

Rosa Lacasella

**Metodología
para el estudio
del desarrollo
infantil desde
la perspectiva
conductual**



Fondo Editorial de
Humanidades y Educación
Universidad Central de Venezuela

Índice general

Presentación	13
Introducción	17
El ambiente natural como campo experimental.	
La investigación de laboratorio con niños	21
Sistemas de observación y registro	31
La confiabilidad y validez de las observaciones	55
Diseños de caso único: Generalidades	63
Consideraciones generales de los diseños de caso único	69
Diseños empleados en el estudio de la conducta infantil desde el punto de vista conductual.....	81
Generalidad de los hallazgos	101
Aprendiendo a vivir con la estadística: algunas recomendaciones para la lectura de artículos basados en diseños de grupo	109
Referencias del capítulo	133
Referencias	137

Presentación

Quando al aproximarnos y casi por alcanzar a la cima de una montaña, observamos que ante nuestra mirada comienza a brotar un pequeño arroyo, que, siendo desconocido en su íntimo origen, empieza a cobrar fuerza y a sortear los roquedales para hendir con su curso la tierra y seguir por la pendiente en busca de otros arroyos, hasta quizá confluir con un gran río; podríamos ante tales circunstancias sentirnos exaltados en la contemplación de un hecho simple y natural, pero también preocupados si suponemos que eso que vemos, lejos del mundanal ruido, está acompañado del presagio de una tormenta.

Supongamos que se trata de la segunda de las percepciones señaladas anteriormente, y que nuestra preocupación se debe a que en la cima del valle hay personas que allí viven, y cuyas vidas y pertenencias pueden verse afectadas si la tormenta se transformara en una inundación. Ante tales circunstancias—y no teniendo el tiempo o la oportunidad suficientes para avisarles acerca del peligro—optamos por intentar cambiar el curso del arroyuelo hacia una de las pendientes de la montaña que, desplazándolo por la diversa topografía del valle, lo alejará de las personas en abajo viven ignorando tal peligro; y para ello bastaría simplemente con que desplazáramos

algunas de las rocas que se encuentran ante el inicial curso del arroyo que ha brotado.

Imaginemos también por un momento, que luego que las circunstancias hayan sucedido según nuestra intención y nuestros actos, felizmente las personas del valle cambian de ubicación después de la tormenta y viven en una región menos amenazadora, al otro lado del valle; queriéndose entonces restituir al arroyo a su cauce inicial, porque de esa manera volverían a irrigarse algunos sembradíos que a fortiori tuvieron que ser abandonados... ¿Qué debemos entonces hacer para restituir al arroyo a su cauce inicial? ¿Qué dejamos de hacer y nos preocupa ahora, tiempo después, para conocer exactamente cómo eran las cosas antes de la tormenta?

Como será obvio para el lector, muchas de esas preguntas y dificultades ahora hubieran quedado de lado si hubiésemos registrado la ubicación de las rocas antes de cambiarlas para alterar el curso del arroyo que apenas brotaba, y cualesquiera otras anotaciones que hubiéramos hecho del curso que al agua siguió antes y después de nuestras actuaciones. Pues, bien, algo semejante a esto es lo que se sugiere en la ciencia, en cualquier ciencia, con el nombre de Metodología. El método, sea aplicado a la búsqueda de relaciones entre variables, dentro del complejo proceso denominado la investigación científica; o sea empleado en la actuación cotidiana que como profesionales podemos hacer para asegurarnos que nuestras actuaciones como tales han sido efectivas; cumple, en ambos casos, con la misma finalidad. No es extraño entonces que la profesora Rosa Lacasella haya insistido en ello, y que, además, haya hecho un esfuerzo consciente por ofrecerle a los estudiantes de Psicología, en particular a quienes cursan la Mención Psicología Clínica en la Universidad Central de Venezuela, los materiales que a

continuación ella expone. Por todo eso y por lo que el lector encontrará en sus páginas, «Metodología para el estudio del desarrollo infantil desde la perspectiva conductual» no lo dejará defraudado, porque, además de ser un texto condensado y breve, tiene además la virtud y el encanto de las explicaciones sencillas y la nitidez de los ejemplos.

Henry Casalta C.

Introducción

Este libro está dedicado a los estudiantes de Psicología, en especial a los estudiantes de Psicología Clínica, aquellos que de alguna manera no sólo están relacionados con la clínica sino también con la investigación.

Cuando comencé a dictar la asignatura Introducción a la Metodología del Análisis Experimental de la Conducta en el año 1989, me llamó la atención el hecho de que los estudiantes no entendían el para qué y el por qué de esa asignatura. Parecía que había un salto importante entre la metodología y la puesta en práctica de esa metodología. No entendían (y no entienden aún) por ejemplo, la relevancia de un diseño de investigación no sólo para la indagación científica sino para la práctica cotidiana de su profesión. Evidentemente, había algún problema y no podía detectar dónde se hallaba.

Esto me llevó a varias acciones:

- Revisar el programa de la asignatura y con ello cambiar, objetivos, contenidos, bibliografía básica y complementaria.
- Cambiar y probar diferentes estrategias de enseñanza de la asignatura.
- Verificar diversas formas o técnicas para facilitar la transferencia de aprendizaje de contenidos importantes.

- Escribir los contenidos que creo son relevantes para los estudiantes y que como mínimo deben ser manejados por ellos.

Ésta es la intención de este libro. Proporcionar a los estudiantes herramientas fundamentales para el manejo apropiado de un método de abordaje de su objeto de estudio. Posiblemente está más dirigido a aquellos estudiantes que siguen la perspectiva conductual, sin embargo creo que como método de trabajo podría servir para cualquier estudiante y profesional del área, ya sea que se dedique a la clínica o la investigación.

A través de su contenido se podrá observar que puede ser empleado por diversas personas.

El Capítulo 1 hace referencia al ambiente natural como campo experimental, a los diferentes tipos de estudios que podemos llevar a cabo, sus ventajas y desventajas.

El Capítulo 2 se adentra al problema de la observación como método y como técnica. Posteriormente se pasea por interrogantes importantes tales como: ¿Qué deberá ser observado? ¿Cómo deberá observarse? ¿Cuándo y dónde observar? ¿Quién observará? y por último ¿Qué métodos se emplearán para el registro de los fenómenos que se han definido?. Finalizando este capítulo se comentan los aspectos relacionados con la representación gráfica de las observaciones, tema fundamental para los psicólogos y en el cual cometemos inadvertidamente muchos errores conceptuales.

El Capítulo 3 introduce el problema de la confiabilidad y validez de las observaciones, su definición, factores que atentan contra las mismas y maneras de prevenir algunos de estos factores.

El Capítulo 4 hace referencia a algunos antecedentes históricos que justifican la importancia y las características de los diseños de caso único para la Psicología.

El Capítulo 5 profundiza sobre las consideraciones generales de los diseños de caso único a la luz de cómo pueden garantizar la validez interna y la validez externa.

El Capítulo 6 resume los diseños empleados en el estudio de la conducta infantil desde la perspectiva conductual. Estos diseños obedecen a una metodología intrasujeto y no se restringen al ámbito de la conducta infantil, sólo se ha usado ésta como paradigma de su empleo tanto en la práctica como en la investigación.

El Capítulo 7 habla de la generalidad de los hallazgos, diferentes tipos de generalidad dependiendo del objetivo del investigador y cómo podemos garantizar la misma.

El Capítulo 8 expuesto por la Prof. Cristina Vargas-Irwin nos presenta, de una manera amena, cómo adentrarse a la lectura de un artículo científico, qué aspectos deberán enfatizarse, cómo extraerlos y lo más importante, cómo interpretar datos estadísticos de manera adecuada, más cuando venimos de un enfoque que por años los ha soslayado.

Como se puede observar hemos intentado a través de este texto resumir algunos de los temas que son de interés en el abordaje de nuestro objeto de estudio: la conducta. Esperamos que realmente sea de utilidad para todos aquéllos que de alguna u otra forma nos encontramos inmersos en el quehacer de la Psicología, ya sea como estudiantes, como docentes, como investigadores y como profesionales.

El ambiente natural como campo experimental. La investigación de laboratorio con niños.

Hablar del método que podemos utilizar para abordar el estudio del desarrollo infantil implica algunos aspectos esenciales. Esos aspectos dependerán en gran medida del objeto de estudio y de la definición de ese objeto, lo cual guiará la forma de recolectar los datos y su interpretación. Esto quiere decir que existe una estrecha relación entre la teoría que se supone trata de explicar un fenómeno y la metodología que permite la observación del mismo (Bijou, 1972).

En el caso que nos atañe, el desarrollo infantil, deberíamos partir en primera instancia del concepto de nuestro objeto de estudio. Desde la perspectiva conductual, el Desarrollo es definido como «... los cambios progresivos en las interacciones entre la conducta de los individuos y los eventos en su medio ambiente» (Bijou y Baer, 1978; p.2). En consecuencia, nos interesa analizar la conducta en función de los eventos presentes en la situación, los eventos pasados asociados y la constitución genética (Bijou y Baer, 1978). Por ello, tal como lo resalta Bijou (1972), los datos brutos de esta teorización del desarrollo son las interacciones entre un individuo activo y un medio activo y, por tanto, la metodología que se despliega es de índole empírica, inductiva e individual

(Guevara, 1994), puesto que se ocupa del estudio de conductas observables y enfatiza la búsqueda de relaciones funcionales entre el comportamiento y el ambiente, lo cual, a su vez, indica que el investigador puede manipular variables que le permitan confirmar determinadas relaciones.

Cuando un científico del comportamiento intenta abordar el análisis de una conducta específica lo hace en términos de preguntas experimentales o de observaciones acerca de la misma, y no sobre la base de hipótesis o planteamientos previos de carácter deductivo (Guevara, 1994); de allí que su papel sea activo en el sentido de que trata de manipular variables que se suponen (a partir de la observación) tienen alguna relación con el fenómeno que se pretende estudiar.

Bijou, Peterson y Ault (1968) resaltan que la evolución de la Psicología del Desarrollo al igual que otras ciencias, depende en gran medida de la precisión con la cual se obtengan los datos, ya sea a partir de descripciones de conducta realizadas en ambientes naturales o del análisis funcional de los eventos ambientales con la conducta. Según esos autores (1968), los primeros estudios darían idea acerca del **Cómo**, y los segundos acerca del **Por qué**. Precisamente por ello, la metodología desarrollada a partir de la perspectiva conductual ha permitido integrar los estudios descriptivos con los experimentales. Esto se logra debido a los presupuestos que le subyacen, tales como: a) los datos brutos son las interacciones entre un organismo que se conduce y los eventos ambientales con los cuales se producen esas interacciones; b) los conceptos y principios se derivan de la observación directa de esas interacciones y c) los estudios de campo habilitan la descripción de las relaciones entre la conducta y el ambiente (Bijou, Peterson y Ault, 1968). Por tanto, los estudios descriptivos dan la

posibilidad de describir las interacciones entre la conducta y el ambiente en el cual se dan, mientras que los estudios experimentales permiten establecer las relaciones funcionales entre los eventos estudiados. Por ello, los tipos de datos que se obtienen de ambas investigaciones deben ser continuos, intercambiables y mutuamente interrelacionables (Bijou, 1972).

Posiblemente, sobre la base de una interpretación errónea de este planteamiento en ocasiones se han dividido a las investigaciones en básicas y aplicadas, entendiendo también de manera inadecuada que la investigación básica es aquella que se realiza en los laboratorios altamente controlados y la aplicada, la que se lleva a cabo en ambientes naturales. Según Etzel y LeBlanc (1982), tal dicotomía no parece ser apropiada y proponen que existe un continuum en la investigación científica: el continuum básico-aplicado y que existe una fuerte relación entre ambos extremos del continuo (Ver tabla página siguiente).

Las diferencias entre un tipo de estudio y otro son posibles en función de la relevancia social de la pregunta experimental o de investigación y de la aplicabilidad inmediata de los resultados en el ambiente social (Baer, Wolf y Risley, 1987). También Etzel y LeBlanc (1982) ponen de relieve el hecho de que los diversos tipos de investigación aportan a la construcción de la teoría, puesto que permiten obtener relaciones empíricas sobre los eventos que analizan.

Como se vislumbra de lo anteriormente expuesto, podríamos tener una clasificación de diferentes tipos de estudios para abordar el desarrollo infantil, basados no sólo en los ambientes en los cuales se trabaja (Achenbach, 1978) sino también sustentados en el objetivo de la investigación. Por su parte, Bijou (1972) destaca diferentes tipos de investigaciones, que toman en cuenta los aspectos

CONTINUUM BASICO-APLICADO EN LA INVESTIGACION CONDUCTUAL (*)

	Investigación de Laboratorio con sujetos infrahumanos	Investigación de Laboratorio con sujetos humanos	Investigación Experimental en ambiente natural	Investigación Aplicada en el ambiente natural
Objetivo	Básico	Básico	Básico-Aplicado	Aplicado - Evaluativo
Relevancia social del problema	Por lo general no constituye un factor prioritario.	A veces se toma en consideración.	Usualmente forma parte de los criterios que determinan la utilidad de un estudio.	Los estudios siempre toman en cuenta este criterio
Contribución de los resultados que derivan hacia el bienestar del hombre	No inmediatamente pero sí en última instancia para que tenga valor.	A veces	Usualmente, pero no siempre de inmediato.	Ya están siendo aplicados

Tomado de: Guevara, M.T. (1994). La observación como estrategia básica para el estudio de la conducta infantil. Caracas: Fondo Editorial de Humanidades y Educación, UCV p.59.

antes señalados, tales como los Estudios ecológicos o descriptivos, los Estudios de campo experimentales y los Estudios experimentales de laboratorio.

Los estudios ecológicos son aquellos que permiten la obtención de información acerca de los eventos conductuales, los ambientales, y de su ocurrencia. Posibilitan la observación de variables que podrían tener alguna relación con el evento estudiado y que, posteriormente, serán retomadas en una investigación más controlada de laboratorio con la finalidad de establecer su relación funcional con la conducta.

Para llevar a cabo este tipo de estudio deberían seguirse algunas pautas, tales como: a) especificar la situación bajo la cual se realizará la observación; situación que puede involucrar tanto los aspectos físicos como los sociales del contexto donde se conducirá el estudio (Bijou, Peterson, Harris, Allen y Johnston, 1969), b) definir las conductas que deberán ser observadas y basar la observación sobre una medición de frecuencia de esas instancias de conductas. Por ejemplo, se pueden emplear listas de categorías conductuales ya sean amplias o restringidas en relación con el fenómeno que les atañe directamente (Bijou, 1972), c) registrar las instancias de conductas a través de algún tipo de sistema que permita obtener toda la información relevante. Algunos de estos sistemas podrían involucrar equipos especiales de video grabación o sonido (Bijou, 1972), y d) obtener la confiabilidad sobre los datos obtenidos, con la finalidad de especificar posibles problemas de registro, de definición y en última instancia, de validez de los datos.

Los estudios de campo experimentales son investigaciones que se llevan a cabo en los ambientes naturales y que aportan información acerca de las relaciones funcionales entre la conducta y el ambiente,

por cuanto permiten establecer los determinantes de los repertorios conductuales del niño (Guevara, 1994). Este tipo de estudio combina algunas características del estudio experimental (con menos control) y del estudio ecológico (realizado en el ambiente natural).

En estas investigaciones se aplican los diseños experimentales de caso único, en los cuales: a) se definen los eventos conductuales por ser observados, b) se elige el sistema de registro más adecuado o representativo de la dimensión conductual por ser observada, c) se toman medidas del comportamiento previo a cualquier intervención (línea base), d) se introduce alguna variable experimental con la finalidad de evaluar sus efectos sobre la conducta, e) en ocasiones, se retira la variable con el fin de establecer su verdadera relación funcional con el evento bajo estudio, y f) se reintroduce la variable experimental. Este tipo de diseño se basa sobre: a) medidas repetidas del fenómeno en el tiempo con el propósito de monitorear constantemente los cambios en el comportamiento y poder descubrir fuentes de variabilidad así como a qué se deben dichas fuentes, b) diseños «rápido-cambiantes», en el sentido de buscar inmediatamente la causa de la tendencia conductual alterando el diseño empleado (Barlow y Hersen, 1988).

Estos estudios aun cuando puedan estar plagados de dificultades metodológicas y prácticas, sin embargo representan un avance extraordinario en la investigación conductual y, en especial, en lo referente al desarrollo infantil; pues tal como lo resalta Guevara (1994):

- «a) Permiten el estudio de sujetos difíciles de ubicar en situaciones de laboratorio.
- b) Preservan las relaciones íntimas entre sujetos y contextos, las cuales a menudo son importantes.
- c) Estudian fenómenos multidimensionales y conductas complejas bajo condiciones naturales.

d) Hacen posible validar datos o informaciones obtenidas mediante estudios en contextos restringidos, en situaciones más naturales.» (p.62-63)

Por su parte, los estudios experimentales de laboratorio son aquellos que intentan establecer relaciones funcionales entre la conducta y el ambiente (Bijou, 1972). Estos estudios se caracterizan por un mayor control experimental, en el sentido de mantener constantes algunas variables y/o manipular otras, con la finalidad de resaltar el efecto de la variable independiente que se está analizando y, por tanto, aportan información más precisa sobre la relación que pueda existir entre las variables bajo estudio. Pueden llevarse a cabo en diversos ambientes, ya sean naturales, semiestructurados, como totalmente estructurados. El valor de estas investigaciones se concentra en el hecho de poder explorar a profundidad las condiciones que se suponen controlan determinada conducta (Bijou, 1972).

Un ejemplo de este tipo de investigación es la que se desarrolla en el Laboratorio Infantil adscrito al Instituto de Psicología de la Universidad Central de Venezuela. Su creación obedeció a la necesidad de estudiar procesos relacionados con conductas denominadas complejas tales como la solución de problemas, el autocontrol, la seriación, la creatividad, y otras que por su dificultad exigían un ambiente propicio para su análisis. A su vez este laboratorio permitiría no sólo llevar a cabo investigaciones sino también el desarrollo de la docencia en el ámbito de Pre y Post-Grado y prestar un servicio a la comunidad.

Por las características que implican los estudios experimentales, el laboratorio quedó conformado por una infraestructura que posee un aula experimental para niños en edad preescolar, y un cubículo experimental anexo al aula donde se llevan a cabo investigaciones con sujetos

individuales. Además, está dotado de material didáctico especialmente elaborado para cubrir las necesidades de las investigaciones y otros equipos tales como aparatos de video filmación así como computadoras, todos los cuales facilitan la observación, registro y análisis de comportamientos.

Dentro del marco de este laboratorio se han llevado a cabo estudios de naturaleza tanto básica como aplicada. En los estudios básicos se ha tratado de indagar procesos relacionados con la adquisición de discriminaciones finas (Villalba, 1983), la evaluación de las variables más importantes en la generalización de la seriación (Guevara, 1985), el análisis de errores y el efecto de distintos tipos de correcciones sobre la adquisición de una discriminación (Englert, 1985), el estudio de la creatividad (Lacasella, 1987), la evaluación de procedimientos de adquisición de la clasificación (Rangel, 1990), entre muchos otros.

El denominador común en todas estas investigaciones fue mantener un riguroso sistema de observación y registro de conducta que a su vez garantizara la confiabilidad y validez de los datos obtenidos así como un exhaustivo control experimental, con la finalidad de poder detectar certeramente las variables que afectaban los comportamientos estudiados.

En lo anterior se ha considerado con especial énfasis que un aspecto que surge cuando se hace este tipo de investigación con sujetos humanos y en especial, con niños, es el de tipo ético. En muchas ocasiones, por el simple hecho de hacer ciencia, se ha despreciado la posibilidad de algún tipo de daño que pueda ser infligido a los sujetos, objeto de las investigaciones. Por ello, algunos autores, como Davison y Stuart (1975) y Bijou (1982), entre otros, han resaltado la necesidad de instaurar el *Consentimiento Explícito* y de otras salvaguardas para la participación en

una investigación, las cuales no sólo garantizarían una protección vital al sujeto sino que podrían proporcionar un mayor grado de entendimiento y beneficio para las partes involucradas. Por tanto el científico, que desea realizar investigaciones experimentales deberá ponderar:

- El nivel de beneficio potencial de la investigación para el niño.
- El nivel y tipos de riesgo.
- La validez del procedimiento.
- El grado en el cual el sujeto es capaz de dar su consentimiento a sabiendas; de no ser este el caso, pedirlo a su representante legal.
- La idoneidad del diseño de investigación por utilizar, de manera que no incida negativamente en los sujetos objeto de la investigación, con especial referencia a los diseños de reversión y los que involucren grupo control y grupo experimental.

Evidentemente, todos los aspectos antes señalados dependerán del objetivo del estudio, por lo cual el científico deberá ser cuidadoso de todos los detalles involucrados de forma de respetar los derechos del niño. Por ello, debemos recordar *siempre* que el niño es un ser humano, no importando su sexo, edad, raza, religión, y que como tal debe ser respetado en sus derechos, buscando siempre un equilibrio entre el bienestar del mismo y el beneficio que se pueda obtener de ello para el desarrollo de una Ciencia del Comportamiento.

Según Guevara (1994), la anterior clasificación de estudios que resumimos no estaría completa si no añadimos las investigaciones Aplicado-Evaluativas, las cuales se centran principalmente en la aplicación y evaluación de programas de enseñanza o intervención. El objetivo básico de las mismas es aplicar un conocimiento derivado de estudios de campo o experimentales a una

conducta o situación determinada. Los resultados de estos estudios contribuirán no sólo a la solución de problemas prácticos sino a nutrir el cuerpo de conocimientos de nuestra ciencia.

A manera de ejemplo, también en el Laboratorio Infantil, se llevan a cabo indagaciones de esta índole. Se han elaborado y validado diversos curricula relacionados con diferentes áreas contempladas en la educación pre-escolar: Motricidad Fina (Guevara, Barreat y Rosso, 1984), Discriminación (Guevara, Dembo, Rosso, Badell y García, 1985), Número (Guevara, Rangel, Lacasella, Gómez y Ladrón de Guevara, 1987; Guevara, Rangel, Lacasella, Gómez y Khan, 1988; Lenguaje (Ladrón de Guevara, Rangel, Guevara y Moreno, 1990), Auto-cuidado (Lacasella, Moreno, Rangel y Villalba, 1991). Otras investigaciones tienen que ver con la evaluación de la generalización de los comportamientos adquiridos en el aula experimental a otros ambientales educativos regulares (Guevara, Villalba, Rangel, Lacasella y Girón, 1987).

De todo lo anterior podemos extraer una primera aproximación al método de estudio y es que tomando en cuenta que el objetivo fundamental de los estudios del desarrollo infantil es precisar las relaciones funcionales entre la conducta observable y el ambiente, y siendo la conducta el evento dependiente por excelencia, es evidente que la observación y registro del mismo se convierten en el centro de toda metodología que involucre su estudio.

Sistemas de observación y registro.

El problema de la observación

Si consideramos a la Psicología como una ciencia natural es porque, de hecho, estamos aceptando que su objeto de estudio es un evento natural, observable, registrable y mensurable que está sometido a leyes. Por tanto, también aceptamos el método científico para el abordaje de ese evento natural en la búsqueda de las leyes que lo regulan.

El instrumento científico por excelencia es la Observación y puede tener varios significados, según se le defina y según el empleo que se haga de la misma. Fernández-Ballesteros y Carroble (1983) sintetizan estos significados en dos: la Observación como método y la Observación como técnica. La observación como método tiene entidad propia para erigirse como una forma de obtención de conocimiento, es decir, a través de ella no sólo se puede lograr describir situaciones en las cuales se puede dar un comportamiento, sino que nos permite establecer relaciones funcionales, lo que equivaldría a explicar algún fenómeno. La observación como técnica significa que ésta se convierte en una estrategia subordinada a toda una directriz de investigación, es decir, funciona como un instrumento de recolección de información y/o de datos complementario a otros métodos que se empleen.

La definición de observación depende en última instancia de la concepción del autor que la esgrime y por ende de su marco conceptual. Por ejemplo, para Claude Bernard, observar era ascultar a la naturaleza sin perturbarla, planteamiento también compartido por Darwin (Chávez, 1992) y la diferencia de la experimentación en el sentido de que esta última se refiere a la manipulación de eventos. Según Casalta (1989) es un contrasentido considerar separadamente la Observación y la Experimentación, puesto que manipular sin observar obviamente no tendría finalidad alguna. Tal como lo refiere el mismo autor, la Observación puede significar:

- a) el contacto directo entre el investigador (sus sistemas sensoriales o sistemas artefactuales de registro) y los fenómenos, con o sin la manipulación activa de los últimos
- b) el contacto directo del investigador (sus sistemas sensoriales o registros por artefactos) con los productos o huellas significativas de los fenómenos con o sin manipulación de los últimos» (p.23).

En síntesis, la observación y la experimentación no serían extremos de un continuo (Casalta, 1989).

«Observar es una propiedad genérica de cualquier pesquisa científica, introspectiva o artística. Se refiere a aquella parte de la búsqueda en la cual sólo somos contemporáneos con los datos que van transcurriendo bajo nuestra mirada y frente a los cuales podemos, si tenemos posibilidades, alterar su rumbo. Inclusive, cualquier observación posterior o histórica, en el sentido en que constatamos productos de nuestras observaciones, es decir, registros de observaciones» (Casalta, 1989; p.23-24.)

Por lo tanto, la discusión de si debemos utilizar la observación o la experimentación para el abordaje del estudio del desarrollo infantil carece de fundamento al considerar que la observación es el método que engloba nuestro quehacer científico y que a través de ella, podemos

no sólo estudiar relaciones funcionales sino también precisar, a partir de registros minuciosos, tales relaciones.

Lo anterior ha sido destacado con el objetivo de resaltar y aclarar algunos puntos que siempre han surgido en las discusiones acerca de la observación y la experimentación, como métodos epistémicos o como técnicas. De aquí en adelante haremos mayor énfasis en la observación como técnica, sus características y especificaciones.

La observación en la psicología del desarrollo infantil

Tomando en cuenta que los datos de la psicología del desarrollo son las interacciones que se dan entre un organismo que se conduce y el ambiente en el cual se encuentra, durante períodos extendidos de la vida de ese organismo es evidente que la observación se erige como el método y la estrategia más idónea para abordar el estudio del desarrollo infantil, puesto que además es congruente con algunos supuestos teóricos subyacentes a la evaluación conductual; a saber, que sea conductual y empírica (Baer, Wolf y Risley, 1987; Fernández y Carrobbles, 1983; Nelson y Hayes, 1981). En tal sentido, las primeras investigaciones en este campo se llevaron a cabo a través de la observación directa de los comportamientos de los niños en sus contextos naturales, aunque también es cierto, tal como lo resaltan Bijou, Peterson, Harris, Allen y Johnston (1969), que esas observaciones poco aportaban puesto que no cumplían con dos condiciones básicas: la objetividad y la confiabilidad. Nuestro objeto y método de estudio debe tener ciertas características genéricas tales como: a) que el hecho o evento sea definible en términos objetivos y b) se cuente con los medios o estrategias para observar, medir y registrar tal evento, de forma que podamos

obtener datos confiables y válidos. En consecuencia si deseamos cumplir con los cánones de una investigación científica, en el sentido de alcanzar objetividad y verificabilidad del conocimiento (Arnau, 1979), es obvio que la precisión y rigor con que se lleven a cabo nuestras observaciones deberá ser uno de los puntos centrales de nuestra metodología.

Problemas relacionados con la observación

Arnau (1979) señala tres interrogantes primordiales que concentran la problemática de la observación:

- ¿Qué deberá ser observado?
- ¿Cómo deberá observarse?
- ¿Qué técnicas de registro habrán de emplearse?

Guevara (1989) ha agregado a estas preguntas otras que tienen gran importancia a la hora de llevar a cabo estudios sobre el desarrollo infantil, las cuales son:

- Cuándo deberán realizarse las observaciones,
- Quiénes deberán realizarlas,
- Cuáles procedimientos podrían garantizar la objetividad, precisión, confiabilidad y validez de las observaciones.

¿Qué deberá ser observado?

La pregunta de lo que deberá ser observado podría encontrar respuesta en la definición del objeto de estudio, es decir, la conducta. Su definición permitirá su medición y cuantificación y por ende, verificar las relaciones funcionales entre ella y los eventos ambientales.

Según Johnston y Pennypacker (1980):

«La conducta de un organismo es aquella porción de la interacción de ese organismo con su ambiente, caracterizada por un desplazamiento observable en el espacio, a través del tiempo, de alguna parte del organismo

y que tiene como resultado un cambio mensurable en, cuando menos, un aspecto del ambiente» (p. 48).

Es decir, que la conducta es una interacción, lo cual en palabras de Ribes (1990) es una construcción teórica. Pero es evidente que necesitamos unidades susceptibles de ser sometidas al análisis experimental. Por ello, se acude al término respuesta, que designa «... una sola instancia de cualquier clase de evento bajo consideración» (Johnston y Pennypacker, 1980; p.96).

Por otra parte, Ribes (1980) destaca un aspecto importante: la conducta se da en un continuo temporal, es decir, no existen fragmentos de conducta, es el científico quien segmenta el continuo conductual para su estudio. Por tanto, el analista conductual se concentra en investigar:

«...las relaciones entre propiedades discretas de la conducta y el medio... Los fragmentos o unidades discretas en que se divide el medio son conocidos con el nombre de estímulos; las unidades discretas en que se fragmenta la conducta han sido llamadas respuestas». (p.192).

Como se puede observar, las unidades de análisis deben incluir propiedades del medio y de la conducta que puedan reproducirse reiteradamente en una investigación. En consecuencia, se emplean definiciones de clases de eventos, ya sea del estímulo o de la respuesta, las cuales están conformadas por múltiples instancias particulares (Guevara, 1994).

Una vez que tenemos aparentemente claro cuál es la unidad de análisis que se pretende estudiar, todavía nos queda el problema de cómo medirla. Es decir, la definición de una conducta debe conllevar no sólo un aspecto conceptual sino uno de medida. Según Johnston y Pennypacker (1980) hay dos estrategias para definir conductas: hacer definiciones funcionales o definiciones topográficas.

La definición funcional consiste en: «...delinear una porción del repertorio del organismo mediante la especificación de las relaciones funcionales entre la acción y los eventos ambientales» (p. 106). Tal definición tiene la ventaja de clarificar posibles fuentes de influencia ambiental que podrían ser manipuladas ulteriormente, pero a su vez, presenta la desventaja de su gran complejidad, máxime cuando se observa en ambiente natural, donde es muy difícil cumplir con todas las especificaciones que podría tener una definición de esta índole. La definición topográfica se refiere a: «...la especificación de una clase de respuesta sólo en términos de la forma que adoptan todas las instancias en espacios tridimensionales» (p. 108). Su ventaja se centra en la facilidad de manejo y de uso obligado cuando la topografía de la conducta es la dimensión relevante. La desventaja es que ensombrece la acción de posibles variables que pueden tener relación con la conducta, debido a que diversas clases funcionales se encuentran reunidas en una sola definición topográfica. También Johnston y Pennypacker (1980) sugieren algunos refinamientos que deberían ser tomados en cuenta a la hora de hacer definiciones tales como: a) Agregar a los requerimientos topográficos o funcionales, requisitos de tiempo; b) Eventualmente, podrían emplearse los productos de las respuestas y c) la observación de individuos en grupo.

Jackson, Della-Piana y Sloane (1973) por su parte mencionan tres modelos que pueden emplearse para la elaboración de definiciones: 1) La especificación de características generales; 2) La descripción de sub-conductas específicas y 3) La determinación de resultados de conducta.

Hawkins y Dobes (1977) sugieren que una táctica útil para medir comportamientos es tratar de definirlos de

manera operacional, es decir, descomponer una conducta en sus componentes observables y mensurables. Recomiendan que una definición operacional debe poseer tres características básicas: a) debería ser objetiva en el sentido de que debería referirse a aspectos observables de la conducta; b) debería ser clara, precisa y entendible; c) requerirá de ninguna inferencia o interpretación por parte del observador.

Todos los métodos que se han revisado tienen sus ventajas y desventajas. Lo relevante, por un lado, es aclarar la unidad de medida que deberá utilizarse, no importando la estrategia empleada para llegar a ella; y por otro, que la definición y, en consecuencia, la medición sea objetiva, válida y confiable (Sulzer-Azaroff y Mayer, 1983). Objetiva, en el sentido de que los observadores no infieren o hacen interpretaciones, es decir, sólo se alude a características observables, tanto del ambiente como de la conducta, en términos inequívocos. Válidas, porque miden en efecto, en forma directa, la conducta que se suponen miden. Confiables porque aluden al acuerdo entre observadores, es decir, la confiabilidad se refiere a que las mediciones son estándares con independencia de la persona que las realiza.

Tal vez uno de los aspectos más complejos para cuantificar los hechos conductuales aun cuando han sido definidos, es el establecer la dimensión o propiedad relevante del comportamiento por medir. Dicha dificultad proviene precisamente de la definición que se haga del comportamiento. Dependiendo de cuáles características observables se hayan resaltado en la definición emergerá la dimensión cuantificable o mensurable de la conducta.

Según Fernández-Ballesteros y Carrobbles (1983), las características observables de la conducta más comunes son:

- a) frecuencia, o número de ocasiones de aparición de la conducta

- b) latencia, o tiempo que tarda en aparecer una conducta desde que se produjo la causa que la originó
- c) ritmo, o frecuencia de aparición por unidad de tiempo
- d) intensidad, o esfuerzo en la ejecución de la conducta
- e) duración, o tiempo en que se mantiene un determinado elemento de conducta...» (p.229).

Craighead, Kazdin y Mahoney (1981) resaltan que los datos conductuales pueden ser reducidos a cuatro grandes tipos: temporales a saber: de magnitud, de frecuencia y de categorización. Los datos temporales «...implican la medición del tiempo» (p.56-57). Los datos de magnitud aluden «...a cualquier medida que implique la fuerza o intensidad de la respuesta» (p.56). Los datos de frecuencia «... incluyen respuestas cuyas incidencias pueden contarse y utilizarse para comparaciones objetivas» (p.57). Los datos de categorización «... abarcan las medidas que diferencian una respuesta de otra» (p.57).

Doke (1976) así mismo destaca otras dimensiones importantes como: a) Tiempo entre respuestas: cantidad de tiempo que se separa una respuesta de otra y b) Topografía: especificación de los elementos tanto cinéستicos como corporales involucrados en la ejecución de una respuesta determinada.

Como se puede observar, los autores antes revisados coinciden en muchas de las dimensiones que se toman en cuenta para cuantificar la conducta y en lo que considerarían en cada una de ellas.

Un último aspecto que debemos mencionar es el de la selección de la dimensión relevante. En ella inciden el objetivo de la investigación, factores de orden práctico, la definición de la respuesta y las limitaciones de carácter físico y sociales implicados. Todas estas consideraciones son de gran importancia, puesto que todos los esfuerzos que se hagan están dirigidos a garantizar la sensibilidad, precisión y relevancia de los datos (Guevara, 1994).

¿Cómo observar?

El punto de cómo observar se refiere a las modalidades que se podrían adoptar para llevar a cabo una observación. Tales modalidades pueden incluir: a) registros automáticos, b) registro de productos permanentes, c) observaciones directas con métodos de registro de papel y lápiz y d) observaciones a través de cintas magnetofónicas y de video-grabación (Guevara, 1994).

Guevara (1994) resalta que no hay una modalidad que sea superior a otra, en términos absolutos, sino que cada uno tiene sus ventajas y desventajas dependiendo de las disponibilidad de equipos, costo económico y humano, así como también la definición de conducta que podrá incidir en la facilidad o no del registro, el grado de entrenamiento que podrían necesitar los observadores, en los casos de observaciones directas, y el rango de aplicabilidad. Muchas de las deficiencias de un tipo de registro, en especial en situaciones naturales donde no es conveniente introducir equipos, podrían solventarse a través de otros aparatos menos obvios, o también con un entrenamiento exhaustivo de los observadores. En última instancia, el investigador deberá tomar en cuenta todos estos aspectos a la hora de decidir cuál podría ser la modalidad más adecuada que le garantice la objetividad y precisión de los datos.

¿Cuándo y dónde observar?

Este aspecto alude al periodo de observación y lugar donde se llevará a cabo la misma. Es muy difícil determinar a priori cuándo se deberá hacer un registro y en qué lugar, puesto que también estos puntos están pautados por el objetivo de la investigación y la definición de la conducta. No obstante, será conveniente obtener información a través de observaciones más informales acerca del lugar donde se produce con mayor frecuencia el comportamiento objeto

de investigación y de cuáles periodos podrían ser los medios más provechosos para la observación; puesto que no tendría sentido hacer mediciones a una muestra durante todo el día por ejemplo, cuando la conducta se presenta en las tardes solamente. Obtener una muestra tanto del periodo más apropiado como del lugar más probable incidirá sobre el registro de conducta que se realice. (Guevara, 1994). También sería conveniente, cuando se realicen observaciones en ambientes naturales, introducir sistemas de observación lo menos conspicuos posible y tomar en cuenta los informes de las personas que forman parte del ambiente social (Kazdin, 1979).

¿Quién observará?

Cuando la observación se realiza a través de medios mecánicos, el registro se garantiza por medio de la calibración de los aparatos empleados (Guevara, 1994). Pero cuando la observación debe ser llevada a cabo por sujetos humanos, los problemas de calibración pueden volverse altamente escabrosos.

Tal como lo resaltan Hartmann (1988) y Hartmann y Wood (1990), los observadores humanos pueden mostrar ó tener un conjunto de errores sistemáticos en las observaciones, los cuales deberán considerarse de manera precisa. Algunos de estos efectos son: la reactividad, el sesgo, la tendencia y las trampas. La reactividad se refiere a que los individuos, objeto de la observación, respondan de manera inusual al percatarse que están siendo observados. El sesgo alude a un error sistemático relacionado con las expectativas del observador sobre lo que se supone está observando. La tendencia del observador se produce cuando se pierde precisión y consistencia en la medida; por ejemplo, un cambio imperceptible en la definición de conducta que pueda hacer el observador, el cual será evidente a lo largo de la

investigación; o el poco dominio de las definiciones. Las trampas aluden a cambios intencionados en los registros con la finalidad de obtener altos índices de confiabilidad.

Para reducir en parte, estos efectos deberán tomarse algunas precauciones, las cuales se pueden resumir en la selección y entrenamiento adecuado de los sujetos que servirán como observadores. Según Hartmann (1988), el entrenamiento debería considerar aspectos tales como: dar una orientación general de la observación (observaciones piloto, sensibilización hacia los aspectos de la investigación), aprender el manual de instrucciones realizado al efecto, llevar a cabo prácticas en situaciones estructuradas y en situaciones reales, y, constatemente hacer la revisión y la calibración de las observaciones. Todo esto debería garantizar que se minimicen al máximo las expectativas que el observador tenga sobre la situación; permitirá que se aprenda perfectamente las definiciones de forma tal que no incidan sus apreciaciones y que dependiendo de las previsiones que se hayan tomado para hacer el registro se deberá reducir la posibilidad de realizar trampas en la observación. En última instancia, lo que se desea asegurar es un registro objetivo, válido y confiable.

¿Qué métodos se emplearán?

La decisión acerca de cuál tipo de registro será el más idóneo para una observación determinada deberá considerar diversos aspectos, muchos de los cuales ya han sido mencionados; como por ejemplo: el objetivo de la investigación, la definición del comportamiento que se haya realizado, las propiedades o dimensiones de esa conducta, las modalidades posibles de registro, las limitaciones espaciales y de tiempo, y otras consideraciones de orden práctico y ético. Realmente, la selección de un procedimiento de registro determinará cuáles características de la respuesta son registradas como

una función que a su vez depende de cómo el flujo conductual ha sido segmentado o dividido (Hartmann y Wood, 1990).

En la literatura se encuentran diferentes clasificaciones de los sistemas de observación y registro. En este apartado mencionaremos las que a nuestro criterio son las más completas, y luego pasaremos a realizar una breve descripción de cada uno de los métodos más usados en el Análisis Conductual.

Hall (1974) propone la siguiente clasificación:

- Registros automáticos
- Registros de productos permanentes
- Registros de observaciones:
 - Registros continuos
 - Registros de eventos
 - Registros de duración
 - Registros de latencia
 - Registro de intervalos
 - Muestreo temporal
- Placheck (registro de actividades planificadas o planned activities checklist)

Casalta y Penfold (1981) presentan una clasificación similar, al igual que Sulzer-Azaroff y Mayer (1983) y Ruggles y LeBlanc (1979). Guevara (1994) aporta una división algo diferente basada en los medios que se usan para hacer las observaciones, distinguiendo las siguientes según los instrumentos empleados:

- a) Observaciones directas del comportamiento, empleando técnicas de registro de papel y lápiz...
- b) Observaciones a través de video-tapes o películas
- c) Observaciones a partir de material fotográfico
- d) Observaciones a través de grabaciones de audio
- e) Observación y registro de productos permanentes
- f) Registros automáticos de eventos» (p.141).

Para fines de este trabajo, tomaremos la clasificación de Hall (1974) por ser una de las más completas y repetida en toda la literatura revisada.

Registro Automático: En este tipo de registro, el organismo acciona un aparato, ya sea de índole mecánica o eléctrica, el cual emite una señal que anota automáticamente la respuesta. En este caso, la definición de la respuesta viene dada por el aparato. Tiene como ventajas el que casi no produce efectos reactivos sobre los sujetos observados y proporciona datos muy precisos y válidos. Su desventaja es la limitación tecnológica (Casalta y Penfold, 1981; Johnston y Pennypacker, 1980).

Registro de productos permanentes: Este se utiliza cuando la conducta que se pretende medir deja «... *evidencia física en la forma de un producto duradero*» (Sulzer-Azaroff y Mayer, 1983; p.85). Sus ventajas consisten en que los observadores no deben dedicar tiempo en observar las conductas de interés: los productos pueden ser medidos de forma exacta y precisa, y su implementación es sencilla. Desventajas: Deben ser validados y no tenemos acceso a la conducta que derivó en ese producto, y, por tanto, desconocemos qué variables están relacionadas con el proceso conductual que originó el producto (Guevara, 1994).

Registros Continuos: «*Consisten en narraciones anecdóticas que recogen la secuencia de eventos conductuales de un organismo y las condiciones y efectos ambientales que anteceden y siguen a las conductas*» (Guevara, 1994; p.144). Ventajas: Es un registro muy flexible y permite observar diferentes conductas en un mismo lapso. Además, nos dan la posibilidad de evaluar las condiciones que mantienen un comportamiento determinado (Bijou, Peterson y Ault, 1960). Desventajas: Exigencia de tiempo y dedicación por parte del observador para registrar durante un periodo completo de observación, limitada objetividad y precisión

en los datos, pueden incluir apreciaciones por parte de los observadores, dificultad para resumir la información y dispersión de la observación (Guevara, 1994).

Registro de eventos: Denominado registro de frecuencia, consiste en anotar la ocurrencia de cada instancia de conducta durante un lapso establecido de observación. Comúnmente, se emplea cuando la dimensión relevante es la frecuencia de la conducta. Este procedimiento se utiliza cuando las respuestas que deben registrarse son discretas, en el sentido de que se puede determinar cuándo comienza y cuando termina cada instancia de conducta (Jackson, Della-Piana y Sloane, 1973). Según Kelly (1977), éste es uno de los métodos más usados en Análisis Conductual Aplicado. También, debido a que este tipo de registro rompe la continuidad del flujo conductual, a menudo debe ser complementado con registros narrativos (Hartmann y Wood, 1990). Ventajas: Es un registro sencillo, produce medidas sensibles de conducta, generalmente son confiables, y expresan la cantidad de conducta. Desventajas: Requieren que la respuesta tenga un inicio y final definido y pueden resultar difíciles de llevar a cabo cuando se tienen múltiples conductas o cuando tenemos eventos infrecuentes y de poca prominencia (Hartmann y Wood, 1990).

Registro de Duraciones: Este tipo de registro se define como el tiempo que perdura o permanece una conducta (Guevara, 1994). Se utiliza cuando la característica relevante de la respuesta es la duración. Ventaja: Suministra información acerca de la proporción o porcentaje de tiempo que ocupa la respuesta en relación con el periodo total de la observación. Desventaja: Exige observación continua sobre la conducta objeto, lo cual a su vez dificulta la recolección de información de otros aspectos importantes (Sulzer-Azaroff y Mayer, 1983).

Registro de Latencias: Consiste en medir el tiempo que transcurre entre la finalización de un evento y el inicio de la respuesta (Guevara, 1994). Sus ventajas y desventajas dependerán del objetivo de la investigación, de la definición del comportamiento y los equipos disponibles para realizarlo.

Registro de Intervalos: Este tipo de registro es posiblemente uno de los más populares (Kelly, 1977) y es particularmente útil cuando interesa tanto la duración como la frecuencia de la conducta y/o cuando es difícil discriminar el inicio y finalización de cada instancia de respuesta. Con esta técnica, se divide la sesión de observación, en periodos más pequeños (todos iguales) de tiempo, dentro de los cuales se anotará la ocurrencia o no de la respuesta. Según Powell, Martindale y Kulp (1975) se pueden distinguir varios tipos de Registro de Intervalo: El registro de intervalo total y el parcial. El registro de intervalo total es aquel donde la conducta se registra sólo cuando ha ocurrido durante todo el intervalo de observación. El registro de intervalo parcial, el observador marcará la ocurrencia de la conducta cuando por lo menos haya acaecido en algún lapso del intervalo.

Otra modalidad de registro de intervalo es el denominado Registro Momentáneo de Tiempo o Flash, cuya característica relevante es que la anotación de la ocurrencia de la conducta se realiza al final del intervalo: es decir, no es necesario observar durante todo el intervalo ni exigir que la conducta ocurra durante el mismo, sino que se esté manifestando al final del intervalo seleccionado para observarla. Dependiendo de la modalidad de registro de intervalo elegido se pueden vislumbrar diferentes ventajas y desventajas, las cuales se resumen en: Ventaja: Permite registrar varias respuestas concurrentemente. Desventaja: Exigente para

de la observación se convierte en una dificultad álgida. Sin embargo, en la medida en que la tecnología avanza, en el sentido de permitirnos desarrollar sistemas más automatizados de registro, en esa misma medida podremos realizar análisis más sofisticados y podremos abordar con mayor propiedad los llamados fenómenos complejos (Hartmann y Wood, 1990).

Representación gráfica de las observaciones

Posiblemente, unas de las tareas fundamentales que debe manejar un científico es la elaboración correcta de tablas y gráficos puesto que los mismos le permitirán interpretar de manera apropiada sus datos y comunicar sus resultados.

Según Aragón y Nava (1992):

«... una gráfica es una representación visual de los resultados de un estudio, con la finalidad de comunicar descripciones de los datos que faciliten el análisis de los hechos ... las tres principales funciones de las gráficas son: a) representar medidas sumarias de los datos, b) proporcionar descripciones detalladas de los mismos, y c) operar como análisis completo de los datos» (p. 127)

Los mismos autores aclaran que tras ese concepto no existe la intención de decir que las gráficas sustituyen el tratamiento estadístico de los datos, pero tal como lo resaltan Parsonson y Baer (1978), en los diseños de caso único, la representación gráfica incrementa la capacidad de evaluación de la investigación.

Una vez destacada la importancia de las gráficas pasaremos a detallar algunos tipos de graficación más comunes. No entraremos en los problemas técnicos como por ejemplo, cuánto debe medir el eje de las abscisas en relación con las ordenadas, ni si se hacen con papel milimetrado o utilizando sistemas automatizados, sino

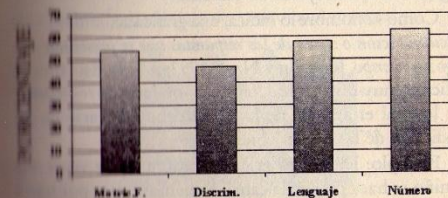
cuáles son los gráficos más empleados y las razones por las cuales se usa uno y no otro tipo de gráfica.

Gráficas de barras

Este tipo de gráfica se usa para representar datos discretos que se encuentran en escalas de medida ordinales o nominales. Para su construcción, comúnmente se consideran los siguientes pasos:

- En el eje de las abscisas se colocan los nombres de las categorías que se van a graficar.
- La amplitud de las barras será igual para cada una de las categorías por ser representadas.
- La distancia entre las barras también deberá ser igual y no mayor de la mitad del tamaño de la barra.
- En el eje de las ordenadas se coloca la escala de magnitud elegida para la medida, sea ésta una de frecuencia o porcentaje.

Ejemplo: Si deseamos representar la ejecución de un niño en diversas áreas de rendimiento, por ejemplo: Motricidad Fina, Discriminación, Lenguaje y Número, podemos elaborar un gráfico de barras, en el cual, el eje de las abscisas estará conformado por las diversas áreas y el eje de las ordenadas, por la escala elegida para medir el rendimiento (porcentaje de respuestas correctas en este caso). El gráfico sería de la siguiente forma:



Gráfica Circular

El principio de esta representación es el mismo de la anterior sólo que en este caso los datos deberán ser llevados a una escala diferente (grados). La única ventaja que tiene sobre la de barras es que permite vaciar mayor cantidad de información sin que se confundan las categorías conductuales (Aragón y Nava, 1992).

Ejemplo: Si deseamos representar la proporción o porcentaje de sujetos que se dedican a una actividad en particular, podemos hacer una gráfica de pastel de la forma siguiente:

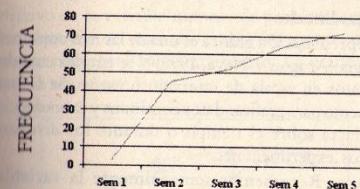


Gráficas acumulativas

Las gráficas acumulativas tienen su origen en las distribuciones acumuladas que en estadística se denominan ojivas. Los analistas conductuales pioneros, tales como Skinner y Ferster dieron mucha importancia a este tipo de registro, puesto que las curvas acumulativas fueron los instrumentos que le permitieron interpretar los datos surgidos en el trabajo de laboratorio.

Como su nombre lo indica, una gráfica acumulada «...es la acumulación o suma de las respuestas que se presentan a lo largo del tiempo» (Aragón y Nava, 1992; p.138). Este tipo de gráfica es muy útil cuando los datos son tasas de respuestas, por lo cual el análisis de la curva debe centrarse en las pendientes de las curvas obtenidas.

Ejemplo: Deseamos graficar el número de veces que un niño se hace pipí en la cama. Podemos optar por hacer una gráfica de este tipo.

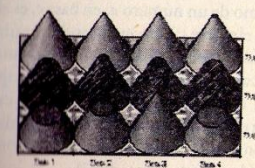


De esta manera a la frecuencia simple del evento, que en la primera semana fue de cinco (5) veces, se le suma la frecuencia de la segunda semana para obtener la frecuencia acumulada de 45 y así sucesivamente.

Gráficas Tridimensionales

Cuando nos encontramos con el caso de que poseemos dos variables independientes y una dependiente, es evidente que un gráfico bidimensional poco aportará al análisis de los datos. En consecuencia, se acude a un eje ortogonal de coordenadas, donde pueden ser graficados más de dos variables. Tal vez por su dificultad no son tan comunes aunque con la irrupción de las computadoras este problema ya ha sido solventado.

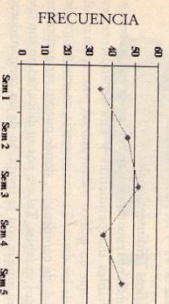
Ejemplo: Cuando se desea considerar en un sólo gráfico, los sujetos que ejecutan una conducta, el número de sesiones y la frecuencia de conducta podemos optar por un gráfico de esta índole, quedando de la siguiente manera:



Gráficas lineales

Esta representación gráfica es una de las más comunes en Psicología (Aragón y Nava, 1992) y se emplea cuando tenemos datos en escala de intervalo o razón. Se le usa principalmente para graficar datos continuos y la tendencia de la conducta sobre el tiempo o durante los diversos tratamientos experimentales.

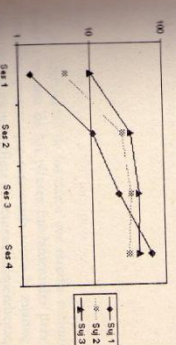
Ejemplo: Representar semanalmente la variable fluidez (entendida como el número de combinaciones diferentes que realiza el niño en un diseño) en un estudio de creatividad. El gráfico podría quedar de la siguiente manera:



Gráficas Semilogarítmicas

En las gráficas semilogarítmicas, uno de los ejes (abstracto) se encuentra en escala aritmética mientras que el otro (ordenado) está en escala logarítmica. Cabe destacar que el logaritmo de un número a , en base c , es el exponente b , al que debe elevarse c para que nos dé el número a . Este tipo de gráfica se emplea usualmente cuando interesa conocer los cambios relativos de esa de respuesta en el tiempo y no cambios absolutos. Además, estas gráficas facilitan la predicción, ya que evitan la pérdida de datos que son la consecuencia de los posibles techos o suelos de las variables (Aragón y Nava, 1992).

Ejemplo: Cuando queremos evaluar el cambio relativo de la variable número de problemas aritméticos resueltos correctamente a través del tiempo de un niño que se encuentra bajo tratamiento. La gráfica obtenida podría ser de este tipo:



En síntesis, la representación gráfica de los datos va a depender de varios factores como por ejemplo los objetivos de la investigación, el tipo y número de datos recolectados, la escala en que se encuentran estos datos, y otros aspectos de orden práctico.

La confiabilidad y validez de las observaciones

La objetividad de los datos que estudia cualquier ciencia es uno de los pilares fundamentales para el científico, puesto que sustentada sobre la misma, el investigador logra verificar de manera independiente, el evento estudiado por otros. Por tanto, es indispensable observar, registrar e interpretar los fenómenos de tal forma que otros observadores obtengan datos similares. Este desideratum es válido, tanto para aquellos métodos en los cuales se utilizan instrumentos de medida ya contruídos y validados como en aquéllos donde se emplean la observación directa de la conducta como estrategia de recolección de datos. A la par de esto, la confiabilidad de un instrumento de medida que indica que cada vez que se aplica el instrumento a lo que se va a medir ofrezca el mismo resultado, es una condición necesaria para la validación del mismo.

Precedente a cualquier discusión sobre la fiabilidad de los datos será conveniente aclarar el concepto que se maneja sobre este punto y su diferencia con otras acepciones del término. Fernández-Ballesteros y Carrobes (1983) afirman que:

...los teóricos de la medida denominan fiabilidad a la consistencia con que mide un instrumento...la seguridad

es el grado de acuerdo o correspondencia entre medidas, que deberían ser iguales si no existiesen errores de medición» (p. 165).

Por su lado, Kerlinger (1988) define confiabilidad como «... la ausencia relativa de errores de medición en un instrumento... es la exactitud o precisión de un instrumento de medición.» (p. 459). Si analizamos estas definiciones apreciamos que Confiabilidad es igual a consistencia y exactitud. Estas concepciones son resultado de las teorías de medición y psicométricas clásicas, las cuales difieren de lo que los científicos de corte conductual conciben como confiabilidad, consideradas por éstos como acuerdo entre observadores, tal como lo expone Anguera (1983):

«En principio, podríamos entender por fiabilidad la similitud o concordancia en el registro de dos observadores que perciben simultáneamente una conducta o evento, o varios, y con la clara convicción de que es un concepto distinto de exactitud o precisión y de estabilidad, aunque tanto la precisión como la exactitud en este caso están formando parte de la compleja medida que es la fiabilidad» (p. 94).

Visto así, en Análisis Conductual estamos aludiendo a una parte de la definición de confiabilidad, es decir, el grado en que dos observadores acuerdan registrar la ocurrencia del segmento conductual, pero hemos soslayado el problema de la exactitud y estabilidad con que registran el evento (Silva y Nava, 1992), y tal como lo resaltan Heyns y Zander (1978) son dos problemas distintos puesto que la confiabilidad del observador es un problema más metodológico mientras que la confiabilidad de la conducta es un problema más sustantivo. Johnston y Pennypacker (1980) destacan que los aspectos de la precisión o exactitud y la estabilidad de la medida deben ser retomados por los analistas conductuales porque aseguran la objetividad de los datos. Pero el problema es

cómo evaluarlos, ya que en la Teoría Clásica de la Confiabilidad (Magnusson, 1977), estos aspectos se juzgan en relación con los valores verdaderos de la variable aunque en nuestro caso, donde desconocemos estos valores, esta táctica llevaría a razonamientos incorrectos.

Una vez realizada esta breve revisión acerca de los problemas de definición en relación con el concepto de Confiabilidad, el cual depende del marco de referencia al que se atiene cada autor, aclararemos que nos detendremos en la definición más consensual en Análisis Conductual, es decir, la de concebirla como acuerdo entre observadores y los diversos métodos para calcular ese acuerdo.

Como ya se mencionó anteriormente, la Confiabilidad se refiere al acuerdo entre observadores y para llevar a cabo esto se comparan los registros simultáneos de dos o más observadores que ven la misma conducta (Bijou, Peterson y Ault, 1968; Hawkins y Dotson, 1975; Romanczyk, Kent, Diamant y O'Leary, 1973).

Los procedimientos para calcular los índices de confiabilidad dependerán de las características, propósitos y métodos de cada investigación (Guevara, 1994). A continuación veremos algunos de los métodos descritos por diversos autores (Bijou, Peterson y Ault, 1968; Casalta y Penfold, 1981; Hawkins y Dotson, 1975; Guevara, 1994).

1. Cuando la dimensión relevante de los datos es la frecuencia, se emplea el índice siguiente:

$$\text{Porcentaje de acuerdos} = \frac{\text{Número menor de observaciones}}{\text{Número mayor de observaciones}} \cdot 100$$

Este índice sólo refleja el grado de acuerdo sobre el número total de ocurrencias pero no si los dos observadores estaban registrando el mismo evento al

mismo tiempo. Para ello, suele dividirse el intervalo de observación en periodos más pequeños y detectar para cada uno de ellos acuerdos y desacuerdos en la observación.

- Cuando tenemos medidas tanto de frecuencia como de duración tal como es el caso de los registros de intervalo, el índice es el siguiente:

$$\text{Porcentaje de acuerdos} = \frac{\text{Acuerdos}}{\text{Acuerdos} + \text{Desacuerdos}} \cdot 100$$

En este tipo de cálculo, se puede precisar un poco más qué conductas fueron observadas y si fueron observadas aproximadamente en el mismo momento.

- Cuando tenemos datos que obedecen a una dimensión como la duración se utiliza comúnmente la siguiente fórmula:

$$\text{Porcentaje de acuerdos} = \frac{\text{Duración acumulada menor}}{\text{Duración acumulada mayor}} \cdot 100$$

Este índice procura la confiabilidad total de la sesión, pero si se han obtenido las duraciones de cada una de las instancias de respuesta observadas puede obtenerse un coeficiente de confiabilidad más riguroso (Kelly, 1977).

- En el caso de que se haya utilizado un registro de actividades planificadas (Placheck), la fiabilidad, según Casalta y Penfold (1981) puede obtenerse a través del siguiente cálculo:

$$\text{Porcentaje de acuerdo} = 1 - \frac{\Sigma \text{ de las diferencias obtenidas de las personas en tarea}}{\text{Número menor de presentes}} \cdot 100$$

Estos son algunos de los métodos más empleados en Análisis Conductual pero hay algunos autores que abogan por el uso de coeficientes de correlación más complejos, cuya elección está asociada con el segmento conductual

de interés, el cual incluye tanto las dimensiones relevantes de la conducta y el nivel de las unidades en el sistema de codificación (Silva y Nava, 1992). Algunos de estos coeficientes son: el Chi Cuadrado, el coeficiente Phi, el coeficiente Kappa, el Coeficiente Pi, Pearson.

Factores que pueden atentar contra la confiabilidad.

El grado de acuerdo entre observadores depende en gran medida de: a) la adecuación del sistema de registro empleado; b) la frecuencia con la que se realicen las observaciones a través del tiempo y c) los métodos para calcular la fiabilidad. Generalmente se asume que el proceso de evaluación de la confiabilidad no es reactivo y que el índice obtenido refleja la consistencia de las observaciones (Guevara, 1994). Sin embargo, existe un conjunto de factores que podrían invalidar la recopilación de datos y su consiguiente confiabilidad, tales como: sesgos del observador, reactividad del observador, complejidad del sistema de registro y expectativas del observador (Hartmann, 1988).

1. Los sesgos del observador: Se refieren a la tendencia de los observadores a cambiar la definición de conducta que se les dió inicialmente, es decir, aplican la definición de manera diferente.
2. La reactividad del observador: Alude a reacciones por parte de los observadores cuando saben que se les está evaluando a través de cálculos de confiabilidad.
3. La complejidad del sistema de observación: La cantidad de categorías conductuales por ser observadas y la variedad de las mismas puede afectar altamente los registros y por ende, la confiabilidad.
4. Las expectativas del observador: Esta fuente de invalidez tiene relación con errores sistemáticos de

evaluación comúnmente asociados a la información que pueda tener el observador sobre el experimento que se lleva a cabo.

Algunas precauciones para combatir estos factores

Todo lo expuesto anteriormente pone de manifiesto la necesidad de tomar en cuenta tales factores y tratar de limitarlos al máximo, con la finalidad de garantizar la confiabilidad de las mediciones y contribuir a la validez de los datos obtenidos. Hartmann (1988), Anguera (1983), Kazdin (1977) y Guevara (1994) se encuentran entre los autores que han realizado un recuento de algunas precauciones que deberían considerarse para combatir los factores que pueden atentar contra la obtención de índices de confiabilidad adecuados. Estas recomendaciones se concentran principalmente en:

- Seleccionar y entrenar exhaustivamente a los observadores.
- Hacer menos conspicuo al observador, utilizando aparatos, espejos unidireccionales u otros que hagan menos visible al observador.
- Intercambiar parejas de observadores.
- No informar cuando se harán las evaluaciones de confiabilidad ni quién será el cotejador.
- Estimar por separado el grado de acuerdo para las categorías conductuales medidas.
- Incorporar diversos observadores o grabar las sesiones con el fin de presentarlas de forma aleatoria.

Hasta los momentos se ha resaltado la importancia de obtener datos confiables, y en consecuencia del proceso involucrado en la confiabilidad del instrumento, pero no podemos dejar de lado a la validez de las observaciones y el proceso que permite establecerla.

El tema de la validez es complejo. Posiblemente es la esencia de cualquier ciencia, puesto que tiene que ver con «... la naturaleza de la 'realidad' y con la naturaleza de las propiedades que se están midiendo» (Kerlinger, 1988; p.489). No es nuestro objetivo hacer una amplia disertación acerca de todo lo que implica el concepto de validez, sólo deseamos tocar algunos puntos que resalten la importancia de tomar en cuenta este aspecto cuando se intenta llevar a cabo una investigación.

En primera instancia trataremos de definirla. Según Fernández-Ballesteros y Carrobbles (1983), la validez de un instrumento implica:

«... adecuación con la que se mide la conducta bajo estudio, capacidad para diagnosticar la conducta real de una persona, sensibilidad para detectar el objetivo, seguridad de las decisiones y utilidad» (p. 176).

Kerlinger (1988) resume la definición en la siguiente pregunta: «¿Se está midiendo lo que se piensa que se está midiendo?» (p.471). Para Silva y Nava (1992), «La validez de un instrumento se refiere a la certeza con que el mismo sirve para la finalidad que su aplicación persigue» (p. 94). Bulser-Azaroff y Mayer así mismo (1983) dicen:

«Las mediciones válidas son aquéllas que miden en efecto de forma directa a la conducta que se supone que miden, en vez de representar esa conducta indirectamente en formas sujetas a distorsión mediante la interpretación individual» (p. 80).

En síntesis, la mayor parte de los autores revisados coinciden en considerar que la validez se refiere al hecho de si los datos proporcionados por el instrumento de medida son realmente los que se esperan y no representan otro aspecto de la realidad estudiada. Para el analista conductual, la determinación de la validez de sus observaciones es un aspecto crucial, puesto que deberá

tener cierto grado de certeza de que los datos recogidos en su indagación expresan correctamente las conductas elegidas y no otra cosa. Según Silva y Nava (1992), la manera como se aborda el problema de la validez y cómo se evalúa es un proceso generalmente empírico que si bien por una parte puede involucrar la adecuación de un instrumento para el propósito elegido, por otra puede incluir a un constructo particular así como conceptos teóricos subyacentes.

En conclusión, visto que el Análisis Conductual se basa en la generación de un conocimiento válido y confiable al igual que cualquier otra ciencia, no está de más resaltar la importancia que tiene para la Psicología Conductual, la confiabilidad y validez de las observaciones y por ende, es capital examinar con precisión, de manera crítica y empírica, ambos aspectos. Para Haynes (en Silva y Nava, 1992), la evaluación de un sistema de registro conductual involucra la evaluación de su utilidad, aplicación, exactitud, comprensión, generalidad y sensibilidad. En otras palabras, esto incluye la evaluación de la validez de contenido, relacionada con el criterio, y la validez de constructo. Es tarea de los analistas conductuales sumergirnos en esta empresa, pues ello le dará mayor relevancia a los datos obtenidos y en consecuencia, a las relaciones funcionales encontradas, con la finalidad de construir un cuerpo teórico coherente que permita la explicación del comportamiento.

Diseños de caso único: Generalidades

Una de las metas de la Ciencia es hacer teoría, entendida ésta como:

«... un conjunto de constructos (conceptos) interrelacionados, definiciones y proposiciones que presentan un punto de vista sistemático de los fenómenos mediante la especificación de relaciones entre variables, con el propósito de explicar y predecir los fenómenos» (Kerlinger, 1988; p. 10).

Llevado este planteamiento a la Psicología, en este caso, una Psicología de enfoque conductual, encontramos que nuestro interés está centrado en la explicación, predicción y control del comportamiento de los organismos individualmente considerados, tal como lo resaltó Skinner (1971). En consecuencia, estamos interesados en establecer las posibles relaciones funcionales entre dos eventos: primero, la conducta, nuestra variable dependiente por excelencia y segundo, el ambiente, nuestra variable independiente. En la medida que conozcamos cómo se relacionan ciertos eventos con otros, podremos manipular variables que nos permitan a su vez controlarlos haciendo que permanezcan constantes, o variándolos y en última instancia, explicarlos y predecirlos. Por tanto, la investigación científica es una forma de indagar acerca

de los fenómenos pero de una manera sistemática, controlada, empírica y crítica (Kerlinger, 1988).

Los procedimientos de investigación utilizados en Psicología se resumen prácticamente en dos grandes rubros: los métodos experimentales y los no experimentales. La observación, la descripción y la medición de los cambios conductuales per se no define a la Psicología. Es una parte del problema. Pero si nuestro interés es el estudio de la relación entre eventos con el mayor control posible de fuentes de error y el control directo de las variables bajo estudio, es imperativo llevar a cabo experimentos, debido a que los mismos demostrarán si los procedimientos empleados fueron los que produjeron los cambios que estamos observando en la conducta o son causados por otras fuentes de variabilidad. En síntesis, una de las formas de establecer relaciones controladas funcionales entre variables es a través de la experimentación.

Y es precisamente a través del diseño experimental que podremos lograr esta meta. Es evidente entonces que tendremos que definir lo que es un diseño de investigación. Para Kerlinger (1988) es: «... el plan, la estructura y la estrategia de investigación para obtener respuestas a preguntas de investigación y controlar la varianza» (p.312). Según Kirk (1972, citado por Castro, 1979), un diseño experimental es un plan de acuerdo con el cual se asignan los sujetos a los diferentes grupos o condiciones experimentales. Craighead, Kazdin y Mahoney (1981) lo definen como la forma en que se disponga la situación con el propósito de evaluar el efecto de la intervención de la variable independiente; es decir, el diseño estructura la situación de manera tal que la causa del cambio conductual puede ser atribuida a la variable que se manipuló. Otros autores como Robinson (citado por Arnau, 1981) consideran que el diseño involucra tanto actividades relacionadas con la

asignación de los sujetos a las diferentes condiciones experimentales o grupos, como otras relacionadas con el planteamiento de la pregunta experimental, la selección de las variables dependientes e independientes, el control de variables extrañas, la manipulación de variables y así como los sistemas de observación y registro, el análisis de datos y las conclusiones.

En resumen, podríamos afirmar que no importando la concepción del autor que esgrime en la definición, hay un gran acuerdo en que un diseño de investigación en particular, un diseño experimental, tiene dos funciones básicas: responder una pregunta de investigación y controlar fuentes de varianza (Kerlinger, 1988).

Será nuestro objetivo en este capítulo aludir a los antecedentes históricos de los diseños de caso único. En los posteriores se referirán sus consideraciones generales en relación específicamente a cómo trata de alcanzar validez interna y validez externa y, por último, se describirán algunos de los diseños más empleados en la investigación de la conducta infantil.

Antecedentes históricos

Aunque la investigación desde la perspectiva conductual con $N=1$ (N es el número de sujetos o unidades experimentales empleadas) ha cobrado gran auge durante las cuatro últimas décadas, ya en los años 1830 y 1860, la fisiología y la psicología respectivamente, iniciaban el estudio de la conducta de organismos individuales (Barlow y Hersen, 1988).

En el dominio de la Psicología, la mayor repercusión del uso de estudios de organismos individuales se concentró en la Psicología Experimental. Los intentos pioneros en ésta estuvieron dirigidos hacia la investigación de la relación entre los cambios en los estímulos y los

cambios en las sensaciones, campo fundamental de la psicofísica. Fechner en sus estudios psicofísicos (1860, citado por Barlow y Hersen, 1988), fue el primero en emplear mediciones repetidas de una respuesta para diferentes intensidades o ubicaciones de un estímulo dado, utilizando un mismo individuo. También se preocupó por la variabilidad (error) observada de ensayo a ensayo. Encontró que los errores eran aleatorios y respondían a una distribución normal, cuyo promedio era cero. Ebbinghaus también empleó el método de las mediciones repetidas en un sólo individuo para el estudio de la memoria utilizando sílabas sin sentido. Pavlov utilizó el estudio intensivo de un sólo sujeto y la réplica con otros individuos (Barlow y Hersen, 1988).

Darwin, Quetelet, Galton y Pearson iniciaron el interés por las diferencias individuales y su medida. Ellos insistieron en que las características psicológicas o rasgos forman parte de alguna propiedad que los hombres poseen en común pero en diferentes grados. Esta diferencia de grado explicaría, valga la redundancia, las diferencias individuales; de aquí el uso extensivo e intensivo de la estadística para abordar este objetivo, y es bastante irónico que el estudio de las diferencias individuales haya llevado al interés por el promedio y la descripción de grupos (Barlow y Hersen, 1988). Este punto de vista llegó a ser tan predominante que marcó casi toda la investigación realizada en Psicología. Pero el uso de esta metodología trajo dos grandes consecuencias: a) el aumento de los estudios de grupos con N muy grandes, puesto que se pensó que era este principio el que podía asegurar la generalidad de los resultados y b) el rechazo y disminución consiguiente de los estudios con pocos sujetos o con $N=1$ (Silva, 1991).

A partir de la segunda mitad del siglo XX se vuelve a incrementar la investigación con $N=1$. Allport, Chassan, Shapiro y otros comienzan a argumentar sobre lo positivo

de la investigación de caso único. De hecho, Skinner en sus primeros trabajos, también hizo énfasis en las medidas repetidas en un mismo organismo bajo condiciones controladas (Barlow y Hersen, 1988). ¿Por qué empiezan a ser cuestionados los estudios de grupo? El problema básico en los estudios de grupo sobre todo en las investigaciones clínicas es que en el grupo experimental hay pacientes que mejoran y otros que empeoran y, por tanto, aparentemente, los resultados promediados no son significativos puesto que se cancelan entre sí al utilizar el promedio como medida de cambio. Estudios en los cuales se aplicaban procedimientos muy generales de psicoterapia a grupos heterogéneos de pacientes definidos ampliamente en una categoría, no ofrecerían respuestas a las preguntas básicas en relación con la efectividad de un tratamiento. Además, se presentaban otras limitaciones (Barlow y Hersen, 1988), tales como las referidas a la falta de tratamiento del grupo control; los problemas éticos, en el sentido de reunir un gran número de clientes homogéneos; el uso del promedio que oscurece los resultados individuales; la limitada generalidad de los resultados, puesto que al trabajar con casos muy homogéneos poco podría inferirse a los casos que no coincidieran con las características de estos sujetos; así como no toman en cuenta la variabilidad intra-sujeto, que es principalmente la responsable por los débiles resultados en los estudios grupales. En consecuencia, todas las dificultades producidas por la comparación de grupos condujeron a dos alternativas: los estudios naturalistas y la investigación de procesos, ambos basados en el método correlacional así bien que el experimental.

Los estudios naturalistas no requieren diseño experimental ni manipulación de variables. Son «experimentos de la naturaleza» (Kiesler, 1971 citado en Barlow y Hersen, 1988). Éstos no son muy diferentes de

la práctica usual, requieren medición de variables consideradas pertinentes (según el enfoque teórico) del terapeuta y del paciente. Además en ellos solamente se interrelacionan las variables. Los problemas asociados a estos estudios están relacionados con la falta de control que de lo contrario permitiría especificar las variables responsables de algún cambio conductual, por ejemplo: si son variables relacionadas con procedimientos terapéuticos, con las características del paciente, con las características del terapeuta o personal auxiliar, o las interacciones entre las características del paciente o las del terapeuta, o cualquier otra (Barlow y Hersen, 1988).

Por otra parte, las investigaciones de proceso se interesan por lo que sucede durante la terapia y no después de la misma. Este tipo de estudio condujo a una infortunada división entre investigaciones de proceso y de resultado; división entre el científico y el profesional, posiblemente debido a que este último sentía que lo que el científico producía no lo ayudaba en su ejercicio profesional (Barlow y Hersen, 1988).

A partir tanto de las críticas a los diseños grupales como a las soluciones intentadas (Estudios naturalistas e investigación de proceso), el estudio de caso intensivo empieza a tomar relevancia, tal como lo resalta Sidman (1975), puesto que éste permitía un mayor control de variables y contribuía a dotar de confiabilidad los datos experimentales. El redescubrimiento del estudio del organismo individual a través de las medidas repetidas en el tiempo, bajo condiciones controladas, el énfasis en las mediciones directas de conducta meta en terapia y la evaluación de las variables independientes en forma experimental, marcan el comienzo de un nuevo énfasis en el estudio experimental del caso individual.

Consideraciones generales de los diseños de caso único.

La concepción del campo de estudio determina en gran medida la metodología que deberá ser empleada para su abordaje. Esto se explica porque existe una gran relación entre la concepción teórica y la metodología de investigación. Debido a que el fenómeno por estudiar en la Teoría del Desarrollo Infantil se define en términos de *los cambios progresivos en las interacciones entre la conducta de los individuos y los eventos en su medio ambiente* (Bijou y Baer, 1978; p.2), nos interesa analizar la conducta en función de los eventos presentes en la situación, los eventos pasados asociados y la constitución genética (Bijou y Baer, 1978). En consecuencia, tal como lo resalta Bijou (1972), los datos brutos de esta teorización son las interacciones entre un individuo activo y un medio activo y por tanto, la metodología que se sigue es de índole empírica, inductiva e individual (Guevara, 1994), puesto que se ocupa del estudio de conductas observables y enfatiza la búsqueda de relaciones funcionales entre el comportamiento y el ambiente, lo cual a su vez, indica que el investigador puede manipular variables que le permitan confirmar determinadas relaciones.

Cuando un científico intenta abordar el análisis de una conducta específica lo hace en términos de preguntas

experimentales o de observaciones acerca el mismo y no sobre la base de hipótesis o planteamientos previos (Guevara, 1994); de allí, que su papel sea activo en el sentido de que trata de manipular variables que se suponen (a partir de la observación) tienen alguna relación con el fenómeno que se pretende estudiar y mide los efectos de esa manipulación.

La metodología del Análisis Conductual se orienta hacia la observación y registro de conductas de un individuo a lo largo de un periodo de tiempo prolongado. Evidentemente, las prevenciones metodológicas se dirigen al desarrollo de procedimientos y métodos que garanticen la observación sistemática y la cuantificación de variables conductuales y ambientales. Para ello, se utiliza un enfoque experimental (Skinner, 1975) que estudia organismos individuales. Los principios se derivan de la observación de relaciones funcionales entre eventos conductuales y ambientales observadas en un sólo organismo, de allí que cada sujeto actúe como su propio control (Castro, 1979). A través de la replica se estudia la generalidad de los hallazgos experimentales.

Características esenciales de los diseños de $N=1$

El rótulo de $N=1$ ha sido utilizado para designar a las investigaciones en las cuales un sólo sujeto o un grupo de sujetos son sometidos a varias condiciones experimentales en puntos sucesivos en el tiempo (Silva, 1992). Kratochwill (1978) asegura que un diseño de series de tiempo es aquel que incluye la investigación de individuos y/o grupos, medidos repetidamente en el tiempo.

En este tipo de diseño se considera:

- La *medición repetida* o periódica de la variable dependiente, en este caso, la conducta. Esto con la finalidad de monitorear constantemente los cambios

en el comportamiento y poder descubrir las posibles fuentes de variabilidad y a qué se deben dichas fuentes (Barlow y Hersen, 1988; Sidman, 1975).

La obtención de un registro de *Línea Base* representativa (Sulzer-Azaroff y Mayer, 1983). Esta proporciona una medida estandar con la cual se pueda comparar los efectos de la variable independiente manipulada constituyéndose así en una forma de ponderar la intervención. Tal como lo resaltan Barlow y Hersen (1988), Sidman (1975), Sulzer-Azaroff y Mayer (1983), una de las características fundamentales de la línea base y que permite su comparación con los efectos de la variable independiente es su Estabilidad. Esto implica que cuanto más estable sea la Línea Base, en el sentido de que su rango de variabilidad es pequeño, en esa medida será posible afirmar que una intervención tuvo algún efecto.

El *sujeto actúa como su propio control*: esta característica se relaciona con la posibilidad de eliminar la variabilidad inter-sujeto típica de los estudios grupales. En los estudios tradicionales, las diferencias individuales se eliminan introduciendo gran cantidad de sujetos con la finalidad de que se cancelen entre sí. Pero esto afecta la posibilidad de generalización de los datos. Utilizando un sólo sujeto es posible ejercer un gran rigor experimental, que a su vez permite el establecimiento de relaciones funcionales, eliminamos la variabilidad inter-sujeto y solamente trabajamos con la variabilidad intraindividual (Castro, 1979).

La posibilidad de aislar las causas del cambio o falta de cambio conductual a través de la *modificación rápida* de la variable bajo estudio, es decir, la búsqueda inmediata de la causa de una tendencia conductual alterando el diseño experimental (Barlow y Hersen, 1988).

Consideraciones metodológicas de los diseños N=1

Cualquier tipo de diseño que se considere sea de grupos o de caso único se enfrenta a dos aspectos fundamentales que le dan validez o que dichos en términos negativos, su no consideración podría invalidarlos, en consecuencia, desestimar los datos obtenidos en una investigación. Tales aspectos como los resalta Campbell y Stanley (1978) son la Validez Interna y la Validez Externa.

Validez Interna

Según Campbell y Stanley (1978), la Validez Interna se refiere a «...la mínima imprescindible, sin la cual es imposible interpretar el modelo: ¿Introducían, en realidad, una diferencia los tratamientos empíricos en este caso concreto?» (p.16). Kratochwill (1978) apunta que: «La validez interna se refiere al grado de certeza de que una manipulación de la variable independiente es la responsable de los cambios observados en la variable dependiente» (p.11). Silva (1992) resume que la validez interna:

«...se refiere al grado en que se logra minimizar el efecto de variables extrañas que puedan influir, conjuntamente con la variable independiente, en los resultados obtenidos; en otras palabras, que los cambios observados en la variable dependiente realmente estén en función de las manipulaciones hechas en la variable independiente» (p.30).

Entonces podemos concluir que la Validez Interna se refiere a la exactitud con la cual podemos afirmar que determinada variable independiente produce efectos sistemáticos sobre la variable dependiente, y que dicho efecto no se encuentra fuertemente afectado por variables extrañas que podrían oscurecer los resultados y dar lugar

a explicaciones alternativas, no debidas a la variable objeto de investigación.

Con respecto a la Validez Interna, Campbell y Stanley (1978) presentan 8 factores que puedan afectar la misma:

- A. *Historia*: considerándola como aquellos hechos que pueden ocurrir entre una medición y otra, no atribuibles a la variable independiente o eventos extraños a la variable independiente pero que ocurren concurrentemente con ésta, pudiendo producir cambios en la variable dependiente.
- B. *Maduración*: procesos inherentes a los sujetos experimentales tales como cambios físicos y/o psicológicos que operan como resultado del paso del tiempo y que pueden afectar la ejecución en la variable dependiente.
- C. *Administración de tests*: influencia que puede ejercer la aplicación de una prueba o medición sobre los resultados de sucesivas administraciones de la misma.
- D. *Instrumentación*: variaciones en los instrumentos de medida o en los observadores que podrían producir cambios en la variable dependiente no atribuibles a la acción de la variable independiente.
- E. *Regresión estadística*: podría presentarse en investigaciones donde se hayan seleccionado los sujetos de manera no aleatoria o con puntuaciones extremas.
- F. *Selección*: sesgos producto de la escogencia no aleatoria de los sujetos experimentales.
- G. *Mortandad experimental*: pérdida de sujetos durante la experimentación y que afectan posibles comparaciones.
- H. *Interacción entre la selección y la maduración*: interacciones que en algunos diseños podrían afectar la acción de la variable independiente.

Kratochwill (1978) coincide en varios de estos factores y considera otros tales como:

- A. *Interferencia de intervenciones múltiples*: en los casos cuando se consideran dos o más tratamientos sobre la misma serie de datos, el resultado de la segunda intervención puede ser debido a ésta como a una combinación de los diversos tratamientos usados. En síntesis se alude a factores de orden, efectos acumulativos, de interacción que podrían desvirtuar los resultados.
- B. *Inestabilidad*: Esta se refiere a la variabilidad de una serie de datos. Cuando esta es muy grande tal vez sea riesgoso anticipar el efecto de la variable independiente.

Según Campbell y Stanley (1978), Kratochwill (1978) y Silva (1992), en los diseños que implican series de tiempo, la historia es el principal problema, puesto que estos experimentos se extienden en el tiempo; de esta manera se incrementa la probabilidad de que un evento extraño a la variable independiente oscurezca su efecto.

Validez Externa

Campbell y Stanley (1978) definen la Validez Externa como «...la posibilidad de generalización: ¿A qué poblaciones, situaciones, variables de tratamiento y variables de medición puede generalizarse este efecto?» (p.16). Achenbach (1978) la refiere como: «... la exactitud con la que se pueden generalizar las conclusiones a otros sujetos y a otras situaciones, cuando se cree que se aplican los mismos determinantes» (p.89). Kratochwill (1978) concibe a la Validez Externa como el grado en el cual los resultados de un experimento pueden ser generalizables a diferentes sujetos, ambientes, experimentadores y algunas pruebas.

Entonces podemos concluir que la Validez Externa hace alusión a lo que Johnston (1979) denomina Generalidad que no es otra cosa que la posibilidad de universalizar los resultados de una investigación, es decir, que puedan ser aplicables a otros sujetos, situaciones o agentes de cambio.

Campbell y Stanley (1978) también aluden a ciertos factores que pueden atentar contra la Validez Externa tales como: el efecto reactivo de las pruebas, la interacción entre los sesgos de selección y la variable experimental, efectos reactivos de los dispositivos experimentales e interferencias de tratamientos múltiples. Pero tal como lo resaltan Castro (1979) y Kratochwill (1978), los factores de los cuales hablan Campbell y Stanley (1978) se refieren mayormente a obstáculos para obtener Validez Interna y no Validez Externa, por lo cual nos acogemos a los aspectos que expone Kratochwill (1978) como posibles amenazantes de la Validez Externa: Consideraciones acerca de la población y de la muestra (Validez Poblacional) y posibles fuentes de invalidez ecológica o del medio ambiente (Validez Ecológica).

Estos tipos de validez apuntan a la consideración de ciertos aspectos tales como: Descripción explícita de la variable independiente, interferencia de intervenciones múltiples, efectos de la información, efectos novedosos y perturbadores, efectos del experimentador, sensibilidad al pretest, sensibilidad al posttest, interacción de los efectos de historia e intervención, generalidad del referente, entre otros. No es nuestro objetivo describir cada una de ellas sino se quiere señalar la importancia de tenerlas en cuenta cuando se hace investigación y se planifica la misma, con el fin de poder asegurar el máximo control posible y que en consecuencia, los resultados puedan ser universales o sujetos a la posibilidad de generalización.

Criterios para garantizar la validez interna y la validez externa en los diseños de caso único.

Los criterios que nos sirven para garantizar la validez interna y externa de un experimento dentro del Análisis Conductual no son tan disímiles a los de los diseños tradicionales. Según Birnbrauer, Peterson y Solnick (1974), los criterios son:

- Medición confiable: Los datos deben ser registrados de manera objetiva y confiable.
- Medición repetida: Los datos deben registrarse periódicamente en el tiempo.
- Descripción de procedimientos y contextos: Los procedimientos de intervención así como el contexto en el que se lleva cabo la investigación deben ser descritos exhaustivamente, de manera que puedan ser replicados.
- Manipulación sistemática de la variable independiente: Esto con la finalidad de asegurar que los cambios observados en la conducta son producto de la manipulación experimental y no de otras variables extrañas.
- Posibilidad de réplica tanto directa como sistemática (Sidman, 1975), con la finalidad de establecer la generalidad de los datos.

Otras consideraciones metodológicas

En las investigaciones de caso único también hay otras dificultades que deberán ser tomadas en cuenta, entre las cuales se destacan: La alteración de fases o condiciones durante la experimentación, la comparación de intervenciones alternadas y la confiabilidad entre observadores (Silva, 1992).

La alteración de fases o condiciones durante la experimentación: Según Silva (1992) no existen reglas

claras en relación con la alternación de fases. Una que generalmente se emplea es la de cambiar de fase cuando se haya observado estabilidad en los datos. La estabilidad (problema altamente escabroso) se refiere principalmente a la ausencia de tendencia y a una variabilidad relativamente pequeña en una ejecución dada. Por consiguiente, uno debería obtener por ejemplo, líneas de base estables antes de introducir cualquier variable independiente. Sin embargo, podemos encontrar ocasiones en las cuales esto puede soslayarse como cuando tenemos una línea base con tendencia opuesta al cambio que deseamos introducir (Kazdin, 1980). Pero cuando es el caso opuesto, es preferible esperar a obtener cierta estabilidad en los datos, ya que lo contrario no nos permitiría afirmar que los cambios observados en la variable sean debido a la variable independiente sino a cualquier otra cosa. Además, la excesiva variabilidad de los datos interferirá en las conclusiones que se obtengan del experimento (Barlow y Hersen, 1988). Como lo expone Silva (1992), el análisis de la variabilidad es básico en Análisis Conductual, puesto que nos brinda información acerca de las posibles variables que controlan la conducta.

Otra dificultad en este tipo de metodología es la relacionada con la duración de cada fase. Este punto no se puede pasar por alto puesto que el investigador necesita datos a partir de los cuales hacer predicciones. Tampoco en este aspecto hay reglas. Algunos autores como Barlow y Hersen (1988) establecen que tres puntos de datos es el mínimo admisible para observar tendencias y determinar patrones de ejecución. Johnston (1972) señala que es necesario que cada fase sea lo suficientemente larga para demostrar estabilidad y disipar cualquier duda del investigador de que los datos de la investigación son creíbles y representativos de lo sucedido en el proceso experimental. Kazdin (1980) recomienda que se puede

cambiar de fase cuando la variabilidad se acerca al nivel medio de ejecución y cae dentro de un rango específico durante cinco días o también cuando la tendencia observada en la ejecución va en dirección contraria a la que se espera en el tratamiento.

La comparación de intervenciones alternadas: En este caso, el investigador está interesado en demostrar la efectividad relativa de diferentes variables independientes sobre el comportamiento. La mayoría de las veces es difícil determinar los efectos de la intervención porque éstos se pueden confundir por la secuencia o el orden de presentación de las variables implicadas. Estos efectos, según Silva (1992) se pueden atenuar de dos maneras. a) introduciendo mediciones de línea base entre las intervenciones y b) variar de manera sustancial el diseño, por ejemplo, empleando uno de línea base múltiple, multielemento, etc.

La confiabilidad entre observadores: Debido a que en la mayor parte de las investigaciones sobre el desarrollo infantil se utilizan observadores es muy probable que sucedan problemas en el registro de las respuestas que puedan poner en duda la veracidad de los procedimientos de evaluación y la consistencia con la que se realizan las observaciones. Aquí la recomendación fundamental va dirigida a varios aspectos: asegurar un buen entrenamiento de observadores, calcular la confiabilidad en puntos diferentes del experimento y con diferentes cotejadores (Silva, 1992).

La selección de un diseño de investigación

Antes de finalizar el apartado sobre consideraciones generales de los diseños de caso único es conveniente aludir a los factores que deben tomarse en cuenta para elegir un diseño. La selección de un diseño obedece a diversas consideraciones entre las que podemos destacar:

- El propósito u objetivo de investigación
- El tipo de conducta que se investiga
- Las posibilidades de manipulación de la variable independiente
- El tiempo y recursos de que se disponga
- El ambiente donde se realiza el estudio
- El grado de control experimental que se desea obtener
- Las garantías de validez interna y externa del estudio
- Y demás consideraciones de orden práctico y ético

Es evidente que no existe un diseño ideal, que su selección se basa en los aspectos antes señalados. En última instancia, el mejor diseño es aquel que permite un máximo de control de variables, un mínimo de interferencia de variables extrañas y en consecuencia, el establecimiento de relaciones entre las variables independientes y dependientes estudiadas.

En el diseño AB, el sujeto es observado en su conducta antes de la introducción de cualquier tratamiento experimental. Este periodo de observación se denomina línea base o nivel operante del comportamiento. Después de la observación de la línea base, se introduce el tratamiento experimental y se observa la conducta del sujeto durante un periodo de tiempo determinado.

La principal ventaja del diseño AB es que permite observar la conducta del sujeto en su estado natural, sin la influencia de ningún tratamiento experimental. Sin embargo, la principal limitación de este diseño es que no permite controlar las variables extrañas que pueden influir en la conducta del sujeto. Por lo tanto, el diseño AB es más adecuado para estudios de exploración que para estudios de confirmación.

En el diseño AB, el sujeto es observado en su conducta antes de la introducción de cualquier tratamiento experimental. Este periodo de observación se denomina línea base o nivel operante del comportamiento. Después de la observación de la línea base, se introduce el tratamiento experimental y se observa la conducta del sujeto durante un periodo de tiempo determinado.

El diseño AB es el más sencillo de los diseños experimentales.

El diseño AB es el más sencillo de los diseños experimentales. Este diseño se caracteriza por la observación repetida bien controlada del tiempo. Se toman datos de la variable dependiente durante un periodo de tiempo en el cual no se introduce ningún tipo de tratamiento o variable. Esto se constituye en la línea base o nivel operante del comportamiento.

Diseños empleados en el estudio de la conducta infantil desde el punto de vista conductual

Existen muchas clasificaciones de los diseños de caso único. Birnbrauer, Peterson y Solnick (1974) asoman dos categorías de clasificación: los diseños de series temporales y los diseños múltiples. Hay otros como Barlow y Hersen (1980) que lo hacen sobre la base del número de variables contenidas en el diseño. No es objetivo de este trabajo tratar de hacer una clasificación de los diseños de caso único, ni apuntar las posibles consideraciones al respecto. Sólo trataremos de describir algunos de los diseños más empleados en el Análisis Conductual en relación con el estudio de la conducta infantil.

Diseño básico AB

El diseño AB representa el fundamento de los diseños experimentales (Kratonchwill, 1978). Este diseño se caracteriza por la observación repetida bien controlada del tiempo. Se toman datos de la variable dependiente durante un periodo de tiempo en el cual no se introduce ningún tipo de tratamiento o variable. Esto se constituye en la línea base o nivel operante del comportamiento. (Thompson y Grabowski, 1972) diferenciaron entre la línea base y el nivel operante, dependiendo de si la respuesta ya se encuentra en el repertorio del sujeto o si

es condicionada experimentalmente]. Luego de haber obtenido estabilidad en esta medición se introduce la variable independiente y se sigue registrando durante este lapso (B). Este diseño puede ser conceptuado como una forma mejorada del diseño de pre-test/posttest (Kratochwill, 1978). Según Risley y Wolf (1972), el diseño AB permite apreciar un cambio en el nivel de la conducta y la magnitud aproximada de ese cambio.

Sin embargo, con algunas consideraciones la afirmación anterior es cierta. Barlow y Hersen (1988) apuntan algunas limitaciones importantes de este diseño que minan su validez interna, por ejemplo: es difícil afirmar que el cambio observado en la VD sea debido única y exclusivamente a la VI manipulada, puesto que pueden haber ocurrido una serie de eventos independientemente del tratamiento que tienen o mayor preponderancia o interactúan con la VI. Por otro lado, en este diseño como en todos los demás, la exigencia de la estabilidad de la Línea Base es fundamental (Birnbauer, Peterson, Solnick, 1974). La aparición de tendencias en los datos antes de introducir una VI puede acarrear problemas. Hay varias tendencias que amenazan las conclusiones que se puedan obtener de unos datos:

- Tendencia en los datos en dirección contraria a la esperada en la intervención.
- Tendencia en los datos en la misma dirección a la esperada en la intervención.
- Gran variabilidad en los datos de la Línea Base que no permite apreciar cambios.

De todas ellas, tal como lo apunta Kazdin (1989) podemos trabajar con la primera, puesto que permite apreciar cambios aún no siendo estable. Pero con las dos últimas es bastante problemático, ya que esto lo que

demuestra es que no hemos podido aislar posibles fuentes de varianza que están controlando el comportamiento (Kerlinger, 1988; Sidman, 1975) y por tanto, cualquier modificación no podría ser atribuible a la variable manipulada. En síntesis, podemos resumir estas debilidades a factores de historia y maduración (Campbell y Stanley, 1978; Kratochwill, 1978).

En resumen, el diseño AB es un diseño muy limitado con respecto a la obtención de conclusiones de relaciones funcionales entre eventos. Su aplicación se restringe a las situaciones en las cuales no es posible otra opción, por ejemplo situaciones clínicas, donde las consideraciones prácticas o éticas exceden en mucho la deseabilidad de control experimental. Pero también cabe destacar, que con el propósito de superar algunas de las dificultades propias de este diseño y por su amplia aplicación en la clínica se han introducido una serie de controles que los hacen más utilizables (Barlow y Hersen, 1988). Se han introducido medidas de seguimiento y medidas concurrentes de conductas, lo cual permite un mayor control de algunas fuentes de variación.

Diseños A-B-A

Con la finalidad de superar las deficiencias de control del diseño AB se agrega una fase de línea base (A). Es decir, se realiza una medición de la conducta bajo condiciones naturales (A), luego se introduce la VI (B) y posteriormente se retira el tratamiento (A). Estos diseños conjuntamente con todas las combinaciones son denominados en forma genérica, diseños de Inversión (Barlow y Hersen, 1988). Es conveniente en este punto hacer una aclaratoria que puede tenerse presente. Cuando se habla de este tipo de diseño se alude, en muchas ocasiones, tanto al procedimiento para obtener el efecto

como al efecto mismo. Y esto es un contrasentido. Será relevante mantener presente, que cuando uno habla de inversión o reversión, se refiere al procedimiento para obtener el efecto (Barlow y Hersen, 1988). Este diseño evidentemente permite afirmar con un mayor grado de certeza que los cambios observados en la VD son producto de la VI y si se realizan diversas repeticiones en distintos sujetos con resultados similares refuerza el poder explicativo de la variable (Barlow y Hersen, 1988).

A pesar de las ventajas, desde el punto de vista experimental que tiene este diseño, sin embargo ofrece algunas limitaciones tales como: a) Cuando se utiliza en contextos clínicos, el paciente terminará el tratamiento en la fase A, negándosele los beneficios del programa de intervención. b) Tal vez la limitación más importante se refiera a la imposibilidad de aislar con este diseño fuentes de interferencias de tratamiento múltiple (Kratowill, 1978; Barlow y Hersen, 1988). Es decir, no permite una comparación inequívoca de las diferentes intervenciones debido a posibles efectos de secuencia o de orden (Kazdin y Hartmann, 1978).

Aún no siendo uno de los más empleados, quisiéramos presentar un ejemplo de la aplicación de este diseño. Walker y Buckley (1968) utilizaron un diseño de la forma A-B-A para analizar los efectos de un programa educacional individualizado en un niño de 9 años que presentaba una alta tasa de conductas inatentas, lo cual interfería con su rendimiento académico. Durante la Línea Base (A), se registró el porcentaje de conducta de atención que presentaba el pequeño en condición natural. Posteriormente, se introdujo un programa de puntos con la finalidad de reforzar la conducta de atención (B). La fase (A) implicó el retorno a las condiciones originales, luego de varias sesiones de entrenamiento. Los resultados apuntan a

la posibilidad de controlar dicho comportamiento a través del programa de puntos empleado.

Diseño A-B-A-B

Este, al igual que otras combinaciones posibles es una extensión del diseño A-B-A. Observemos que aquí se supera la dificultad de terminar la intervención en Fase A y también permite evaluar en dos ocasiones los efectos del tratamiento. Usualmente se le denomina diseño de replicación (Kratowill, 1978).

Según Kratowill (1978) presenta por lo menos tres variantes:

- 1 Diseños de reversión: La segunda fase de Línea se puede obtener por dos procedimientos: a) aplicando el tratamiento a una conducta alterna pero incompatible o b) se altera la contingencia aplicándola sobre cualquier conducta menos la que estaba siendo reforzada (RDO).
- 2 Diseños de retirada: La segunda fase de Línea base consiste en la discontinuación de la variable de tratamiento o simple retiro de la misma.
- 3 Aplicación no contingente: En la segunda medición de línea base, las contingencias siguen en efecto pero son aplicadas de manera no contingente sobre la ejecución.

Este diseño conserva algunas de las deficiencias del diseño A-B-A, en el sentido de no poder controlar los efectos de intervención múltiple, de difícil aplicación en conductas irreversibles (Kratowill, 1978; Orth-Lopes, s/f) y de considerables problemas prácticos y éticos.

Algunos ejemplos del uso de este diseño son los siguientes: Hall, Fox, Willard, Goldsmith, Emerson, Hagen, Davis y Porcia (1971) llevaron a cabo un estudio con un niño retrasado de 10 años cuya conducta de hablar

fuera de tiempo durante la clase se había convertido en un grave problema. Durante la Línea Base, el profesor registró la conducta problema. Posteriormente, en la fase B, se le dieron instrucciones al profesor para que ignorara la conducta problema y prestara atención a la conducta productiva del niño. Las sesiones de Línea base 2, consistieron en un regreso a las condiciones de la primera medición. Por último, se reinstala el procedimiento experimental empleado. Los resultados fueron altamente exitosos y se demostró en dos ocasiones que la variable manipulada fue la responsable del cambio.

Así mismo, otro ejemplo, Handrickson, Strain, Tremblay y Shores (1982), utilizaron el mismo diseño pero con la finalidad de estudiar los efectos de la atención desplegada por compañeros para incrementar conductas sociales en un niño de cuatro años que presentaba aislamiento social. También en este caso los resultados fueron exitosos.

Extensiones del Diseño A-B-A-B

Barlow y Hersen (1988) manifiestan que la literatura conductual aplicada está plagada de diversos ejemplos en los cuales se han utilizado extensiones y variantes del Diseño básico A-B-A. Según estos autores, estas extensiones y/o variantes pueden ser divididos en cinco grandes categorías.

La primera categoría alude a la replicación en varias ocasiones del modelo básico AB, es decir, el diseño se presenta en la forma A-B-A-B-A-B. Esta planificación de la administración de la variable independiente presenta ciertas ventajas, tales como: a) comprobación del efecto de la VI, b) prolongación del estudio hasta finalizar el tratamiento clínico.

La segunda categoría se refiere a la evaluación de diversas técnicas o VI sobre un comportamiento determinado. Sigue la forma A-B-A-C-A con algunos refinamientos dependiendo del objetivo del estudio y del control que se puede introducir. En este tipo de diseño es difícil comparar la efectividad diferencial de los tratamientos involucrados, pero pueden determinarse los efectos controladores.

La tercera categoría estudia variaciones concretas de un mismo procedimiento sobre la variable dependiente. Toma la forma de A-B-A-B'-B''-B'''. Estas investigaciones se han denominado comúnmente paramétricas.

La cuarta categoría examina los efectos interactivos de dos o más variables sobre una conducta determinada. La forma más común es A-B-A-BC-B-BC. En estos diseños se analiza tanto los efectos de cada variable por separado como en combinación.

La quinta categoría alude al Diseño de Criterio Cambiante, del cual hablaremos específicamente en el apartado siguiente.

Diseño de Criterio Cambiante

Este diseño puede constituir una estrategia útil para ponderar el moldeamiento de programas con el fin de acelerar o disminuir conductas. Hartmann y Hall (1976) proponen el uso de este diseño para evaluar los efectos de procedimientos de moldeamiento; es decir, cuando los niveles de conducta inicial son tan bajos que no permiten al experimentador regresar a la línea base y además se desea incrementar o disminuir paulativamente a los comportamientos. La estrategia que se sigue es la siguiente: inicialmente, se obtiene una línea base para determinar el nivel pre-tratamiento de conducta. Considerando el

promedio de esta fase, se establece un criterio de consecución de los efectos esperados de las contingencias. Posteriormente, en la condición de intervención, las contingencias se administrarán con base en el criterio preestablecido. Una vez alcanzado éste, se cambia el criterio haciéndose cada vez más exigente hasta alcanzar el nivel terminal de comportamiento. Cada una de las fases del diseño se convierte en la línea base de la próxima. En síntesis, el Diseño de Criterio Cambiante se concibe como una serie de diseños AB, en los cuales el efecto es replicado consecutivamente con cambios en el criterio (Hall y Fox, 1977).

El control se demuestra si el cambio conductual se corresponde cercanamente al nivel del criterio establecido y si se obtiene una ejecución estable. Tal como lo expresa Kazdin (1980), este diseño presupone «...una estrecha correspondencia entre el criterio y la conducta a lo largo del curso de la fase de intervención» (p.160). Por su parte, Hall y Fox (1977) han resaltado un aspecto importante de este diseño y que afecta su validez interna. En el caso de que no se logre una estrecha correspondencia entre la conducta y el criterio establecido es muy difícil decir que se ha alcanzado el control y, en consecuencia, las interpretaciones de los datos pueden ser equívocas. Pero esto podría solventarse en parte con el empleo de una reversión en el sentido de regresar a un criterio más bajo ya logrado y volviendo después al criterio más estricto.

Otros aspectos por considerar en este diseño son: la longitud de la línea base y de las diversas fases de tratamiento, la magnitud de los cambios en el criterio y el número de fases o cambios en el criterio (Hartmann y Hall, 1976).

En cuanto a la longitud de la línea base y de las diversas fases se recomienda que las fases sean lo suficientemente

largas para asegurar que los cambios conductuales no se deban a efectos de la historia, maduración o factores de medida. También se aconseja que las fases tengan diferente longitud, y, si son todas iguales, incorporar una Línea Base más larga que las fases experimentales para garantizar que los cambios no ocurren por los cambios en el criterio. También se exige estabilidad de la Línea Base o a lo sumo que su tendencia sea opuesta a la del cambio que se pretende, y que la ejecución en cada fase sea suficientemente estable con el fin de poder pasar a la fase siguiente.

En relación con la magnitud de los cambios de criterio, se estima que el tamaño de los cambios en el criterio deben ser lo suficientemente grandes como para que sean detectables.

En cuanto al número de fases o cambios en el criterio, ello depende del número de replicaciones que se consideren necesarias para demostrar el control experimental.

Ejemplos de la aplicación de este diseño los reportan Hall y Fox (1977). En un caso se trató de disminuir las conductas disruptivas presentadas por dos niños en edad escolar. Estas conductas interferían fuertemente el aprendizaje de los niños, en especial en el área de matemáticas. Los autores tomaron registros de línea base del número de problemas de aritmética (Sustracción y División), posteriormente elaboraron los criterios sobre los cuales administrarían los reforzadores (Receso y jugar basketball), partiendo de un criterio inicial sustentado en el promedio de problemas resueltos correctamente en la Línea Base. Los resultados fueron exitosos y se demostró de manera altamente satisfactoria el control experimental puesto que los cambios conductuales se acercaban fuertemente a los criterios establecidos. En el otro caso, los autores trabajaron con un niño de 10 años de edad,

diagnosticado como retardado educable y con espina bífida. El problema fundamental era que este niño no podía completar la cantidad estimada de trabajo, a menos que el maestro trabajase con él. Por ello se obtuvo la Línea base del número de asignaciones realizadas de manera completa. Posteriormente, se introdujo la variable independiente, que en este caso era una economía de fichas con la finalidad de incrementar el número de asignaciones completas aunque siguiendo diferentes criterios. Luego se incorporó una fase de reversión y por último se reinstauró el procedimiento de economía de fichas. También en este caso, los resultados fueron exitosos y se demostró control experimental.

Diseños de Línea Base Múltiple

Este tipo de diseño es posiblemente uno de los más usados en Análisis Conductual Aplicado (Kratochwill, 1978; Baer, Wolf y Risley, 1987). Se emplea cuando se trata de lograr cambios permanentes, cuando se supone que la conducta es irreversible como por ejemplo en el caso de las conductas conceptuales (Orth-Lopes, s/f) y cuando consideraciones prácticas o éticas limitan el uso del reversión (Barlow y Hersen, 1988).

Existen tres variantes básicas: entre conductas, entre sujetos y entre situaciones:

Diseño de Línea Base entre conductas: En este caso, se toman Líneas Base de dos o más conductas, las cuales se suponen son independientes funcionalmente. Una vez obtenida estabilidad, se introduce la variable independiente en una de ellas y se continúa el registro en las otras conductas. Si se observa un cambio en el comportamiento tratado y ninguno en los otros se aplica el tratamiento en la segunda conducta y así sucesivamente hasta que la variable independiente ha sido aplicada a todas

las conductas. Se debe alcanzar estabilidad de la Línea Base antes de cada aplicación.

Barlow y Hersen (1988) presentan una lista de ejemplos de diferentes investigaciones en las cuales se ha empleado el diseño de Línea Base Múltiple. A manera de ilustración presentaremos una aplicación de este diseño. Berler, Gross y Drabman (1982) utilizaron un diseño de Línea Base Múltiple entre conductas para evaluar la efectividad de un entrenamiento en destrezas sociales aplicado a tres niños, con edades comprendidas entre los 8 y 10 años, que fueron etiquetados como niños con problemas de aprendizaje. El entrenamiento incluía modelaje, ensayos conductuales y feedback sobre la ejecución. Las conductas entrenadas fueron: contacto visual durante la conversación, respuestas verbales adecuadas y duración de la conversación. También se tomaron medidas sociométricas. Los resultados indican que niños con problemas de aprendizaje pueden ser entrenados y adquirir destrezas sociales apropiadas en situaciones de juego. Sin embargo, no se obtuvo generalización de estos comportamientos a otras situaciones.

Diseño de Línea Base Múltiple entre sujetos: El principio es el mismo considerado anteriormente; es decir, se despliegan en paralelo los datos de una conducta particular en diferentes sujetos. Posteriormente se aplica la variable independiente sobre el comportamiento de uno de ellos. Cuando se observen cambios en esta conducta en uno de los sujetos y no en la de los otros sujetos, se aplica el tratamiento al individuo siguiente y así sucesivamente se van interviniendo los demás.

Ejemplo de este tipo de diseño lo presentan Jones, Kazdin y Haney (1981). El estudio tenía como propósito fundamental evaluar los efectos de un entrenamiento en

habilidades de escapar del fuego en una emergencia, en 5 niños cursantes del tercer grado de educación básica. El entrenamiento consistía de instrucciones, moldeamiento, modelamiento, retroalimentación y reforzamiento externo y autoaplicado. La variable dependiente era la respuesta de escape correcta medida como porcentaje. Se utilizó el diseño de Línea Base Múltiple entre sujetos para evaluar la efectividad de este paquete de intervención. Los resultados indican que los niños adquirieron las habilidades de escape del fuego cuando en cada uno de ellos se produjo la intervención, pero luego de un seguimiento de 5 meses se encontró que tales conductas habían disminuido; por lo que los autores recomiendan hacer seguimiento de las intervenciones, puesto que la durabilidad del cambio conductual no debe ser una medida secundaria de los efectos del entrenamiento sino un aspecto de suma importancia por considerarse en cualquier estudio.

Diseño de Línea Base Múltiple entre situaciones. Siguiendo el mismo principio, un tratamiento es aplicado secuencialmente a un único sujeto o a un grupo a través de diversas situaciones independientes entre sí.

Ollendick (1981) empleó una estrategia de Línea Base Múltiple entre situaciones para evaluar la efectividad del automonitoreo y autosobrecorrección sobre la conducta de tics (guiños) que presentaba un niño de 11 años de edad. Tal conducta comenzó cuando el niño tenía 5 años. Las situaciones consideradas para el estudio fueron la escuela y el hogar, iniciándose el tratamiento en el colegio. Los resultados obtenidos dan confirmación positiva del efecto del tratamiento sobre la disminución de los tics tanto en la escuela como en el hogar. Además, resaltan en el estudio los índices de confiabilidad y la forma en que fueron obtenidos.

En resumen, en el Diseño de Línea Base Múltiple entre conductas, una variable independiente es aplicada secuencialmente a conductas independientes de un mismo sujeto. En el Línea Base Múltiple entre sujetos, una variable independiente es aplicada secuencialmente a la misma conducta a través de sujetos distintos que comparten las mismas condiciones experimentales. En el Línea Base Múltiple entre situaciones, una variable independiente es aplicada secuencialmente a la conducta de un mismo sujeto a través de situaciones diferentes e independientes.

Sin embargo, debemos destacar las bondades y limitaciones de este tipo de diseño. Si bien es cierto que este diseño se convierte en una alternativa viable para cuando tenemos casos de conductas irreversibles, cuando deseamos un cambio conductual permanente o cuando las consideraciones éticas y prácticas no nos permiten la aplicación de controles más depurados, sin embargo no es menos cierto que es un diseño débil, puesto que demuestra su control de manera indirecta; es decir, se infiere el resultado en la variable dependiente a partir de lo sucedido en las otras conductas, sujetos o situaciones tratadas (Barlow y Hersen, 1988). Además, la necesidad de obtener líneas de base estables hace que se emplee demasiado tiempo y esfuerzo, en especial en el monitoreo de datos (Kratowill, 1978). Por otro lado, es imprescindible que las conductas y situaciones sean independientes, puesto que si se observan cambios en alguna de ellas cuando no han sido tratadas, el control experimental es difícil de demostrar y la interpretación de los datos se oscurece (Kazdin, 1980). Por último, en este tipo de diseño tampoco se eliminan los posibles efectos de secuencia de los tratamientos (Kazdin y Hartmann, 1978; Orth-Lopes, s/f), y Kratowill (1978) apunta algunas consideraciones que podrían tomarse en cuenta

para mejorar el control en estos diseños; por ejemplo: introducir cuando sea posible, fases de reversión, asegurar independencia funcional de conductas y situaciones, tomar un mínimo de tres o cuatro líneas base.

Diseño Multielemento o de Condiciones Alternantes

El primero en asomar este tipo de diseño fue Sidman (1975) llamándolo «Diseño Línea Base Multielemento», pero tal como lo exponen Ulman y Sulzer-Azaroff (1975), el nombre Diseño de Condiciones Alternantes puede ser igualmente descriptivo. Éste se utiliza cuando se desea evaluar el efecto de dos o más variables independientes sobre un comportamiento determinado, y según Higgins y Baer (1989) podría ser considerado como un tipo de diseño de reversión (muchas reversiones en tiempo mínimo). Resumidamente consiste en la presentación alternada de condiciones de línea base y experimentales, ya sea en forma aleatoria o siguiendo algún programa sistemático, dentro de una misma sesión o entre sesiones. El diseño involucra la medición repetida de la conducta bajo condiciones discriminables alternadas de las variables independientes. Según Kratochwill (1978), el diseño multielemento es un tipo de diseño de programa múltiple, debido a que la única diferencia es técnica: en un diseño de programa múltiple el empleo de una contingencia de reforzamiento se hace bajo un programa específico asociado a cada estímulo, mientras que en el caso del multielemento no se realiza esto. Por otro lado, el control experimental se demuestra cuando se desarrollan patrones diferenciales de respuesta ante cada situación experimental particular.

Ulman y Sulzer-Azaroff (1975) resaltan una serie de ventajas que presenta este diseño:

- A. Permite trabajar con conductas irreversibles o cuya tasa sea menos propensa a la reversión después de presentaciones prolongadas de una variable independiente.
- B. Es apropiado para analizar conductas complejas y especialmente aconsejable para aislar los efectos de variables controlantes interrelacionadas.
- C. Permite evaluar el efecto de dos o más variables independientes sobre el comportamiento.
- D. Habilita la posibilidad de concluir el estudio con relativa rapidez.
- E. Es conveniente su uso cuando tenemos Líneas de Base muy inestables.
- F. Tiene mayor aceptación por parte del personal de una institución, sea hospitalaria, escolar o empresarial.
- G. Minimiza los efectos de variables extrañas tales como los de secuencia o el contraste conductual.

Orth-Lopes (s/f) destaca que aún teniendo todas esas ventajas, cuando en muchas ocasiones se estudian conductas conceptuales o donde el aprendizaje requiere de cierto tiempo, este tipo de diseño no ofrece todas las garantías de control; pues tendrá que asegurarse que el sujeto adquirió el comportamiento y podrían confundirse los procedimientos más efectivos. Tanto él como Kazdin y Hartmann (1978) recomiendan introducir una medida de pre-test y post-test, además de finalizar la intervención con el tratamiento que haya demostrado ser más efectivo. También será conveniente no sólo contrabalancear los tratamientos, sino hacerlos bien discriminativos para disminuir efectos acumulativos (Barlow y Hersen, 1988).

Glendick, Shapiro y Barrett (1981) llevaron a cabo un estudio donde pretendían evaluar los efectos de dos tratamientos (restricción física y sobrecorrección con práctica positiva) versus el no tratamiento, en la

disminución de conductas estereotipadas presentadas por tres niños con retardo en el desarrollo y trastornos emocionales. Adicionalmente, se introdujo el tratamiento más efectivo al final de la intervención. Los resultados indicaron patrones diferenciales de respuesta ante cada condición de tratamiento, aunque todos siguieron la misma función; pero la magnitud de cambio favoreció a uno u otro tratamiento dependiendo del caso. En uno de los niños, la variable independiente más efectiva fue la práctica positiva mientras en los dos restantes fue la restricción física.

Alfonzo, Bermúdez y Calleja (1990) emplearon un diseño de condiciones alternantes con la finalidad de evaluar los efectos de dos tipos de contingencias de reforzamiento orientadas al grupo (Interdependiente y Dependiente) frente al reforzamiento no contingente sobre la conducta creativa. El factor de Creatividad reforzada en todos los casos fue Fluidez. Los resultados indican que las contingencias de reforzamiento produjeron cambios sustanciales sobre la conducta medida pero no se observaron efectos diferenciales ante cada tipo de contingencia. Por otro lado, se evidenció la generalización de un factor entrenado (Fluidez) a otro no entrenado (Originalidad).

Diseños balanceados y contrabalanceados

Estos diseños se emplean cuando el investigador está interesado en evaluar la ejecución de varios sujetos, pero manteniendo el principio fundamental de que cada sujeto actúa como su propio control. Permite controlar factores de historia, secuenciación de condiciones experimentales y orden de los mismos. También se pueden hacer comparaciones tanto intra como inter-individuales (Castro, 1979).

Diseños balanceados

Este diseño se caracteriza porque cada sujeto (o cada grupo) es sometido a todas las condiciones experimentales, solamente que los diferentes grupos o sujetos pasan por dichas condiciones en diferente orden. Este diseño no sólo permite evaluar el efecto de las variables independientes sobre la variable dependiente sino también los efectos residuales que produce un tratamiento en el siguiente y los efectos de orden en que se presentan las condiciones experimentales. Además, permite comparaciones intra e inter-individuales (Underwood, 1972).

Según Castro (1979), pueden presentarse de la siguiente manera:

A A-B-A ó A-B-A-B
B A-B B-A-B-A

En este caso, el diseño puede extenderse en forma horizontal, lo cual indica que el mismo sujeto o grupo de sujetos son sometidos a la secuencia correspondiente.

B ABA-BAB ó ABAB-BABA

Así se está asegurando una replicación intrasujeto y el guión indica que deberá transcurrir un periodo de tiempo lo suficientemente prolongado para que la misma unidad experimental pueda ser sometida a la otra secuencia del diseño y evitar posibles efectos acumulativos.

Castro (1979) señala varias razones para preferir el primer tipo de balanceo sobre el segundo: a) realmente ningún diseño balanceado cancela los efectos acumulativos b) pone de manifiesto los efectos de orden, y b) es preferible utilizar sujetos independientes en cada secuencia, precisamente para resaltar las posibles relaciones funcionales que se estén investigando.

Rangel (1990) llevó a cabo un estudio en el cual evaluó los efectos de dos estrategias de enseñanza de la indicación en niños en edad pre-escolar. Para ello empleó

un diseño balanceado con dos grupos experimentales, donde el orden de presentación de las condiciones (Enseñar una dimensión vs Enseñar dos dimensiones) fue diferente para cada grupo. Los resultados de la investigación mostraron la posibilidad de que es más efectivo enseñar primero una dimensión y luego dos, en vez de la secuencia contraria.

Por su parte, Lacasella (1987) realizó una investigación cuyo objetivo fundamental era evaluar los efectos de dos tipos de contingencias de reforzamiento (Reforzamiento Descriptivo-Social vs Reforzamiento Social) sobre la conducta creativa en la actividad de collage. Utilizó un diseño balanceado con dos grupos experimentales, en donde el orden de presentación de las variables fue diferente para cada grupo. También se incluyó un Grupo Control al cual no se le administró ningún tipo de contingencia. Aunque los resultados no permiten evidenciar la supremacía de un tipo de reforzamiento sobre el otro, hubo un mayor incremento de las conductas cuando estaba vigente el Reforzamiento Descriptivo-Social. Por otro lado, se concluyó que las contingencias de reforzamiento produjeron incrementos sustanciales en las conductas reforzadas.

Diseños Contrabalanceados

Según Castro (1979) se pueden clasificar en diseños contrabalanceados completos e incompletos, dependiendo de si se pueden presentar todas las combinaciones posibles de las condiciones. Están caracterizados porque cada sujeto o grupo de sujetos es sometido a todas las condiciones experimentales, aunque deban seguirse las siguientes consideraciones:

- A. Cada unidad experimental es sometida una vez a cada condición.

- B. Los diferentes sujetos pasan por cada una de las condiciones en diferentes órdenes.
- C. Los grupos o sujetos deben ser asignados al azar a las diversas condiciones.

Realmente en estos diseños y por lo general no se emplea la tecnología de Línea Base y no hay replicación intra-sujeto o intra-grupo.

Villalba (1983) evaluó el efecto de dos tipos de instrucciones (En referencia sólo al aspecto relevante del estímulo vs En referencia a los aspectos relevantes y no relevantes del estímulo) sobre la adquisición de una discriminación visual en niños preescolares. Para ello, utilizó un diseño contrabalanceado, donde se tomaron en cuenta tanto los sujetos, las condiciones experimentales y las tareas por enseñar. La autora justifica la selección de tal diseño debido a que le permitía controlar los efectos del orden de entrenamiento, la interacción de las tareas, el tipo de instrucciones utilizadas para su enseñanza y las posibles diferencias en dificultad de las tareas. Asimismo, las variables edad y sexo fueron controladas al balancear los grupos de forma que los sujetos quedasen equiparados en estas características. Los resultados de esta investigación demostraron que es más efectiva la instrucción, que incluye información tanto del aspecto relevante como el irrelevante, en relación con aquella que sólo incluye información sobre el aspecto relevante del estímulo, puesto que instiga la comparación entre dichos estímulos.

Por otra parte, Englert (1985) empleó un diseño contrabalanceado con el fin de evaluar el efecto de las instrucciones y la extinción como técnicas de corrección en el logro de una discriminación visual en niños preescolares. En el diseño se consideraron tanto los sujetos, las tareas y los procedimientos de corrección. Igualmente, los sujetos fueron balanceados con respecto al sexo y la

edad. Los resultados confirmaron que la extinción funciona como un procedimiento más efectivo en el logro de una discriminación visual. Por otro lado, el uso de la instrucción detallada era necesaria solamente para la adquisición de la respuesta deseada, pero posteriormente podía mantenerse con una mínima instrucción.

No podríamos finalizar este punto sin hacer referencia a que los diseños antes expuestos son sólo una demostración. Ellos se erigen como la estrategia básica de abordaje en Análisis Conductual. También en aras de la búsqueda de un mayor control experimental que nos permita afirmar con mayor precisión las relaciones encontradas entre eventos, los diseños experimentales se han hecho cada vez más complejos pero siempre conservando los principios básicos de la investigación de caso único entre ellas: las medidas repetidas en el tiempo, el sujeto como su propio control, y el cambio rápido de variables cuando se identifica la fuente de varianza. No podrían hacerse afirmaciones acerca de cuáles son los diseños más empleados en los últimos tiempos puesto que exigiría una inversión de esfuerzo que excede este trabajo, pero sí se pueden resaltar los intentos de algunos investigadores por alcanzar un mayor control experimental, como por ejemplo Kazdin y Hartmann (1978), quienes sugieren el empleo del diseño de Tratamiento Simultáneo, donde se combinan estrategias de los diseños de Línea Base Múltiple, de Criterio Cambiante y Programas múltiple. Así mismo, otras posibilidades las asoma Orth-Lopes (s/f). En este mismo sentido, en los últimos años, en el Instituto de Psicología y la Escuela de Psicología de la Universidad Central de Venezuela se han llevado a cabo estudios en los cuales se han combinado diseños de grupos tradicionales o cuasiexperimentales con diseños intrasujeto (Lacaille,

1996; Martínez y Mucci, 1997; Llindis y Messina, 1998; Stefanile, 1999). Todo lo expuesto anteriormente no hace otra cosa que destacar nuevamente la importancia del control experimental para demostrar relaciones funcionales entre eventos y así permitir la construcción de una teoría del comportamiento.

Generalidad de los hallazgos

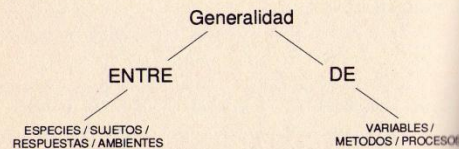
Tal como lo resalta Kerlinger (1988), la generalidad de los hallazgos es una de las metas implícitas de la Ciencia. Sin ella, la labor científica quedaría excesivamente circunscrita y no permitiría en última instancia la construcción de una teoría. Posiblemente deberíamos aclarar el término.

Para Johnston (1979), la generalidad

«...se refiere a la universalidad o replicabilidad; formalmente hablando, ésta puede ser definida como las características de los datos numéricos o interpretaciones verbales de los datos, los cuales describen algún significado (efecto) más allá de las circunstancias de su origen» (p. 3).

Es decir, la generalidad tiene que ver con la posibilidad de que los resultados obtenidos sean aplicables a otras circunstancias que difieren en algún grado de la situación original (Johnston y Pennypacker, 1980) y por lo tanto en alguna medida es equivalente a la validez externa. Johnston (1979) resalta sobre la base de la definición anterior, diversos tipos de generalidad en los cuales puede estar interesado el investigador, dependiendo de su objetivo. Por ello distingue la Generalidad *entre* y la Generalidad *de* sobre la base de si se refiere a la generalidad de las relaciones funcionales entre las variaciones de respuestas del sujeto y el ambiente experimental

(generalidad entre) o si se refiere a las características propias de la relación (generalidad de). Expondremos el esquema que propone el autor, por considerar que aclara esta distinción.



La *Generalidad entre Especies* se refiere a la posibilidad de que los hallazgos experimentales encontrados en una especie biológica, sean generalizables a los sujetos de otras especies (Sidman, 1975).

La *Generalidad entre Sujetos* se refiere a la representatividad de los hallazgos entre sujetos, es decir, «...¿hasta qué punto es este resultado, representativo de otros organismos de esa especie?» (Sidman, 1975; p. 54).

La *Generalidad entre Respuestas* alude a la búsqueda de universalidad de una relación funcional entre clases de respuestas diferentes a la investigada (Johnston, 1979).

La *Generalidad entre Ambientes* también apunta a la posibilidad de generalizar los hallazgos encontrados en una situación, por ejemplo desde el ambiente experimental a otros que difieren en algún grado, como el ambiente natural (Johnston, 1979).

La *Generalidad de las Variables* se refiere a si «... una variable o clase de variables dada es relevante más allá de los confines de un experimento concreto» (Sidman, 1975; p. 64).

La *Generalidad de los Métodos* hace referencia a la demostración de la generalidad de las técnicas

experimentales; es decir, a la replicabilidad de los efectos de los procedimientos sobre una conducta. (Guevara, 1994). Tal como lo resalta Johnston (1979) tanto la generalidad de variables y la de métodos alude a la parte de la relación funcional relacionada con la variable independiente.

La *Generalidad de los Procesos* apunta hacia el lado de la variable dependiente en una relación funcional, y se refiere a la generalidad de una interacción de variables, la cual podría ser llamada proceso conductual, o a la generalidad de la relación funcional obtenida (Johnston, 1979).

Sidman (1975) ha formulado dos procedimientos para garantizar la búsqueda de generalidad de los hallazgos desde una perspectiva conductual. Estos son la réplica directa y la réplica sistemática. Barlow y Hersen (1988) además asoman otro tipo de réplica que es muy común en la investigación aplicada, la cual es la réplica clínica.

La *réplica directa* es definida como «...la repetición, efectuada por el mismo experimentador, de un experimento» (Sidman, 1975; p. 79). Ésta, según el autor, puede llevarse a cabo de dos formas: repitiendo el experimento con otros sujetos o repitiendo las observaciones en el mismo sujeto. Es evidente, que la primera de ellas puede asegurar la generalidad de los resultados mientras que la segunda da más pruebas acerca de la fiabilidad de los datos.

Barlow y Hersen (1988) resumen una serie de consideraciones que deberán tomarse en cuenta para realizar una réplica directa en la investigación aplicada:

- A. Mantener los terapeutas y contextos constantes a lo largo de la réplica.
- B. Tratar que los trastornos abordados sean lo suficientemente similares, desde el punto de vista topográfico por lo menos.

- C. Mantener una correspondencia lo más cercana posible de las historias de los pacientes.
- D. Aplicar de manera uniforme el procedimiento a cada sujeto involucrado en la réplica.
- E. Se recomienda un experimento y tres réplicas exitosas con la finalidad de dar por terminada la réplica directa y pasar a la réplica sistemática.

A partir de la réplica directa, tal como se mencionó anteriormente, puede obtenerse la generalidad de los resultados entre sujetos; pero no podemos responder preguntas relacionadas con la aplicabilidad de los hallazgos a otros contextos o por otros agentes de cambio. Para ello, Sidman (1975) propone el segundo tipo de réplica: La réplica sistemática. Ésta permitiría replicar los resultados de una réplica directa, pero bajo diferentes contextos, agentes de cambio, conductas o cualquier otra combinación; es decir,

«Allí donde la réplica directa ayude a establecer la generalidad de un fenómeno entre los miembros de una especie, la réplica sistemática también puede lograrlo, a la vez que extiende la generalidad del fenómeno en cuestión a una gran variedad de situaciones distintas» (Sidman, 1975, p.114).

Igualmente, Barlow y Hersen (1988) realizan un conjunto de recomendaciones para llevar a cabo con éxito una serie de réplicas sistemáticas:

- A. Que la réplica sistemática sea la consecuencia de una réplica directa sólida.
- B. Deberán especificarse todos los detalles en relación con los sujetos, las conductas, los agentes de cambio y cualquier otra variable que se haya introducido, que hagan a la investigación diferente de la original.
- C. El procedimiento experimental debe ser susceptible al registro de cualquier resultado, sea a favor de la

réplica o no, puesto que la búsqueda de las excepciones es parte esencial de la réplica sistemática.

- D. A diferencia de la réplica directa donde los autores, junto con la opinión de Sidman (1975) asoman la posibilidad de finalizar una réplica directa luego de tres experimentos exitosos; ellos no proporcionan en este caso recomendación alguna. El número de réplicas dependerá del fenómeno, del objetivo de investigación, de la habilidad del investigador.

Por último, aludiremos brevemente a lo que Barlow y Hersen (1988) denominan réplica clínica, ya que ésta es de suma importancia en la investigación aplicada. Los autores la definen como un proceso avanzado de réplica donde se evalúan dos o más procedimientos de intervención o un paquete de intervención, sobre la conducta de varios sujetos con trastornos de conducta más o menos similares. Ésta podría ser muy poderosa si los resultados son exitosos, pero podría ser muy costosa de hallarse resultados inconsistentes. Acerca de ella valen las mismas recomendaciones hechas para la réplica directa.

Como hemos visto el problema de la generalidad de los hallazgos no es sencillo ni tiene abordajes sencillos. De cara a la demanda diaria en relación con técnicas y métodos, no sólo los analistas conductuales aplicados, sino los científicos que desean construir cuerpo de datos coherente, deben considerar seriamente que la puesta en marcha de réplicas tanto sistemáticas como directas darán una ruta consistente para la evaluación de la generalidad de los datos y en última instancia para la construcción de una teoría del comportamiento.

**Aprendiendo a vivir
con la estadística:
algunas recomendaciones
para la lectura de artículos
basados en diseños de grupo.**

Cristina Vargas-Irwin

Una de las alegrías más generalizadas al asumir una posición teórica conductual en psicología es poder decirle adiós a la estadística y a su pariente cercano, el diseño de grupos. Este gozoso distanciamiento del análisis estadístico no se limita a los estudiantes de pregrado, si no que ha sido expresado en forma abierta y jubilosa por varias personalidades del análisis conductual como Branch (1999) o Ator (1999). El conductista goza no sólo de la aprobación social de sus colegas al utilizar diseños experimentales sin estadística, si no que también cuenta con un amplio bagaje de literatura metodológica conductual que le ofrece muy buenas razones para no hacer aquello que de todas formas resulta tan desagradable. ¿Porqué hablar entonces de estadística, si podemos gozar de lo mejor de todos los mundos?

Existen varias razones para ello. Por una parte, el asumir un rechazo generalizado a los diseños estadísticos sitúa al Análisis Conductual de buena parte de la disciplina psicológica, tal y como lo han puesto de manifiesto autores como Huitema, (1986). Difícilmente existe algún área dentro del análisis conductual que no tenga un paralelo dentro de la investigación psicológica en general con varias décadas de investigación utilizando diseños de grupo. En

el ámbito clínico o educativo esto es especialmente cierto, ya que el análisis conductual es sólo una pequeña minoría en cuanto a investigaciones realizadas en cualquiera de estas áreas. Basta dar un vistazo a cualquier ejemplar de revistas de psicología clínica, como el *Journal of Clinical Psychology and Experimental Therapy* o aún a revistas de Terapia Conductual como *Behavior Therapy* para darse cuenta del aplastante predominio de la metodología estadística. Esta situación sólo se acentúa en las publicaciones psicológicas de interés general, como *American Psychologist*, *Annual Review of Psychology*, *Psychological Bulletin* y *Psychological Review*, donde el número de publicaciones de orientación conductual no parece haber pasado de un 6% en los últimos 20 años (Robbins, Gosling y Craik, 1999), lo cual evidencia la influencia limitada que la metodología del diseño intrasujeto puede tener en la actualidad. El descalificar automáticamente todo el inmenso cúmulo de información que ha generado la metodología de grupo en psicología a lo largo de los últimos cien años luce cuando menos arrogante y cuando más temerario. Con toda seguridad debe existir aunque sea una mínima porción de esta literatura que resulte valiosa, aún asumiendo una posición centrada en la conducta del organismo individual.

Una segunda razón para preocuparnos por los diseños de grupo son las limitaciones intrínsecas de los diseños intrasujeto. El valor de la metodología conductual para el estudio y la modificación del comportamiento del organismo individual es indiscutible. No obstante, su poder y validez interna derivan de un estricto control de variables que permiten el estudio de los efectos de un muy reducido número de factores simultáneamente. De hecho, buena parte de los diseños tradicionales del Análisis Conductual, tales como las distintas variantes de los diseños de reversión o los diseños de línea base múltiple,

han sido concebidos para evaluar el efecto de una sola variable independiente. Aún los denominados diseños conductuales de interacción (Hersen y Barlow, 1976) resultan terriblemente engorrosos cuando el número de variables independientes pasa de dos o tres. En cambio, los diseños de grupo y la metodología analítica de la moderna estadística multivariada permiten manejar simultáneamente decenas de variables dependientes e independientes, así como modelar complejas relaciones entre ellas más allá de la simple relación lineal del tipo «a mayor inteligencia mayor rendimiento académico.»

Una tercera y última razón para ocuparnos de la estadística es que no podemos evitarla. Los alcances de esta modalidad analítica rebasan los límites de la vida académica, invadiendo cada rincón de nuestro quehacer cotidiano. El valor argumentativo de la estadística ha sido reseñado por autores como Abelson (1995), quienes nos alertan sobre sus usos inadecuados. Otra de las razones por las cuales el analista conductual debe saber manejar el argot estadístico, es entonces el evitar caer víctima de él.

La utilidad y ubicuidad de la estadística y de los diseños de grupo no los eximen, sin embargo, de los peligros que tan agudamente han señalado los teóricos conductuales como Sidman (1960) y Skinner (1956). También es cierto que no todas las investigaciones que utilizan diseños de grupo son igualmente buenas o valiosas desde una perspectiva conductual. El objetivo del presente capítulo radica en identificar esos puntos claves en las investigaciones basadas en diseños de grupo que pueden resultar más importantes desde la perspectiva conductual del desarrollo infantil. Más específicamente, buscamos llamar la atención acerca de ciertos aspectos metodológicos de los reportes de investigación que suelen pasar desapercibidos en una lectura guiada principalmente por

intereses sustantivos. Para ello, formulamos una serie de ocho preguntas que deberían ser respondidas en un examen de las características metodológicas de una investigación.

¿Qué buscar en un artículo?

La lectura de un artículo de investigación no es un proceso unívoco, si no que debe relacionarse con los objetivos de la investigación o intervención para la cual se está recabando información. Dependiendo de dichos objetivos, ciertos aspectos se tornan más relevantes que otros. Es por ello que resulta difícil el formular prescripciones específicas para la lectura de artículos. No obstante, el asumir una perspectiva conductual conlleva necesariamente el poner en relieve algunos elementos particulares de los reportes de investigación, tanto en el ámbito sustantivo como en el metodológico.

¿Qué constituye un buen artículo, es decir, un artículo útil, desde una perspectiva conductual? Johnston y Pennypacker (1990) han dado respuesta parcial a esta pregunta en su formulación de los lineamientos de la investigación conductual pura en contraposición a lo que estos autores denominan «investigación cuasi conductual»¹. Por investigación conductual pura Johnston y Pennypacker entienden aquellas investigaciones que preservan la gama completa de propiedades de la conducta a través de todo el proceso de investigación, «(...) sin diluir, abreviar o distorsionar su manifestación en los datos.» (p. 32)

¹ Para los jóvenes estudiantes de psicología que no se formaron bajo el texto clásico de Campbell y Stanley «Diseños Experimentales y Cuasiexperimentales en Psicología y Educación» vale la pena señalar la travesura de Johnston y Pennypacker. En el mencionado libro Campbell y Stanley denominaban «cuasi experimentales» a los diseños conductuales, de allí la venganza de utilizar el término «cuasi conductual» para hacer referencia a los diseños de grupo.

Para que una investigación sea considerada como investigación conductual pura debe cumplir con una serie de criterios que pueden agruparse en dos grandes categorías: las consideraciones con respecto a la medición de las variables y las consideraciones con respecto a las estrategias de inferencia.

Las Trampas de la Medición Psicológica

En lo que respecta a las consideraciones relacionadas con la medición de las variables, el primer aspecto por tomar en cuenta es la clara definición de las unidades de análisis en términos del comportamiento del sujeto individual que se desea estudiar. Pareciera que en principio, la medición en psicología suele adherirse a este primer criterio tanto dentro como fuera del análisis conductual, ya que las técnicas clásicas de recolección de información psicológica, tales como el autorreporte, la entrevista o las pruebas psicométricas parten siempre del comportamiento del organismo único (aun cuando estas medidas de organismos individuales se combinen luego durante el análisis de datos, pero sobre esto ahondaremos más adelante). No obstante, existen a este respecto tres peligros o trampas fundamentales referentes a la medición de las variables presentes en los reportes de investigación ante las cuales debe estar alerta, a saber: las mediciones indirectas del comportamiento, la técnica de recolección de información como fuente espúrea de correlación y las medidas adimensionales del comportamiento. En primer lugar, es frecuente encontrar que si bien las mediciones se recogen para cada sujeto en particular, la información no se recaba directamente del comportamiento del sujeto de interés, sino indirectamente a través del comportamiento de los sujetos en contacto próximo con aquel cuya conducta es el foco de la investigación. Un ejemplo puede contribuir a aclarar este punto.

Evans, Shejwal y Palsane (1998) estudiaron las repercusiones del hacinamiento crónico sobre una serie de variables relacionadas con el bienestar infantil, tales como el rendimiento académico, el ajuste escolar, la presión arterial y la desesperanza aprendida. Pareciera que todas estas variables hicieran alusión directa al comportamiento de los niños en condición de hacinamiento doméstico crónico. No obstante, una revisión más cercana de la definición de las variables reportadas por los autores revela, por ejemplo, que el «ajuste escolar» no se mide directamente a través del comportamiento de los niños en el colegio sino mediante un cuestionario aplicado a las maestras. El «ajuste escolar» así medido, constituye una muestra directa sólo del comportamiento verbal de las maestras, estando relacionado cuando más en forma indirecta con el comportamiento de los niños. No es de extrañarse entonces que esta variable resultara menos sensible que las demás a los efectos del hacinamiento doméstico, ya que el hacinamiento debía producir no sólo diferencias en el ajuste escolar de los niños si no también diferencias del tipo que pudiesen afectar directamente el reporte verbal de las maestras. En función de las consideraciones anteriores, podemos entonces formular una primera pregunta a plantearnos en la revisión de un artículo de investigación:

¿Se mide realmente el comportamiento individual del sujeto de interés?

Una segunda versión más sutil del problema anterior viene dado por aquellas mediciones del comportamiento del sujeto particular que es el foco de la investigación, pero que no miden aquella porción de su repertorio conductual que nos interesa sustantivamente. Este suele ser el caso de las investigaciones basadas en el autorreporte

como técnica de recolección de información, donde se pretende acceder bien sea a otras porciones del repertorio conductual del sujeto o a porciones importantes de su historia de vida mediante su reporte verbal.

Un primer ejemplo especialmente ilustrativo a este respecto es el que mencionan Leland Wilkinson y la Comisión sobre Inferencia Estadística de la Asociación Psicológica Norteamericana, APA (1999). Este artículo, que recoge una serie de recomendaciones metodológicas para los reportes de investigación en psicología y presenta coincidencias, a veces sorprendentes, con algunas de las críticas que el Análisis Conductual ha formulado a la forma tradicional de hacer investigación en psicología. Esto es especialmente cierto en lo que atañe a las recomendaciones sobre la medición. Los autores enfatizan en este punto la importancia del nombre que se le da a las variables. No es lo mismo, nos advierten los autores, hablar de «abuso sexual infantil» que del «autorreporte retrospectivo del abuso sexual infantil». En el primer caso se le atribuye mayor fiabilidad a la ocurrencia del incidente de abuso, mientras que en el segundo caso se resalta la influencia del recuerdo y la elaboración verbal del sujeto sobre el incidente particular. Podríamos entonces formular la una segunda pregunta por tomar en consideración en la lectura de un artículo de la siguiente manera:

¿Hasta qué punto las variables reportadas en el artículo constituyen medidas directas del comportamiento?

Una tercera trampa de la medición psicológica estrechamente vinculada con la anterior, ya que suele ser también una consecuencia del uso indiscriminado del autorreporte, es lo que autores como Gottman (1998) denominan «efecto glop». El «efecto glop» hace referencia a la presencia de correlaciones significativas entre variables

conceptualmente distintas entre sí, siendo estas correlaciones un artefacto de la forma en que dichas variables son medidas, y no un reflejo de relaciones funcionales entre los eventos estudiados. El efecto glop es especialmente común en aquellas investigaciones en las cuales varias variables tanto dependientes como independientes se miden mediante el autorreporte de una única fuente, o de fuentes altamente relacionadas entre sí; esto hace que las correlaciones encontradas entre dichas variables respondan más a la fuente de información utilizada que a la relación sustantiva bajo estudio.

Gottman (1998) menciona como ejemplo paradigmático del efecto glop diversas investigaciones sobre relaciones de pareja que arrojaban correlaciones inversas importantes entre la satisfacción matrimonial y la frecuencia de conflictos de pareja. Ambas variables solían ser medidas mediante el autorreporte de cada uno de los cónyuges. No fue sino hasta que los conflictos de pareja fueron medidos directamente mediante observación sistemática, que se encontró que las parejas funcionales y disfuncionales no diferían en la cantidad de conflictos entre los cónyuges sino en la dinámica particular de las peleas. Mientras que en las parejas disfuncionales las peleas siguen un patrón de escalada, donde cada queja o agresión es respondida mediante otra queja o agresión más fuerte, en las parejas funcionales cualquiera de los cónyuges es capaz de hacer bajar el tono emocional de la discusión, respondiendo de forma no conflictiva. Este tipo de hallazgo sólo fue posible mediante medidas completamente independientes de cada una de las variables en juego, utilizando incluso medidas psicofisiológicas del grado de activación emocional de las parejas. Es por ello que un tercer punto por considerar en la revisión de los artículos de investigación es la siguiente:

¿Garantizan los procedimientos de medición la evaluación independiente de las variables estudiadas?

Un segundo criterio enunciado por Johnston y Pennypacker para considerar a una investigación como investigación conductual pura es la medición de alguna cantidad dimensional del comportamiento (Johnston y Pennypacker, 1990). Latencia, duración, frecuencia, tasa, celeración y tiempos entre respuesta son todas cantidades dimensionales de la conducta que permiten la interpretación sustantiva no sólo de valores individuales sino también de diferencias entre dos o más valores. Esta última propiedad de la medición en términos de cantidades dimensionales es especialmente relevante en el momento de evaluar la efectividad de las intervenciones. Si un tratamiento es capaz de producir una disminución en el número de cigarrillos fumados por día de 20 a 5 cigarrillos, la diferencia de 15 cigarrillos tiene un significado claro y unívoco con respecto al comportamiento del sujeto. A diferencia de lo anterior, si un tratamiento logra mejorar el rendimiento en matemáticas de un niño del percentil 60 al percentil 75 en una prueba estandarizada, la diferencia de 15 puntos percentilares nada nos dice acerca de los cambios concretos operados en el comportamiento del niño. Dependiendo de la dispersión de la ejecución del grupo normativo, esta diferencia de 15 puntos puede ser tan trivial como una pregunta adicional contestada correctamente.

Lamentablemente, el uso de medidas adimensionales del comportamiento parece ser un mal igualmente común tanto dentro como fuera del análisis conductual. En una revisión de los artículos del *Journal of Applied Behavior Analysis* publicados durante la década de los setenta, Springer, Brown y Duncan (1981) encontraron que un 41%

de las medidas reportadas en los artículos eran adimensionales, producto de procedimientos de observación discontinuos, tales como el muestreo de tiempo. Esta situación parece haberse mantenido durante los últimos 20 años, a juzgar por la reiterada insistencia de amplios sectores del Análisis Conductual Aplicado en la necesidad de volver a las medidas de tasa de respuesta, por contraposición a las medidas porcentuales del comportamiento (Binder, 1993).

Otro tanto puede decirse de la medición en la psicología en general. Reputadas figuras dentro de la estadística como John Tukey (el de la T de Tukey) y más recientemente, Jacob Cohen han cuestionado seriamente la proliferación de medidas adimensionales en psicología. «Estar tan poco interesados por nuestras variables,» nos advierte Tukey, «como para que no nos importen sus unidades, difícilmente puede ser algo deseable.» (Tukey, 1969, p. 89). Cohen por su parte exhorta a los psicólogos a «(...) comenzar a respetar las unidades con las que trabajan o a generar nuevas unidades lo suficientemente respetables como para que los investigadores del área puedan concordar en su utilización.» (Cohen, 1994, p. 1001). Estas consideraciones pueden entonces traducirse en una cuarta pregunta a responder en la lectura de un artículo:

¿Cuáles son las unidades en que se miden las variables de la investigación?

Un último requisito referente a la medición enunciado por Johnston y Pennypacker hace referencia a la necesidad de garantizar su fiabilidad y validez. En este sentido, las estrategias de estimación de fiabilidad y validez utilizadas en los diseños de grupo suelen diferir de aquéllas empleadas en la investigación conductual, ya que en este último caso la medición suele realizarse mediante la observación directa, mientras que en los diseños de grupo suelen

predominar los cuestionarios, autorreportes y pruebas estandarizadas. A este respecto, es conveniente resaltar otra de las recomendaciones de Wilkinson y la Comisión sobre Inferencia Estadística de la APA (1999), quienes nos recuerdan que la confiabilidad y la validez son propiedades de las puntuaciones en la prueba para una población en particular y no del instrumento en sí mismo. Por consiguiente, toda investigación debe reportar datos sobre la fiabilidad y validez del instrumento utilizado para la población particular estudiada. Nuestra quinta pregunta puede entonces formularse de la siguiente manera:

¿Qué tan válidas y confiables son las mediciones para la población estudiada?

Puede resultar extraño el énfasis dado a aquí a las operaciones de medida, sobre todo porque nuestro objetivo manifiesto consiste en ofrecer una guía para la interpretación de los resultados de investigación de los diseños estadísticos. No obstante, insistimos en que el carácter concreto de las variables del estudio y su distancia con respecto al comportamiento del organismo individual deben constituir el primer punto a abordar en la lectura de un artículo no conductual. No importa que el diseño de investigación sea impecable o cuán elaborado sea el análisis estadístico de los datos si las variables de fondo nada nos dicen sobre la conducta del organismo único. Tal y como lo señalan Johnston y Pennypacker (1990), si bien la metodología del diseño de grupo no hace prescripciones específicas sobre los procedimientos de medición, el sólo hecho de requerir grandes muestras de sujetos ejerce demandas sobre el investigador que le alejan de las medidas directas, dimensionales, válidas y confiables del organismo individual. Es por ello que una lectura cuidadosa de los procedimientos de medición empleados se hace

especialmente importante. Adicionalmente, sólo sobre la base de un adecuado proceso de medición tiene sentido la evaluación de los resultados de investigación. Sobre el respaldo que brindan los datos a las hipótesis y conclusiones de investigación versan los siguientes puntos focales en la lectura de artículos basados en diseños de grupo.

Tratamiento Estadístico

El Análisis Conductual, tanto básico como aplicado, pero especialmente el segundo, ha insistido a lo largo de su desarrollo en la importancia de que el análisis de datos sea determinado por el objeto de estudio; el comportamiento del investigador debe estar controlado por los datos (Michael, 1974). Como consecuencia de lo anterior, para utilizar las palabras de Killeen (1978), el análisis de datos es visto como conducta gobernada por contingencias, no como conducta gobernada por reglas, siendo la tradición conductual poco prescriptiva sobre el procedimiento analítico, aún más de 20 años después del artículo de Killeen (al respecto, véase Davidson, 1999). Por contraposición, la tradición del diseño estadístico ha desarrollado un amplio caudal de estrategias de análisis y toma de decisiones con respecto a los resultados experimentales. Estas estrategias analíticas han constituido parte de la formación metodológica de los psicólogos por generaciones y no en poca medida por su carácter cuantitativo estas estrategias han tendido a ser sacralizadas dentro de la disciplina. El aspecto detectivesco del análisis de datos ha sido sustituido poco a poco por la repetición mecánica de rituales estadísticos, que como la letanía en Latín, son realmente entendidos solamente por un pequeño grupo de «ungidos». No obstante, mientras los fieles rasos continuábamos repitiendo nuestro Rosario de

«Rechazo la hipótesis nula al nivel de 0.05» y el por demás aberrante «Acepto la hipótesis nula al nivel de 0.01»², en el cónclave de los Sumos Sacerdotes las pruebas de significancia estadística eran sometidas a durísimas críticas, dando origen a una avalancha de artículos tanto de literatura estadística especializada como de artículos de divulgación.

El movimiento crítico en torno a las pruebas de significancia alcanza un nuevo pico en la literatura psicológica con la publicación del artículo de Jacob Cohen «La Tierra es Redonda ($p < 0.05$)» (Cohen, 1994). Como consecuencia de esta tendencia, la Asociación Psicológica Norteamericana nombró en 1996 una comisión especial conformada por Cohen, Robert Rosenthal (padre del meta-análisis) y Robert Abelson para estudiar el abuso de las pruebas de significancia en la investigación psicológica. Esta comisión se amplió para incluir a nueve personas más, así como a figuras de la talla de Lee Cronbach, Paul Meehl, Frederick Mosteller y John Tukey dentro del Comité asesor, produciendo su primer documento oficial a finales de 1999 (Wilkinson y la Comisión sobre Inferencia Estadística, 1999). Tan importante es el movimiento que se viene gestando que el próximo objetivo de la Comisión es una revisión completa del Manual de Publicación de la APA, lo cual tendrá repercusiones más allá de las fronteras de nuestra disciplina, dado que dichas normas constituyen un patrón de referencia también para la publicación de investigación en otras áreas.

² Aun cuando la probabilidad de Error Tipo I sea inferior al nivel de significancia, es un error decir que se acepta la hipótesis nula; lo más que nos permite este resultado es «no rechazarla». El evitar esta desafortunada expresión es ya una recomendación explícita de la Comisión Sobre Inferencia Estadística de la APA y con toda seguridad será incorporada en la nueva edición del Manual de Normas de Publicación.

¿Qué implicaciones tiene esto para quienes nos desenvolvemos dentro del ámbito del Análisis Conductual? Por una parte, debemos preocuparnos por entender el sentido de las pruebas de significancia, de forma de no hacernos partícipes de la perpetuación de malos entendidos en torno a ellas. En segundo lugar, tenemos que revisar nuestras críticas tradicionales al análisis estadístico a la luz de las nuevas tendencias en el análisis de datos, so pena de terminar atacando a un hombre de paja. En tercer lugar, este reciente cuestionamiento de las prácticas analíticas debe hacernos lectores más críticos tanto de las investigaciones ajenas como propias, con el fin de mantenernos al día sobre los nuevos cánones de evaluación de la investigación psicológica.

Es en este sentido que queremos llamar la atención sobre tres puntos claves por considerar en la lectura de investigaciones que utilizan diseños estadísticos: el peligro de los tamaños muestrales, la importancia de magnitud de los efectos y la importancia de la réplica. Para un adecuado tratamiento de estos puntos es sin embargo importante el aclarar la expresión «estadísticamente significativo».

Significativo pero irrelevante: Significancia clínica y estadística

Si bien en castellano la palabra «significativo» quiere decir que «significa o indica claramente una cosa» (Diccionario Larousse), constituye un americanismo común el entender la significación como importancia. El invertir además a la significación de carácter estadístico como en la expresión «significancia estadística» no hace más que aumentar el sesgo de considerar a lo significativo como importante, trascendente o relevante. Nada más

alejado de la verdad. Tal y como ha sido repetido *ad nauseam* dentro de la literatura estadística, diferencias minúsculas y absolutamente irrelevantes desde el punto de vista tanto teórico como práctico pueden ser «estadísticamente significativas». Lo único que nos indica un resultado «estadísticamente significativo» es que los efectos observados probablemente no son nulos, pero no que sean grandes, importantes o relevantes. De hecho, esta «ausencia de nulidad» es todo lo que podemos afirmar al rechazar el eje de la inferencia estadística Fisheriana: la hipótesis nula.

¿Qué significa realmente la hipótesis nula en términos de las relaciones funcionales entre variables? Básicamente una de dos cosas: que los grupos que estamos comparando son iguales en lo que respecta a la variable estudiada o bien que la asociación entre dos variables es nula. Si nos detenemos un momento en esta última afirmación podemos darnos cuenta de algo que autores como Cohen (1990) han estado predicando en el desierto desde hace muchos años: la hipótesis nula siempre es falsa. Nunca encontraremos que las medias de dos grupos (o de dos pares de tratamiento, para los efectos) sean exactamente iguales, y de encontrarse, esto sólo hablaría de la poca sensibilidad de nuestros instrumentos de medida. Otro tanto puede decirse de los coeficientes de correlación, ya que éstos siempre serán distintos de cero, aunque sea en el décimo decimal. Por lo tanto, al rechazar la hipótesis nula lo único que podemos afirmar es que la diferencia entre los grupos es distinta de cero, lo cual no quiere decir que sea grande, independientemente de que el nivel de significancia utilizado sea de 0.05, 0.01 o 0.001.

Si la magnitud de la relación funcional no es el principal determinante de la significancia estadística, entonces ¿qué la determina? Uno de los factores con más peso al establecer la significancia estadística es el tamaño

muestral. Quienes nos formamos a la vieja usanza, utilizando tablas para determinar los valores críticos de la F de Snedecor o la t de Student y no la probabilidad de Error Tipo I reportada por los programas estadísticos contemporáneos, pudimos ver como los valores críticos de estos estadísticos se hacían más pequeños en la medida en que aumentaban los grados de libertad, para los efectos, el tamaño muestral. Esto evidencia que no importa cuán pequeña sea la diferencia entre los grupos, ella puede ser significativa si se utiliza un tamaño muestral lo suficientemente grande.

Cohen (1990), cita un ejemplo especialmente ilustrativo de cómo el tamaño muestral puede hacer significativas asociaciones minúsculas. Este autor reporta unas investigaciones publicadas en el periódico *New York Times*, en las cuales se encontró una asociación significativa entre la estatura y la inteligencia en una muestra de 14.000 niños entre 7 y 17 años. Cohen, con el escepticismo característico del buen investigador, calculó el valor mínimo del coeficiente de correlación que resultaría significativo con este tamaño muestral y un nivel de significancia de 0.001. El valor resultante fue de 0.027%. Esto quiere decir que la varianza común entre las variables (el coeficiente de determinación, o cuadrado del coeficiente de correlación) podría ser tan pequeña como del 0.07% y todavía la correlación resultar estadísticamente significativa.

Mencionamos en primer lugar el peligro de las conclusiones extraídas a partir de tamaños muestrales sumamente grandes ya que el entrenamiento estadístico tradicional nos enseña a considerar a las muestras grandes como incondicionalmente buenas, sin advertirnos sobre lo triviales que pueden resultar en estos casos las pruebas de significancia. No obstante, si los tamaños muestrales grandes prácticamente garantizan el rechazo de la hipótesis

nula, lo inverso también es cierto: con tamaños muestrales lo suficientemente pequeños resulta muy difícil que las relaciones encontradas resulten estadísticamente significativas.

¿Cuándo es entonces una muestra lo suficientemente grande como para permitir detectar relaciones importantes pero no tan grande como para que todas las posibles relaciones resulten significativas? La respuesta exacta a esta pregunta requiere adentrarnos en el mundo del cálculo del poder de las pruebas estadísticas, rebasando por consiguiente el objetivo de la presente discusión. No obstante, podemos señalar que la prueba de hipótesis estadística tiene en realidad cuatro parámetros, cada uno de los cuales viene determinado por los otros tres (Cohen, 1988). Estos cuatro parámetros son el nivel de significancia o α (α = probabilidad de rechazar la hipótesis nula siendo esta verdadera), la magnitud de la relación en la población (tamaño del efecto), el tamaño muestral y el poder de la prueba estadística ($1 - \beta$), la probabilidad de rechazar la hipótesis nula. De estos cuatro parámetros, sólo tres se encuentran bajo el control del experimentador, ya que la magnitud de la relación en la población viene dada por el fenómeno estudiado, aun cuando se pueda estimar a partir de las investigaciones previas (ahondaremos más en esto en las consideraciones sobre la replicabilidad). Esto nos deja con tres parámetros "libres": al fijar cualquier par de ellos, automáticamente estaremos determinando el valor del tercero. Tradicionalmente, el nivel de significancia α se elige entre uno de dos valores: 0.01 y 0.05 y consideraciones prácticas llevan a fijar el tamaño muestral, con lo cual se fija automáticamente el poder de la prueba estadística. Esto exime al investigador de tomar decisiones sobre β , la probabilidad de rechazar la hipótesis nula, pero es muy importante destacar que esto

no quiere decir que b sea el inverso de a . De hecho, una estrategia metodológicamente más sana sería decidir sobre los valores de a y b , estimar la magnitud del efecto a partir de las investigaciones anteriores, y en función de estos tres parámetros determinar el tamaño óptimo de la muestra. ¿Qué hacer, sin embargo, mientras no contemos con las destrezas del cálculo de poder? Como regla práctica, debemos estar alertas y tomar con cautela cualquier diferencia significativa encontrada con muestras superiores a 500 sujetos, así como cualquier ausencia de diferencias con muestras inferiores a 50; por más allá de esto, debemos asumir un escepticismo ilustrado sobre las pruebas de significancia. Resumiendo entonces las consideraciones anteriores sobre poder y tamaño muestral, podemos formular nuestra siguiente pregunta por contestar en la revisión de un artículo de investigación como sigue:

¿Es el tamaño muestral lo suficientemente grande como para detectar las relaciones buscadas, pero no tan grande como para hacer trivial la significancia estadística?

El tratamiento anterior de las pruebas de significancia pareciera confirmar todas nuestras sospechas acerca de la irrelevancia de la estadística, justificando así el rechazo de las mismas por parte de los analistas conductuales y más recientemente (aun cuando repitiendo en forma menos elaborada las críticas de los conductistas) el rechazo de los enfoques cualitativos en ciencias sociales. No obstante, basta con pasearse por cualquiera de las publicaciones sobre métodos cuantitativos en psicología publicadas en los últimos 20 o 30 años para notar cómo estas críticas han sido respondidas desde dentro de la comunidad estadística y la variedad de las soluciones cuantitativas que tenemos a la mano hoy por hoy para

responder a las preguntas de investigación. Una de estas familias de estrategias que resulta especialmente relevante desde la perspectiva de la práctica clínica es la estimación de la magnitud de los efectos.

Existe un consenso importante entre los Analistas Conductuales y los metodólogos cuantitativos (por llamar de alguna forma a aquéllos que perpetran la metodología estadística en psicología) de que la ciencia psicológica y sus aplicaciones tecnológicas no pueden avanzar mucho sobre la base exclusivamente de relaciones de orden, sino que resulta imprescindible el especificar relaciones de magnitud. Ya para 1969, John Tukey planteaba el ejemplo interesante de lo que sería la física de la elasticidad si los físicos se hubiesen limitado a enunciados del tipo «*Si lo halas, se hace más largo.*» (Tukey, 1969, p. 86). Con cuánta fuerza se hala y qué tanto se alarga son aspectos imprescindibles para llegar a enunciados como la ley de Hooke o el límite de elasticidad. Es precisamente este tipo de representación formal de los datos por la cual abogaba Skinner (1950) como la forma idónea de hacer teoría en el análisis conductual, tal y como lo expresó en su célebre artículo ¿Son Necesarias las Teorías de Aprendizaje?. A pesar de lo anterior, rara vez encontramos referencias a la magnitud de los efectos de las intervenciones tanto experimentales como clínicas en la literatura psicológica, aun cuando sea una recomendación explícita de la cuarta edición del Manual de Normas de Publicación de la APA. En vez de ello, conseguimos al anodino y poco informativo valor de F o t junto con su consabido nivel de significancia, 0.05 o 0.01.

No es de extrañarse entonces el desdén hacia la estadística cuando todo lo que nos ofrece es esa F enclenque que nada nos dice acerca de las relaciones sustantivas que realmente nos interesan. ¿Fue efectivo el tratamiento?

¿Qué tanto mejoró la ejecución de los sujetos? ¿Se justifica la inversión en tiempo y dinero para la magnitud de cambio obtenida? ¿Qué tan significativo fue el cambio desde el punto de vista clínico? El problema está en buscar la respuesta a estas preguntas en el lugar equivocado, es decir en las pruebas de significancia, cuando lo que realmente nos interesa es la magnitud del efecto.

El ejemplo de Cohen acerca de la relación entre estatura e inteligencia es especialmente ilustrativo a este respecto. Como recordarán, Cohen reporta cómo la relación entre la estatura y la inteligencia resultó ser estadísticamente significativa al nivel de $\alpha = 0.001$ para una muestra de 14.000 estudiantes entre 7 y 17 años (Cohen, 1990). La correlación entre ambas variables era mayor que el valor significativo mínimo para ese tamaño muestral calculado por el autor (0.0278), siendo en realidad de 0.11. Ahora bien, la pregunta realmente interesante es ¿cuánto debe crecer una persona para aumentar su inteligencia, digamos, en 30 puntos de C.I.? La respuesta es 87.5 cm. Esto revela que ninguna cantidad de hormona de crecimiento sería capaz de producir cambios en la inteligencia lo suficientemente grandes como para tener alguna importancia práctica. Podemos pensar que el ejemplo de Cohen es un caso extremo, pero desgraciadamente no lo es. Buena parte de nuestra tradición psicológica local bien podría estar basada en hallazgos significativos pero totalmente irrelevantes. Como ejemplo, basta citar la relación entre la autoestima y el rendimiento académico en estudiantes universitarios, donde el duplicar la autoestima de una persona sólo produciría cambios pírricos en su promedio de notas, aun cuando la relación entre las variables resulta ser estadísticamente significativa (Robles, 1992). Este falso sentido de seguridad que da el rechazar la hipótesis nula

se ve reforzado por otra característica común a la investigación psicológica en general que mencionamos con anterioridad: el carácter adimensional de la medición. En el ejemplo de la estatura, la irrelevancia de la asociación es fácil de ver porque tenemos unas unidades unívocas y con significado para medir esta variable. No obstante, si en lugar de estas unidades tuviésemos puntos percentilares (un aumento del Percentil 45 al 75 en estatura), sería mucho

$$r_{xy} = \frac{\text{covarianza}}{S_x \cdot S_y}$$

más difícil juzgar la importancia del hallazgo. Este es precisamente el caso cuando se reportan los resultados de investigación mediante coeficientes estandarizados como el coeficiente de correlación. Al igual que en el caso de los porcentajes, es precisamente su independencia de las unidades de medida lo que los hace fácilmente comparables unos con otros: sabemos que una correlación de 0.8 es mayor que una correlación de 0.5. No obstante, el precio que pagamos por esta facilidad de comparación es el perder de vista las unidades de cambio en la variable dependiente. Esto se debe a que el coeficiente de correlación es el resultado de dividir la covariación entre las variables por sus desviaciones típicas (Glass y Stanley, 1973).

Por ello, el coeficiente de correlación realmente no nos dice cuánto influye una variable sobre otra, al menos no en términos sustantivos si desconocemos las unidades de medida de las variables.

Si la significancia estadística es engañosa, los valores de P poco informativos y los coeficientes de correlación no nos hablan de magnitudes sustantivas, ¿qué nos queda? Principalmente el tamaño del efecto, que no es más que

una forma de expresar las diferencias entre las medias, por lo general entre unidades de desviación típica (Serlin y Lapsley, 1993). No obstante, la diferencia bruta entre las medias resulta de por sí insustituible cuando tenemos variables dimensionales, con unidades de medida con significado sustantivo. En el caso de las correlaciones, los coeficientes de regresión constituyen también buenos indicadores de la magnitud de los efectos. Adicionalmente, un buen reporte de investigación debe proporcionar intervalos de confianza para la magnitud de los efectos (véase Wilkinson y la Comisión sobre inferencia estadística, 1999), aun cuando esta práctica resulte hoy por hoy infrecuente. Esta información sobre la magnitud de los efectos devuelve el foco del análisis y el juicio sobre la utilidad de las intervenciones al plano sustantivo y por tanto a la competencia y responsabilidad del investigador. Resumiendo muy sucintamente la discusión anterior, la pregunta central por responder al leer un reporte de resultados es la siguiente:

¿Qué tan grandes fueron los efectos encontrados?

Una última preocupación que debe tener todo lector curioso de un artículo de investigación es la generalidad de los resultados obtenidos. ¿Hasta qué punto son estos resultados replicables? Una muy difundida malinterpretación de las pruebas de significancia es pensar que éstas nos hablan directamente de la replicabilidad. De hecho, el rechazar la hipótesis nula a un nivel de significancia de 0.05 se interpreta erróneamente como que los resultados obtenidos se replicarán en un 95% de los casos. Esto es un error garrafal. La probabilidad de 0.05 no hace referencia a la frecuencia de resultados distintos a los de la investigación, porque no hace referencia a ninguna frecuencia. El nivel de significancia no es una probabilidad absoluta, sino una probabilidad condicional. ¿Condicional

a qué?, cabría preguntar. Pues condicional a la hipótesis nula. El famoso 0.05 es la probabilidad de obtener unos resultados como los observados en la investigación en cuestión *si la hipótesis nula fuese verdadera*. Para utilizar la expresión de Box (1979) no es un enunciado del tipo «La luna está hecha de queso», sino más bien «Si la luna estuviese hecha de queso, sería un lugar ideal para los ratones». Como cualquier texto de lógica nos diría, el encontrar a la luna plagada de ratones de ninguna forma probaría que estuviese hecha de queso. Erróneamente tratamos de interpretar las pruebas de significancia como la probabilidad de la hipótesis nula en vista de nuestros datos cuando la información que realmente nos dan es la probabilidad de nuestros datos dada la hipótesis nula.

Tal y como bien lo afirmó Sidman (1960), pero también Fisher (1935) veinticinco años antes que él, la generalidad sólo puede establecerse mediante la réplica. Ni siquiera el tamaño muestral es un sustituto de la réplica independiente, ya que como apuntaba Nunnally, (1960) una muestra de seiscientos sujetos sólo constituye una instancia de los procedimientos, ambientes y condiciones utilizadas. De hecho, todas las consideraciones anteriores con respecto a la significancia estadística, el tamaño muestral y la magnitud de los efectos deben hacernos no sólo más exigentes con la réplica de los hallazgos experimentales, buscando consistencias más allá del rechazo o no de la hipótesis nula sino también con respecto a la magnitud de las diferencias y sus intervalos de confianza. Esto solamente es posible mediante la contrastación de los resultados con las investigaciones anteriores. Resumiendo esto en nuestra última pregunta, esta podría formularse como sigue:

¿Hasta qué punto son consistentes los resultados con la investigación anterior?

A lo largo de estas páginas hemos tratado de identificar una serie de aspectos importantes por considerar al examinar reportes de investigación basados en diseños estadísticos, a la vez que procuramos aclarar algunos malentendidos comunes acerca de las pruebas de significancia. En ninguna de estas cosas pretendemos ser originales y remitimos a los lectores a los excelentes artículos que a este respecto han publicado Cohen (1990, 1994) y Wilkinson y la Comisión sobre Inferencia Estadística (1999). No obstante, pensamos que por el mismo distanciamiento entre los analistas conductuales y la literatura estadística era conveniente llamar la atención sobre estos artículos que bien pudieran haber pasado desapercibidos para una importante proporción de los miembros de nuestro «colegio invisible».

Una última aclaratoria. Buena parte de las críticas al análisis estadístico que aquí se recogen pueden ser perfectamente aplicables al análisis de los diseños conductuales, los cuales presentan a su vez dificultades propias (al respecto, véase Vargas, 1999). El poder y la magnitud de los efectos en los diseños intrasujeto son tópicos que sólo recientemente han recibido atención por parte de los «metodólogos cuantitativos», constituyendo por tanto un área potencialmente fértil de intercambio con el análisis conductual. Lo anterior, aunado a los cambios que se están operando con respecto al análisis de datos en la comunidad psicológica en general, demandan del analista conductual una competencia en el manejo de la estadística. Esperamos que como comunidad científica podamos estar a la altura de este reto.

Referencias del capítulo

- ABELSON, R. (1995). *Statistics as a Principled Argument*. Hillsdale: Lawrence Erlbaum Associates.
- ATOR, N. (1999). Statistical Inference in Behavior Analysis: Environmental Determinants. *The Behavior Analyst*, 22, 93-97.
- BINDER, C. (1993, October). Behavioral fluency: A new paradigm. *Educational Technology*, 8-14.
- BOX, G. (1979). Some Problems of Statistics and Everyday Life. *Journal of the American Statistical Association*, 74, 1-4.
- BRANCH, M. (1999). Statistical Inference in Behavior Analysis: Some things Significance Testing Does and Does Not Do. *The Behavior Analyst*, 22, 87-92.
- COHEN, J. (1988). *Statistical and Power Analysis for the Behavioral and Social Sciences*. Hillsdale: Erlbaum.
- COHEN, J. (1990). Things I have Learned (So Far). *American Psychologist*, 45, 1304-1312.
- COHEN, J. (1994). The Earth is Round ($p < 0.05$). *American Psychologist*, 49, 997-1003.
- DAVIDSON, M. (1999). Statistical Inference in Behavior Analysis: Having my cake and eating it?. *The Behavior Analyst*, 22, 99-104.

- EVANS, G., LEPORE, S. SHEWAL, B. & PALSANE, M. (1998). Chronic Residential Crowding and Children's Well-Being: An Ecological Perspective. *Child Development*, 69, 1514-1523.
- FISHER, R. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.
- GLASS, G. & STANLEY, J. (1973). *Métodos Estadísticos Aplicados a las Ciencias Sociales*. México: Prentice Hall.
- GOTTMAN, J.M. (1998). Psychology and the study of marital processes, *Annual Review of Psychology*, 49, 169-97.
- HERSEN, M. & BARLOW, D.H. (1976). *Single Case Experimental Designs: Strategies for Studying Behavior Change in the Individual*. New York: Pergamon.
- HUITEMA, B. (1986). Statistical Analysis in Single Subject Designs. En: A. Poling y R.W. Fuqua, *Research Methods in Applied Behavior Analysis*, New York: Plenum Press.
- JOHNSTON, J.M. & PENNYPACKER, H. (1990). *Readings in Strategies and Tactics of Behavioral Research*. Hillsdale: Lawrence Erlbaum Associates.
- KILLEEN, P. (1978). Stability Criteria. *Journal of the Experimental Analysis of Behavior*, 29, 17-24.
- MICHAEL, J. (1974). Statistical Reasoning for Individual Organism Research > Mixed Blessing or Curse. *Journal of Applied Behavior Analysis*, 7, 647-653.
- NUNNALLY, J. (1960). The Place of Statistics in Psychology. *Educational and Psychological Measurement*, 20, 641-650.
- ROBLES, J. (1992). *Un Modelo Estructural sobre el Rendimiento Académico en Estudiantes Universitarios*. Documento no Publicado del Centro de Investigaciones del Comportamiento, Universidad Católica Andrés Bello, Caracas, Venezuela.

- ROBBINS, R., GOSLING, S. & CRAIK, K. (1999). An Empirical Analysis of Trends in Psychology. *American Psychologist*, 54, 117-128.
- SERLIN, R. & LAPSLEY, D. (1993). Rational Appraisal of Psychological Research and the Good Enough Principle. En: G. Keren y C. Lewis (Eds.) *A Handbook of Data analysis in the Behavioral and Social Sciences*. Hillsdale: Lawrence Erlbaum Associates.
- SIDMAN, M. (1960/1998). *Tactics of Scientific Research*. Acton: Copley Publishing Group.
- SKINNER, B.F. (1950). Are Theories of Learning Necessary?, *Psychological Review*, 57, 193-216.
- SKINNER, B.F. (1956). A Case History in Scientific Method. *American Psychologist*, 11, 221-233.
- SPRINGER, B. BROWN, T. & DUNCAN, P. (1981). Current Measurement in Applied Behavior Analysis. *The Behavior Analyst*, 4, 9-31.
- TUKEY, J. (1969). Analyzing Data: Sanctification or Detective Work. *American Psychologist*, 24, 83-91.
- VARGAS, C. (1999). Consideraciones Sobre el Análisis Estadístico en los Diseños Conductuales. En: A. Rangel, C. Silva, & L. Sánchez (Eds.) *Contribuciones a la Psicología en Venezuela: Vol III, (250-261)*. Caracas: Facultad de Humanidades y Educación, Universidad Central de Venezuela.
- WILKINSON, L. & LA COMISIÓN SOBRE INFERENCIA ESTADÍSTICA (1999). Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist*, 54, 594-604.

Referencias

- ACHENBACH, T. (1978). *Research in Developmental Psychology: Concepts, Strategies, Methods*. New York: The Free Press, A division of MacMillan Publishing Co. Inc.
- ALFONZO, Y., BERMÚDEZ, N. & CALLEJA, F. (1990). *Efectos de dos tipos de contingencias orientadas al grupo sobre la conducta creativa*. Trabajo Especial de Grado para optar al título de Licenciado en Psicología, Mención Clínica, Escuela de Psicología, UCV.
- ANGUERA, M. (1983). *Manual de Prácticas de Observación*. México: Trillas.
- ARAGÓN, L. & NAVA, C. (1992). Las gráficas en Psicología. En: A. Silva (ed) *Métodos cuantitativos en Psicología*, (127-164), México: Trillas.
- ARNAU, J. (1979). *Psicología experimental. Un enfoque metodológico*. México: Trillas.
- ARNAU, J. (1981). *Diseños experimentales en Psicología y Educación*. Vol. 2, México: Trillas.
- BAER, D., WOLF, M. & RISLEY, T. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 20 (4), 313-327.
- BARLOW, D. & HERSEN, M. (1988). *Diseños experimentales de caso único*. Barcelona: Martínez Roca.

- GUEVARA, M. T. (1994). *La observación como estrategia básica para el estudio de la conducta infantil*. Caracas: Fondo Editorial Facultad de Humanidades y Educación.
- GUEVARA, M., BARREAT, Y. & ROSSO, I. (1984). *Validación empírica de destrezas componentes en un curriculum de motricidad fina para niños pre-escolares*. XXXIV Convención Anual de AsoVAC, Cumaná.
- GUEVARA, M., DEMBO, M., ROSSO, I., BADELL, M. & GARCIA, M. (1985). *Desarrollo y validación de un curriculum de discriminación sensorial para niños pre-escolares*. XXXV Convención Anual de AsoVAC, Mérida.
- GUEVARA, M., RANGEL, A., LACASELLA, R., GOMEZ, K. & LADRON DE GUEVARA, I. (1987). *Validación de un curriculum en el área de número en la enseñanza pre-escolar*. XXXVII Convención Anual de AsoVAC, Maracaibo.
- GUEVARA, M., RANGEL, A., LACASELLA, R., GOMEZ, K. & KHAN, M. (1988). *Validación de un curriculum de número para la enseñanza pre-escolar*. XXXVIII Convención Anual de AsoVAC, Maracay.
- GUEVARA, M., VILLALBA, D., RANGEL, A., LACASELLA, R., & GIRÓN, S. (1987). *Seguimiento de un grupo de niños que pasan de un aula preescolar experimental a otros ambientes educativos*. XXXVII Convención Anual de AsoVAC, Maracaibo.
- HALL, R. (1974). *Managing behavior (Part 1). Behavior Modification: The measurement of Kansas*. Kansas: H & Enterprises, Inc.
- HALL, R. & FOX, R. (1977). *Changing-Criterion Design: An alternative applied behavior analysis procedure*. En: B. Etzel, J. LeBlanc & D. Baer (eds) *New developments in behavioral research (Theory, method*

- and application)(151-166). New Jersey: Lawrence Erlbaum Associates, Publishers.
- HALL, R., FOX, R., WILLARD, D., GOLDSMITH, L., EMERSON, M., OWEN, M., DAVIS, F. & PORCIA, E. (1971). The teacher as observer and experimenter in the modification of disputing and talking-out behaviors. *Journal of Applied Behavior Analysis*, 4, 141-149.
- HARROP, A. & DANIELS, M. (1986). Methods of time sampling: A reappraisal of momentary time sampling and partial interval recording. *Journal of Applied Behavior Analysis*, 19, 73-77.
- HARTMANN, D. (1988). Estrategias de Evaluación. En: D. Barlow y M. Hersen *Diseños experimentales de caso único* (107-131), Barcelona: Martínez Roca.
- HARTMANN, D. & HALL, R. (1976). The changing criterion design. *Journal of Applied Behavior Analysis*, 9 (4), 527-532.
- HARTMANN, D. & WOOD, D. (1990). Observational methods. En: A. Bellack, M. Hersen & A. Kazdin (eds) *International Handbook of Behavior Modification and Therapy* (107-138), New York and London: Plenum Press.
- HAWKINS, R. & DOTSON, V. (1975). Reliability scores that delude: And Alice wonderland trip through the misleading characteristic interobserver agreement scores in interval recording. En: E. Ramp & J. Semb (eds) *Behavior Analysis: Areas of Research and Application* (359-376). Englewood Cliffs, New Jersey: Prentice Hall Inc.
- HAWKINS, R. & DOBES, R. (1977). Behavioral definitions in applied behavior analysis: Explicit or implicit. En: B. Etzel, J. LeBlanc & D. Baer (eds) *New Developments in behavioral research: Theory, method and*

- application. In honor of Sidney Bijou (167-188), Hillsdale, New Jersey: Lawrence Erlbaum.
- HAYNES, R. & ZANDER, A. (1978). Observación de la conducta de grupo. En: L. Festinger y D. Katz (eds) *Los métodos de investigación en las ciencias sociales* (353-385), Buenos Aires: Paidós.
- HENDRICKSON, J., STRAIN, P., TREMBLAY, A. & SHORES, R. (1982). Interactions of behaviorally handicapped children: Functional effects of peer social interactions. *Behavior Modification*, 6, 323-353.
- HIGGINS, A. & BAER, D. (1989). Interaction effects in multielement designs: Inevitable, Desirable and Ignorable. *Journal of Applied Behavior Analysis*, 22 (1), 57-69.
- JACKSON, D., DELLA-PIANA, G. & SLOANE, H. (1973). *How to establish a Behavior Observation System: A self-instructional Program*. University of Utah: Bureau of Educational Research.
- JOHNSTON, J. (1979). On the relation between generalization and generality. *The Behavior Analyst*, Fall, p.1-6.
- JOHNSTON, J. (1972). Punishment of human behavior. *American Psychologist*, 27, 1033-1054.
- JOHNSTON, J. & PENNYPACKER, H. (1980). *Strategies and tactics of Human Behavioral Research*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- JONES, R., KAZDIN, A. & HANEY, J. (1981). Social Validation and training of emergency fire safety skills for potential injury prevention and life saving. *Journal of Applied Behavior Analysis*, 14, 249-250.
- KAZDIN, A. (1977). Artifact, bias and complexity of assessment: The ABC'S of reliability. *Journal of Applied Behavior Analysis*, 10, 141-150.

- KAZDIN, A. (1979). Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis*, 12, 713-724.
- KAZDIN, A. (1980). *Research design in clinical psychology*. New York: Harper and Row, Publishers.
- KAZDIN, A. & HARTMANN, D. (1978). The Simultaneous-Treatment Design. *Behavior Therapy*, 9, 912-922.
- KELLY, M. (1977). A review of the observational data-collection and reliability procedures reported in The Journal of Applied Behavior Analysis. *Journal of Applied Behavior Analysis*, 10, 97-101.
- KERLINGER, F. (1988). *Investigación del Comportamiento*. México: McGraw Hill.
- KRATOCHWILL, T. (1978). *Single subject research (Strategies for evaluating change)*. New York: Academic Press.
- LACASELLA, R. (1987). *Efectos de dos tipos de contingencias de reforzamiento sobre la conducta creativa: Un estudio de Validación Social*. Trabajo de Grado para optar al título de Magister Scientiarum en Psicología, Mención Análisis Experimental de la Conducta. Universidad Central de Venezuela, Caracas.
- LACASELLA, R. (1996). *Análisis de interacciones verbales en niños preescolares: Una réplica sistemática*. Trabajo presentado para optar al ascenso correspondiente a la categoría de Profesor Asistente en el escalafón del personal docente de la Universidad Central de Venezuela.
- LACASELLA, R., MORENO, R., RANGEL, A. & VILLALBA, D. (1991). *Diseño y validación de un currículum para el desarrollo de destrezas de autocuidado*. XLI Convención Anual de AsoVAC, Maracaibo.

- LADRON DE GUEVARA, I., RANGEL, A., GUEVARA, M. & MORENO, R. (1990). *Validación de un currículum en el área de lenguaje para la enseñanza pre-escolar*. XL Convención Anual de AsoVAC, Cumaná.
- LLINDIS, V. & MESSINA, P. (1998). *Adquisición y transferencia de expresiones temporales en niños preescolares*. Trabajo Especial de Grado para optar al título de Licenciado en Psicología, Mención Clínica. Universidad Central de Venezuela, Caracas.
- MAGNUSSON, D. (1977). *Teoría de los tests*. México:Trillas.
- MANSELL, J. (1985). Time sampling and measurement error. The effect of interval length and sampling pattern. *Journal of Behavior Therapy and Experimental Psychiatry*, 16, 245-251.
- MARTÍNEZ, J. & MUCCI, L. (1997). *Variables del contexto y su influencia en la adquisición de repertorios lingüísticos en niños preescolares*. Trabajo Especial de Grado para optar al título de Licenciado en Psicología, Mención Clínica. Universidad Central de Venezuela, Caracas.
- NELSON, R. & HAYES, S. (1981). Nature of Behavioral assessment. En: M. Hersen & A. Bellack (eds). *Behavioral Assessment. A practical handbook*, New York: Pergamon Press.
- OLLENDICK, T. (1981). Self-monitoring and self-administered overcorrection. *Behavior Modification*, 5, 75-84.
- OLLENDICK, T., SHAPIRO, E. & BARRETT, R. (1981). Reducing stereotypic behaviors: An analysis of treatment procedures utilizing and alternating treatments design. *Behavior Therapy*, 12, 570-577.

- ORTH-LOPES, L. (s/f). *Una revisión de diseños experimentales*. Material mimeografiado-Draft.
- PARSONSON, B. & BAER, D. (1978). The analysis and presentation of graphic data. En: T. Kratochwill (ed) *Single subject research. Strategies for evaluating change*, New York: Academic Press.
- POWELL, J., MARTINDALE, A. & KULP, S. (1975). An evaluation of time-sample measures of behavior. *Journal of Applied Behavior Analysis*, 8, 463-469.
- RANGEL, A. (1990). *La conducta de clasificación: dos estrategias de enseñanza*. Trabajo de Grado para optar al título de Magister Scientiarum en Psicología, Mención Análisis Experimental de la Conducta, UCV.
- RIBES, E. (1980). Teoría de la Conducta. En E. Ribes, C. Fernández, B. Rueda, M. Talento y F. López (eds). *Enseñanza, ejercicio e investigación de la Psicología. Un modelo integral*. México:Trillas.
- RIBES, E. (1990). *Problemas conceptuales en el Análisis del Comportamiento Humano*. México:Trillas.
- RISLEY, T. & WOLF, M. (1972). Strategies for analyzing behavioral change over time. En: J. Nesselroade y H. Reese (eds) *Life-span developmental psychology: Methodological issues (175-183)*, New York: Academic Press.
- ROMANCZYK, R., KENT, R., DIAMENT, C. & O'LEARY, K. (1973). Measuring the reliability of observational data: a reactive process. *Journal of Applied Behavior Analysis*, 6, 175-184.
- RUGGLES, T. & LeBLANC, J. (1979). *Observation methods in Applied Behavior Analysis*. Kansas Research Institute for Early Childhood Education of the Handicapped.

- SIDMAN, M. (1975). *Tácticas de investigación científica*. Barcelona: Fontanella.
- SILVA, A. (1992). Diseños de investigación con N=1 o de series de tiempo. En: A. Silva (ed) *Métodos cuantitativos en Psicología* (45-70). México: Trillas.
- SILVA, A. & NAVA, C. (1992). Confiabilidad y validez en Psicología. En: A. Silva (ed). *Métodos cuantitativos en Psicología* (71-98), México: Trillas.
- SKINNER, B. F. (1971). *Ciencia y Conducta Humana*. Barcelona: Fontanella.
- SKINNER, B. F. (1975). Tendencias actuales en Psicología Experimental. En: B.F. Skinner *Registro Acumulativo*, (329-350). Barcelona: Fontanella.
- SPRINGER, B., BROWN, T. & DUNCAN, P. (1981). Current measurement in Applied Behavior Analysis. *The Behavior Analyst*, 4, 19-31.
- STEFANILE, A. (1999). *Efectividad de un programa de entrenamiento para facilitar la transferencia de aprendizaje de la seriación*. Trabajo Especial de Grado para optar al título de Licenciado en Psicología, Mención Clínica. Universidad Central de Venezuela, Caracas.
- SULZER-AZAROFF, B. & MAYER, G. (1983). *Procedimientos del Análisis Conductual Aplicado con niños y jóvenes*. México: Trillas.
- THOMPSON, C. & GRABOWSKY, J. (1972). *Reinforcement schedules and multioperant analysis*. New York: Appleton Century Crofts.
- THOMSON, C., HOLMBERG, M. & BAER, D. (1974). A brief report on a comparison of time-sampling procedures. *Journal of Applied Behavior Analysis*, 7, 623-626.
- ULMAN, J. & SULZER-AZAROFF, B. (1975). Multielement baseline design in educational research.

- En: E. Ramp & G. Semb (eds) *Behavior Analysis. Areas of research and application*. New Jersey: Prentice Hall, Inc. Englewood Cliffs.
- UNDERWOOD, B. (1972). *Psicología experimental*. México: Trillas.
- VILLALBA, D. (1983). *Efecto de las instrucciones sobre la adquisición de una discriminación visual en niños preescolares*. Trabajo de Grado para optar al título de Magister Scientiarum en Psicología, Mención Análisis Experimental de la Conducta, UCV.
- WALKER, H. & BUCKLEY, N. (1968). The use of positive reinforcement in conditioning attending behavior. *Journal of Applied Behavior Analysis*, 1, 245-250.