiento, siendo aconsejable partir de grupos que presenten diferencias de puntajes iniciales. Es más conveniente admitir dichas diferencias y aplicar análisis estadísticos que las tengan en cuenta (como el análisis de la covariancia) que pretende partir de grupos inicialmente equivalentes.

#### Validez externa.

Como hemos indicado, la validez externa se refiere a la posibilidad de generalización de resultados. Es posible que surja alguna interacción entre la selección y el tratamiento, debido a las propias características de los grupos. Puede esperar, lógicamente, que algunos grupos reaccionen mejor al tratamiento que otros. Ahora bien, puesto que en este diseño se parte de grupos ya formados, y nos implica la formación artificial de los mismos, se pueden obtener más fácilmente conjuntos comparables, y por tanto, resultados capaces de ser ampliados a contextos y poblaciones más extensas.

Puede darse también el efecto reactivo de los procedimientos experimentales. No obstante, dado que se parte de clases ya formadas como unidades experimentales, el experimentador tendrá una mayor facilidad de disimular el carácter de experiencia de su investigación. Sobre todo si se presenta la tarea como algo propio de la actividad escolar. Este diseño, más que ningún otro, logra controlar los posibles efectos que comprometan su validez externa.

#### 2.4.4.b.- Diseños contrabalanceados.

En los experimentos en los que se pretende conseguir una máxima precisión o un gran control, todos los sujetos son sometidos a cada una de las condiciones experimentales. Esta sería, en principio, la lógica del diseño contrabalanceado, llamado, también, "alternativo" (cross-over, desing, Cochran y Cox, 1957), o bien, "conmutativo" (switch-over desing, Kempthorne, 1952).

Los diseños contrabalanceados siguen la típica disposición del diseño Cuadrado latino. Es decir, los tratamientos son designados al azar a cada uno de los grupos o bien sujetos individuales. Hemos de tener en cuenta, como ya hemos indicado, que cada uno de los grupos o sujetos son sometidos a todas las condiciones experimentales; lo único que se aleatoriza es el orden de aplicación de los tratamientos.

El diseño contrabalanceado es especialmente apropiado para aquellas situaciones en las que se cuenta con un número limitado de sujetos y para aquellos contextos en los que la toma de medidas antes no es apropiada.

Esquema de diseño contrabalanceado con cuatro tratamientos.

|         | 1ª Sesión | 2ª Sesión | 3ª Sesión | 4ª Sesión |
|---------|-----------|-----------|-----------|-----------|
| Grupo A | $X_1O$    | $X_2O$    | $X_3O$    | $X_4O$    |
| Grupo B | $X_2O$    | $X_4O$    | $X_1O$    | $X_3O$    |
| Grupo C | $X_3O$    | $X_1O$    | $X_4O$    | $X_2O$    |
| Grupo D | $X_4O$    | $X_3O$    | $X_2O$    | $X_1O$    |

Este diseño contiene tres clasificaciones: sesiones o momentos en que se aplican los tratamientos; los grupos de sujetos (pueden ser individuos), y, los tratamientos experimentales ("X"). Cada dimensión de variación o clasificación es "ortogonal" con las restantes. Así por ejemplo, en el esquema propuesto, cada "tratamiento" ocurre una vez y sólo una vez en cada fila y sólo una vez en cada columna.

Una vez realizado el experimento se computa la media de puntajes por columnas, lo cual nos indicará la posible influencia sobre el "momento de realizar la prueba"; la medida por filas, lo cual indicará la posible relación entre los grupos y los puntajes; y por último, podemos calcular la media por tratamientos, para averiguar la posible eficacia de los mismos.

Validez del diseño.

Este diseño proporciona un control difícil de lograr, como veíamos en el diseño anterior, cuando los grupos no están formados al azar. Efectivamente, en este diseño, puesto que los tratamientos son aplicados a cada uno de los grupos en las distintas sesiones quedan compensadas las posibles diferencias iniciales entre ellos, quedando así compensadas. Dado, por tanto, que los tratamientos se administra a todos los grupos (o individuos) queda controlado al posible efecto de contaminación de la selección.

Una posible fuente de contaminación puede darse como consecuencia de la interacción maduración-selección. Ahora bien, si este efecto de maduración que puede tomar la forma de cansancio o fatiga, se produce en todos los grupos por igual, su acción no interfiere con la del tratamiento. Sólo afectaría, en el caso que el fenómeno de maduración se produjera sólo en algunos grupos según las diferentes sesiones.

También puede ocurrir, y éste es uno de los grandes peligros de este diseño aplicado en el campo educativo, que los grupos al ser sometidos a un tipo de tarea, ésta por efecto de transferencia facilite la realización de las poste-





riores. Esto puede producirse, sobre todo, si la acción del tratamiento tiene un efecto duradero. En tal caso, la acción del tratamiento anterior puede interferir con el siguiente. Para estos casos es aconsejable dejar pasar un tiempo entre sesión y sesión a fin de que desaparezcan los efectos de los tratamientos anteriores. No obstante, en investigación educativa, muchas veces es totalmente imposible evitar que desaparezcan el fenómeno de transferencia. Debido a ello este tipo de diseño tiene un uso bastante limitado.

Por otra parte, puesto que la disposición del diseño contrabalanceado, tal como lo hemos presentado sigue el esquema del diseño de Cuadrado latino, mediante la utilización del análisis de la variancia se pueden probar, respectivamente, el efecto principal de cada una de las fuentes principales de variación: grupos, sesiones, y tratamientos. No obstante, dado que existen tres fuentes de variación, un diseño completo implicaría "4 x 4 x 4" posibles combinaciones. Es decir, un total de 64 "unidades experimentales", extremo que no se cumple. De ahí, se deduce, que este diseño, desde el punto de vista del análisis de variancia es una réplica incompleta, ya que sólo se trabaja con "16 unidades experimentales". De ello se deduce, que si se demuestra que uno de los tres efectos principales es significativo, queda, por la misma razón, demostrado la significación de la interacción entre los otros dos. Así, si se demuestra que es significativa la acción de los tratamientos, puesto que cada uno de éstos se aplica a una específica combinación entre los valores de las otras dos restantes variables (sesiones y grupos), se infiere la significación de la interacción compleja entre aquellas dos fuentes de variación. De todo ello se deduce la necesidad de replicar este tipo de diseños, con lo que se convierte el diseño "casi-experimental" en un "diseño experimental", siempre y cuando puedan asignarse al azar los grupos a cada una de las condiciones experimentales.

#### 2.4.4.c.- Diseño de series cronológicas

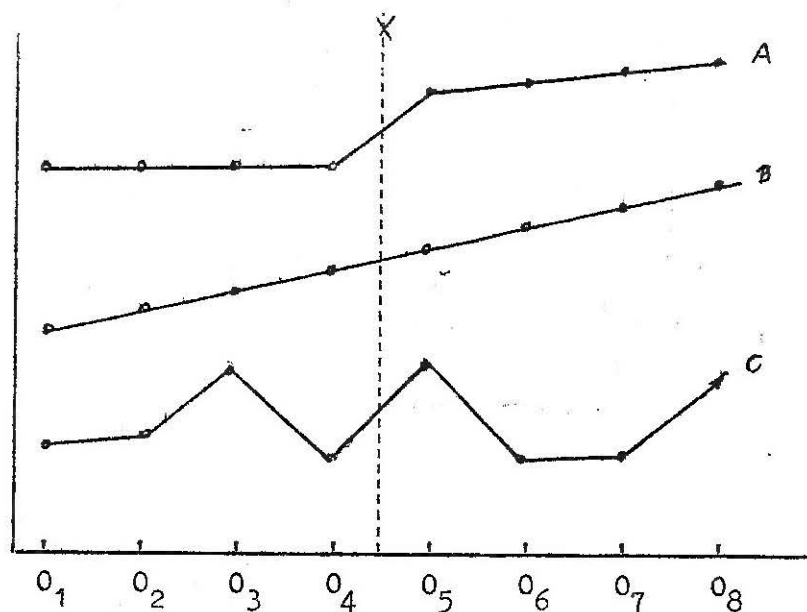
La base de este diseño consiste en tomar una serie de medidas de un sujeto o grupo de sujetos, antes y después de la aplicación de un tratamiento. De ahí el porque a estos diseños se los conoce, también, con el nombre de "diseño de medidas repetidas". El posible efecto de "X" (o tratamiento) puede ser estimado en función de la discontinuidad que presentan las medidas registradas después de su administración. Las medidas repetidas registradas antes y después de la aplicación o presencia del tratamiento presentarán determinada tendencia o estructura regular. La cuestión consiste en probar hasta qué punto dicha estructura ha quedado alterada como consecuencia de "X", en el intervalo donde ha sido aplicado.

### Esquema del diseño de series cronológicas.

| <u>Grupo</u> | <u>Observaciones antes</u> |       |       |       | <u>Tratamiento</u> | <u>Observaciones después</u> |       |       |       |
|--------------|----------------------------|-------|-------|-------|--------------------|------------------------------|-------|-------|-------|
| E            | $O_1$                      | $O_2$ | $O_3$ | $O_4$ | $X_1$              | $O_5$                        | $O_6$ | $O_7$ | $O_8$ |
| C            | $O_1$                      | $O_2$ | $O_3$ | $O_4$ | $X_1$              | $O_5$                        | $O_6$ | $O_7$ | $O_8$ |

### Validez del diseño.

La validez del diseño depende, fundamentalmente, de la estructura que presentan las medidas. Siempre que se produzca un cambio en la misma en el intervalo en que ha sido aplicado el tratamiento, puede llegarse a inferir la hipótesis. Si analizamos el gráfico nº 3, podemos concluir que X es la causa del cambio que se observa entre  $O_4$  y  $O_5$  en los resultados observados en "A". En cambio, no se



Graf. 3. Posibles estructuras de medidas repetidas después de la introducción de un tratamiento "X".

puede inferir que X sea la causa de la variación observada entre  $O_4$  y  $O_5$ , para las series de medidas en B o C. Probablemente, se puede concluir que el factor maduración sea el responsable del progreso habido en B, y que la fluctuación del azar haya sido la causa de la variación en C.

Sin duda alguna, mediante la introducción del grupo control quedan controlados, en este diseño, las posibles variables extrañas como la historia, maduración, puesto que ambos grupos se hallan sometidos a la misma influencia. También, la posible interacción "selección-maduración" puede ser controlado, o al menos medida, dado que de cada grupo se obtiene más de una medida. En el caso que, en las medidas antes, un grupo progresara más rápidamente que el otro, éste se atribuiría a este efecto. Siguiendo este razonamiento se puede también

observar, a partir de las puntuaciones antes, los posibles efectos de los tests e instrumentos de medición. Si éstos ejercen algún posible efecto facilitador o distorsionantes no hay razón alguna para esperar que éste se manifieste en el cambio operado entre  $O_4$  y  $O_5$ . Todo ello hace que este diseño sea muy eficaz, sobre todo en situaciones educativas, ya que permite valorar la acción de los tratamientos no sólo inmediatamente sino a largo plazo.

Con respecto a la validez externa, la posible interacción entre el pre-test y el tratamiento puede neutralizarse en la medida que se empleen prueba que normalmente se utilicen en la Escuela para valorar los rendimientos de los alumnos. Es decir, en la medida que las pruebas aplicadas sean las que normalmente se utilizan en los programas de enseñanza. Con ello aseguramos la presentatividad de los resultados.

Por otra parte, puede producirse un efecto de interacción entre selección y tratamiento, en el sentido que el investigador seleccione aquellos grupos que por su regularidad en la escuela o especiales condiciones sea más idóneo para su estudio experimental. En este caso, los datos obtenidos en situaciones demasiado preparadas puede llegar a comprometer la significación y extensión de los resultados.

## 2.5.- Valor del diseño de investigación en el estudio educacional.

El paradigma de investigación experimental utilizado por las ciencias naturales y su eficaz aplicación, entusiasmó a principios de siglo a los investigadores de las ciencias educativas. Ahora bien, pronto comprobaron que la realidad educativa era totalmente diferente, y que implicaba una serie de dificultades que no se hallaban presentes en las ciencias naturales. Era evidente que el procedimiento experimental, en lo que implica de manipulación y control era, en muchos casos, totalmente inviable para las ciencias de la conducta. Esta fue una de las causas del paulatino abandono de las técnicas experimentales.

No obstante, a partir de los años cincuenta, se ha revitalizado un inusitado interés hacia los estudios experimentales. Ello debido en parte al desarrollo de una metodología que ha permitido el análisis de los aspectos educativos en situaciones o contextos naturales. Esto ha animado a muchos educadores a realizar, de nuevo, trabajo de carácter experimental, utilizando, para ello, el auténtico método científico.

Nuestro intento ha sido, pues, el presentar unos esquemas de investigación que permitieran al educador superar las dificultades que implica el trabajo en situaciones reales, pero que a su vez, garantizan los mínimos exigidos para la obtención de resultados válidos y representativos. Todo experimentador ha de tener conciencia de la complejidad propia de los fenómenos conductuales y que, por tanto, en la mayoría de los casos es totalmente imposible realizar experimentos se-

gún el paradigma clásico. En efecto, son muchas las variables que el investigador no podrá llegar a controlar totalmente, de ahí la necesidad de usar del grupo control. Con el grupo control se neutralizan los posibles efectos de historia y maduración. No obstante, dado que existen otras muchas variables extrañas, sólo con una adecuada técnica de distribución de sujetos al azar, será capaz de formar dos grupos equivalentes, condición fundamental, para la inferencia de la hipótesis.

En otros casos, el investigador podrá formar los grupos en base a una técnica de apareamiento. Con ello se consigue grupos más homogéneos, y la posibilidad de trabajar con menor número de sujetos. Ahora bien, en este caso, deberá tener en cuenta los posibles peligros de dicho procedimiento, sobre todo con respecto a las variables que no han sido objeto de apareo y los posibles efectos de regresión estadística. En estos casos, sería aconsejable que junto a la técnica de apareo se utilizara la de distribución de los sujetos a los respectivos grupos según un criterio al azar. De esta forma, se logrará superar los inconvenientes que implica al apareo, y se reforzará la formación de grupos equivalentes. Ahora bien, actualmente, gracias a los desarrollos de las técnicas de análisis (como el de la covariancia), se puede evitar el árduo trabajo del apareo de los sujetos.

Dada la importancia que tiene el control de las variables extrañas para el diseño experimental, y las dificultades que implica su utilización en los estudios educacionales, vamos a presentar a continuación el cuadro, propuesto por Campbell y Stanley (1963), en la que se nos presenta la lista de los posibles factores que pueden llegar a comprometer la validez interna y externa de los diseños. Nosotros, adaptaremos dicho cuadro a los diseños que hemos presentado a fin de que se pueda evaluar la eficacia con que dichos diseños controlan los factores contaminadores.



Cuadro nº 1. Factores que pueden comprometer la validez de los diseños experimentales.

| Factores contaminantes                                              | Diseños Experimentales                                    |   |   |   |   |   |
|---------------------------------------------------------------------|-----------------------------------------------------------|---|---|---|---|---|
|                                                                     | 1                                                         | 2 | 3 | 4 | 5 | 6 |
| <u>Validez interna</u>                                              |                                                           |   |   |   |   |   |
| Historia                                                            | +                                                         | + | + | + | + | + |
| Maduración                                                          | +                                                         | + | + | + | + | + |
| Efecto de medida                                                    | +                                                         | + | + | + | + | + |
| Instrumentos de medida                                              | +                                                         | + | + | + | + | + |
| Regresión estadística                                               | +                                                         | + | + | ? | + | + |
| Selección diferencial de sujetos                                    | +                                                         | + | + | + | + | + |
| Mortalidad experimental                                             | +                                                         | + | + | + | + | + |
| Interacción entre selección-maduración,<br>selección-historia, etc. | +                                                         | + | + | - | ? | + |
| <u>Validez externa.</u>                                             |                                                           |   |   |   |   |   |
| Efecto reactivo o interactivo del<br>pretest.                       | -                                                         | + | + | - | ? | - |
| Interacción entre selección y "X"                                   | ?                                                         | ? | ? | ? | ? | - |
| Interacción de los procedimientos<br>experimentales.                | ?                                                         | ? | ? | ? | ? | ? |
| Interferencia del tratamiento múltiple.                             | /                                                         | / | / | / | - | / |
| 1.- Diseños antes y después.                                        | 4.- Diseño antes y después con grupo control equivalente. |   |   |   |   |   |
| 2.- Diseño solamente después.                                       | 5.- Diseños contrabalanceados.                            |   |   |   |   |   |
| 3.- Diseño de Solomon.                                              | 6.- Diseño de series cronológicas.                        |   |   |   |   |   |

(Aclaración: el signo más indica el control del factor; el signo menos, la ausencia de control; el de interrogación indica que el factor debe ser examinado; la barra indica que el factor no es pertinente).

## 2.6.- Resumen :

Se han presentado, en este escrito, los fundamentos básicos del diseño experimental. Para ello se ha insistido en la importancia que tiene el diseño para el control de los factores extraños que pueden comprometer su validez.

La validez del diseño tiene dos aspectos: la interna y la externa. La validez interna es la exigencia mínima para inferir la acción efectiva de los tratamientos; la validez externa se refiere a la posible extensión y generalización de los resultados. En la medida pues que se intenta aumentar la validez interna queda restringida la validez externa. El ideal de todo diseño de investigación es aquel en la que puedan garantizarse ambos tipos. No obstante, éste es un ideal que, en las condiciones actuales, está aún lejos de conseguirse. En el momento de tomar una opción, la respuesta, según Campbell (1957) es clara, deberemos dar prioridad a la validez interna. Una de las exigencias básicas de toda ciencia es la prueba de la hipótesis, y ello deberá conseguirse, aunque, para ello se sacrifiquen las posibilidades de generalización de resultados.