

EL PLAN DE CUATRO GRUPOS DE SOLOMON: EVALUACION EDUCATIVA TRANSHISPANOAMERICANA

Klaus-Dieter Gorenć¹, Luis Castro Bonilla,
Sandra Peredo Rivera, Luis Felipe Abreu Rivera,
Luis Armando Oblitas Guadalupe y Roberto Llanos Zuloaga

Se aplicaron dos modalidades de talleres de investigación –corta y larga– para habilitar por espacio de diez años a 658 investigadores hispanoamericanos, divididos en 30 grupos. La eficiencia terminal de los talleres fue medida con un diseño de pre y post-test y las diferencias con una prueba t-Student –para muestras dependientes–. Los resultados mostraron que los valores del post-test eran más elevados – $p \leq 0.001$ – que los del pretest. Contrastando las dos versiones del taller, el promedio del largo fue significativamente mayor – $p \leq 0.001$ – que el del corto; esto, bajo la tutela de un diseño múltiple de pre y post-test y la prueba t-Student, pero para muestras independientes. Con el diseño de cuatro grupos de Solomon se estableció que el incremento significativo en el conocimiento sobre diseños de investigación no fue intervenido por los factores tratamiento y pretest –interacción– y no se observaron diferencias significativas entre ambas versiones.

Palabras claves: entrenamiento, diseños de investigación, pre-post-test, pre-post-test múltiple, diseño de cuatro grupos de Solomon, interacción del tratamiento-pretest.

The Solomon four groups plan: A cross-Hispanamerican educational evaluation

Two modalities of research design's workshops –long and short– were developed to qualify 658 Hispanamerican research workers, whom were divided into 30 groups, and were applied along ten years. The terminal efficiency of the workshops was measured with a pre-post-test design and the differences were assessed with a t-test –for dependent samples–. The results showed that the values of the post-test were higher – $p \leq 0.001$ – than the pretest averages. Contrasting the two workshop versions, the average of the long workshop was significantly higher – $p \leq 0.001$ – than the short one, by means of a multiple pre-post-test design and the t-test for independent samples. Using Solomon four-group design it was feasible to establish that the significant increase of knowledge about research designs was not influenced by the interaction of the pretest and treatment factors and there was not observed any significant difference between both versions.

Keywords: training, research designs, pre-post-test, multiple pre-post-test, Solomon four-group design, pretest-treatment interaction.

1 Investigador Nacional (Nivel II) del Sistema Nacional de Investigadores y Profesor titular B, tiempo completo, Facultad de Medicina, Secretaría de Planeación y Desarrollo Institucional, Edificio B, 1er. piso; Circuito Escolar, Ciudad Universitaria. 04510 Coyoacán, D.F. México; Teléfono y Fax: 623-21-52.

La planeación se conjuga con la evaluación y en consecuencia, al constituirse esta interacción en un binomio *p-e*, semejante a la expresión algebraica formada por dos términos, se estaría aproximando, a nivel burdo, a la investigación empírica conocida en la literatura como de comprobación de hipótesis. Esto, en un sentido estricto no es factible, ya que las hipótesis únicamente pueden ser refutadas o no, que se fundamenta en la asimetría existente en la comprobación hipotética: las hipótesis no pueden verificarse a nivel empírico, evidenciándolas como verdaderas, pero puede refutarse mediante una observación contraria; esto es, a través del mecanismo de la falsificación. Este, en su concepción más amplia, se refiere a la refutación empírica de una teoría: las observaciones realizadas no concuerdan con el pronóstico efectuado a través de una cierta estructura teórica (Popper, 1982), por lo que una modalidad de investigación no está dirigida a examinar un supuesto teórico subyacente. En el presente caso, se intenta ponderar dos versiones de un taller de diseños de investigación orientado a difundir la cultura científica básica (Gorenć y cols., 1988 y 1993) derivado de la teoría del aprendizaje por ensayo y error; es decir, comparar a muestras sometidas a un mismo contenido y duración, pero esta última, dicotomizada en modalidades de extendida (versión larga) y comprimida (versión corta) y establecer el efecto que pudiera tener o no la capacidad de fijar en forma discontinua o continua la atención respectivamente. Esto, en contraste con aquellas investigaciones empíricas descriptivas que son de tipo exploratorio, enfocadas a especificar en forma precisa un fenómeno y con aquéllas dirigidas a desarrollar y evaluar los productos provenientes del ámbito tecnológico (Wellenreuther, 1982).

Ubicado el presente estudio en el terreno de la contrastación empírica de la teoría del ensayo-error y para seguir el *hilo rojo mitológico* del ritual científico (Oblitas y Gorenć, 1990), la tendencia de la investigación concordó con la fórmula propuesta en 1953 por Kurt Lewin (citado en Wellenreuther, 1982):

$$E_p = f(P, O)$$

donde:

E_p = evaluación parcial, orientada en este caso al número de probandos sometidos a las dos versiones de un taller, así como su rendimiento. Mientras que la evaluación integral (E_i) se dirige al análisis objetivo de los factores que inciden en la educación y cómo está inmersa en la sociedad que le da origen y es razón primordial de su formación, abarcando los elementos de mayor peso que intervienen en un proceso de estudio, así como el medio en el que se desarrolla; es decir, con este tipo de evaluación es factible valorar la intervención de indicadores sociales, culturales, económicos, ideológicos y políticos. La evaluación social (E_s), que permite objetivizar la realidad, en sí misma hipotética (Popper, 1982), en cuanto a necesidades y realizaciones (Uribe, 1990).

f = (en) función (de)

P = planeación

O = operacionalización

Evaluación

Independientemente de las peculiaridades de la evaluación apuntadas –parcial, integral y social–, el auge del desarrollo de métodos o técnicas –modelos o ritos– de esta actividad científica, se inició en la década de los cincuentas. Sobre todo, en el sector de proyectos industriales de desarrollo y en las empresas orientadas a producir bienes. A pesar de los esfuerzos realizados, la evaluación de los sistemas muestra enormes distorsiones y omisiones, impidiendo tomar decisiones efectivas. Siendo la distorsión más evidente cuando se intenta valorar los programas o proyectos educativos. Esto indica, que si se desea actuar conforme a los resultados obtenidos por medio de la evaluación, se requiere tomar decisiones, que deberán ser evaluadas posteriormente (Sánchez-Guerrero, 1990). Por lo tanto, a nivel general, este autor expresó que la evaluación debe considerarse como el acto o resultado de juzgar la importancia, calidad, significado, cantidad, grado o condición de la educación. Al desarrollar esta actividad evaluativa se compara y califica para obtener un juicio acerca de

una situación encaminada al cumplimiento de ciertos fines. El juicio se define como el resultado de llevar a cabo una comparación racional; la calificación alude a la asignación de la calidad a un ente y el ordenamiento de estas asignaciones es un proceso preferencial. Por lo anterior, y desde un enfoque sistémico, la evaluación es un insumo para la toma de decisiones y éstas, a su vez, insumos para la solución de problemas o para la autocorrección o refinamiento del propio sistema.

En 1990, Keeves informó que la evaluación educativa está enfocada a realizar comparaciones con un patrón o a través de un criterio que deriva de los objetivos explícitos o implícitos en el binomio conformado por la enseñanza y el aprendizaje. Señaló, asimismo, que la evaluación es una actividad estrechamente relacionada con la investigación y el desarrollo de estrategias educativas con sostén empírico, siendo necesario efectuar la medición de los resultados –*outcomes*– educacionales tanto de individuos como de grupos.

No obstante, al margen de la definición anotada, Sowden y Keeves (1990) expresaron que gran parte de la evidencia asequible acerca de la evaluación educacional ha sido coleccionada en forma de documentos publicados, transcripciones de entrevistas, observaciones de prácticas, notas de campo, grabaciones de presentaciones orales e informes escritos. Este tipo de datos es de naturaleza cualitativa y evidentemente tienen un enorme valor, debido a que son altamente ricos, personales y cercanos al mundo real, conteniendo además un profundo significado de las formas más abstractas. Pero les falta evidencia, debido a que esta forma de coleccionar y usar los datos se asocia con dificultades sustanciales en la investigación educativa de corte empírico.

La colección de tales datos es una labor intensa, que en algunos casos consume muchos años, aunado a que el análisis correspondiente implica un gran dispendio de tiempo y en ocasiones es muy difícil y frecuentemente condenado al rechazo de la comunidad científica, ya que la evidencia fue coleccionada sin una estructura teórica reconocible. También en este tipo de ejercicios, se omite constantemente el empleo de las técnicas de muestreo y se ignora si las muestras representan *bien* al universo de donde fueron extraídas o si fueron probabilísticas y como consecuencia, la pregunta importante acerca de la generalización de los resultados debe de permanecer

abierta. Ante esta limitación, se ha intentado efectuar generalizaciones con los resultados provenientes de las investigaciones cualitativas a través de la posibilidad de replicarlos con el análisis de datos cuantitativos. Igualmente, se han realizado esfuerzos para ofrecer al público usuario cánones avanzados para realizar exámenes de evidencia de los datos coleccionados en investigaciones conocidas como robustas que permitan incrementar la consistencia y la solidez de los resultados obtenidos en el plano de la evaluación. Por tal circunstancia, se han entablado discusiones entre diversos autores, siendo necesario subrayar que ni las perspectivas de las investigaciones científicas ni de las humanísticas tienen un mayor peso, debido a que conforman una unidad epistemológica de la investigación educativa. Así, se ha reconocido que la línea dura divisoria que se ha hecho entre los métodos de colección de datos cualitativos y cuantitativos es meramente artificial. La diferencia entre datos cualitativos y cuantitativos se encuentra en el nivel de abstracción y en un límite de la simplificación. Se reconoce que las cantidades son cualidades –*quantities are of qualities*– existiendo el clamor que uno de los métodos es antiético o alternativo al otro, que se deduce, ha sido concebido en forma dolosa. Por ejemplo, se ha logrado establecer la influencia que tienen los valores –axiología– en la investigación educativa, debido a que éstos tienen implicaciones en la replicabilidad de los resultados obtenidos en la investigación, donde otro científico con valores diferentes, conduce un estudio semejante y el único camino para mantener controlado el relativismo subjetivo en la investigación consiste encarar los valores e introducir las premisas valorativas en forma explícita y lo suficientemente concretizadas. En este caso, la pregunta está dirigida a cómo fundamentar o medir los valores a nivel empírico.

En la evaluación educativa se emplean tanto estrategias deductivas como inductivas. Los procedimientos son deductivos en cuanto algunas construcciones de orientación son ganadas por el conocimiento anterior y la experiencia, así como por los valores del científico, que han sido operacionalizados y conjugados en un cuerpo de datos obtenidos en algún campo de estudio. Esto es conveniente para focalizar y reducir los datos que pudieron ser coleccionados. La inducción se emplea en la medida en que los datos se acumulan y se emplean para modificar y reconstruir la estructura original. Debido a que la deducción y la inducción son diferentes y conceptos separados, la disimilitud entre los procesos deductivos

e inductivos es imprecisa si se reconoce que a menudo, la concepción de la orientación de las construcciones teóricas utilizadas en el proceso deductivo son producto de la inducción. La interacción entre los procesos deductivos e inductivos en las estrategias de evaluación se logra cuando la investigación se fundamenta en la evidencia empírica.

Examinando lo manifestado con respecto a la dicotomización de los datos en cualitativos y cuantitativos y de procesos en deductivo e inductivo, así como la concreción efectuada con la fórmula apuntada, el siguiente modelo permitió efectuar la operacionalización de la variable endógena dirigida a evaluar las modalidades del taller descrito con 26 ítems –definiciones–. Este modelo está orientado a derivar la hipótesis nula con base en la contradicción encontrada tanto dentro de la literatura sobre un tópico dado, como en la contradicción existente entre los resultados obtenidos:

El modelo expuesto y adaptado para evaluar los 30 talleres de diseños de investigación utilizado, con un promedio anual de tres aplicaciones durante diez años en diversos escenarios hispanoamericanos -México, Ecuador, Perú y Argentina-, muestra cómo las cuatro etapas de la evaluación y los tres componentes clave del análisis de los datos forman un proceso interactivo y cíclico.

Es necesario advertir, que este modelo no difiere en forma importante de un modelo que puede ser construido para representar la estrategia empleada en estudios educacionales cuantitativos. La mayor diferencia radica en que la investigación cuantitativa frecuentemente parte de una sólida base de conocimiento y tiende a proceder secuencialmente en etapas bien definidas, más que en la forma interactiva e iterativa, que caracterizan a la investigación cualitativa. Esta es más fluida, pero conlleva una base de conocimiento debilitada. Por tal circunstancia, la investigación cuantitativa implica un enfoque altamente estructurado, mientras que la cualitativa descansa sobre una mínima preestructuración y sobre un modelo impreciso, construyendo la teoría a partir de las observaciones efectuadas en el mundo real, mostrando una estructura imprecisa y sin límites fijos –sin resultados previsibles, no limitado de antemano u ¿holístico?: *open-ended*– con respecto a la colección de los datos. Luego entonces, no sorprende, que ninguna descripción proveniente de la investigación cualitativa sea apropiada o suficientemente general.

Planeación

Wing (1972) y Uribe (1990) coincidieron en que la planeación a nivel general y la educativa en particular, consiste en anticipar las dificultades que posiblemente surjan cuando se intente lograr un cierto propósito y desarrollar en forma anticipada algunas medidas para resolverlas. Los problemas educacionales son posiblemente complejos, variados, de surgimiento continuo y frecuentemente no esperados; por ello, la estructura organizativa para tratar con ellos debe ser flexible y a largo plazo, como en este ejercicio científico de corte experimental. Por lo tanto, esta tarea corresponde a un proceso dinámico que facilite establecer cualitativa y cuantitativamente, hasta donde los recursos lo permitan, la magnitud de la situación o problema, pudiendo desprender de ese conocimiento inicial metas y objetivos claros, precisos y factibles, que puedan ser logrados a través de planes y programas –de estudio–. Sujetos éstos a la cuantificación, supervisión y sobre todo a la evaluación, que otorguen la experiencia –conocimientos ganados– usando el método científico, que no es infalible, pero minimiza en forma importante la influencia de los aspectos subjetivos emanados del sentido común –*common sense*– o de las personas que controlan el poder –educativo–, necesario para modificar a su vez, el punto o problema del cual se partió (Uribe, 1990).

En este terreno es indispensable distinguir entre la planeación racional, en contraste con la *ad hoc* o instantánea, que requiere de la adquisición y el uso de la información más relevante y actualizada –al día– concerniente a los probables efectos de cada uno de los cursos alternativos de las posibles acciones.

Las decisiones de los planeadores en parte son políticas, administrativas y educativo-científicas y esta compleja malla de interacción, ha llevado a los expertos de la educación a mostrar una clara tendencia a permanecer anclados en lo que ellos mismos denominan investigación operacional, siendo por definición el trabajo operacional más descriptivo que analítico y evidentemente no apoya a la parte administrativa, aunado a que los científicos miran con desdén este tipo de trabajo. Manteniendo al margen la inclinación apuntada, estos expertos han logrado desarrollar líneas de investigación que tienen un elevado valor en la planeación. Este tipo de quehacer científico educativo podría corresponder al término de la planeación

estratégica, debido a que esta tarea concierne a la aplicabilidad, siendo por ello *táctica* (Wing, 1972). Abreu, en 1990, senaló la necesidad de hallar la combinación de elementos estratégicos y tácticos en un programa evaluativo orientado directamente a captar ciertos problemas involucrados en el *desgaste educativo*, que hoy por hoy persisten, ya que los planeadores de la educación frecuentemente desconocen las implicaciones de tales líneas de investigación y por ello, no se encuentran en una posición para juzgar la aplicabilidad de los resultados. Con respecto a la pregunta qué problemas son *investigables*, se requiere de una compleja valoración integral. En consecuencia, la relación entre planeación y evaluación es crucial; verbigracia, los planeadores toman las decisiones, pero deben de prepararse para recibir los resultados de una evaluación independiente: ellos no pueden evaluar sin sesgo sus propias decisiones acerca de cierta política educativa. A los evaluadores, como se anotó, concierne establecer la calidad de tales decisiones. No obstante, no están capacitados para tomar decisiones y tal vez en este doble amarre evitativo —*determant interblock*— radique el meritorio mecanismo de salud sistémica. En resumen, la filosofía de la planeación es fundamentalmente política, mientras que la filosofía de la evaluación es fundamentalmente científica (Wing, 1972). Lo apuntado indica, que la administración de los fondos para la investigación evaluativa es materia tanto de planeadores —que deben estar aptos para decidir sobre áreas de elevada prioridad educativa—, como de evaluadores —que deben de insistir en una independencia científica—. Por ejemplo, en el Reino Unido una gran parte de este tipo de investigaciones tienen una base de largo alcance —*long-term basis*—, con lo que se logra establecer un balance entre investigaciones estratégicas y tácticas.

A nivel específico, la planeación de cualquier tipo de enseñanza tiene como principios fundamentales adecuar algunos componentes del acto educativo, tales como (Uribe, 1990):

- naturaleza de los materiales de estudio;
- métodos de estudio;
- tipo de enseñanza —individual y grupal—;
- actividades extracurriculares de los educandos;
- tipo de evaluación;
- utilidad de las evaluaciones;

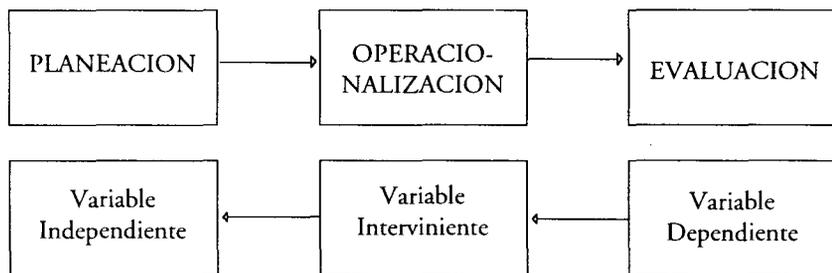
- función del docente de acuerdo al sistema de enseñanza seleccionado y,
- limitaciones temporales para el aprendizaje propuesto.

Y estos actos se encuentran enmarcados por tres premisas:

- que la intervención sea siempre racional y programada;
- que se involucren no sólo los aspectos tecnológicos, o de forma, en las relaciones estructurales internas y externas del sistema generador y multiplicador del conocimiento sustantivo y,
- que se haga el replanteamiento del problema, de las alternativas sugeridas, de los medios empleados y del conocimiento cuantificable de los resultados obtenidos.

Operacionalización

Para examinar alguna hipótesis, en 1982, Wellenreuther propuso que mediante la operacionalización es factible concretizar una situación específica –planeación– a partir de un concepto teórico –*basic term*–, que pueda ser medible y transportado a un escenario concreto a través de las reglas de correspondencia que permitan realizar las tareas de traducir la teoría en operaciones de investigación (Friedrichs, 1981). Por ello, se requiere interpretar los conceptos teóricos de un determinado estudio de corte científico para que deriven en los conceptos operacionalizados, a los que se denominarán variables, siendo factible articularlas en el siguiente esquema, el cual, se desprendió de la fórmula presentada previamente.



En 1927, el físico Bridgman (citado en Bortz, 1984) situó el concepto de *definición operacional* como sinónimo de *operacionalización* de una característica basada en:

- la definición operacional corresponde a un complejo de maniobras; por ejemplo, el concepto de *longitud* contiene ni más ni menos una serie de cálculos, mediante los cuales se expresa una longitud;
- un concepto no debe de ser definido de acuerdo a sus características, sino en relación con las operaciones de asociación;
- el significado verdadero –hipotético, según Popper (1982)– de un concepto no se integra con base en la observación acerca de lo que se expresa de él, sino que se debe de registrar lo que se hará con él y,
- el conocimiento todo tiene una correspondencia con las operaciones que fueron elegidas para medir y relativizar los conceptos científicos. Si existen varias ecuaciones de operación, señala que se encuentran fundamentadas en una multitud de conceptos y cada una debe orientarse a un concepto específico.

Por lo expresado, la definición operacional exige un extenso análisis del significado del concepto a definir; no obstante, este tipo de examen, que peculiariza a una cierta variable, no necesariamente muestra cómo fue operacionalizado el concepto. Verbigracia, si se establece una contradicción en las diversas formas de operacionalizar el mismo concepto, entonces indica que éste no ha sido analizado en forma precisa –cfr. cuarto punto bridgmaniano– y por ende, presenta un contenido empírico pobre (Friedrichs, 1981). Esto es, únicamente es posible determinar con el análisis del significado o definición analítica de cierto concepto si los indicadores utilizados en el proceso de la operacionalización tienen sentido o no. Al respecto, se anota que también las definiciones conceptuales precisas permiten la presencia de diversas operacionalizaciones. Ahora bien, no existe ningún inconveniente en que cierta variable funga como independiente o como dependiente dentro del concepto en una investigación dirigida a probar hipótesis. Si se trata de una variable exógena, que puede ser manipulada por el investigador, basta ofrecer a nivel experimental la expresión –*Ausprägung*– de la variable, cuyo significado se evidencia mediante

un grupo control –versión corta *versus* versión larga del taller–. Durante la operacionalización de la variable endógena es necesario cuidar que ésta pueda ser medida o referida con precisión a través de una estratificación muy diferenciada. Lo manifestado, se puede resumir con lo expresado por el filósofo sudafricano Johnson (citado en Steeb, 1991): *Define your terms, gentleman, define your terms. It saves arguments!*

Conjugando los problemas no resueltos en otros ejercicios (Gorenć y cols., 1988; Gorenć, 1991 y Oblitas y Gorenć, 1990) y de la circularidad (Castro, 1984) con el atentado a la lógica que da sustento al diseño factorial de cuatro grupos de Solomon (Bortz, 1984), considerando además que ambas restricciones se encuentran enmarcadas por una submuestra sesgada, se procedió a operacionalizar la evaluación en forma *post hoc* conforme a lo representado en la fórmula de Lewin (expuesta en Wellenreuther, 1982). Se hizo necesario concretar la variable endógena con apoyo en el modelo de evaluación propuesto por Miles y Huberman en 1984, para proceder a evaluar parcialmente –*Ep*–(Uribe, 1990) la planeación del evento educativo operacionalizado bajo el auspicio de la siguiente construcción hipotética:

$$H_0: (T_{vl} = T_{vc})^{[ePRE (T_{vl}, T_{vc}, POST)]}$$

donde:

- T_{vl} = taller versión larga
- T_{vc} = taller versión corta
- ePRE = efecto del pretest –primer registro– sobre ...
- T_{vl}, T_{vc} = ambos talleres
- POST = post-test –segundo registro–

Material y métodos

Probandos

En cada uno de los 30 talleres impartidos fueron entrenados en el manejo de los diseños de investigación un promedio de 21.93 clientes. Para cuantificar el efecto del primer registro sobre las dos modalidades de manipulación y sobre el segundo registro, la muestra fue estratificada como a continuación se esquematiza:

Figura 1: Modelo de evaluación interactivo adaptado para datos cuantitativos concernientes a establecer la efectividad educativa. (Miles y Huberman, 1984)

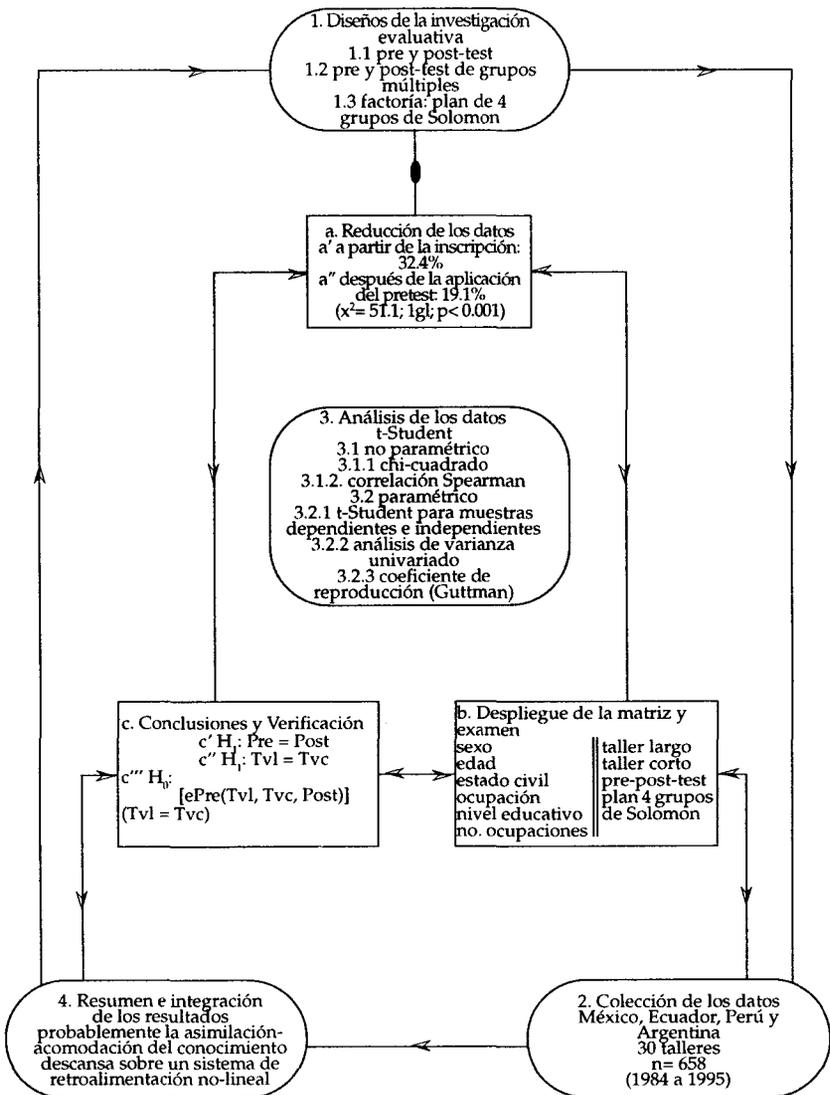
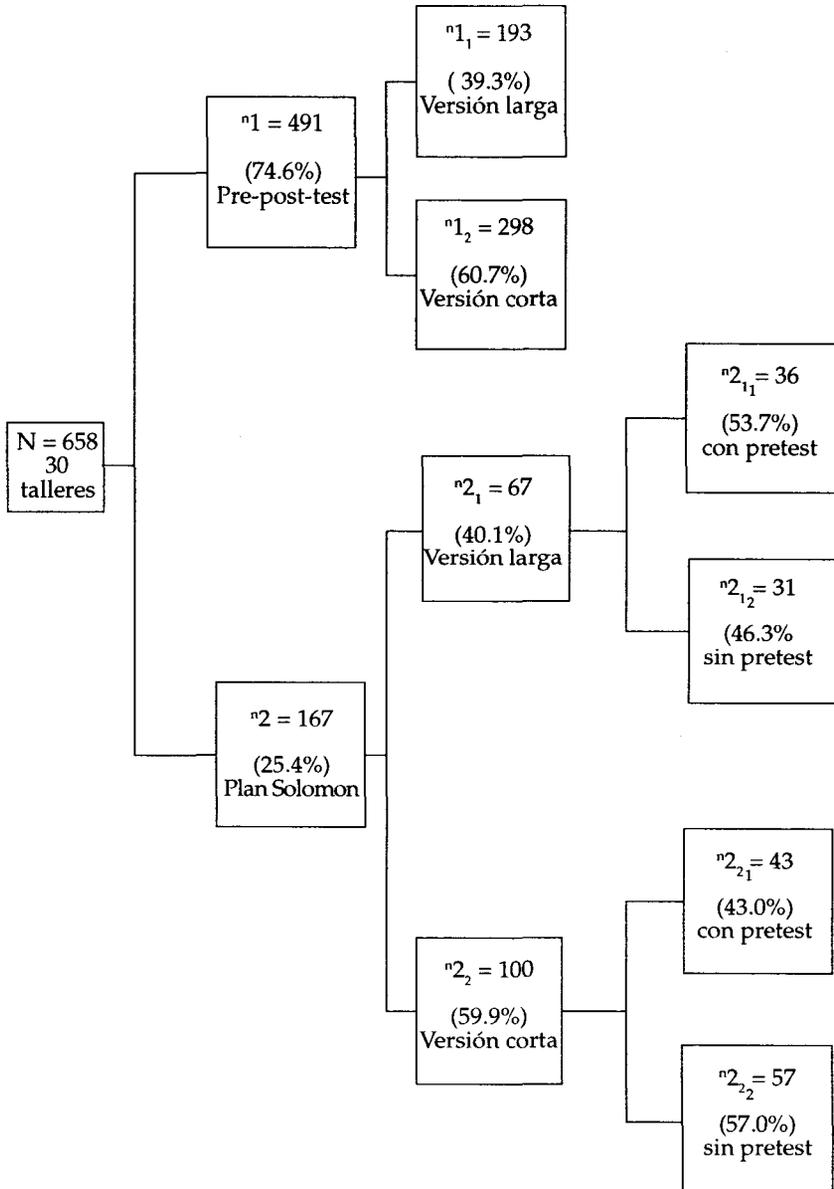


Tabla 1: Estructura de la muestra



Apartándose de la mera comparación por cociente entre el número de probandos de las submuestras expuestas en la primera tabla y el número de participantes de la muestra a la que pertenecen las apuntadas submuestras, a través de la medida de resumen que se refiere a la comparación por cociente entre dos cifras de similar naturaleza, se demuestra que este ejercicio corresponde a los denominados *post hoc*, debido a que por cada probando destinado al plan de cuatro grupos de Solomon, 2.51 participantes fueron evaluados con el diseño de antes y después. Sin embargo y por efecto de la casualidad, esta disimilitud se redujo a 1:1.54 y 1:1.49 entre las versiones largas y cortas correspondientes a los diseños de pre y post-test y cuatro grupos de Solomon respectivamente. A partir de este momento se observa la acción de la técnica de distribución aleatoria simple (Gorené y cols., 1986) entre los probandos que fueron o no fueron expuestos a la primera medición y sometidos a las versiones larga (1:1.16) y corta (1:1.32), que nuevamente se amplía cuando por cada probando adiestrado con la versión larga del taller que no fue medido con el pretest, 1.83 corresponden a la versión corta. Este desbalance no se registró entre los participantes de ambas versiones a los cuales se les aplicó la primera medición (1:1.19). La dinámica muestral descrita, indicó la necesidad de establecer el efecto sobre la representatividad dentro y entre los diseños y modalidades del taller empleadas a través de los descriptores de la muestra.

Considerando que la pérdida de probandos limita la validez interna, se decidió examinar dicho fenómeno dentro de este rubro, que debería situarse en el contexto de los resultados, con el propósito de fijar la atención del lector sobre el efecto que pueda tener o no el pretest sobre la reducción de la muestra. Por ende y sin controlar la intervención de las versiones, las pérdidas serán analizadas desde dos perspectivas: a partir de la inscripción –32.7%– y después de la aplicación del pretest –19.1%–, diferencia que está asegurada a nivel estadístico $\chi^2 = 51.1$; 1 gl.; $p \leq 0.001$ –, indicando que para reducir el costo de operación del total de las inscripciones registradas, deberán reducirse los recursos materiales en una tercera parte. Partiendo del principio que el todo oculta el comportamiento de las partes, el resultado anteriormente anotado fue desglosado según las versiones del taller, encontrándose en la modalidad larga una pérdida significativa tanto desde la inscripción $\chi^2 = 38.7$; 6 gl.; $p \leq 0.001$ – como a partir del pretest

$-x^2 = 18.8$; 5 gl.; $p \leq 0.001$ -. Igualmente persistió la pérdida en la versión corta cuando la pérdida se examinó desde la inscripción $-x^2 = 54.3$; 5 gl.; $p \leq 0.001$ -, pero el efecto del pretest fue mínimo $-x^2 = 7.6$; 4 gl.; $p = 0.20$ -. No obstante, comparando las pérdidas entre ambas versiones a partir de la inscripción, se observó una mayor proporción en la reducción de la muestra en la versión corta (61.3%) que en la larga y ésta se encuentra mínimamente influida por el azar $-x^2 = 14.3$; 1 gl.; $p = 0.005$ -, mientras que la disminución registrada a partir de la aplicación del pretest entre las versiones no fue significativa $-x^2 = 1.2$; 1 gl.; $p = 0.20$ -. En consecuencia, el efecto del pretest sobre el tamaño de la muestra en la versión larga del taller, pierde su peso al establecer que el reducido monto de la pérdida es semejante entre ambas versiones, contrastando estos resultados con las extraordinarias reducciones registradas a partir de las inscripciones, donde también es mayor la diferencia de pérdida en la versión corta (78.3%) que en la larga $-x^2 = 47.0$; 1 gl.; $p \leq 0.001$ -, que a su vez está intervenida por la reducción que se observó entre la inscripción y la aplicación del pretest de la versión corta del taller. Lo anotado puede estar relacionado con la actitud de los participantes ante una mayor presión y atención sostenida implícitas en dos días de trabajo intenso durante el tiempo libre –fines de semana–, que los cinco días destinados para la versión larga durante la jornada laboral. Este resultado visto desde la perspectiva de la teoría de la motivación, presenta un riesgo de confusión entre la aparente falta de motivación de la muestra de la versión corta del taller y su menor rendimiento registrado en el post-test, debido a que se infiere que existe un mayor estado motivacional de aquéllos que sí aceptaron trabajar durante los días de asueto. Esto no fue posible establecer, ya que no se midió la motivación, como variable interviniente (Wellenreuther, 1982), de los participantes en ambas versiones y en el caso de no hallarse la diferencia esperada según los resultados expuestos previamente, se hablaría de una irrelevancia de la motivación sobre el rendimiento de ambas versiones del taller.

Descriptores de la muestra

Con respecto a la variación del número de probandos según la estructuración de la muestra apuntada en la primera tabla, de las seis variables exógenas utilizadas para peculiarizar a la muestra y submuestras,

tres fueron recodificadas debido a que sus valores estaban fuertemente estratificados, incrementando su dispersión dentro de las celdillas, cuando estaban medidas a nivel de escala nominal –estado civil y profesión u ocupación– o de rango –posgrados: especialidad, maestría, doctorado posdoctorado–, impidiendo el análisis de los datos cuando fueron sometidos a la prueba de significancia de chi-cuadrado, considerando las restricciones implícitas en esta prueba para refutar o no la hipótesis nula:

- si los grados de libertad –gl– = 1; es decir, $k = 2$, entonces ninguna frecuencia esperada –por celdilla o intervalo– debe ser < 5 (Siegel, 1976), preferible 10 (Nesbitt, 1966);
- si los grados de libertad –gl– > 1 ; es decir, $k > 2$, entonces no más del 20% de las frecuencias esperadas –por celdilla o intervalo– debe ser < 5 (Siegel, 1976), preferible 10 (Nesbitt, 1966).

El estado civil se dicotomizó en participantes sin pareja –solteros, divorciados, separados y viudos– y con pareja –casados y que viven en unión libre–. Para taxomizar la variable profesión, se utilizó el parámetro de clasificación de la ciencia según el objeto de estudio: ideas o hechos; esto es, formal –lógica, matemática– y fáctica –antrópica: antropología, sociología y psicología; biótica: ecología, genética, fisiología y física: química y física– respectivamente (Oblitas, 1987). La variable posgrado también se dividió en con o sin este atributo.

Diseños

Referente a la estratificación de la muestra expuesta en la primera tabla, cada taller, independientemente de las versiones y diseños –de los probandos sometidos al de cuatro grupos de Solomon, únicamente se contrastaron a los que fueron medidos con el pretest–, fue evaluado en forma individual bajo el siguiente planteamiento hipotético:

$$H_0: \text{PRE} = \text{POST}$$

donde:

PRE = pretest –primer registro–

POST = post-test –segundo registro–

operacionalizándose, como se mencionó en el preámbulo, mediante un diseño de pre y post-test:

PRE T POST

donde:

T = taller –manipulación–;

que posteriormente sufrió una duplicación, debido a que el taller original –versión larga– fue ajustado a las circunstancias hispanoamericanas –peruanas– para difundir la cultura científica básica a uno de bajo costo de operación, modificándose para obtener una versión corta (Gorenć, 1991):

PRE Tvl POST

PRE Tvc POST

donde:

Tvl = taller –manipulación– versión larga

Tvc = taller –manipulación– versión corta

y esta circunstancia, facilitó la salida del escenario de la hiperrealidad baudrillardiana (1987) que había sido empleado en dos reportes previos (Gorenć y cols, 1988 y Oblitas y Gorenć, 1990), así como abocarse, en el terreno empírico, al problema de la circularidad implícita en el diseño de pre y post-test –de un grupo– expuesto por Castro (1984).

Finalmente, y en forma tardía, se intentó establecer el efecto del primer registro sobre la manipulación larga y el segundo registro (Gorenć y cols., 1989). Debido al sesgo y al reducido tamaño de la muestra no fue posible generalizar los resultados dentro de las muestras sometidas a esta versión del taller ni a los probandos que integraron las muestras adiestradas en el manejo de los diseños de investigación mediante la versión corta de la manipulación. El experimento fue ampliado para ambas versiones y en consecuencia, la estructura del diseño anterior fue factorizada:

	PRE	Tvl	POST
		Tvl	POST
PRE		Tvc	POST
		Tvc	POST

señalando que aproximadamente la mitad del número de probandos sometidos tanto a la versión larga como a la corta de la manipulación se mantuvieron controlados de la influencia del primer registro –cfr. primera tabla–. Con este diseño –cuatro grupos de Solomon– se operacionalizó la hipótesis colocada al final de la introducción.

Medidas de seguridad del pre y post-test

Ambos instrumentos de medición fueron sometidos a dos procesos de validación: criterio exterior simulado (Baudrillard, 1987) en su modalidad de concurrente y por construcción, así como por dos técnicas de confiabilidad contenidas en el grupo donde se utiliza una sólo medición: división por mitades y consistencia interna (Carmines y Zeller, 1982).

Materiales

El taller, independientemente de las versiones, fue estructurado mediante la combinación de un material diadáctico con uno de apoyo y para evaluar la eficacia se construyó un pre y un post-test, ambos derivados de los 26 temas del material didáctico que originó un número igual de reactivos, mostrando la siguiente distribución: el 8% de los items estaba destinado a explorar el conocimiento de los participantes acerca de la utilidad de los diseños de investigación; el mayor porcentaje –42%– cubría el área de los conceptos básicos con los cuales se *movilizan* los diseños no experimentales –19% de los items– y experimentales –31% de los reactivos–, cuyas respuestas de opción múltiple, para efectuar el análisis, fueron transformadas en dicotómicas –acierto/error–. Posteriormente, este material, probado en el campo empírico, se ofreció en el mercado como un texto autoadministrable (Gorenć y cols., 1986). Para controlar la influencia del docente, se preparó una serie con 22 diapositivas y dos guías –formatos– para realizar los ejercicios correspondientes. Todo el material se encuentra dispuesto para los lectores interesados sin costo alguno, así como los programas para la evaluación y la base de datos emplada en este ejercicio.

La duración del taller se fijó arbitrariamente en 22.30 horas durante cinco días/4.46 horas diarias –versión larga–, que formalmente se debería haber establecido con apoyo en los resultados obtenidos de la literatura y/o mediante ejercicios científicos. Verbigracia, con un diseño de pre y

post-test de grupos múltiples (Spector, 1982), donde las manipulaciones –talleres– estarían caracterizadas por diversas duraciones, eligiendo aquélla que hubiera mostrado una correlación inversamente proporcional entre la eficacia del aprendizaje obtenido bajo la tutela de la teoría del ensayo y el error y el tiempo: días/horas –mayor eficacia en el menor tiempo–. Ambos momentos hubieran incrementado el costo del taller; no obstante, la duración apuntada fue ajustada a las circunstancias hispanoamericanas –disminuir el costo de operación–, reduciéndose a 2 días/11.15 horas diarias –versión corta–. Para evaluar la eficacia del taller, se utilizó el diseño de pre y post-test; esto, con el propósito de establecer el nivel de conocimientos acerca de los diseños de investigación antes de someter a los participantes a una de los dos modalidades del taller correspondiente, que asiste al docente para situar los pesos durante la exposición oral. Este primer registro fue contrastado con la segunda medición, que es semejante a la primera, pero para controlar la influencia de la memoria, se modificó la secuencia de las preguntas y de las cinco opciones de las respuestas empleando el principio de la aleatoriedad mediante la técnica simple (Rojas, 1982).

Procedimientos estadísticos

Para establecer si las submuestras incluidas en el plan de cuatro grupos de Solomon representaban *bien* a la muestra de donde fueron extraídas, así como para comprobar la efectividad de la aleatoriedad empleada para someter o no a los probandos a la primera medición, los descriptores medidos tanto a nivel de escala nominal como ordinal fueron examinados mediante la prueba de significancia de chi-cuadrado de doble entrada (Nesbitt, 1966 y Siegel, 1976) y cuando éstos estaban medidos a nivel de escala de razón, se utilizó la prueba de significancia t-Student para muestras independientes.

Para evaluar los talleres, así como cada reactivo en forma individual, siguiendo la estructura del diseño correspondiente, el primer registro se contrastó con el segundo a través de la prueba de significancia t-Student, pero en esta ocasión, para muestras dependientes; mientras que para comparar ambas modalidades de manipulación –larga y corta–, se empleó nuevamente esta prueba de significancia en su modalidad para muestras

independientes. Para cuantificar el efecto del pretest sobre las manipulaciones y el post-test se combinaron las dos estratificaciones anotadas de la prueba de significancia t-Student con el análisis de varianza, requeridas por la estructura de este diseño incompleto desde la perspectiva estadística (Gorenć y cols., 1986). Con apoyo en el algoritmo de Guttman (Nie y Hull, 1983) se examinó la variación del grado de dificultad de los reactivos antes y después de la manipulación; y para determinar la diferencia entre el orden por rangos de los items entre ambos momentos, se empleó la correlación de rango, según la siguiente fórmula (Haseloff y Hoffmann, 1965):

$$s = 1 - \frac{6 \sum d^2}{N * (N^2 - 1)}$$

donde:

- s = correlación de rango
- d = diferencia entre los rangos
- $\sum d^2$ = suma de todas las diferencias al cuadrado
- N = número de reactivos

Para efectuar el análisis estadístico previamente anotado, se utilizó una computadora SUN (Sparc Server 10) con un sistema operativo UNIX (Solaris 2.2) donde se encuentra implantado el paquete estadístico denominado *Statistical Package for the Social Sciences* –SPSSx: versión 4.0–, del cual se utilizó el subprograma CROSSTABS para determinar la representatividad dentro y entre las muestras; el subprograma T-TEST para muestras dependientes –PAIRS– e independientes –GROUPS– para evaluar el taller antes y después de las manipulaciones, así como el subprograma ANOVA para establecer el efecto del pretest sobre el post-test.

Para determinar la diferencia del grado de dificultad de los reactivos entre ambos registros, y para establecer la magnitud del coeficiente de reproducción, fue necesario utilizar una computadora CYBER-173 –sistema operativo NOS 2.7.1 688-716– donde aún se encuentra situada la versión 8.3 del SPSS, de la cual se utilizó el subprograma GUTTMAN SCALE, que fue excluido de la versión 4.0 del paquete SPSSx.

Con el propósito de establecer la validez de ambos instrumentos, se activaron los subprogramas CROSSTABS –coeficiente de contingencia de Cramer– y FACTOR y la precisión de la medición se calculó mediante el subprograma RELIABILITY con los modelos de SPLIT y ALPHA.

Resultados

Comportamiento del pre y post-test

Ambos instrumentos mostraron restricciones con respecto a su construcción, mientras que con una técnica, considerada en la literatura como endeble (Carmines y Zeller, 1982 y Holtzman, 1980), dirigida a establecer el nivel de asociación entre el criterio de los instrumentos con el denominado como exterior, en este caso, simulado (Baudrillard, 1987), se obtuvo un coeficiente aceptado como adecuado para demostrar en el terreno empírico que ambos instrumentos miden lo que pretenden medir (Carmines y Zeller, 1982): conocimiento sobre la cultura científica básica. La tabla 2 muestra además que no existen diferencias significativas entre los valores, independientemente de las técnicas de validez empleadas, de ambas mediciones.

Los valores expuestos en la parte inferior de la tabla anterior, complementan el binomio que todo instrumento de medición también debe de exhibir: la confiabilidad, que sistemáticamente fue inferior en el pretest, en contraste con el post-test, pero sin refutar en ningún momento la hipótesis nula implícita en esta comparación. El comportamiento de los coeficientes apuntados mostró un entrecruzamiento con respecto a los reactivos dicotomizados en dos colas. El perfil de estos valores señala, que ambos instrumentos muestran una precisión ligeremante por debajo de los parámetros exigidos por la literatura (Carmines y Zeller, 1982). Esto significa, que los reactivos deberán ser sometidos a modificaciones y a la evaluación correspondiente.

Tabla 2: Comparación de los diversos errores en la medición del pre y post-test

	P r e t e s t	P o s t e s t	X ²	P (1 gl)
Validez criterio exterior simulado, modalidad concurrente (coeficiente Cramer)	0.86	0.87	0.0001	>20%
Validez por construcción: número factores (valor propio ≥ 1.0)	10	9	0.53	>20%
Porcentaje de la variancia	52.2	52.5	0.0001	>20%
Número reactivos ≥ 0.5000	19	17	0.1	>20%
Confiabilidad consistencia interna: Coeficiencia a de Cronbach	0.62	0.74	0.01	>20%
Confiabilidad división por mitades:				
Correlación entre las formas	0.35	0.55	0.05	>20%
Coeficiente Spearman-Brown	0.35	0.55	0.03	>20%
Coeficiente α primera cola items > 0.60150 / 0.42231	2	3	/	
Coeficiente α segunda cola items > 0.35273 / 0.70560	9	0	/	>20%

/ se violó primera regla de X² (Nesbitt, 1966)

Consideraciones acerca de la validez interna y externa

En un primer plano, se observó que la submuestra correspondiente a los cuatro grupos de Solomon no representó en forma adecuada a la submuestra que fue examinada mediante el diseño de pre y post-test. Esto es, de las cuatro variables exógenas dicotomizadas y medidas a nivel de escalas nominal y ordinal, una mostró diferencia significativa: el descriptor de la ocupación/profesión en relación al objeto de estudio de la ciencia, donde se encontró una imagen en espejo entre la diferencia encontrada en la versión corta del taller –por cada probando de las ciencias formales, 1.78 se ubicó en las fácticas en la submuestra del pre y post-test. Lo opuesto se halló en la muestra que constituyó la de los cuatro grupos de Solomon: 1:1.32– y al no controlar las versiones: en la submuestra perteneciente al diseño de antes y después la razón fue de 1:1.24 –formal:fáctica– y en la de los cuatro grupos de Solomon fue de 1:1.61 –fáctica:formal–. Esta diferencia refutó la hipótesis nula implícita.

La siguiente limitación se observó al comparar los dos indicadores sociodemográficos medidos a nivel de escala de razón, lo que demuestra que a menor estratificación de las variables menor precisión en los resultados. Así, por ejemplo, el mayor promedio de ocupaciones hallado en la submuestra que integró el diseño cuatro grupos de Solomon, se debió preponderantemente al más elevado promedio de ocupaciones que arrojó la submuestra adiestrada con la versión corta del taller lo que habla a favor de un incremento significativo a la *fuga interna de cerebros* (Falcony y cols., 1986), lo cual asociado a la elevada *fuga externa de cerebros* durante las dos últimas décadas, han *descapitalizado* a nivel intelectual y en forma brutal a este subcontinente. Como ejemplo de la última modalidad de *fuga* anotada es el Premio Nobel de Química otorgada en el mes de octubre de 1995 al mexicano Mario José Molina, egresado de la Facultad de Química, de la Universidad Nacional Autónoma de México y actualmente es investigador del Departamento de Ciencias Atmosféricas y Planetarias del Instituto Tecnológico de Massachusetts, Estados Unidos de América (Gaceta UNAM, 1995).

Tabla 3: Análisis de la representatividad de las muestras utilizadas para estructurar el diseño de cuatro grupos de Solomon

Diseños	M a s c u l i n o		F e m e n i n o		S i n p a r e j a		C o n p a r e j a		F a c t i c a s		F o r m a l e s		S i n p o s g r a d o		C o n p o s g r a d o	
	Sin controlar las versiones															
Pre-post test (n = 491)	317 (311)	174 (179)	235 (242)	256 (248)	272 (251)	219 (240)	440 (436)	51 (55)								
Cuatro grupos Solomon (n = 167)	101 (126)	66 (61)	90 (83)	77 (84)	64 (85)	103 (82)	144 (148)	23 (19)								
chi-cuadrado *significancia	0.89663 0.34369		1.81315 0.17813		14.53747 0.00014		1.43094 0.23161									
Versión larga																
Pre-post test (n = 193)	102 (97)	91 (96)	102 (104)	89 (89)	81 (76)	112 (117)	168 (168)	25 (25)								
Cuatro grupos Solomon (n = 67)	29 (34)	38 (33)	36 (36)	31 (31)	21 (26)	46 (41)	58 (58)	9 (9)								
chi-cuadrado *significancia	1.82062 0.17724		0.00048 0.98254		2.35536 0.12485		0.01006 0.92011									
Versión corta																
Pre-post test (n = 298)	215 (214)	83 (83)	131 (139)	167 (160)	191 (175)	107 (123)	272 (268)	26 (30)								
Cuatro grupos Salomon (n = 100)	72 (72)	28 (28)	54 (47)	46 (54)	43 (59)	57 (41)	86 (90)	14 (10)								
chi-cuadrado *significancia	0.00081 0.97727		3.03416 0.08153		13.75168 0.00021		2.30478 0.12898									

* Grados de libertad = 1

() Distribución de los valores esperados

Tabla 4: Examen de la representatividad de las muestras sometidas al diseño de cuatro grupos de Solomon

Grupos	Número de casos	Promedio	Desviación estándar	Error estándar	Valor t	Grados de libertad	Prueba de dos colas
--------	-----------------	----------	---------------------	----------------	---------	--------------------	---------------------

Edad: Sin controlar las versiones

Pre-post	491	31.73	8.38	0.37	0.09	369.86	0.930
Solomon	167	31.79	6.45	0.49			

Número de ocupaciones: sin controlar las versiones

Pre-post	491	1.40	0.67	0.03	4.49	287.70	0.000
Solomon	167	1.66	0.66	0.05			

Edad: versión larga

Pre-post	193	31.20	8.22	0.59	1.50	157.52	0.134
Solomon	67	32.61	5.98	0.73			

Número de ocupaciones: versión larga

Pre-post	193	1.28	0.63	0.05	1.99	122.05	0.049
Solomon	67	1.45	0.58	0.07			

Edad: versión corta

Pre-post	298	32.08	8.47	0.49	1.01	212.72	0.312
Solomon	100	31.24	6.72	0.67			

Número de ocupaciones: versión corta

Pre-post	298	1.47	0.68	0.04	4.30	171.34	0.001
Solomon	100	1.81	0.67	0.06			

La primera y tercera tablas señalan que los ejercicios científicos *post hoc* limitan gravemente la validez externa; es decir, que el efecto del pretest, en sus dos modalidades, medido a través del diseño de cuatro grupos de Solomon, presenta restricciones con respecto a la generalización de los resultados hacia la submuestra evaluada con el diseño de pre y post-test –uno y dos grupos–.

En la siguiente tabla se observa el efecto de la utilización en forma adecuada de la técnica de selección aleatoria simple, propuesta entre otros por Rojas (1982). En contraste con la tercera tabla, las submuestras sometidas o no al primer registro, por estructura del diseño factorial empleado e independientemente de las versiones –larga y corta–, son comparables con aquéllas que no lo fueron. Esto indica que la influencia que podrían tener estas variables sobre ambas mediciones es mínima o al menos, no significativa.

Con el propósito de obviar espacio, se omitió la tabla de las variables exógenas medidas a nivel de escala de razón –edad y número de ocupaciones–, que tampoco mostraron diferencias significativas entre los probandos que integraron las dos versiones en sus modalidades de haber sido o no medidos con el pretest.

$$H_0: \text{PRE} = \text{POST}$$

La hipótesis nula se refutó en los 30 eventos, duplicándose el conocimiento sobre la cultura científica básica en el 50% de las aplicaciones del taller, independientemente de los escenarios hispanoamericanos, correspondiendo el 50% tanto de las 16 versiones largas como de las 14 cortas, como se muestra en la subsiguiente tabla.

No obstante, por la estructura del diseño empleado, al no superar la circularidad, lo único que es factible externar al emplear este diseño de un grupo es que se produjo un cambio entre el primer y segundo registro. Por tanto, fuera de esta limitada afirmación, toda especulación acerca del papel que desempeña la manipulación –talleres: versión larga y corta– en la determinación del cambio sería aventurada (Castro, 1984). A lo cual, cabe preguntar, si la consistencia de los resultados durante 30 momentos sin controlar las variables exógenas y las fuentes de invalidación, permite romper con la afirmación circular, ya que sería altamente improbable que

Tabla 5: Examen de la aleatoriedad de los sujetos que fueron y no sometidos a la primera medición

Diseños	M a s c u l i n o	F e m e n i n o	S i n p a r e j a	C o n p a r e j a	F a c t i c a s	F o r m a l e s	S i n p o s g r a d o	C o n p o s g r a d o
Sin controlar las versiones								
Pre-test (n = 79)	47 (48)	32 (31)	49 (43)	30 (36)	28 (30)	51 (49)	69 (68)	10 (11)
Post-test (n = 88)	54 (53)	34 (35)	41 (47)	47 (41)	36 (34)	52 (54)	75 (76)	13 (12)
chi-cuadrado *significancia	0.06090 0.80508		3.99092 0.04575		0.52621 0.46821		0.15673 0.69219	
Versión larga								
Pre-test (n = 36)	14 (15)	22 (20)	21 (19)	15 (17)	10 (11)	26 (24)	31 (31)	5 (5)
Post-test (n = 31)	15 (13)	16 (18)	15 (17)	16 (14)	11 (10)	20 (21)	27 (27)	4 (4)
chi-cuadrado *significancia	0.61213 0.43399		0.66282 0.41557		0.45965 0.49779		0.01391 0.90609	
Versión corta								
Pre-test (n = 43)	33 (31)	10 (12)	28 (23)	15 (20)	18 (19)	25 (25)	38 (37)	5 (6)
Post-test (n = 57)	39 (41)	18 (16)	26 (31)	31 (26)	25 (25)	32 (33)	48 (49)	9 (8)
chi-cuadrado *significancia	0.84222 0.35876		3.75285 0.05272		0.03997 0.84154		0.35256 0.55267	

* Grados de libertad = 1

() Distribución de los valores esperados

Tabla 6: Diferencias entre el primer y segundo registro utilizando la prueba de significancia t-Student para muestras dependientes

n's	$\sum fx/n$ R ₁	Patrocinadores	Versiones		$\sum fx/n$ R ₂	Valor t	gl (n-1)	Prueba de dos colas p
			Larga	Corta				
16	10.19	Secretaría de Gobierno, estado de Tabasco, México	*		17.06	7.69	15	0.001
10	10.30	Instituto Nacional de Ciencias Penales, Procuraduría General de la República, México	*		21.10	6.45	9	0.001
12	9.50	Coordinación de la Licenciatura en Derecho, División de Ciencias Sociales y Humanidades, Universidad Autónoma Metropolitana (Unidad Azcapotzalco), México	*		19.25	9.89	11	0.001
13	8.00	Instituto de Administración Pública de Tabasco, México	*		19.92	14.97	12	0.001
23	9.65	Asociación Ecuatoriana de Psiquiatría, sede provincia del Azuay, Ecuador	*		19.61	12.66	22	0.001
15	12.87	Centro de Investigaciones Jurídicas, Universidad Nacional de Córdoba, Argentina (taller matutino)	*		21.07	8.04	14	0.001
33	11.82	Centro de Investigaciones Jurídicas, Universidad Nacional de Córdoba, Argentina (taller vespertino)	*		21.27	13.93	32	0.001
7	11.00	Instituto Nacional de Ciencias Penales, Procuraduría General de la República, México (Maestría en Criminología, cohorte 13)	*		21.85	11.29	6	0.001
19	8.73	Maestría en Derecho Penal y Especialidad en Sistemas Penitenciarios, División de Estudios de Posgrado de la Facultad de Derecho Universidad Autónoma de Querétaro, México	*		19.63	9.61	18	0.001
22	13.81	Dirección de Estadísticas Demográficas y Sociales, Instituto Nacional de Estadística, Geografía e Informática, estado de Aguascalientes	*		20.54	5.89	21	0.001
<3>	8.66	Fundación Mexicana de Asistencia a Víctimas, México	*		17.33	9.83	2	0.010
<7>	16.42	Maestría en Educación Médica, División de Estudios de Posgrado e Investigación, Facultad de Medicina, Universidad Nacional Autónoma de México	*		22.14	3.09	6	0.021
13	9.15	Seminario de Investigación Jurídica, Maestría en Administración de Justicia, Facultad de Derecho, Universidad Autónoma del estado de México	*		13.46	3.16	12	0.008
57	7.56	Procuraduría General de Justicia, estado de Tabasco, México	*		17.32	17.94	56	0.001
18	7.57	Colegio y Barra de Abogados del Ayuntamiento de Macuspana, estado de Tabasco, México	*		16.13	8.49	17	0.001
27	8.10	Colegio de Bachilleres, plantel 1, Villahermosa, estado de Tabasco, México	*		15.47	9.57	26	0.001
18	7.69	Colegio y Barra de Abogados del Ayuntamiento de Cárdenas, estado de Tabasco, México	*		17.14	14.43	17	0.001
40	8.22	Colegio de Bachilleres, plantel 6, municipio de Cunduacán, estado de Tabasco, México	*		16.92	15.58	39	0.001
18	7.76	Colegio y Barra de Abogados del Ayuntamiento de Comalcalco, estado de Tabasco, México	*		14.23	5.99	17	0.001
25	7.64	Colegio y Barra de Abogados del Ayuntamiento de Huimanguillo, estado de Tabasco, México	*		12.72	6.58	24	0.001
9	12.88	Posgrado de Psiquiatría, Hospital Psiquiátrico Fray Bernardini Alvarez, Secretaría de Salud, México	*		17.11	3.17	8	0.013
29	9.00	Maestría en Ciencias Penales con sede en el puerto de Acapulco, estado de Guerrero, México	*		16.68	7.16	28	0.001
20	9.65	Maestría en Derecho Penal, División de Posgrado de la Escuela de Derecho, Universidad Autónoma de Hidalgo, México	*		18.68	7.16	28	0.001
18	11.44	Asociación Psiquiátrica Peruana, sede Lima Metropolitana, Perú	*		19.39	8.60	17	0.001
10	13.60	Secretaría de Educación Médica, Facultad de Medicina Universidad Nacional Autónoma de México	*		20.10	6.21	9	0.001
9	10.44	Sociedad de Psicología de Guanajuato, sede Celaya, México	*		18.22	4.19	8	0.003
10	7.23	Coordinación de la Maestría en Política Criminal, Escuela Nacional de Estudios Profesionales Acatlán, Universidad Nacional Autónoma de México	■		19.52	7.83	9	0.001
17	5.94	Maestría en Ciencias Penales, Universidad Autónoma de Campeche, México	■		12.82	11.14	16	0.001
9	9.11	Diplomado de Metodología, Dirección de Extensión Universitaria, Universidad Intercontinental, México	■		17.33	5.63	8	0.001
43	8.23	Instituto Nacional de Ciencias Penales, Procuraduría General de la República, México (Maestría en Ciencias Penales, cohortes 14 y 15)		■	17.25	13.07	42	0.001
570	946	30	13/3	13/1	17.69	45.06	569	0.000

■ Muestras dicotomizadas para someterlas al diseño de cuatro grupos de Solomon

<3> Unicamente con fines descriptivos, debido a que siete participantes no presentaron el post-test (ver pérdidas)

<7> Unicamente con fines descriptivos, debido a que dos participantes no presentaron el post-test (ver pérdidas)

Tabla 7: Dinámica de los reactivos del pre y post-test utilizando la prueba de significancia t-Student para muestras dependientes; se excluyeron las muestras utilizadas para conformar el diseño de cuatro grupos de Solomon

Areas	Reactivos del pre-test	Orden en R2	Dificultad* R1 R2		Registros	Σ fx/n	Valor t	Prueba de dos colas**	
Diseños (introducción)	1 Definición de diseño	22	(26)	(26)	R1 R2	7.42 8.61	5.34	0.000	
	2 Control de variables	21	(23)	(17)	R1 R2	5.54 7.38	7.05	0.000	
	3 Definición de variables	18	(22)	(1)	R1 R2	5.24 2.19	10.37	0.000	
Conceptos y lógica básica	4 Niveles de medición de las variables	5	(12)	(8)	R1 R2	3.12 5.93	10.21	0.000	
	5 Errores en la medición	20	(15)	(25)	R1 R2	3.98 8.61	19.05	0.000	
	6 Definición de confiabilidad y técnicas	6	(21)	(15)	R1 R2	4.93 7.07	7.97	0.000	
	7 Definición de validez	16	(16)	(12)	R1 R2	4.12 6.73	10.08	0.000	
	8 Definición de control y técnicas	23	(19)	(10)	R1 R2	4.77 6.29	5.81	0.000	
	9 Definición de aleatoriedad y técnicas	14	(18)	(23)	R1 R2	4.56 8.36	16.23	0.000	
	10 Definición de validez interna	8	(3)	(5)	R1 R2	1.66 4.82	12.56	0.000	
	11 Definición de validez externa	19	(8)	(6)	R1 R2	2.28 4.93	9.37	0.000	
	12 Secuencia lógica de una investigación	3	(17)	(4)	R1 R2	4.19 4.73	1.78	0.076	
	13 Fuentes de invalidez	24	(08)	(22)	R1 R2	2.56 7.91	22.85	0.000	
	Diseños no experimentales (de un grupo)	14 Definición diseños no experimentales	1	(1)	(7)	R1 R2	1.01 5.91	20.87	0.000
		15 Estructura del diseño pre y post-test	25	(13)	(18)	R1 R2	3.40 7.42	15.83	0.000
		16 Estructura diseño de serie de tiempo interrumpida	9	(14)	(21)	R1 R2	3.50 7.89	17.39	0.000
17 Estructura diseño de sección cruzada		2	(2)	(2)	R1 R2	1.42 4.10	10.50	0.000	
18 Estructura diseño longitudinal		4	(6)	(11)	R1 R2	2.10 6.63	17.91	0.000	
19 Estructura diseño de dos grupos		13	(9)	(14)	R1 R2	2.89 6.96	14.91	0.000	
Diseños de grupos múltiples	20 Estructura diseño de post-test de grupos múltiples	26	(10)	(13)	R1 R2	3.00 6.82	15.15	0.000	
	21 Estructura diseño de pre y post-test de grupos múltiples	10	(24)	(19)	R1 R2	5.63 7.63	7.87	0.000	
	22 Estructura diseño ex post-facto	15	(05)	(16)	R1 R2	1.80 7.17	22.31	0.000	
	23 Estructura diseño de serie de tiempo en grupos múltiples	7	(20)	(24)	R1 R2	4.82 8.45	15.25	0.000	
	24 Gráfico del diseño de tipo factorial	12	(11)	(3)	R1 R2	3.07 4.73	6.11	0.000	
	Diseños factoriales	25 Estructura diseño de 4 grupos Solomon	17	(4)	(20)	R1 R2	1.68 7.78	26.70	0.000
		26 Gráfico del diseño factorial M * N	11	(25)	(9)	R1 R2	5.84 6.19	1.31	0.191
	cuatro	veintiseis	26	0.7127 ***		R1 R2	9.37 17.64	45.07	0.000

*Ordenados por rango (1) más difícil... (26) más fácil

** n = 563 - 1 = 562 (g1 = n - 1)

*** Coeficiente de reproducción

las n variables extrañas –factores de confusión (Méndez, 1992) que tendrían la capacidad de competir con la manipulación– tuvieran una influencia sostenida para determinar el cambio observado en los valores del segundo registro. Una probable respuesta de Castro (1984) sería: ... *ante la imperfección* –del diseño– *la repetición* –producir múltiples observaciones– *tampoco la suple*.

El cuadro situado al final de la sexta tabla, oculta tanto el comportamiento de cada uno de los 30 talleres impartidos, como el de los 26 reactivos, por lo que en la parte superior de la anotada tabla fue examinada la diferencia del promedio entre ambas mediciones para cada taller y en la séptima tabla, salvo los ítems 12 y 26, el complemento mostró una diferencia significativa a favor del segundo registro, probablemente por efecto de la manipulación; traduciéndose en una correspondencia del 92.3% entre el comportamiento del todo y las respectivas partes. Sin embargo, mediante la oscilación del grado de dificultad entre ambos registros –cuarta columna de la séptima tabla–, es factible, indirectamente, establecer el efecto de la manipulación. Esto es, en el 7.6% de los reactivos el docente no logró mejorar el conocimiento acerca de la secuencia lógica del quehacer científico –ítem 12– ni de la representación gráfica del diseño factorial $M * N$ –reactivo 26–. En el 42.3% de los ítems, inclusive, se elevó el grado de dificultad entre 1 y 16 grados y el porcentaje complementario refleja la efectividad del docente en mejorar únicamente el 50% del conocimiento acerca de los diseños de investigación y ésta varía entre 1 y 21 grados.

La diferencia entre ambos ordenamientos por rango arrojó un coeficiente de asociación de rango de Kendall 0.21 – $p = 0.128$ –, demostrándose empíricamente la afirmación circular implícita en la estructura de este diseño de un solo grupo señalado por Castro (1984) y por tal situación, se procedió a duplicarla, modificándose con ello la hipótesis.

$$H_0: Tvl = Tvc$$

La parte inferior de la octava tabla refuta la hipótesis nula planteada, siendo necesario aceptar la alterna:

$$H_1: Tvl \neq Tvc$$

asegurando a nivel estadístico, que existe una asociación directamente proporcional entre la eficacia del aprendizaje y el costo de operación, ya que el segundo registro de la versión corta mostró un promedio menor que el de la modalidad larga, lo que podría estar relacionado con el decremento paulatino de la atención durante las 11.15 horas diarias por dos días consecutivos.

El resultado previamente expuesto se encuentra *controlado*, debido a que, como se observa en la tabla anterior, independientemente de las modalidades –larga/corta– de la manipulación, el promedio del segundo registro fue significativamente mayor que el del primero y además, no se estableció una diferencia significativa en el primer registro entre las versiones larga y corta; por lo tanto, las muestras son comparables con respecto a un conocimiento homogéneamente bajo antes de las manipulaciones. No obstante, la pregunta acerca de la circularidad (Castro, 1984) debe permanecer nuevamente abierta, ya que únicamente se puede despejar con la siguiente hipótesis nula:

$$H_0: C = Tvl = Tvc,$$

indicando, en este caso, que deben de contrastarse dos muestras experimentales –Tvl y Tvc: versiones larga y corta respectivamente–, con una muestra control –C –; esto es, al triplicarse la estructura del diseño de pre y post-test, una submuestra seleccionada en forma aleatoria no deberá estar sujeta a ninguna manipulación.

La diferencia hallada en el promedio del segundo registro en contra de la versión corta, abre la pregunta si el pretest presentó un mayor efecto sobre los probandos que fueron sometidos a la versión larga del taller, la que será nuevamente hipotetizada.

$$H_0: (Tvl = Tvc)^{[e^{PRE} (Tvl,vc, POST)]}$$

La estructura del diseño que permite operacionalizar esta hipótesis nula está orientada a precisar si los reactivos incluidos en la primera variable dependiente –pretest– pueden fungir como disparadores del aprendizaje, ya sea reactivando los conocimientos olvidados o estimular la reflexión acerca de las estructuras que presentan los diseños. Por ello, se supuso que

Tabla 8: Examen comparativo entre dos versiones -larga y corta- de la manipulación, siguiendo la estructura del diseño de pre y post-test de dos grupos, se excluyeron los probandos sometidos al plan de cuatro grupos de Solomon.

Grupos	Número de casos	Promedio	Desviación estándar	Error estándar	Valor t	Grados de libertad	Prueba de dos colas
--------	-----------------	----------	---------------------	----------------	---------	--------------------	---------------------

Versión larga

Pre-post	193	11.06	4.04	0.29	27.36	192	0.000
Solomon		19.65	3.90	0.28			

Versión corta

Pre-post	298	8.84	4.37	0.25	32.02	297	0.000
Solomon		16.41	3.20	0.18			

Pretest

Pre-post	193	11.06	4.04	0.29	6.44	343.21	0.000
Solomon	298	8.84	3.20	0.18			

Post-test

Larga	193	19.65	3.90	0.28	8.56	441.96	0.000
Corta	298	16.41	4.37	0.25			

el pretest por sí mismo se comporta como una manipulación, modificando la segunda variable endógena; es decir, los resultados obtenidos en el post-test no únicamente pueden atribuirse a las dos formas de manipulación programadas.

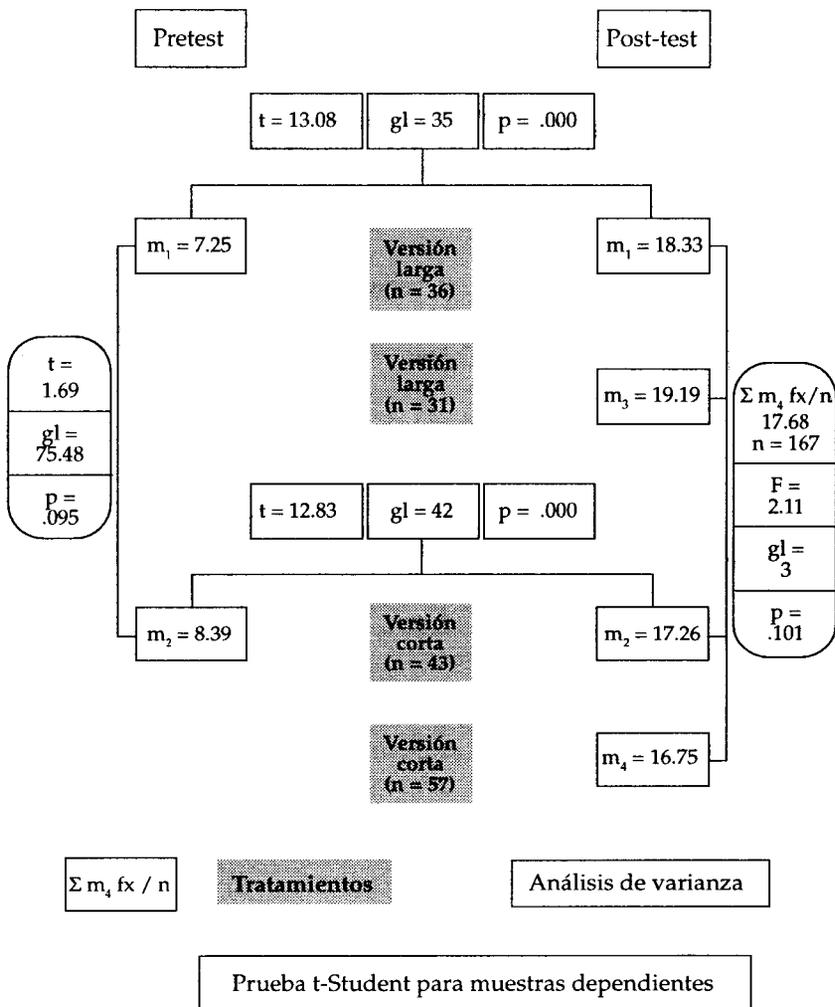
Para controlar este tipo de efectos imputados al pretest, se estructuró un esquema de investigación bifactorial, que en la literatura especializada se le conoce como *plan de cuatro grupos de Solomon* (Bortz, 1984) esquematizado a continuación:

	Versión larga	Versión corta
Con pretest	muestra ₁	muestra ₂
Sin pretest	muestra ₃	muestra ₄

El primer factor está dirigido a diferenciar los distintos modelos de manipulación –versiones larga y corta del taller– o bien, a una condición de manipulación y a otra de control sugerida por Castro (1984) para resolver el problema de la circularidad. El segundo factor está enfocado a distinguir a las muestras que fueron sometidas al efecto del pretest, de las que no lo fueron.

Como se exhibe en la novena tabla, en una primera instancia se examinó con la prueba t-Student para muestras independientes si se diferencian o no los resultados arrojados por los pretests aplicados a las muestras – m_1 y m_2 –, debido a que únicamente es posible examinar las disimilitudes después de las manipulaciones cuando los valores del pretest, en promedio, sean semejantes. Para operacionalizar el esquema propuesto por Bortz (1984), por un lado fue necesario dividir a los probandos en forma aleatoria para someterlos o no el primer registro. Y, por el otro extremo, se aplican los pretests para determinar si ambas submuestras presentan un promedio semejante del conocimiento acerca del tópico anotado o, fijar si a través de la aleatoriedad se obtuvieron dos muestras que se diferencian respecto del saber sobre los diseños utilizados en el quehacer científico empírico.

Tabla 9: Operacionalización del diseño de cuatro grupos de Solomon con respecto a los niveles de comparación



m_1 , m_2 , m_3 , y m_4 = muestra₁, muestra₂, muestra₃, y muestra₄; cfr. esquema de investigación bifactorial propuesto por Bortz (1984).

Al no refutarse esta primera parte de la hipótesis $-H_0$: PRE (Tvl) = PRE (Tvc)–, se aseguró, a nivel estadístico, la comparabilidad de ambas submuestras $-m_1$ y m_2 apuntadas en el esquema anterior y en la parte izquierda de la novena tabla–, permitiendo continuar analizando los valores obtenidos mediante los post-tests siguiendo el esquema de un análisis de varianza bifactorial, encontrando que el efecto del factor pretest –contrastación entre las muestras m_1 y m_2 y las m_3 y m_4 – no fue significativo.

En caso contrario, una interacción asegurada a nivel estadístico entre el factor de manipulación y el de pretest significaría que el primer registro de la manipulación con la versión larga del taller desencadena efectos diferentes que la manipulación con la versión corta. Indicando además, que los probandos medidos y sometidos a la manipulación larga tendrían mejores posibilidades de aprendizaje que los seleccionados en forma aleatoria para la manipulación corta.

Asimismo, a través del diseño de cuatro grupos de Solomon es posible ofrecer claridad acerca de la efectividad absoluta de las dos modalidades de manipulación, pero únicamente cuando se demuestra que no existe una intervención del factor de pretest y una interacción significativas. En caso que el factor de pretest sea significativo, pero no el de la interacción, el factor de manipulación informa acerca de una efectividad relativa de ambas manipulaciones.

En conclusión, al observarse en la parte derecha de la novena tabla, que no se presentó el efecto del pretest sobre ambas modalidades de la manipulación y el post-test, así como una similitud entre los cuatro promedios del segundo registro, no se refutaron la segunda ni la tercera partes de la hipótesis nula. Sin embargo, existen limitaciones con respecto a la validez externa –generalización de los resultados hacia la submuestra evaluada únicamente con el diseño de pre y post-test de uno y dos grupos–, debido probablemente a las disimilitudes apuntadas en la tercera y cuarta tablas, así como en la parte inferior de la octava tabla, donde se observaron diferencias significativas en el 33.2% de los seis descriptores sociodemográficos de las submuestras y un mayor promedio en el pre y post-test de la manipulación larga respectivamente.

Discusión

La imposibilidad de generalizar los resultados ganados con la utilización de la combinación de un diseño factorial 2×2 con uno de pre y post-test, 2 grupos hacia el 73% de los 658 clientes, se asocia fuertemente con el condicionamiento *post hoc* de este experimento, producto de la disociación entre el *deber ser* y *ser* del quehacer científico, ya que la hipótesis de salida indicaba que sería evaluada únicamente la versión original –larga– de la manipulación y que meramente podría sujetarse a una *quasifactorización* si la mitad de la muestra no hubiese sido sometida al primer registro, posibilidad que apareció por casualidad cuando se capitalizó la carencia de suficientes pretests durante la aplicación de la versión larga de la manipulación (Gorenć y cols., 1989). De esta forma y sin un propósito comparativo, se redujo de cinco a dos días la duración del taller, que permitió abandonar el terreno de los diseños no experimentales –de un sólo grupo– y penetrar en el de los experimentales –de grupos múltiples– (Gorenć y cols., 1986), pero con fallas en la aleatoriedad y en el control, que peculiariza el escenario de la investigación extramural. No obstante, a la estrecha exactitud con que se trabajó, se contrapone un mayor acercamiento a la realidad, que es en sí hipotética (Popper, 1982). Concluyendo al igual que E.T.A. Hoffmann –1776-1822– (Bravo-Villasante, 1973), que toda teoría –*deber ser*– es una sombra en comparación con la práctica viva –*ser*–. Con base en la relativa facilidad con que es posible conducir este diseño en el laboratorio *abierto*, no apoya lo expresado por Spector (1982) con respecto a que tiene una limitada aplicabilidad en este campo. Por ejemplo, mediante una tabla de números aleatorios se hubieran seleccionado tanto a las submuestras para someterlas a las versiones larga y corta, como a las submuestras para aplicarles o no el pretest, ejerciendo un mayor control sobre las variables extrañas (Castro, 1984 y Méndez, 1991). Para establecer qué variables deberían mantenerse constantes, únicamente se hubiera requerido efectuar un sencillo ejercicio empleando el diseño de variables concomitantes (Spector, 1982). Con estas maniobras, la condición de extramuralidad se hubiera acercado a la de intramuralidad, sin perder el potencial de tener una mayor pronpiciudad a la amelga fáctica. Asimismo, el análisis con el soporte de los paquetes estadísticos que se ofrece en el mercado es sumamente sencillo, debido a que la reagrupación de las submuestras se logra mediante simples recodificaciones.

La estructura de este diseño, tanto en su forma restringida como en la amplia (Bortz, 1984 y Spector, 1982) podría explotarse en las instituciones de educación superior. Por ejemplo, cuando el docente pretenda medir el conocimiento de los educandos antes y después de impartir la cátedra, pero además, si desea establecer el grado de reacción al instrumento —en los dos sentidos anotados—, la aplicación del diseño de dos grupos de Solomon resulta óptima, aunado a que de acuerdo a las cargas de los reactivos, se le facilita ubicar el nivel de su exposición, así como los aspectos que deben reforzarse (Gorenć y cols., 1989).

El plan de cuatro grupos de Solomon también es posible aplicarlo en planos multifactoriales complejos —ampliación a tres o más modalidades de tratamientos simultáneamente—; siendo decisiva la inclusión de otro factor más, cuyos probandos se dicotomizan también con respecto a la aplicación o no del primer registro.

En 1977, Huck y Chuang ofrecieron un plan, semejante al de los cuatro grupos de Solomon, pero que intenta establecer el efecto del post-test. A través de este diseño se examinan las modificaciones provocadas por el efecto de la manipulación, que están relacionadas con el análisis de lo esperado después de haber finalizado la manipulación. Considerando que no es factible determinar la efectividad de la manipulación sin el post-test, aunado a que no es posible efectuar una comparación entre probandos con y sin post-test, los autores citados recomiendan aplicar a una parte de los probandos un doble post-test, mientras que los complementos —con y sin pretest— únicamente se le somete a una medición después de la manipulación, que puede tener la siguiente estructura:

PRE	Tvl	POST	
	Tvl	POST	
PRE	Tvl	POST ₁	POST ₂
	Tvl	POST ₁	POST ₂
PRE	Tvc	POST	
	Tvc	POST	
PRE	Tvc	POST ₁	POST ₂
	Tvc	POST ₁	POST ₂

y podría ampliarse utilizando el diseño de grupos múltiples que deriva del pre y post-test ≥ 2 grupos, la serie de tiempo interrumpida de grupos múltiples, que además permite establecer el momento en que el efecto de las manipulaciones empieza a extinguirse, como se ejemplifica a continuación:

PRE	Tvl	POST					
	Tvl	POST					
PRE	Tvl	POST ₁	POST ₂				
	Tvl	POST ₁	POST ₂				
PRE	Tvl	POST ₁	POST ₂	POST ₃	POST ₄	...	POST _n
	Tvl	POST ₁	POST ₂	POST ₃	POST ₄	...	POST _n
PRE	Tvc	POST					
	Tvc	POST					
PRE	Tvc	POST ₁	POST ₂				
	Tvc	POST ₁	POST ₂				
PRE	Tvc	POST ₁	POST ₂	POST ₃	POST ₄	...	POST _n
	Tvc	POST ₁	POST ₂	POST ₃	POST ₄	...	POST _n

En caso que se desee controlar el efecto del pretest, pero bajo la óptica de la factorización de Solomon, sería posible adaptarla al diseño de post-test de grupos múltiples:

Tvl	POST	
Tvl	POST ₁	POST ₂
Tvc	POST	
Tvc	POST ₁	POST ₂

Los dos últimos ejemplos indican, que los dos elementos estructurales de los diseños –PRE/POST y Tvl/Tvc– al mostrar una semejanza con los números binarios, las posibilidades de combinación se conducen dentro del sistema abierto propuesto por Bertalanffy (1976) y por ende, al igual que con los números binarios, es posible construir un universo de diseños.

La evaluación operacionalizada a través de la fórmula exhibida por Lewin (expuesta en Wellenreuther, 1982) y con apoyo en el modelo

desarrollado por Miles y Huberman en 1984, se realizó con lo que se conoce en el mercado como los modelos lineales. Sin embargo, la asimilación-acomodación del conocimiento (Piaget y cols., 1977) al parecer no descansa sobre los supuestos que reclama la linealidad, ya que el binomio apuntado sigue un sistema de retroalimentación no-lineal (Meinhardt, 1991) y para esta orientación hipotética, se ofrece una formulación matemática alternativa, que permite calcular la tasa de conocimiento a partir de las diversas fuentes educativas y su influencia recíproca, así como el desarrollo de estrategias de aprendizaje. Con ello es factible realizar una comparación fina entre la hipótesis y el experimento educativo –material de enseñanza–, que consecuentemente facilita el *mejoramiento* de la hipótesis. Esta formulación matemática posibilita incluir un criterio con respecto a la determinación de las estrategias educativas –oral y paquete de software (Fernández-Rivera-Río y Gorené, 1995a y b)– que conducen o no a un conocimiento estable. La siguiente ecuación cumple con la necesidad de asentar las modificaciones que se obtienen a través de la activación o de la inhibición –*a* y *h* respectivamente en la ecuación– del conocimiento:

$$(1a) \quad \frac{sa}{st} = \frac{a^2}{h} - ua + Da \frac{s^2 a}{sx^2} + r_0$$

$$(1b) \quad \frac{sh}{st} = a^2 - uh + Dh \frac{s h}{sx^2} + r_1$$

donde:

a^2 = autoaceleradores y entrecruzamiento de los impulsores del conocimiento

$1/h$ = efecto de la inhibición:

$\left. \begin{matrix} ua \\ vh \end{matrix} \right\}$ = tasa de disociación del conocimiento

$\left. \begin{matrix} D_a \ s^2a/sx^2 \\ D_h \ s^2h/sx^2 \end{matrix} \right\}$ = difusión del conocimiento

$\left. \begin{matrix} r_0 \\ r_1 \end{matrix} \right\}$ = pequeños activadores independientes de la enseñanza

Este sistema ecuacional permite, antes de introducir el nuevo material de enseñanza –planes de estudio; por ejemplo, el Plan Unico impuesto en la Facultad de Medicina (1993)– al mercado, efectuar simulaciones con el comportamiento de la malla del aprendizaje como medida de control de calidad, debido a que con este sistema-activador-inhibidor de impartición de conocimiento en relación con el manejo del método científico, no únicamente puede clarificar la formación de fuentes locales y de gradientes de aprendizaje, sino que también se logra comprender los resultados experimentales paradójicos a través de las peculiaridades de este tipo de sistema.

Referencias

- Abreu, L.F. (1990). El futuro de la educación médica. Una frontera cambiante. *Revista de la Facultad de Medicina*, 33, 279-288.
- Baudrillard, J. (1987). *Cultura y simulacro*. Barcelona: Editorial Kairos.
- Bertalanffy von, L. (1976). *Teoría general de los sistemas*. México, D.F.: Fondo de Cultura Económica.
- Bortz, J. (1984). *Lehrbuch der empirischen Forschung*. Berlin: Springer Verlag.
- Bravo-Villasante, C. (1973). *El alucinante mundo de E.T.A. Hoffmann*. Madrid: Nostromo.
- Carmines G.C. y Zeller, R.A. (1982). *Reliability and validity assessment*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-017. Beverly Hills: Sage Publications.
- Castro, L. (1984). *Diseño experimental sin estadística*. México, D.F.: Editorial Trillas.
- Falcony, C., González, G., Hernández, I., Mendoza, J., Peña, J.L. y Sánchez, F. (1986). Sugerencias sobre los criterios de evaluación del Sistema Nacional de Investigadores (SNI). *Ciencia*, 37, 3-7.
- Fernández Rivera-Río, L. y Gorenc, K-D. (1995a). Evaluación del paquete de enseñanza asistida por computadora: los aminoácidos. *Revista de la Facultad de Medicina*, 38:8-21.
- Fernández Rivera-Río, L. y Gorenc, K-D. (1995b). Medición de la efectividad del aprendizaje mediante un paquete de software. *Revista Expresiones*, en prensa.
- Friedrichs, J. (1981). *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag.

- Gaceta Universidad Nacional Autónoma de México.* (12.10.1995). Número 2,960:1,3-5.
- Gorené, K-D., Herrera, M.E. & Reynaga, J. (1986). *Manual de diseños de investigación: Texto autoevaluativo.* Manual No. 3 del Instituto Nacional de Ciencias Penales. México, D.F.: Instituto Nacional de Ciencias Penales.
- Gorené, K-D., Romero, G., Zurita, M.A. y Cadena, D.A. (1988). Evaluación de seis talleres de diseños de investigación auspiciados por la Procuraduría General de Justicia del estado de Tabasco, México. *Revista Universidad y Ciencia*, 10, 31-54.
- Gorené, K-D., Alday, E. & Gorenc, J. (1989). Medición del efecto de la reacción al instrumento mediante el diseño de dos grupos de Solomon: un ejercicio científico. *Revista Intercontinental de Psicología y Educación*, 1-2, 259-277.
- Gorené, K-D. (1991). La utilización del pre- and post-test para evaluar un taller de diseños de investigación aplicado en América Latina: reporte de los resultados. *Revista de Neuro-Psiquiatría*, 1-2, 29-39.
- Gorené, K-D., Peredo, S. y Abreu, L.F. (1993). El plan de cuatro grupos de Solomon: evaluación transhispanoamericana de un taller de diseños de investigación. *Revista Mexicana de Educación Médica*, 2:111-140.
- Haseloff, O.W. y Hoffmann, H.J. (1965). *Kleines Lehrbuch der Statistik.* Berlin: Walter de Gruyter Verlag.
- Holtzman, W.H. (1980). Projective techniques (1980). En: H.C. Triandis & J.W. Berry (Eds.). *Handbook of cross-cultural psychology.* Vol. 2 Methodology. (pp.245-278). Boston: Allyn & Bacon.
- Huck, S.W. y Chunag, I.C. (1977). A quasi-experimental design for the assessment of posttest sensibilization. *Educational and Psychological Measurement*, 37: 409-416.
- Keeves, J.P. (1990). The improvement of measurement for educational research: Introduction. En J.P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook* (pp. 244-245). Oxford: Pergamon Press.
- Meinhardt, H. (1991). Bildung geordneter Strukturen bei der Entwicklung höherer Organismen. En B-O. Küppers (Ed.). *Ordnung aus dem Chaos* (pp.215-268). München: Piper Verlag.
- Méndez, I. (1992). *Valoración estadística en la investigación.* México, D.F.: Universidad Nacional Autónoma de México.

- Miles, M.B. y Huberman, A.M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills: Sage Publications.
- Nesbitt, J.E. (1966). *Chi-square*. Manchester: Manchester University Press.
- Nie, N.H. y Hull, C.H. (1983). *SPSS-9 Statistikprogrammssystem für die Sozialwissenschaften*. Stuttgart: G. Fischer Verlag.
- Oblitas, L. (1987). *Metodología de la investigación*. Lima. Biblioteca Peruana de Psicología.
- Oblitas, L.A. & Gorenc, K-D. (1990). El ritual científico: evaluación del modelo de iniciación. En L.A. Oblitas & K-D. Gorenc (Eds.). *Psicología mexicana contemporánea* (pp. 203-226). México, D.F.: Publicaciones de la Universidad Intercontinental.
- Piaget, J., Grize, J-B., Szeminska, A. y Bang, V. (1977). *Epistemologie und Psychologie der Funktion*. Stuttgart: Klett-Cotta Verlag.
- Plan Unico de Estudios de la Carrera de Médico Cirujano. (1993). *Revista de la Facultad de Medicina*, 36: 3-40.
- Popper, K-R. (1982). *Logik der Forschung*. Tübingen: J.C.B. (Paul Siebek).
- Rojas, R. (1982). *Guía para realizar investigaciones sociales*. Stuttgart: Fischer Verlag.
- Sánchez-Guerrero, G. (1990). *Un marco teórico para la evaluación*. Cuadernos de planeación y sistemas No. 8. Facultad de Ingeniería. México, D.F.: Universidad Nacional Autónoma de México.
- Siegel, S. (1976). *Nichtparametrische statistische Methoden*. Frankfurt/Main: Fachbuchhandlung für Psychologie Verlagabteilung.
- Sowden, S. y Keeves, J.P. (1990). Analysis of evidence in humanistic studies. En J.P. Keeves (Ed.). *Educational research, methodology, and measurement: An international handbook* (pp. 513-526). Oxford: Pergamon Press.
- Spector, P.E. (1982). *Research designs*. Sage University Paper series on Quantitative Applications in the Social Sciences series 07-023. Beverly Hills: Sage Publications.
- Steeb, W-H. (1991). *A handbook of terms used in chaos and quantum chaos*. Mannheim: Wissenschaftsverlag.
- Uribe, R. (1990). *Reflexiones sobre educación médica*. México, D.F.: Secretaría de Salud.
- Wellenreuther, M. (1982). *Grundkurs: Empirische Forschungsmethoden*. Königstein/Ts.: Athenäum Verlag.
- Wing, J. K. (1972) *Evaluating a community psychiatric service*. London: Oxford University Press.