

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

E.N.E.P. IZTACALA

PSICOLOGIA



U.N.A.M. CAMPUS
IZTACALA

VALIDACION DE CONSTRUCTOS: LA CONDUCTA
HOMOSEXUAL COMO EJEMPLO.

001
31921
N1
1984-3

TESIS PARA OBTENER EL TITULO DE:

LICENCIADO EN PSICOLOGIA.

PRESENTA:

CARLOS N. NAVA QUIROZ.

1984



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

A mis padres:

Amalia y Vicente.

A mis hermanos.

Paty, Luis, Kochitl.

A mi sobrino.

Adan.

A Laura y Arturo.

Por ser buenos amigos,

por toda la ayuda.

A:

Perla, Enrique, Javier, Chucho, José (el muerto), Berna,
Pepe V., Arturo, Octavio, Alex, Victor, Felipe...

Indice

Introducción	IZT. 1000310	I-III
I.- Los constructos.		I
2.- Indicadores de la homosexualidad.		24
3.- Validez de los instrumentos.		33
4.- Análisis de los resultados terapéuticos.		61
5.- Conclusión.		75
Bibliografía.		83

Introducción

La evaluación de la terapia del comportamiento ha sido enfocada desde distintas perspectivas, tales como los principios teóricos que respaldan algunas técnicas específicas, mediante la determinación de la efectividad terapéutica para resolver algún problema particular, o bien mediante críticas metodológicas.

Un aspecto al cual se le ha prestado poca atención, en la evaluación de la terapia, es la relevancia de las medidas, es decir, ¿realmente se está midiendo lo que se quiere? ¿Las variables escogidas son las más relevantes del fenómeno de interés?

Estas preguntas están estrechamente relacionadas con aspectos de medición y teóricos. Dichos aspectos pueden enmarcarse dentro de lo que se conoce como validez de constructo. De esta forma, la congruencia y validez existente entre los conceptos teóricos y la medición es el interés básico de la validación de constructos.

Por otra parte, existe además de la validación de constructos, dos tipos adicionales de validez, la primera se denomina validez de contenido y se enfoca a estudiar si la variable o variables escogidas, de entre todo el universo de variables que componen a un constructo, son las más representativas de dicho constructo; y la segunda se conoce como validez de criterio, que en términos generales se refiere a los puntajes de una prueba, los cuales son comparados con

otra variable (criterio) que mide el atributo deseado.

Sin embargo, la validez de constructo tiene una importancia capital ya que, primero, se enfoca directamente sobre la teoría, y segundo, por que en el momento en que se llega a validar algún constructo, automáticamente se está implicando que la variable elegida o variables elegidas fueron las apropiadas (validez de contenido) y por consiguiente puede servir para predecir alguna ejecución futura de los sujetos (validez de criterio).

Por otro lado, existen cuando menos tres formas diferentes para validar un constructo, la primera es mediante análisis numéricos (e.g. mediante correlaciones, análisis factorial, etc.); segundo, mediante métodos experimentales (se someten a la comprobación experimental las hipótesis de un constructo o teoría, que es lo que más comunmente se hace en la psicología conductual) y tercero, mediante procedimientos lógicos (e.g. el recopilar, integrar, y analizar la información pertinente de un constructo, de manera tal que dicho análisis permita llegar a conclusiones que indiquen el grado de validez del constructo en cuestión).

A partir del análisis de los anteriores aspectos, se decidió como objetivo de este trabajo, analizar la posible validez de constructo de los conceptos utilizados en la medición de la conducta homosexual, así como también, la congruencia de los instrumentos usados en la medición de dicho constructo,

III

con las dimensiones de interés. Por último, estimar los logros terapéuticos en la investigación de dicha conducta, a partir de tres criterios: a) criterio estadístico, b) criterio terapéutico, y c) criterio de seguimiento.

Para llevar a cabo el cumplimiento de dicho objetivo se decidió adoptar el procedimiento de validación de constructo por medios lógicos. La razón por la que se escogió este método, por un lado, es porque se facilita el integrar los estudios tanto experimentales como numéricos y por otro, el que permite poner de relieve la congruencia de los instrumentos usados en la medición de dicho constructo con las dimensiones de interés y además, si se llega a validar el constructo homosexualidad se validarán automáticamente los otros dos tipos de validez (de contenido y de criterio), por lo cual se propone la siguiente estructura:

En el primer capítulo se expondrá lo que es la validez de constructo y algunos conceptos inherentes a éste, en el segundo capítulo se centrará el constructo de interés a analizar (conducta homosexual) y se identificarán los indicadores de interés, en el tercer capítulo se intentará establecer la posible validez de los instrumentos de evaluación, en el cuarto capítulo se evaluarán los logros terapéuticos en el comportamiento homosexual, finalmente, en el último capítulo (conclusiones), se sustentará una discusión de la posible validez del constructo conducta homosexual y se sugerirán líneas de acción general.

LOS CONSTRUCTOS

CAPITULO I

En este primer capítulo se intentará presentar un bosquejo general de la validez de constructo, así como algunos conceptos inherentes a éste, para lo cual se definirá lo que es la validez de constructo y su importancia, enmarcando las diferentes formas de validación de constructos.

Enseguida se expondrán los elementos que están implicados en un constructo (indicadores e instrumentos de medición) la validez del instrumento de medición y su relación con la validez de constructo, finalmente se tratarán otros conceptos que están estrechamente vinculados con la validación de constructo.

Los Constructos

La validación de constructos ha sido por mucho tiempo dominio casi exclusivo de la teoría psicométrica; sin embargo, esto no quiere decir que en la psicología conductual no pueda llevarse a cabo la validación de éstos, ya que, "la validez de constructos también es útil en las técnicas de evaluación conductual, ya que el uso de constructos no tiene porque limitarse a la psicología de los rasgos, puesto que su significado es mucho más amplio." (Ballesteros y Carrobles, 1981, pag. 183.)

Profundizando, la validez de constructo puede ser entendida mediante una pregunta: ¿El evento identificado corresponde realmente al dominio que se cree ó pertenece a otro? (Nunnally, 1967; Ballesteros y Carrobles, 1981; Kerlinger, 1973; Haynes, 1978.)

Concretamente, la anterior pregunta puede ser comprendida con mayor claridad a partir de la siguiente exposición: En principio un constructo puede entenderse como cualquier fenómeno o variable "siempre que clasificamos situaciones, personas ó respuestas, se están utilizando constructos. Una opinión similar la encontramos en Johnson y Bolstad (1973) ésta es cuando hablamos de categorías amplias de comportamiento, tales como conductas apropiadas versus conductas no apropiadas ...estamos poniendo nombres a dimensiones amplias de conducta. en ese nivel estamos hablando de constructos, nos guste admitirlo o no..." en Ballesteros y Carrobles, (1981, pag. 183.)

La cita anterior enmarca lo que es el paso inicial, esto es, la conceptualización del fenómeno de interés (ubicar el constructo).

El siguiente paso a tratar es lo que se va a entender por validación de constructo. Concretamente, se dice que un constructo se valida cuando se logra demostrar que las variables elegidas como relevantes, pertenecen realmente al dominio del constructo en cuestión. Por ejemplo, Masters y Johnson (1966) postularon a la erección peneana como parte del constructo respuesta sexual, y mediante investigaciones experimentales demostraron que tal erección sí pertenecía a dicho constructo, esto es, estos autores demostraron que cuando un estímulo actúa como sexual provoca cambios a nivel fisiológico, tales como: vasocongestión superficial y profunda...y la erección se da como resultado de esa estimulación y de la vasocongestión.

Hay que aclarar que la erección como tal puede darse sin que necesariamente exista un estímulo (e.g. priapismo), pero aquí lo importante es la demostración de que la erección sí pertenece al constructo respuesta sexual.

De lo anterior se deduce que, cuando se habla de validación de constructos, se está haciendo referencia a las variables que efectivamente pertenecen al constructo de interés y no se están tomando como tales a aquellas que pertenecen a otro dominio no contemplado.

Ahora bien, particularmente en la psicología conductual la validación de constructos se ha hecho regularmente mediante métodos experimentales (Ballesteros y Carrobles, 1981), y en mucha menor medida por medios lógicos y estadísticos, que son dos formas también útiles en la validación de dichos constructos.

Una vez que se ha validado el constructo, por cualquier método, su importancia radica esencialmente en la confianza que da al investigador para la deducción y uso de las hipótesis que de ese constructo se deriven, es decir, al validar un constructo se estará interesado directamente en el valor de la teoría que del constructo surja y además por las investigaciones científicas de carácter empírico que suponen la comprobación de las posibles relaciones. (Ballesteros y Carrobles, 1981; Haynes, 1978; Kerlinger, 1973.)

En resumen, se dice que la validez de constructo se interesa por la validez de la teoría.

Validez de contenido.

Sin embargo, la sola validación de constructo no garantiza de ninguna manera que al someterse a estudio dicho fenómeno, se hayan escogido las variables más representativas de éste (validez de contenido). Es decir, de entre todos los elementos que componen un fenómeno en particular, aquellos que han sido escogidos para estudiar dicho evento constituyen

realmente la mejor representación del fenómeno que se pretende medir? (Nunnally, 1967; Ballesteros y Carrobles, 1981; Kerlinger, 1973; Haynes, 1978.)

De aquí se desprende que si por un lado, la validez de constructo permite analizar y validar la teoría subyacente al constructo mismo, por otro lado, la validez de contenido permite tener un parámetro de comparación de las operaciones del investigador, puesto que si esta persona elige el total de variables apropiadas de un constructo para su estudio, tendrá mayor validez de contenido que si sus variables elegidas no son representativas del fenómeno.

Operativización de indicadores.

Una vez que se ha validado el constructo en estudio, y se está seguro de que las variables elegidas para ser medidas representan válidamente al constructo, es necesario medir dichas variables de interés y diseñar los instrumentos que se van a utilizar en su medición.

En lo que respecta a la definición de las variables, ésta ha sido citada de diferentes maneras, Ballesteros y Carrobles (1981, pag. 184.) afirman, "En el momento de la construcción de un instrumento de evaluación, es necesario establecer las especificaciones de los objetivos del mismo, como ya hemos señalado antes, el dominio conductual al que han de referirse los elementos del instrumento de evaluación. En esta

primera fase, lo que se hace realmente es una elección entre diferentes formas de operativizar el dominio de conductas en las que estamos interesados. Cuando hacemos esto, la tarea es similar a lo que Kerlinger (1973) denomina definición operativa del constructo." Por otro lado Kerlinger (1973, pag. 303.) opina, "Afirmamos que medimos objetos, pero eso no es del todo cierto. Lo que medimos son sus propiedades o características. No obstante, esta afirmación tampoco es totalmente verdadera. Lo que medimos, son los indicadores de las propiedades, por lo tanto al decir que medimos objetos lo que realmente queremos dar a entender es que medimos los indicadores de sus propiedades... Los indicadores en base a los cuales se infieren las propiedades se precisan mediante definiciones operacionales que especifican las actividades u "operaciones" necesarias para medir las variables o las construcciones hipotéticas."

En concreto, lo que se hace al definir las variables que se consideran como relevantes, es operativizar los indicadores del constructo. A manera de ejemplo, cuando a un niño se le evalúan las conductas ecoicas, lo que se está haciendo es medir el indicador "conducta ecoica" como parte del constructo denominado conducta verbal. Esto es, la conducta verbal comprende un amplio rango de conductas y además una serie de postulaciones teóricas que son las que forman el constructo "conducta verbal" y el repertorio ecoico es solo un indicador de dicho constructo.

Instrumentos de evaluación y su validez.

Una vez que se han operativizado los indicadores del constructo, el siguiente paso es diseñar un instrumento que mida estos mismos, dicho instrumento estará ligado estrechamente a los indicadores definidos, y así mismo fuertemente vinculado al sistema de constructo dentro del cual fué desarrollado o construido, (Haynes, 1978; Kerlinger, 1973.)

La importancia de los indicadores y del instrumento de evaluación, radica esencialmente en la utilidad que estos factores aportan para deducir hipótesis y facilitar la confirmación de teorías que de los constructos se deriven, esto es, la validación del instrumento. Expresado formalmente, la validez del instrumento indicará el grado en que una herramienta de evaluación es una medida adecuada del constructo y en qué medida las hipótesis emanadas del constructo pueden confirmarse mediante el uso del instrumento en cuestión.

Relación entre validez de constructo, de contenido y del instrumento.

La validez del instrumento guarda una estrecha relación con la validez de constructo, aunque hay que aclarar que el hecho de que un instrumento posea validez no implica necesariamente que el constructo sea válido también.

El caso contrario sería cuando un instrumento de evalua-

ción es diseñado para medir un constructo y la validación de ambos se da a la par; en este caso se está tratando con una estrecha relación entre lo que es la validez de constructo y la validez del instrumento.

Ahora bien, si se quiere ver la relación que existe entre la validez de constructo y la del instrumento, se puede pensar que los constructos son dimensiones amplias de comportamiento, las cuales son operativizadas en los indicadores, y que además, los instrumentos de evaluación y su validez, están relacionados íntimamente con aquellos últimos.

A partir de esto se deduce que con la validación del instrumento de evaluación, por lo tanto, se está validando la teoría subyacente al constructo en el cual fué diseñado y es utilizado dicho instrumento.

Así, se tiene que el propósito de la validación de constructos no es otra cosa que el validar la teoría subyacente al sistema de medición, así como al mismo sistema, la validación de un constructo es muy importante, más no puede ocurrir independientemente del instrumento que mide.

En la validez de constructo se está interesado en como un instrumento de evaluación mide el constructo seleccionado y si la extensión de las hipótesis derivadas del constructo son confirmadas por el instrumento de evaluación. Por otra parte, esta validez de constructo se da independientemente de que se haga referencia a una teoría de los rasgos o a una

teoría de la conducta.

Así, cuando se valida un constructo lo que se está haciendo, en muchas ocasiones, es validar a la vez, los instrumentos de medición, esto es, un instrumento de evaluación estará inherentemente ligado al sistema de constructo dentro del cual fué construído. La implicación más significativa del instrumento de evaluación es que un instrumento de medición no puede ser más válido que el constructo en el cual fué diseñado para medir. Por ejemplo, en el constructo eyaculación prematura, en el pasado se pensó que todo el problema radicaba en el tiempo que tardaba en eyacular el hombre. En este caso el instrumento de medición -reloj- tenía un alto grado de validez. Posteriormente se comprobó que el tiempo, aún cuando era una variable importante, no podía explicar ni establecer criterios que definieran de manera precisa lo que era la eyaculación prematura; por el contrario, se consideraron otros factores que no incluían tiempo de eyaculación, como frecuencia de orgasmos, reporte de la compañera, etc.

Nat Lehrman (1979) menciona como muchas de las definiciones dependen de tiempos arbitrarios; por ejemplo, si el hombre resiste un minuto por cronómetro se dice que es un eyaculador precoz, además agrega como Masters y Johnson rechazan estos criterios y menciona como una adecuada estimulación previa es suficiente para que la mayoría de las mujeres puedan satisfacerse en este lapso, lo que significa que sus hom-

bres no son eyaculadores precoces, aunque podrían serlo con otras mujeres. Por lo que la definición de Masters y Johnson es relativa; si el hombre no puede llevar a su pareja al orgasmo cuando menos en el 50% de los casos, independientemente de su rendimiento cronométrico, es un eyaculador precoz.

Este es un ejemplo en el que la validez de un instrumento tiene poco que ver con la validez del constructo, puesto que como se observa en el ejemplo, existen otras fuentes de variación en el constructo "eyaculación precoz" como son la frecuencia, los orgasmos femeninos, los reportes de la compañera, etc. que no son tomados en cuenta.

La identificación de estas fuentes de variación, como parte constituyente del constructo, implica una concepción teórica diferente, ya que la eyaculación precoz no solo debe verse de manera unilateral, sino también en relación a los efectos que produce en la compañera del eyaculador precoz. En una sociedad donde se pensara que las mujeres solo sirven de recipiente de los espermias del hombre, probablemente nunca hubieran sido identificadas estas fuentes de variación como parte constituyente del constructo, ya que la finalidad de las relaciones sexuales estarían limitadas, en esta sociedad hipotética, a la sola eyaculación del macho dentro de la hembra.

Por otro lado, en lo que respecta al criterio de éxito; la influencia de esta aproximación teórica al problema de la

eyaculación precoz, muy probablemente guiaría a considerar resuelto el problema cuando se retardara la eyaculación lo suficiente como para lograr la introducción.

Sin embargo, conforme a la concepción teórica de que las relaciones sexuales en una pareja debe llegar a la satisfacción mutua de ambos miembros, el constructo "eyaculación precoz" no se limita solo a un criterio temporal sino que se identifica otro elemento relevante al fenómeno mismo, como es los orgasmos producidos a la compañera. Por ende el criterio de éxito no se limitaría a la sola demora de la eyaculación, sino también a la forma en que se aporta a la satisfacción sexual mutua.

Es por eso que a partir de esta última concepción teórica, el hecho de que el investigador solo tome como índice a medir la demora temporal como indicador de la eyaculación precoz, sería cuestionado en base a la validación de contenido de las fuentes de variación que el investigador tomó para estudiar el contenido del constructo.

En lo que respecta al instrumento, su validez queda confirmada con el hecho de que se utiliza un instrumento que mide el paso del tiempo.

De las anteriores ideas se desprende en conclusión, que el solo hecho de encontrar validez en el instrumento de medición, por si solo no garantiza la validez de contenido y mucho menos la de constructo.

II

En resumen, si se toma en cuenta que en la psicología conductual existen categorías de conductas, que de acuerdo a lo expuesto, no son otra cosa que constructos, y que éstos (e.g. excitación sexual) han definido sus variables y operativizado sus indicadores (e.g. erección peneana) y que cada uno de éstos ha creado a su vez su propio sistema de medición (pletismografía peneana). Es entonces que se deduce que en la medida en que cada uno de estos constructos ha logrado validarse es que se ha conseguido obtener avances importantes en la psicología conductual. Lathrop (1979) afirma que si todos los posibles constructos hubiesen sido completamente validados, muchos de los problemas comunes de la psicología estarían resueltos.

Confiabilidad, validez e isomorfismo del instrumento de medición.

Confiabilidad.

Una vez expuesto lo que es la validación de constructos, de contenido, del instrumento y la relación que existe entre éstos tres tipos de validez, queda por mencionar algunos puntos que están inherentemente ligados a estos tres conceptos.

Estos puntos se derivan de la siguiente pregunta: ¿Es válido, confiable e isomórfico el instrumento que se está utilizando para medir los indicadores del constructo de interés?

Antes de comenzar a abordar estos puntos, es importante

aclarar lo que se entiende por instrumento de medición.

Al hacer referencia a "instrumento de medición", se entenderá a este como cualquier tipo de aparato (contadores, relojes, pletismógrafo, galvanómetro, etc.) ó procedimiento (entrevista, encuestas, registro conductual, etc.) que proporcionen datos acerca de la conducta o evento de interés.

Una vez aclarado esto, el siguiente paso es centrarse en la pregunta inicial. Así, Ballestros y Carrobbles (1981, pag. 165) afirman que: "los teóricos de la medida denominan fiabilidad a la consistencia con que mide un instrumento...la seguridad es el grado de acuerdo ó correspondencia entre medidas, que deberían ser iguales si no existiesen errores de medición.", mientras que Kerlinger (1973, pag. 311) considera: "la fiabilidad del instrumento dependerá del grado en que produzca errores de medición. Dicho de otro modo, la confiabilidad se puede definir como la carencia relativa de errores de medición del instrumento. La confiabilidad es la exactitud o precisión de un instrumento de medición."

De acuerdo con estas definiciones, la carencia de errores en la medición es igual a la consistencia y la exactitud; es decir, se toman como sinónimos de confiabilidad consistencia y exactitud. Esto causa problemas que se complican aún más cuando se considera que las anteriores definiciones son el resultado de concepciones psicométricas clásicas, que di-

fiere en mucho de lo que los psicólogos conductuales concibe como confiabilidad. La definición es un problema ya que muchos investigadores usan diferentes lenguajes cuando hablan sobre la confiabilidad observacional. El significado dado a términos como "acuerdo", "exactitud", "precisión", "estabilidad", e igual que "confiabilidad" varía en el uso práctico de un autor a otro.

La confusión en la terminología es solo sintomática del fracaso por integrar las teorías psicométricas tradicionales de la confiabilidad con una teoría de la observación, (Hollenbeck, 1978).

Más adelante el mismo Hollenbeck, después de citar una revisión exhaustiva llega a la conclusión de que casi una tercera parte de los autores revisados habían fracasado en la estimación de la confiabilidad, mientras que la mayoría de ellos habían reportado confiabilidad en base a "porcentaje de acuerdo entre observadores" y los coeficientes de exactitud y estabilidad fueron muy poco utilizados. Menciona también que el acuerdo entre observadores no puede ser tomado como índice de confiabilidad mientras no se pruebe que dicho acuerdo esté respaldado por estándares previamente establecidos. Así mismo, cuando se refiere a la teoría psicométrica, pone de manifiesto que la exactitud se logra mediante métodos estadísticos, a saber: análisis de varianza, regresión lineal múltiple y los métodos no paramétricos, mientras que la esta-

bilidad se alcanza mediante la obtención de los mismos resultados usando el mismo instrumento en repetidas ocasiones.

Así, basándose en el análisis de Hollenbeck se puede concluir que hace falta más investigación sobre la confiabilidad, tanto desde el punto de vista de la teoría psicométrica como de la observación conductual. Además, es posible considerar la sugerencia hecha por el mismo Hollenbeck con respecto a que la confiabilidad debe ser definida en términos de exactitud y estabilidad conjuntamente y no solamente como exactitud ó como estabilidad de una medición, ya que ambos son necesarios para la confiabilidad; ninguno es suficiente por si solo.

Asimismo, la exactitud puede ser solo establecida por medio de análisis del error de medición o sus equivalentes no paramétricos. Similarmente, el acuerdo de observadores solo significa acuerdo y no necesariamente exactitud estabilidad o confiabilidad; sin embargo, el acuerdo de los observadores sirve como un índice de confiabilidad cuando una relación previa entre acuerdo y confiabilidad ha sido establecida para un sistema de puntajes particular. Así, los puntajes de acuerdo quizá sean el reporte estadístico final en la literatura, cuando un sistema de puntuaciones ha demostrado su confiabilidad. Esta misma opinión es compartida por Kazdin (1977), con la diferencia de que mientras Hollenbeck menciona que la exactitud en psicometría se alcanza

mediante métodos numéricos, para Kazdin, en psicología conductual, se logra mediante la predeterminación de un estándar de las mismas conductas que se registran, además agrega que, dicho estándar es determinado por otros observadores que alcanzan un consenso sobre los datos, o por la construcción de materiales observacionales, tales como videotapes o audiotapes, con muestras conductuales predeterminadas.

Tomando en cuenta el análisis hecho por Hollenbeck y Kazdin se decidió adoptar la siguiente postura:

La exactitud se entenderá como carencia relativa de errores al medir. Esta carencia de errores es establecida, en la psicometría, mediante métodos numéricos (pruebas paramétricas o no paramétricas o análisis de regresión lineal), mientras que en la psicología conductual se establecerá mediante la comparación con estándares previamente establecidos.

Y la estabilidad se comprenderá como la obtención de los mismos resultados cuando las medidas van repetidamente hechas, usando el mismo instrumento de evaluación y bajo condiciones similares.

De esto se desprende que el acuerdo entre observadores, interpretado generalmente como un índice de confiabilidad, no garantiza ni la exactitud ni la estabilidad de los datos arrojados por el instrumento de medición. Para ejemplificar, supóngase que en una escuela se han identificado ciertas conductas que definen a un niño como agresivo (e.g. golpear a

sus compañeros). Estas conductas han sido observadas repetidamente tanto en el salón de clases como en el patio de juegos, y usando los mismos instrumentos, se encontraron siempre los mismos resultados, esto es, se determinó su estabilidad, mientras que la exactitud se logró mediante la comparación de las conductas de otros niños, considerados por sus profesores como no "agresivos", y la opinión de terapeutas expertos en la materia. Es decir, se ha establecido la confiabilidad. Ahora bien, cuando a dos registradores independientes se les pide registren estas conductas y se obtiene; a) un acuerdo entre observadores alto, se dice entonces que dicho acuerdo sirve como índice de confiabilidad ya que previamente se ha determinado ésta o b) un acuerdo entre observadores bajo, se dice tan solo que dicho acuerdo es bajo y no es útil como índice de confiabilidad.

Supongase el caso contrario en el que no se ha establecido ni la exactitud, ni la estabilidad, es decir, no se ha determinado la confiabilidad. En este caso, a) un acuerdo entre observadores alto, tan solo indicaría que existe un acuerdo alto y nada más, ya que no se ha dado aún la confiabilidad; b) un acuerdo bajo entre observadores indicaría que habría que replantear el problema desde el principio.

Resumiendo, como se vé, la confiabilidad del instrumento se refiere a la exactitud y estabilidad con que una herramienta o instrumento mide algún evento, ya que por ejemplo,

si un instrumento carece de ella, los resultados arrojados por cualquier investigación no tendrán ningún sentido.

En concreto, la confiabilidad del instrumento cumple dos grandes propósitos, primero: cuando se afirma que un instrumento es confiable, se quiere decir que al utilizar dicho instrumento se tendrá exactitud y consistencia al medir, de lo cual se deriva que el investigador tendrá una alta confianza en la utilización de dicho instrumento y en el reporte de los resultados medidos por éste. Como afirma Kerlinger (1973) "la confiabilidad de una prueba es importante para su interpretación, ya que si no se tiene seguridad en el resultado de la medición de las variables, es imposible precisar con confianza las relaciones de las variables."

Segundo: la confiabilidad es una condición necesaria para que se dé la validez del instrumento, (Haynes, 1978; Kerlinger, 1973).

Esto quedará más claro una vez que se haya comprendido lo que es la validez.

Validez.

La validez del instrumento se refiere a la certeza con que el instrumento sirve para la finalidad perseguida con su aplicación. La investigación de la validez del instrumento permite saber si este es útil para medir una conducta en especial. Se debe de tener siempre presente que lo que se valida

no es el instrumento, sino, la interpretación de los datos obtenidos por medio de un procedimiento especificado. Por otro lado; Kerlinger (1973), opina que la definición más generalizada de validez es la que se expresa en la siguiente pregunta, ¿Medimos en realidad lo que nos proponemos medir?.

Dicho de otra forma, ¿los datos proporcionados por el instrumento de medición son lo que en realidad se espera o son algún otro aspecto que no se ha considerado?. Contemplado este concepto desde el punto de vista de su utilidad práctica se tiene que: los procedimientos de evaluación conductual, como la observación, entrevista, cuestionarios, auto-monitoreo, ó evaluación psicofisiológica, etc., permiten al analista conductual hipotetizar acerca de la conducta de los sujetos en el medio ambiente natural. La validez de dichas hipótesis será una función de la validez de los procedimientos de evaluación conductual empleados. (Haynes, 1978).

Así que, si se pretende hacer algún tipo de enunciado hipotético o predictivo, basado en las conductas medidas, es preciso tener presente primeramente si el instrumento tiene validez, o mejor dicho, si realmente está midiendo lo que pretende medir. Esto implica que los instrumentos de medición también tienen que ser evaluados, ya que la naturaleza empírica y práctica de los sistemas conductuales necesitan del uso de instrumentos con los que el investigador tenga un alto grado de confianza.

El analista conductual debe saber si los datos resultantes son una representación exacta de las conductas elegidas y el sujeto, y la forma en que se llega a evaluar la validez de estos instrumentos es generalmente un proceso empírico en el que se intenta evaluar la adecuación con que el instrumento mide un evento o un constructo particular o conceptos teóricos subyacentes. (Haynes, 1978), por lo que solo se podrá evaluar la validez del instrumento en la medida en que este sea coherente con las hipótesis que se derivan de su uso y que deberán de coincidir con los supuestos teóricos que del evento estudiado se tengan. Por ejemplo, supongase que un investigador diseñó una prueba de ansiedad heterosocial, la cual fue aplicada a grupos de sujetos en repetidas ocasiones y situaciones. Los resultados obtenidos fueron los siguientes: se encontró que las calificaciones de la prueba variaron muy poco cuando se hizo la comparación intrasujeto (se tuvo exactitud), y cuando las calificaciones se compararon entre sujetos, y con una prueba de reconocida utilidad, se observó que coincidían con un grado mínimo de error (se tuvo estabilidad), de lo que se concluyó que era confiable dicho instrumento. Además, se predijo con éxito y en base a las puntuaciones de la prueba, qué sujetos tendrían mayor o menor problema para relacionarse heterosocialmente, lo cual sirvió como indicador para asignarles los tratamientos más convenientes (se tuvo entonces validez).

Así pues, particularmente en la psicología conductual, la confiabilidad y validez del instrumento es de gran trascendencia si se toma en cuenta que uno de los objetivos a alcanzar por todo investigador, es el reportar con alto grado de confianza, para que así las hipótesis que se deriven de sus investigaciones sean lo más acertadas posibles, ya que el tener una medición defectuosa puede provocar que se invalide cualquier investigación, por lo cual es preciso examinar la confiabilidad y validez, en forma crítica y empírica, de todos los instrumentos de medición.

A manera de integración, Haynes (1978) opina que la evaluación de la validez de un sistema de observación conductual implica la evaluación de su utilidad, aplicación, exactitud, comprensión, generalidad, y sensibilidad. Esto incluye una serie de evaluaciones cualitativas y empíricas del contenido del instrumento, su capacidad para discriminar sobre diferentes poblaciones, su estabilidad y confiabilidad, la adecuada medición de los constructos subyacentes, y la adecuación de estos constructos.

En otras palabras, se está evaluando la validez de contenido, la validez relacionada con el criterio, y la validez de constructo.

Isomorfismo.

Un concepto más que se relaciona con la confiabilidad y

validez del instrumento, es el denominado isomorfismo.

El isomorfismo puede ser entendido como identidad o semejanza de forma. Si esto es comprendido dentro de la teoría de la medida, es posible preguntarse acerca de la relación existente entre el fenómeno que se está midiendo y el instrumento de medición empleado. Expresado formalmente se diría, ¿existe alguna correspondencia empírica entre la realidad y el instrumento de medición empleado?. Supongase por ejemplo, que un maestro desea medir la estatura de sus alumnos y lo hace mediante el uso de una balanza, la que como es sabido, posee un alto grado de confiabilidad y validez para medir peso más no estatura o longitud; por esta razón se afirma que no existe isomorfismo entre el fenómeno (estatura) y el instrumento de medición (balanza). Este ejemplo, aunque extremo, representa de manera clara y un poco burda, lo que es el isomorfismo.

Ahora bien, retomando el ejemplo anterior, el mismo maestro decide medir a una parte de sus alumnos mediante una vara de árbol, y a la otra con una cinta métrica. En ambos casos, tanto la vara como la cinta miden longitud, entonces, ¿ambos instrumentos de medición son isomórficos con respecto a la estatura?; si ambos son isomórficos entonces ¿cuál es la diferencia entre ellos? ¿es tan bueno el uno como el otro para medir?. En realidad ambos son isomórficos con respecto a la longitud, se puede hablar de tres varas de estatura, ó

cinco varas, ó tres y media varas, lo mismo que de 1.50 metros de estatura ó 1.70 metros, etc.; la diferencia radica en que la cinta métrica detectará con mayor precisión las diferencias de estatura entre los alumnos, mientras que la vara será menos sensible para reportar estos cambios; ésto es, si dos alumnos son reportados como con una y media varas de estatura, lo más probable es que la cinta métrica detecte que existe una pequeña diferencia de tamaño entre ellos (por ejemplo, uno mide 1.50 y el otro 1.52 metros de estatura). De lo anterior se concluye entonces, que la cinta métrica tiene mayor isomorfismo que la vara para medir estatura.

De esta manera, el isomorfismo se entenderá como la coherencia que exista entre el instrumento de medición y la realidad, y además, la precisión o sensibilidad que muestre dicho instrumento para detectar los cambios ocurridos en la variable de interés.

Diferencia entre conceptos.

La diferencia entre los anteriores conceptos es que la confiabilidad está referida a la exactitud y estabilidad del instrumento en sí, mientras que la validez está orientada a si la información proporcionada por el instrumento es la que en realidad se espera ó se está interpretando como cierta; siendo que pertenece a otro dominio ya sea parcial o totalmente; y el isomorfismo se relaciona con la coherencia exis-

tente entre el instrumento y la realidad, así como la precisión con que dicho instrumento detecta los cambios ocurridos en la variable de interés.

Como conclusión, es posible decir que la validación de constructos no es otra cosa que el probar la adecuación de la teoría para explicar los fenómenos de la naturaleza y la importancia de esto es trascendente, si se piensa que en la medida en que se pruebe la validez de los diferentes constructos, en esa misma medida el cuerpo teórico de la ciencia de la conducta ganará terreno en la conceptualización y explicación del comportamiento en general.

INDICADORES DE LA HOMOSEXUALIDAD

CAPITULO 2

Este capítulo pretende ser el paso inicial en la validación de constructos por medios lógicos, esto es, en esta parte se ubicará el constructo conducta homosexual a partir de la exposición, primero, de las definiciones encontradas; en segundo lugar, se considerarán únicamente las que estén enfocadas sobre el análisis experimental de la conducta, para así restringir el análisis a una orientación psicológica exclusiva, finalmente, se identificarán los indicadores que los diferentes autores han considerado como relevantes en el estudio del comportamiento homosexual, esto se hará a partir de la investigación hecha por Bancroft (1977) y mediante la revisión de diferentes artículos de tipo experimental y aplicado, que se enfocan a este tema.

El paso inicial, cuando se pretende hacer un análisis de validación de constructo, es ubicar dicho constructo. El constructo elegido para analizar es el denominado: "comportamiento homosexual", y la razón de su elección es debido a que es un tema sobre el cual se ha escrito y dicho mucho y porque existen un gran número de estudios enfocados a la parte terapéutica y de medición. Así, una breve revisión de distintas definiciones pondrá de manifiesto los puntos relevantes de coincidencia que serán los que ayuden a centrar el constructo de interés.

En principio, es necesario aclarar la génesis del término "homosexualidad", que es abordado por Kinsey, Pomeroy, y Martin, (1948), de la siguiente forma: "Durante casi un siglo ha sido aplicado el término homosexual, en relación con la conducta humana, a las relaciones sexuales francas o psíquicas entre individuos del mismo sexo. Derivada de la raíz griega homo, más que de la latina que designa al hombre, el término resalta la similitud de los individuos que practican un acto sexual."

En cuanto a las definiciones concretas se tiene que: el Diccionario Enciclopédico Básico (1974) dice: "Atracción sexual hacia individuos del mismo sexo; se le da también los nombres de pederastía ó sodomia en los hombres y de amor lesbio ó tribadismo en las mujeres." El Gran Libro de la Salud (1971) menciona: "Es la relación exclusiva con personas del mis-

mo sexo". En la Enciclopedia Familiar de la Medicina y la Salud (1976), se enuncia: "Atracción sexual hacia personas del mismo sexo, más bien que hacia personas del sexo opuesto. En la mujer se llama lesbianismo." Solomon y Patch (1976), definen a la homosexualidad como "El deseo del acto sexual con individuos del mismo sexo."

En cuanto a las definiciones conductuales se tiene a Walen, Hauserman, y Lavin (1977), que dicen "La homosexualidad se refiere a la excitación sexual ó al contacto genital erótico entre dos personas del mismo sexo". Otro autor, Coleman (1976), define a la conducta homosexual como "la conducta sexual dirigida a un miembro de nuestro mismo sexo". Y Ullman y Krasnner (1975), consideran "Homosexualidad: se refiere al deseo o conducta sexual dirigida hacia un miembro del mismo sexo".

Concretamente, centrándose en las definiciones conductuales, se encuentra que en ellas se usaron términos descriptivos, tales como "excitación sexual" (Walen y col. 1977) y "conducta sexual" (Coleman, 1976; Ullman y Krasnner, 1975), mientras que las otras definiciones no usan términos descriptivos, con excepción de Kinsey y col. (1948), que no estando considerados como autores conductuales, sí usan el término "conducta sexual".

Las definiciones sobre las cuales se enfocará el interés a partir de este momento, serán las denominadas como con-

ductuales. De esta forma, el campo de acción se restringe a aquellos estudios que estén orientados en la modificación de conducta.

Ahora bien, en la conducta homosexual se han identificado dos factores, que según algunos autores (Walen y col. 1977; Ullman y Krasnner, 1975; Coleman, 1976; Kinsey y col. 1948) aglutinan todas las fuentes de variación de la conducta homosexual, estos factores son: primero, la conducta sexual; y segundo, el hecho de que esta conducta esté dirigida hacia personas del mismo sexo. De acuerdo a esta conclusión, cabe hacer la pregunta siguiente: ¿estos dos factores identificados son los únicos relevantes para representar cabalmente al constructo "homosexualidad" (validez de constructo)?.

El contestar ésta pregunta es uno de los fines de éste trabajo. El primer paso es la ubicación del constructo "conducta homosexual", la cual puede ser entendida como la conducta sexual entre individuos del mismo sexo. Cabe aclarar que estos conceptos conllevan problemas tales como; ¿qué se va a entender por conducta sexual?, ¿qué clase de criterio se va a tomar para definir mismo sexo?. No es propósito el aclarar éstos términos aquí, sino determinar cuáles son las variables ó indicadores elegidos como importantes para analizar el constructo "homosexualidad", es decir, si se define a la homosexualidad como la conducta sexual entre individuos del mismo sexo, ¿cuáles serían entonces las variables que se

han considerado relevantes para el estudio del comportamiento homosexual?, y si realmente esos indicadores que se han estudiado, son representativos de la conducta homosexual (validez de contenido).

Es pertinente recordar que, como se dijo en el capítulo I, junto con la validez de constructo, existen otros tipos de validez; uno de ellos es el denominado de contenido. En este caso, el demostrar la representatividad de los indicadores que a continuación se desglosan, permitirá demostrar la validez de contenido.

Las variables importantes.

Hay que tener presente que la conducta sexual puede ser entendida de muchas maneras, que van desde el comportarse sociosexualmente (e.g. conductas de cortejo) hasta la respuesta sexual fisiológica. Esta amplia gama de comportamientos pueden ser considerados dentro de una definición de conducta sexual.

Lo mismo ocurre cuando se habla del término "mismo sexo" el cual puede ser definido, por lo menos, de cuatro formas diferentes, a saber: cromosómicamente, hormonalmente, por medios gonadales, y finalmente por sexo de asignación.

Tomando en cuenta la gran cantidad de posibles variables que están implicadas en estos conceptos que definen lo que es homosexualidad, lo pertinente es determinar cuáles son los in-

dicadores escogidos como relevantes para investigar y tratar el comportamiento homosexual, recordando que "Los constructos son los que determinan qué conductas han de seleccionarse para su observación y la situación en la que será aplicada." Ballesteros y Carrobles (1981, pag. 184), lo que será establecido mediante la revisión de diferentes reportes en el área de modificación de conducta.

En principio, es posible resumir los indicadores de interés a partir de la clasificación adoptada por Bancroft (1977), el cual subdivide en cinco categorías dichos indicadores. Estos son: a) Identidad sexual; se refiere a aquellas conductas que son identificadas como masculinas ó femeninas típicamente, b) Preferencia sexual y otras actitudes sexuales; esto es, las actitudes que se guardan con respecto al compañero de relación sexual y conjuntamente con ésto la erección del pene, c) Gratificación sexual; esto es entendido como la calidad e intensidad del placer sexual, d) Excitación sexual; puede ser entendida como todas aquellas respuestas fisiológicas que se presentan durante la relación sexual así como la erección misma, e) Actividad sexual; se refiere a los actos sexuales francos y/o imaginarios con ambos sexos.?

Esta clasificación reúne las principales variables medidas en la homosexualidad. Aún cuando no se pueda afirmar que existen definiciones muy claras para cada categoría, sí es posible afirmar que esta categorización engloba a la inmensa

mayoría de las variables relevantes, a pesar de que no estén contemplados muchos trabajos previos y posteriores a la propuesta de Bancroft, (Maletzky y George, 1973; Herman, Barlow y Agras, 1974a; Tanner, 1973; Tanner, 1974; Freeman y Meyer, 1975; Standford, Tustin y Priest, 1975; Conrad y Wineze, 1976; MacConaghy, 1977; James, Turner y Orwin, 1977; James, 1978.)

Desde otra perspectiva, regularmente no se ha pretendido modificar todas las instancias antes citadas, sino que más bien se han tomado como medidas colaterales, y las variables específicas sobre las que ha dado algún tipo de tratamiento son ante todo, las respuestas erectivas en mayor grado, (Solyom y Miller, 1965; Tanner, 1973; Tanner, 1974; Herman y col. 1974a; Herman, Barlow y Agras, 1974b; MacConaghy, 1975; Conrad y Wineze, 1976; James y col. 1977; James, 1978.), y menormente se han intentado modificar otras conductas como fantasías sexuales durante la masturbación y el uso de estímulos negativos imaginarios, (Maletzky y George, 1973; Maletzky, 1978; Gold y Newfeld, 1965.)

Es conveniente aclarar que las terapias que han utilizado técnicas como el condicionamiento clásico, o la evitación anticipatoria, se han centrado sobre la respuesta erectil generalmente, (Solyom y Miller, 1965; MacCulloch, Fellman y Pinshof, 1965; Tanner, 1973; MacConaghy, 1975; Standford y col. 1975; Freeman y Meyer, 1975; James y col. 1977;

James, 1978.).

Mientras que técnicas diferentes a estas lo han hecho sobre respuestas imaginarias o subjetivas básicamente, (Maletzky, 1978; Maletzky y col., 1973.).

Por otra parte, de forma invariable se encuentra que todos los autores, antes citados, han tomado y considerado que los estímulos pertinentes en el tratamiento de la homosexualidad son aquellos que guardan similitud anatómica o gonadal con el paciente que se está tratando; es decir, consideran como relación homosexual a la interacción sexual de los individuos que tienen similitud gonadal.

En resumen, las variables relevantes a modificar han sido tradicionalmente la respuesta eréctil del pene, así como las fantasías sexuales durante la masturbación (lo que Bancroft, 1977, ha denominado como excitación sexual), siendo las restantes variables tan sólo tomadas como medidas colaterales; es decir, son medidas únicamente y se espera cambien en la dirección deseada, una vez que se ha implementado algún tipo de tratamiento.

El prestar atención especial a las variables manipuladas en el tratamiento es importante, sobre todo cuando se pretende lograr la validación de constructo.

Es pertinente hacer una aclaración final, todos los artículos terapéuticos citados en este capítulo y en los siguientes, no constituyen una revisión exhaustiva de lo hecho hasta

la fecha, sino por el contrario, pretenden ser una muestra representativa.

En conclusión, el constructo conducta homosexual ha sido conceptualizado como la conducta sexual entre individuos del mismo sexo; las variables elegidas por los diferentes autores, es posible resumirlas en cinco categorías que son; a) preferencia sexual, b) gratificación sexual, c) excitación sexual, d) identidad sexual y e) actividad sexual. Enfocados el cambio o la modificación casi exclusivamente sobre las respuestas erectiles peneanas, es decir, se ha intentado modificar únicamente la excitación sexual.

VALIDEZ DE LOS INSTRUMENTOS

CAPITULO 3

Este tercer capítulo tiene como propósito el intentar establecer la posible validez, primero, de los instrumentos de medición usados en la evaluación de la conducta homosexual, y segundo, la posible validez del constructo denominado comportamiento homosexual.

Así, primero se hará la exposición de los instrumentos de evaluación más comúnmente usados en cada uno de los indicadores expuestos por Bancroft (1977), en segundo lugar, se propondrá un nuevo esquema que aglutine a los instrumentos de evaluación, citados en el primer punto, en cuatro grandes categorías: instrumentos usados en el diagnóstico de la conducta homosexual, medidas usadas durante la terapia de homosexualidad, medidas de postratamiento, y medidas empleadas durante el seguimiento. En tercer lugar, se hará un análisis de los instrumentos mencionados en cada uno de los puntos del esquema propuesto.

Por último, se sustentará una discusión con respecto a la posible validez de los datos arrojados por los instrumentos de medición y de igual forma con respecto a la probable validez del constructo "comportamiento homosexual".

Ya que se han determinado los indicadores relevantes en el comportamiento homosexual, el siguiente paso es considerar cómo es que han sido medidos éstos y lo adecuado o inadecuado de los instrumentos de evaluación utilizados, así como la misma validez de constructo.

En primer lugar, se retomará la clasificación expuesta por Bancroft (1977) para ubicar dichos instrumentos, así: la identidad sexual se ha medido mediante tests como: MF de Terman y Miles, escala 5 MMPI, escala MMPI, I6 PF, Rorschach, Slater de vocabulario apropiado, red de repertorio, inventario de personalidad dinámica, EPI.

Las preferencias sexuales se han medido por medio de: inventario de sexo, inventario de valoración diferencial semántica, método de orientación sexual, escala de ansiedad sociosexual, valoración verbal ante hombres y mujeres, pruebas psicofisiológicas (e.g. pletismógrafo), (Bancroft, 1977).

Para la gratificación sexual, solo se reporta un intento de medición (Thorne, 1966, en Bancroft, 1977), fuera de esto, no hay noticias de nuevos intentos de medición. La excitación sexual comúnmente ha sido medida mediante la respuesta peneana, (pletismógrafo), (Bancroft, 1977). La actividad sexual ha sido medida mediante: escala de Bentler, escala de Kinsey, escala de Bancroft, (Bancroft, 1977).

Hay que considerar también algunos sistemas, que más que de medición del comportamiento homosexual, han servido a otros

propósitos; tal es el caso de los exámenes físicos para determinar algún posible problema fisiológico. Su utilidad no va más allá de esto.

Dentro de esta clasificación de indicadores, caen todas las medidas de los trabajos revisados por Bancroft (1977), lo mismo ocurre con los artículos aquí revisados y que no están considerados por Bancroft en su revisión (Solyom y Miller, 1965; Maletzky y George, 1973; Tanner, 1973; Herman, Barlow, y Agras, 1974a; Tanner, 1974; Herman, Barlow y Agras, 1974b; Standford, Tustin y Priest, 1975; Freeman y Meyer, 1975; MacConaghy, 1975; Maletzky, 1977; James, Orwin y Turner, 1977; Conrad y Wineze, 1976; James, 1978.)

En realidad, al no encontrarse diferencia significativa al comparar los artículos arriba citados en términos de los indicadores elegidos como importantes y los que menciona Bancroft (1977), ésto pareciera estar indicando que el camino elegido es el apropiado. Sin embargo, hay que hacer algunas consideraciones antes de aceptar esto como un hecho. Estas consideraciones no son otra cosa que el analizar la pertinencia del uso de estos instrumentos, por lo que el paso siguiente es la proposición de un esquema que ayude y facilite el desarrollo de dicho análisis, y al mismo tiempo, sirva de enlace con el siguiente capítulo. Este esquema comprenderá cuatro categorías: medidas de diagnóstico, medidas durante el tratamiento, medidas postratamiento y medidas de seguimiento.

Así, se decidió clasificar a las medidas como de diagnóstico cuando éstas pretendieran diferenciar a los homosexuales de los heterosexuales en términos de su masculinidad o femineidad, preferencia sexual, actitud sexual, identidad sexual, o identificación sexual. Estas son: MF de Terman y Miles, escala 5 MMPI, I6 PF, MMPI, Rorschach, Slater de vocabulario apropiado, inventario de personalidad dinámica, pletismógrafo, inventario de sexo, registro conductual.

Las anteriores son, básicamente, las medidas más frecuentemente usadas para diagnosticar homosexualidad; ésto fue establecido a partir de lo expuesto por Bancroft (1977), y la revisión de diferentes artículos, ya antes citados (pag. 2).

Entre las medidas más usadas durante el tratamiento se encuentran las siguientes: pletismógrafo, red de repertorios, valoración diferencial semántica, método de orientación sexual, autoreporte.

El criterio que se tomó para clasificar a las anteriores medidas de esta manera, fue que dichos instrumentos sirvieran como un índice que proporcionara datos sobre los avances terapéuticos sesión tras sesión.

Las medidas de postratamiento más frecuentemente encontradas son: escala de Kinsey, escala de Bancroft, escala de Bentler.

El criterio adoptado para considerarlas de postratamiento fue que hubieran sido aplicadas una vez que la terapia

hubiere concluido.

Finalmente, las medidas de seguimiento identificadas fueron , primordialmente: autoreporte y reporte de colaterales.

El criterio adoptado para considerar a estas medidas como de seguimiento fue que intentaran valorar el cambio a largo plazo, una vez terminado el tratamiento.

Es oportuno aclarar que los instrumentos de evaluación antes mencionados, son los que con mayor frecuencia han sido utilizados por los diferentes investigadores. Aún cuando existen otros que también se han usado en la evaluación de la homosexualidad, no se han considerado aquí ya que su uso se restringe, en ocasiones, a un solo reporte y no se vuelven a tener noticias de su uso por otros autores o reportes; tal es el caso, por ejemplo, de la "prueba de tentación" diseñada por Maletzky y George (1973) y que solo aparece en su estudio.

Ahora bien, una vez propuesto el anterior esquema, se pasará a analizar los instrumentos considerados en cada uno de los niveles sugeridos. Dicho análisis se hará basándose principalmente en la investigación llevada a cabo por Bancroft (1977), y complementada por otros investigadores que también abordan el tema. El análisis hecho por Bancroft está enfocado primordialmente sobre la confiabilidad y validez de los instrumentos de evaluación. La pertinencia del uso de este material concuerda con los propósitos de este trabajo, por lo que

es oportuno recordar que la confiabilidad es un requisito necesario para la validez, y que el hecho de que un instrumento tenga validez no necesariamente implica que el constructo se haya validado también, sino que la validez del instrumento, en muchas ocasiones, puede ser un índice que dé pautas de la posible validez de constructo, no la validez de éste. Esto ya fue expuesto en el primer capítulo.

A manera de aclaración, aún cuando Bancroft y los otros autores citados para este análisis pocas veces hacen referencia directa a los conceptos de confiabilidad y validez, éstas pueden deducirse de los estudios y afirmaciones, que dichos autores hacen, por lo cual solo se mencionarán aquellas afirmaciones que pongan de manifiesto dicha confiabilidad y validez.

Pasando directamente al análisis de los instrumentos usados en el diagnóstico de la conducta homosexual, se tiene que: Al analizar el test de Terman y Miles, Bancroft (1977) encuentra que el problema más serio que enfrenta esta prueba es el de confiabilidad, ya que como enuncia este autor, dicho test varía de una cultura a otra, este test probablemente no es más que un reflejo de los intereses y valores sexuales del paciente. Además el mismo Bancroft afirma que no parece tener valor alguno en la investigación de tratamientos. El hecho de que este test presente diferencias transculturales es importante, puesto que esto habla de cómo dicho test fue diseñado



para responder a situaciones éticas y actitudes culturales muy específicas, lo cual es una limitante muy importante si se pretende usar en un contexto cultural diferente.

En el caso de la escala 5 MMPI, Bancroft (1977) reporta en sus conclusiones que esta escala tiene un índice muy bajo de validez y además agrega que esta escala en particular está aún más falta de validación que la mayoría de las del MMPI ya que de nuevo se encontraron importante diferencias culturales.

IZT. 1000310

Las mismas observaciones hechas a la anterior prueba pueden ser retomadas para ésta.

Para el MMPI, I6 PF, y Rorschach, Bancroft (1977), así como Barlow y Abel (1981), ponen de manifiesto lo dudoso de la confiabilidad de estos tests para diferenciar entre homosexuales y no homosexuales; en general, si aportan algún resultado, éste es de menor valía que el obvio procedimiento de preguntar a los sujetos cuáles son sus preferencias sexuales.

En el test de vocabulario apropiado, Bancroft (1977, pag. 253) encuentra que la confiabilidad es muy dudosa e incluso pone en duda la utilidad de dicho test; esto lo enuncia como sigue: "Slater y Slater, (1947)...encontraron que los homosexuales conocían más palabras femeninas que los del control normal...Clarke, (1965) demostró que el test dependía en gran medida de la edad, el C.I. y la clase social, y que cuando se controlaban estas variables, el grado de coincidencia entre

homosexuales y normales era lo suficientemente grande como para afirmar que el test no tenía valor clínico."

El inventario de personalidad dinámica es analizado por Bancroft (1977), citando diferentes estudios en los cuales él afirma que existe un considerable desacuerdo con respecto a lo que realmente mide dicho test; es decir, su validez es dudosa, por tanto, una medición de este tipo no puede encuadrar de un modo útil en la investigación de la modificación del comportamiento homosexual.

Para el inventario de sexo, Bancroft (1977), concluye que esta prueba tiene una confiabilidad aceptable, basándose en estudios de confiabilidad hechos con estudiantes a los cuales se les aplicaron dichos tests y en un periodo posterior se les volvieron a aplicar, obteniéndose correlaciones que variaron de 0.67 a 0.89, estando la mayoría por encima de 0.75. Por otra parte, dicha prueba necesitaría ser estandarizada en poblaciones más representativas con el propósito de ser más ampliamente aplicada. Además, agrega el mismo Bancroft que podría ocurrir lo mismo que con otros test de este mismo corte, y es que sea sensible a los cambios culturales y así sea inoperante o inapropiado rápidamente.

En el registro conductual no se encontraron pruebas de lo útil o inútil del uso de este instrumento de evaluación, aún cuando se ha usado con cierta frecuencia. Algunos puntos importantes a este respecto serán abordados más adelante cuando se

hable de las medidas usadas durante la terapia.

Para la pletismografía peneana se encontró que ésta tiene confiabilidad, validez e isomorfismo como indicador de diagnóstico, estas apreciaciones se justifican cuando se toman en cuenta los siguientes resultados aportados por diferentes autores; Bancroft (1971) reporta un estudio en el cual se evaluó la respuesta erectil en comparación con las medidas de auto-valoración en homosexuales y heterosexuales, siendo significativas las correlaciones. Barlow y Abel (1981) reportan diferentes investigaciones en donde se utilizó el pletismógrafo con diferentes grupos de diagnósticos de hombres con desviaciones sexuales, ante cuadros inanimados de hombres, mujeres y niños. Apoyándose en pequeños cambios subliminales, medidos por el pletismógrafo, categorizó con éxito a esos sujetos como homosexuales, heterosexuales y pedofílicos sobre la única base de sus respuestas de erección. Además, reportan otros estudios en los que se diferenciaron a hombres heterosexuales de pedofílicos, y otro estudio más en el cual se tenía el problema de saber si el paciente era homosexual o transexual, lo que fue establecido con éxito mediante el uso del pletismógrafo.

Bancroft (1977) concuerda con lo anterior al afirmar que este tipo de medición se ha venido utilizando con mucha frecuencia para determinar las preferencias sexuales de los sujetos, incluyendo la discriminación entre homosexuales y hetero-

sexuales y las preferencias de edad.

Todos estos autores coinciden en señalar series de experimentos en los cuales, mediante la proyección de diferentes diapositivas (mujeres, hombres, niños, etc.), cintas magnetofónicas, etc., se podía determinar, midiendo la respuesta penil-pletismográfica, cuál era la orientación sexual de los sujetos (heterosexual, homosexual, pedofílico, etc.). Como todo el anterior análisis lo marca, y retomando las afirmaciones hechas en principio, la pletismografía tiene un alto grado de confiabilidad (exactitud y consistencia), asimismo, su validez es demostrada cuando este instrumento detecta las erecciones ante el estímulo de preferencia del sujeto; ésto se ha puesto de manifiesto en los estudios antes citados que afirman que mediante el uso de este instrumento se han podido detectar las preferencias sexuales. El isomorfismo en este caso también es una realidad ya que los cambios detectados se pueden reportar de manera muy precisa en milímetros.

Es pertinente señalar, que si bien el pletismógrafo tiene su máxima utilidad en la identificación de las preferencias sexuales; su uso más corriente ha sido para evaluar los cambios en la erección a medida que el tratamiento avanza, (Bancroft, 1977; Barlow y Abel, 1981). La conveniencia y validez de este uso será analizado en el siguiente apartado, que corresponde a las medidas usadas durante la terapia.

Como conclusión de las medidas usadas para diagnosticar,

se puede afirmar que la mayoría de los instrumentos usados para este fin son de poca utilidad, ya que como lo muestra el análisis previo, muchos de ellos carecen de confiabilidad y validez; ésto es de gran importancia si se piensa en las implicaciones que puede tener en la elaboración de un diagnóstico; esto es, un buen diagnóstico, además de dar la certeza de que se está abordando el problema real del paciente, puede proporcionar información acerca de los excesos o déficits conductuales, así como de los repertorios actuales que posee el sujeto, la clasificación de las situaciones problema, y además las líneas generales para el tratamiento de conductas blanco específicas y el encontrar modos probables de tratamientos específicos. Todo esto se consigue cuando se ha logrado un diagnóstico apropiado, por lo cual el contar con instrumentos que posean confiabilidad y validez es de suma importancia, ya que de lo contrario las posibilidades de éxito tanto en el tratamiento como en la demostración de los fundamentos teórico-terapéuticos serán inciertos. Por lo mismo, si se abandonan los instrumentos de medición que tienen poca o ninguna confiabilidad y validez, verdaderamente no se está perdiendo nada; por el contrario, se puede ganar más si se piensa en diseñar instrumentos que simplifiquen esta tarea en particular, partiendo de un análisis funcional de las conductas a modificar.

Otro punto importante que es oportuno señalar es el he-

cho de que la mayoría de estos instrumentos fueron diseñados dentro de un marco de referencia de la teoría psicométrica clásica, (e.g. MMPI, I6 PF, Rorschach, etc.), y primordialmente para evaluar personalidad o actitudes y no específicamente homosexualidad, aun cuando se ha supuesto que pueden medir dicha identidad sexual.

Por otra parte, se encontró, de acuerdo al análisis previo, que solo dos instrumentos, de toda la serie revisada, son apropiados para diagnosticar homosexualidad.

El primero de ellos, el inventario de sexo, fue específicamente diseñado para evaluar preferencias y actitudes sexuales; sin embargo, aún cuando se logró determinar un grado aceptable de confiabilidad, todavía hace falta lograr un mayor grado de confiabilidad. Su uso ha sido muy restringido; se podría decir que casi nulo.

En el pletismógrafo se pudo comprobar que el grado de confiabilidad y validez es aceptable; sin embargo, una objeción muy importante a este instrumento es que como un simple indicador de preferencia sexual resulta ser muy costoso, además de su gran tamaño, el cual causa considerable estimulación del pene durante su aplicación.

Habiéndose expuesto lo que es la confiabilidad y validez de los instrumentos de diagnóstico, el siguiente paso es considerar de la misma forma los instrumentos utilizados durante el tratamiento. Así, en principio se analizará el pletismogra-

fo como primer y principal instrumento usado durante el tratamiento de la modificación de la conducta homosexual.

La principal crítica al uso que se le da a este instrumento es la falta de validez que tiene como indicador de modificación de la excitación sexual. Aclarando esta afirmación, hay que recordar que la inmensa mayoría de las investigaciones se han centrado en la modificación de la excitación sexual (decrementándola o incrementándola según sea el caso) y la medida más usada, en base a la cual se han reportado los cambios, ha sido mediante la pletismografía peneana, la cual debido a sus características de confiabilidad y validez en el diagnóstico de preferencias sexuales, es preferentemente usada por los diferentes investigadores. Esto ya fue fundamentado en un apartado anterior.

Puntualizando, el hecho de que la pletismografía posea un alto grado de confiabilidad y validez en un campo (diagnóstico), no necesariamente implica que automáticamente le sea dada la validez en otro (excitación sexual). El validar la pletismografía peneana como medida relevante de la excitación sexual no es tarea fácil.

Si se toma en cuenta que Bancroft (1971, pag. 119), define a la excitación sexual como "una combinación de excitación general y respuestas sexuales específicas", entonces se puede suponer que la "excitación general" son todas aquellas respuestas que están presentes en el momento en que algún es-

tímulo actúa como sexual, es decir, la excitación sexual es un conjunto de respuestas fisiológicas en diferentes niveles, como cambios en la tasa cardiaca, presión sanguínea, miotonía... (esto es, excitación general); así como la erección del pene y las valoraciones subjetivas, (estas son, respuestas sexuales específicas).

Partiendo de este hecho, Bancroft (1977) señala que la presencia de erección peneana no debe tomarse como un índice de excitación sexual generalizada y así, si se quiere medir dicha excitación, se deben recoger medidas tanto de la erección del pene como de las valoraciones subjetivas o respuestas autónomas (e.g. conductividad de la piel, presión sanguínea, ritmo cardiaco, etc.).

Ballesteros y Carrobles (1981, pag. 398), afirman: "1) no es posible utilizar ninguna medida psicofisiológica como índice generalizado de funcionamiento fisiológico global; 2) por ello es conveniente tomar medidas múltiples siempre que sea posible...". Por otra parte, Bancroft (1971) afirma que no se puede suponer que los cambios en la erección reflejarán cambios en la conducta o actitudes.

Estas observaciones son respaldadas por lo que Ballesteros y Carrobles (1981) enuncia como "fraccionamiento de respuesta" o "disincronización de respuesta", lo cual consiste en escasos niveles de covariación observados entre las diferentes medidas (cognitivas, motoras y fisiológicas) ante los

estímulos externos, y de igual forma se refiere a la variabilidad encontrada entre estas diferentes medidas, en respuesta a diferentes tipos de intervenciones terapéuticas conductuales. El tema de fraccionamiento de respuesta surgió, básicamente, como un problema en relación con el análisis específico de diversos constructos clínicos clásicos, como los de ansiedad, miedo, excitación sexual, en los que se observó la escasa covariación existente en algunos casos entre los distintos elementos componentes de estos problemas. Además agregan Ballesteros y Carrobles que el fraccionamiento de respuesta también puede ocurrir entre distintas medidas fisiológicas así como entre estas y otras conductuales y subjetivas.

Un ejemplo de lo anterior lo da Barlow (1977) cuando habla de cómo las mediciones de escalas de clasificación de inconformidad subjetiva, en la conducta sexual, pueden correlacionarse pobremente con mediciones fisiológicas y conductuales y menciona un estudio reciente, (Navissakalian, Blanchard, Abel y Barlow, 1975, en Barlow, 1977), en el que la respuesta sexual de seis hombres heterosexuales y seis homosexuales fue evaluada mientras se les proyectaban diferentes películas eróticas, incluyendo una película con contenido de interacción homosexual.

La respuesta del pene, reportes subjetivos de excitación sexual, y clasificación de actitudes de "placer" fueron medidas, los hombres heterosexuales reportaron a la película homosexual como implacentera, pero la erección del pene mostró

promedios de 20 por ciento de una erección total y también reportaron sentirse, inicialmente, excitados sexualmente. En general, las correlaciones de autoreporte fracasaron con las medidas objetivas de excitación, en una población de no-pacientes.

Un último ejemplo de esto sería el estudio de Conrad y Wineze (1976), quienes usaron una técnica de recondicionamiento orgásmico y midieron el cambio en tres niveles: 1) actitudes, 2) conductuales y 3) fisiológico. Los resultados mostraron que no hubo mejora en los sujetos (registro fisiológico), pero reportaron que sí existió mejora en las otras mediciones, (actitudes y registro conductual).

Recordando la afirmación hecha en principio, la validez de la pletismografía peneana con respecto a la excitación sexual aún está por justificarse, ya que de acuerdo a lo antes expuesto no hay evidencias que marquen al uso de la pletismografía como un indicador aceptable de la excitación sexual porque de hecho, la excitación sexual es una conducta compleja que involucra varios niveles de respuesta, siendo la erección solo uno de ellos y el cambio en éste no necesariamente implica un cambio generalizado de todo el sistema denominado excitación sexual. Y por ende como un indicador confiable del cambio terapéutico.

Otra medida usada en el tratamiento es la red de repertorio, en la cual se encontró una confiabilidad aceptable. Esto es,

Bancroft (1977) reportó que existía correspondencia entre la técnica de red de repertorios y las observaciones clínicas hechas en tres pacientes. Para el primer caso, este autor cita que los cambios sugerían que este paciente era más capaz de aceptar tanto su homosexualidad como su feminidad, lo cual concordaba con las observaciones clínicas; para el segundo paciente reporta como este fue más masculino y menos femenino, y cómo estos cambios coincidieron estrechamente con las observaciones clínicas hechas; finalmente en el tercer paciente reporta cómo los cambios en sus puntuaciones de red variaron conjuntamente con el tratamiento, disminuyendo hacia la mitad de éste y aumentando al final del mismo.

Así esta técnica de red de repertorio puede ser muy útil para valorar los cambios durante el tratamiento, por lo que su confiabilidad es aceptable, de acuerdo con la evidencia dada por Bancroft, más esta técnica ha sido muy poco usada. Aunque en realidad hace falta más investigación para poder confirmar estos resultados de confiabilidad preliminares.

La prueba de valoración diferencial semántica también ha sido usada durante el tratamiento y su nivel de confiabilidad es apenas aceptable. Bancroft (1977) cita como se utilizaron estas escalas para describir la mezcla de conceptos sexuales y no sexuales, y cómo las correlaciones entre los tests, para las puntuaciones de factores, fueron todas superiores a 0.70 y se consideró que eran lo suficientemente estables para

ser consideradas de valor clínico y detectar cambios durante el tratamiento.

Por otra parte, en otro estudio Freeman y Meyer (1975), reportan como diferentes medidas sí cambiaron conjuntamente en la dirección esperada, más no ocurrió lo mismo con el test de valoración semántica, en el cual no se observó cambio alguno.

Este resultado, que es contradictorio a lo reportado por Bancroft, habla acerca de las reservas con que debe de ser tomada esta prueba en lo que respecta a su confiabilidad.

En el método de orientación sexual (SOM), al igual que en el anterior test, se encontraron resultados contradictorios, por lo cual se consideró de confiabilidad apenas aceptable en términos generales. Diferentes autores respaldan esta afirmación; así, Bancroft (1977, pag. 262.), señala cómo se ha informado de un alto grado de coherencia interna, "La validez de esta medida fue verificada comparando las puntuaciones de antes y después del tratamiento en dos grupos, mejorados y no mejorados. Antes del tratamiento no diferían las puntuaciones medias de estos dos grupos. Después del tratamiento las diferencias tanto en las puntuaciones homosexuales como en las heterosexuales eran muy importantes.". Además el mismo Bancroft señala una serie de estudios donde se cuestionan las puntuaciones de incoherencia, ya que señala que la validez interna basada en las puntuaciones de incoherencia es incierta.

Por otra parte, James, Orwin y Turner (1977) afirman que una evaluación del SOM indicó que este no era totalmente satisfactorio y James (1978) cita la forma en que el SOM fue administrado en un estudio, y que a pesar de la incidencia de inconsistencias notadas en los puntajes obtenidos en una investigación previa, los cambios en los puntajes SOM generalmente se correlacionaron con otras medidas de progreso.

Como comentario final sobre esta prueba, está el hecho de que también ha sido usada con cierta frecuencia para diagnosticar la preferencia sexual de los sujetos (e.g. James y col. 1977; Herman y col. 1974a).

Una limitación más de este test es que "es simplemente una medida de las preferencias homosexuales y heterosexuales" (Bancroft, 1977, pag. 263).

En lo que respecta al autoreporte, se encontró que su confiabilidad es muy baja, ya que como menciona Barlow y Abel (1981) desafortunadamente el reporte verbal es fácilmente invalidado por el paciente y en muchas ocasiones es muy difícil que el paciente evalúe adecuadamente su propio avance, puesto que la percepción de su conducta puede ser distorsionada involuntariamente por diversos motivos, "la evidencia reciente demuestra que bajo diversas condiciones los pacientes dirán que están sexualmente excitados cuando no lo están (Barlow, Agras, Leitenberg, Callaghan y Moor, 1972), o informarán que están excitados por un estímulo cuando de hecho, es otro

el estímulo responsable de la excitación (Abel y Otros en prensa)... Estos factores hacen del informe verbal un medio deficiente para evaluar los patrones de excitación sexual." en Barlow y Abel (1981, pag. 371.).

Otros autores han usado el autoreporte como un medio para obtener información del comportamiento del sujeto fuera del gabinete; un ejemplo de esto son Maletzky y George (1973), quienes usan dicho autoreporte como la medida básica para reportar sus datos desde la línea base hasta el seguimiento.

Además dicho registro se correlaciona positivamente con las otras medidas usadas mensualmente (registro de observadores independientes), escala de Kinsey y "la prueba de tentación".

Herman, Barlow y Agras (1974a, pag. 44), citan el avance del tratamiento basados en el reporte de sus sujetos y en el pletismógrafo, concluyendo que: "El incremento en las respuestas peniles ante estímulos femeninos fue también paralela con el incremento en otras mediciones heterosexuales como la medición de actitudes subjetivas de excitación (SOM), reportes de excitación fuera del laboratorio (urgencias y fantasías) y reportes relacionados con las fantasías masturbatorias."

Freeman y Meyer (1975) reportan, en terminos generales, cómo el registro de fantasías heterosexuales, cambió en la dirección esperada y además esta medida covarió con otras

16 tomadas.

Por otro lado, Conrad y Wineze (1976) muestran en sus resultados que no hubo mejora en los sujetos en el registro fisiológico, más en el reporte verbal sí se encuentran cambios significativos.

A manera de aclaración, esta falta de confiabilidad no es propia de todos los registros conductuales; Barlow (1977) y Abel (1977): reportan la existencia de una alta confiabilidad en grupos de sujetos no pacientes.

Basándose en el análisis previo, se puede concluir que ninguna de las medidas anteriores tiene un grado aceptable de validez y por lo tanto las hipótesis derivadas de los datos obtenidos no pueden ser confirmadas por el uso del instrumento en cuestión, ya que hay que recordar que un requisito indispensable para que exista la validez, es que exista primero un grado aceptable de confiabilidad, requisito que muchos de estos instrumentos no cumplen (e.g. valoración diferencial semántico, método de orientación sexual, etc.), en segundo lugar ya que estos instrumentos intentan medir "excitación sexual", es importante tener presente que dicha excitación tiene un gran número de componentes, los cuales no pueden ser evaluados apropiadamente, hasta la fecha, por un solo instrumento; este es el caso de la pletismografía peneana, la cual posee un alto grado de confiabilidad para detectar los cambios del pene, más no se puede afirmar que tenga validez para re-

portar con efectividad los cambios en la excitación sexual. 1

En cuanto al isomorfismo que contiene esta herramienta psicofisiológica, es igual que el mencionado en la sección anterior cuando se habló del pletismógrafo como medida de diagnóstico.

En lo que respecta a las medidas usadas en el postratamiento, se encontraron los siguientes resultados:

Para la escala de Bentler se encontró que esta tiene una aceptable confiabilidad y Bancroft (1977) lo menciona como una forma de análisis no métrico que produce escalas masculinas y femeninas con un alto grado de coherencia interna y de ordenabilidad. Sin embargo, no se encontraron más estudios que confirmen o rechacen la utilidad de esta escala, asimismo, solo se encontró un estudio terapéutico en el que se empleó dicha escala (Tanner, 1973). Finalmente, el mismo Bancroft señala cómo esta misma escala es relativamente insensible a los primeros cambios del tratamiento y no refleja la calidad y frecuencia del comportamiento.

La escala de Kinsey tiene una confiabilidad aceptable; esto fue deducido a partir de sus características y mediante la crítica que Bancroft (1977, pag.271) hace al respecto "tal como fueron descritas en principio, estas valoraciones se referían a un periodo de tiempo definido. Indica poco más que una simple valoración de la proporción relativa de interés y comportamiento homosexual y heterosexual. Hay poco cam-

po para medir el cambio, y por esta razón son de valor limitado para la investigación de tratamiento."

Realmente, esta escala ha tenido su máxima utilidad en la valoración del comportamiento, mediante la comparación del grado de comportamiento homosexual con que ingresan los pacientes al tratamiento y el grado que tienen una vez terminado éste (Maletzky y George, 1973; Freeman y Meyer, 1975; James y col. 1977; James, 1978.)

Para la escala de Bancroft se encontró que ésta tiene un grado más alto de confiabilidad que la de Kinsey, aún cuando su validez está por confirmarse, esto es, hacen falta más investigaciones enfocadas a validar dicha escala, ya que las que existen son tentativas pero no definitivas. Bancroft (1977) afirma con respecto a la confiabilidad de esta escala particular que los coeficientes de confiabilidad fueron muy altos: puntuaciones heterosexuales, r 0.96; puntuaciones homosexuales, r 0.93 y además agrega que se debería de dar mayor atención a la utilización de estas escalas en la investigación de tratamiento.

Esta escala ha sido muy poco usada para valorar los cambios terapéuticos, y no se encontró ningún reporte que usara dicha escala en la investigación de tratamiento.

En conclusión, los datos arrojados por estas escalas particulares de medición son confiables en un plano aceptable, faltaría evaluar el grado de validez de estas escalas, lo mis-

mo que determinar su isomorfismo.

Quedan por analizar las medidas usadas durante el seguimiento.

De hecho, el autoreporte fue la medida preferentemente usada para evaluar seguimiento. No se encontraron estudios que proporcionen una estimación de la confiabilidad y validez de esta herramienta durante el seguimiento.

Sin embargo, al revisar diferentes artículos terapéuticos se encontró que los pacientes reportaban que los efectos del tratamiento se mantenían una vez terminado éste (Gold y Neufeld, 1965; MacCulloch, Feldman y Pinshoff, 1965; MacConnaghy, 1975; James y Col., 1977; Maletzky y George, 1973; Maletzky, 1977.). Esta consistencia no necesariamente implica confiabilidad, ya que dentro de estos autoreportes pueden estar operando un sinnúmero de variables, que no permitan valorar el grado de confiabilidad y validez, tales como; el que algunos de los pacientes fueron remitidos por la corte para tratamiento y los efectos, positivos o negativos, tenían mucho que ver con sus condenas; otro factor podría ser el que los pacientes estimaran inadecuadamente el retroceso o permanencia de los efectos terapéuticos, etc.

En segundo término, para el reporte de colaterales es más difícil establecer la confiabilidad y validez, ya que no se encontraron estudios relacionados con estos aspectos, y solo se logró ubicar dos estudios que usaron esta herramien-

ta (Maletzky y George, 1973; Maletzky, 1977.). Un primer aspecto interesante de este instrumento es que dichos reportes de colaterales coincidían en afirmar que los efectos terapéuticos se mantenían durante el seguimiento.

Segundo, este instrumento fue usado conjuntamente con el autoreporte encontrándose que existía covariación entre ambos instrumentos.

Realmente, analizando en su conjunto a estos instrumentos (autoreporte y reporte de colaterales), la primera conclusión tentativa es que la confiabilidad y validez de estas herramientas es incierta; segundo, hace falta implementar investigación que ayude a establecer el valor real de éstas; tercero, mientras se siga reportando seguimiento con autoreporte y reporte de colaterales y no se haya estimado su confiabilidad y validez, los resultados clínicos deben ser tomados con reserva, ya que de lo contrario se corre el riesgo de no poder valorar, de manera objetiva, los efectos terapéuticos a largo plazo.

Como conclusión general, con la excepción de la pletismografía usada como medida de diagnóstico de preferencia sexual, las restantes medidas empleadas para evaluar el constructo "conducta homosexual" carecen de validez. Esta apreciación está fundamentada en el análisis anterior. Resumiendo, en primer lugar, se logró establecer que la mayoría de las medidas usadas para diagnóstico carecen de confiabilidad y vali-

dez, ya que como los diferentes estudios lo marcan, estas no lograron cumplir con los requerimientos mínimos necesarios para obtener dicha confiabilidad y validez. Por otra parte, muchas de estas herramientas han sido diseñadas dentro de un marco teórico dinámico clásico, lo cual es una limitante ya que no se han desarrollado específicamente para detectar "comportamiento homosexual" sino "personalidad". Además, estos instrumentos tienen la particularidad de haber sido hechos obedeciendo a características etnocéntricas, esto es, han sido elaborados para ciertos grupos sociales en particular, lo cual conlleva la limitante implícita de que los conceptos de masculinidad, feminidad, actitud sexual, etc., varíen de una cultura a otra e inclusive dentro de la misma cultura, razón por la cual, estos instrumentos fácilmente proporcionan datos dudosos, de otra forma, al evolucionar los diferentes grupos sociales estos mismos instrumentos pronto resultan obsoletos, inclusive para el mismo grupo para el cual fueron contruídos.

Otra limitante de estos sistemas de evaluación es que en poco o en nada contribuyen al establecimiento o planeación de formas de intervención terapéutica, e inclusive en muchas ocasiones, son usados independientemente de la técnica terapéutica que posteriormente se implementa.

Realmente el único instrumento que ha logrado probar su validez ha sido el pletismógrafo, aún cuando hay que aclarar

que esta validez es con respecto a las preferencias sexuales y nada más.

En segundo lugar, cuatro de las cinco medidas usadas durante la terapia, tienen cierta confiabilidad para detectar los cambios ocurridos durante la misma y ninguno de estos cinco tiene validez para detectar los cambios de excitación ocurridos durante el tratamiento. La razón fundamental por la que esta validez no se ha dado es debido a que dichos instrumentos han intentado medir los cambios en la excitación sexual durante el tratamiento, y ninguna de estas medidas, usadas individualmente, ha logrado confirmar las hipótesis de cambio de excitación, ya que como antes se expuso, la excitación sexual tiene diferentes niveles de complejidad, los cuales no pueden ser explicados por el uso de uno solo de estos instrumentos.

En tercer lugar, aunque los instrumentos de postratamiento tienen una confiabilidad aceptable carecen de validez, otro aspecto importante es que estas herramientas no miden el comportamiento homo o heterosexual, sino que más bien determinan el grado de homosexualidad.

Cuatro, las medidas usadas durante el seguimiento son particularmente difíciles de valorar ya que no se encuentran estudios que fundamenten el nivel de confiabilidad o validez, por lo que los datos arrojados por éstos deben ser tomados e interpretados con muchas precauciones.

En resumen, con las medidas que en la actualidad se posee no existe razón alguna para usar una medida en lugar de otra, ya sea para diagnosticar, detectar cambios durante el tratamiento, el postratamiento o durante el seguimiento. Realmente, es muy pobre la validez de los instrumentos de medición usados en la evaluación del constructo "conducta homosexual", por lo que sería conveniente empezar a pensar en el diseño de nuevas herramientas que puedan proporcionar información útil de la conducta de los sujetos, tanto dentro del laboratorio como fuera de éste, y que además arrojaran datos pertinentes de los avances terapéuticos logrados, así como del posible mantenimiento de las conductas fuera del laboratorio.

En conclusión, al no encontrarse validez en los instrumentos usados en la evaluación del comportamiento homosexual, se puede decir, tentativamente, que los problemas para validar este constructo son muchos, ya que siempre es una limitante el carecer de instrumentos confiables y válidos con los cuales se respalden las afirmaciones y hallazgos encontrados experimentalmente o en forma aplicada.

ANALISIS DE LOS RESULTADOS TERAPEUTICOS

1960

CAPITULO 4

Análisis de resultados.

El siguiente capítulo tiene como objetivo evaluar los logros terapéuticos, partiendo del seguimiento y los conceptos inherentes a éste, así como el sustentar una discusión de la posible validez del constructo " conducta homosexual ".

Este objetivo se cubrirá a partir, primero, de la exposición de los conceptos de "criterio experimental", "criterio terapéutico y la forma en que éstos se evalúan o determinan. Segundo, mediante la determinación de los conceptos de "mantenimiento de respuestas" y " transferencia de entrenamiento". En tercer lugar, se integrarán los anteriores conceptos y se marcará la relación que con el seguimiento guardan. Cuarto, se analizarán los logros terapéuticos alcanzados por los diferentes autores. Finalmente, se sustentará una discusión sobre la posible validez del constructo, "conducta homosexual".

Una vez que se ha analizado la confiabilidad y validez de los instrumentos usados en la medición del constructo conducta homosexual, el siguiente paso es analizar los logros terapéuticos de este mismo constructo.

Este análisis se hará a partir del seguimiento conductual, el cual está estrechamente ligado a conceptos de suma importancia, como son; a) el criterio experimental o estadístico y el criterio clínico o terapéutico; y b) el mantenimiento de respuesta y la transferencia del entrenamiento.

El seguimiento conductual no es simplemente preguntarle al paciente "¿Cómo se encuentra, todo va bien?" sino por el contrario, las implicaciones teóricas y prácticas son muy importantes y profundas.

Aclarando, se tiene que el seguimiento conductual puede ser entendido como el proceso que se da una vez que se ha terminado el tratamiento, y que consiste en valorar si la conducta modificada persiste o ha sufrido algún cambio. Esta forma de valoración está relacionada, en primer lugar, con el criterio estadístico y terapéutico.

Resaltar las características de estos criterios es una tarea necesaria si se toma en cuenta que son la base de esta crítica de resultados.

En principio, el criterio experimental o estadístico, se

refiere a cómo es que se presenta la conducta durante la línea base y cómo debería de presentarse ésta, si los procedimientos experimentales no se hubieran introducido (Risley, - 1970; Kazdin, 1978).

Los métodos de análisis para el criterio experimental son bien conocidos y practicados (análisis visual, de tendencias, pendientes, pruebas de significancia, etc.).

Por otra parte, el criterio terapéutico o clínico se refiere a si el cambio obtenido durante la terapia es de importancia tal que sea significativo socialmente, es decir, el cambio logrado debe de ser de suficiente magnitud para que el sujeto funcione adecuadamente en su sociedad (Risley, 1970; Kazdin, 1978). Este criterio surge por la necesidad de validar el cambio en la práctica profesional, o lo que es lo mismo, el paso de la investigación básica a una práctica profesional.

Para este criterio clínico se tiene que los métodos de análisis son dos: método de comparación social y método de de evaluación subjetiva (Kazdin, 1978).

Esto es, "Con el método de comparación social, la conducta de un cliente, es comparada, antes y después del tratamiento, con la conducta de sus iguales 'no desviados'.

La pregunta inquisidora, para esta comparación, es de si la conducta de los clientes después del tratamiento se distingue del rango normativo de conductas de sus iguales, de él o ella. Con el método de evaluación subjetiva, la conducta es

evaluada por individuos que están, o tienen probablemente, contacto con el cliente para determinar si el cambio hecho durante el tratamiento es importante" (Kazdin, 1978 pag. 639).

En el uso del método de comparación social para evaluar el tratamiento, es esencial identificar a los individuos que son similares al cliente, en las variables sujeto y demografía, pero que muestran conducta aceptable. La primera es que, los sujetos iguales al cliente, deben diferir marcadamente en la conducta elegida previa al tratamiento.

En el método de evaluación subjetiva Kazdin (1978) ofrece un panorama de cómo se debe de llevar a cabo; esto es, en la aplicación del método de evaluación subjetiva, los individuos que normalmente interactúan con el cliente o que están en una posición especial (e.g. un experto), juzgan una conducta particular proveyendo una valoración global para evaluar que tan bien está funcionando el cliente.

Para el segundo punto (mantenimiento de respuesta y transferencia de entrenamiento), se tiene que el mantenimiento de respuestas puede ser entendido como la permanencia de la conducta una vez que el programa de tratamiento se retira.

El principal problema que afronta el mantenimiento de respuesta es el hecho de que generalmente la conducta se extingue cuando el programa se retira (Kazdin, 1978a).

Por otra parte, la transferencia del entrenamiento se refiere a si la conducta modificada, en una situación determi-

nada, se generaliza a otras situaciones, ya sea que el programa esté corriendo o después que se ha retirado. (Kazdin, 1978a).

De hecho, la principal limitante de esta transferencia es la poca generalización conductual que se da a otras situaciones, ya que en muchas ocasiones el cambio conductual está su peditado a la situación específica donde se lleva a cabo el tratamiento y en presencia de quienes lo administran.

Tanto el mantenimiento como la transferencia de entrenamiento, en la mayoría de los casos, están estrechamente relacionados. Por ejemplo, cuando un paciente es tratado por poseer un bajo nivel en conductas sociales, es importante que la respuesta se mantenga y de la misma forma, que dicho entrenamiento se transfiera a la situación natural. De otra forma, la transferencia del entrenamiento puede presentarse sin que necesariamente se de el mantenimiento, (por ejemplo, cuando un niño es tratado por conductas agresivas en salón de clases) o puede darse el mantenimiento de respuesta sin que se transfiera el entrenamiento, (por ejemplo, el paciente que es tratado por bajo nivel social mediante entrenamiento asertivo puede presentar conductas asertivas apropiadas en el gabinete más no fuera de este (situación natural)).

Como aclaración, estos temas (criterio experimental y clínico; mantenimiento de respuesta y transferencia de entrenamiento) son muy extensos y no es propósito dar una amplia exposición de ellos aquí, sino por el contrario, ubicarlos y ha-

cer explícita la relación que con el seguimiento guardan, por lo que el siguiente paso, es marcar la importancia de dichos criterios en el seguimiento.

Así, la forma en que el criterio terapéutico puede ayudar a establecer el seguimiento, está dada por el hecho de que tanto en el método de evaluación social como en el de evaluación subjetiva, se propone comparar la conducta del sujeto antes y después del tratamiento, por lo cual, estos mismos criterios pueden ser usados para evaluar la conducta a largo plazo, sin que por esto se cambien las medidas usadas en el tratamiento y seguimiento, ya que como se vio en el caso de la conducta homosexual, las medidas de tratamiento difieren de las empleadas en los seguimientos, factor importante si se toma en cuenta que estas últimas son reportes verbales o telefónicos, en contra de los cuales existen muchas objeciones de confiabilidad. Por lo que al utilizar siempre las mismas medidas a lo largo de la pre-evaluación, post-evaluación y seguimiento, es importante -hablando tanto desde la perspectiva práctica como desde el punto de vista clínico y experimental- ya que el adoptar un enfoque de este tipo permite la completa integración de reportes confiables a largo plazo, que además puedan ser comparados con otros involucrados en el mismo tema, independientemente de que las técnicas de intervención difieran cualitativamente. Esto no implica que se tenga que abandonar otras medidas que dan información de los progresos terapéuticos sino que pueden complementarse

mutuamente.

Por otra parte, cuando no se adopta un criterio terapéutico y sí uno de tipo experimental, los problemas para implementar un seguimiento adecuado crecen, y una opción válida para tener seguimientos de calidad sería el que se evaluara al sujeto con los mismos instrumentos que se usaron durante dicho tratamiento para así intentar determinar, por lo menos, la posible permanencia de la respuesta, aún cuando se tenga serias limitaciones para poder afirmar la generalización al medio ambiente natural del paciente.

Ahora bien, una vez visto el seguimiento y la relación que guarda con el criterio terapéutico, es oportuno enmarcar la relación que existe con los conceptos de mantenimiento de respuesta y transferencia de entrenamiento.

Así, cuando el seguimiento se establece, fundamentado en un criterio terapéutico, la tarea de evaluar el mantenimiento de la respuesta se facilita, ya que las conductas que se espera se mantengan están ya especificadas y previamente evaluadas al tratamiento, lo cual permite establecer comparaciones que den índices claros del posible mantenimiento de dicha conducta, de la misma forma, la transferencia del entrenamiento puede ser igualmente establecida, esto es, un seguimiento basado en un criterio terapéutico permite evaluar con mayor precisión la transferencia del entrenamiento, ya que las conductas establecidas como importantes están identifica-

das y evaluadas con antelación, lo cual permite que dicho comportamiento pueda ser identificado y evaluado en otras situaciones (es decir, si se ha generalizado). Por ejemplo, cuando un paciente con falta de habilidades sociales se le implementa un entrenamiento asertivo, lo que se espera es que tal conducta asertiva persista al final del tratamiento (mantenimiento de respuesta) y además se generalice a situaciones nuevas que se le presenten al sujeto y en diferentes ambientes (transferencia de entrenamiento).

Una implicación importante que se relaciona tanto con el criterio terapéutico como con el mantenimiento de respuesta y la transferencia del entrenamiento, es que el seguimiento en su proceso puede contribuir a establecer o enriquecer los criterios terapéuticos basándose en la retroalimentación que proporciona éste (el seguimiento), cuando de alguna forma también evalúa la pertinencia de la elección de las conductas que se escogieron como centrales a medir.

Finalmente una aclaración es oportuna: se prefirió analizar los reportes terapéuticos del comportamiento homosexual partiendo del concepto de criterio de éxito como tal, más que de los reportes de porcentaje de éxito, mejorados, ligeramente mejorados, etc. La justificación para adoptar este criterio es que existe un vínculo lógico entre lo que es el criterio terapéutico y la definición de los criterios de éxito, los cuales quedan automáticamente establecidos en el momento

que se adopta el método de comparación social y/o el método de evaluación subjetiva. Por ejemplo, si el objetivo de una terapia es que los pacientes al final se comporten como sus iguales en el medio ambiente natural, entonces el éxito se medirá de acuerdo al cumplimiento de éstos objetivos y no únicamente en base a los criterios y medidas experimentales.

De esta forma, ya que se han establecido los criterios de éxito, es entonces que el seguimiento puede ser implementado fácilmente por medio de la extensión de registros de comparación social y evaluación subjetiva para así no alterar los instrumentos usados durante la línea base, fase de tratamiento y seguimiento; con lo cual se tiene una medida igual y confiable que permita estimar los cambios de conducta en forma sistemática y objetiva.

En concreto, el paso siguiente es analizar los logros terapéuticos alcanzados en el constructo enunciado como conducta homosexual.

Así, en la revisión de artículos se encontró que todos los autores reportan algún grado de éxito, que va desde "ligeramente mejorado" hasta "mejorados" (Gold y Neufeld, 1965; Herman, Barlow y Agras, 1974 b; Standford, Tustin y Priest, 1975; Maletzky, 1977.). También se reportó éxito en términos de porcentaje, (MacCulloch, Feldman y Pinshoff, 1965; Tanner, 1974; MacConaghy, 1975; Freeman y Meyer, 1975; James, Orwin, y Turner, 1977; Maletzky George, 1973; James, 1978.).

A pesar de lo divergentes que puedan parecer estos artículos, guardan un punto en común, y es que los criterios para considerar el éxito en el tratamiento son muy ambiguos o están mal definidos; por ejemplo Standford y col. (1975), establecieron como criterio para dar por terminada la terapia que ya no hubiera cambios en la circunferencia del pene ante estímulos homosexuales; Freeman y Meyer (1975), enuncian que los sujetos al final del tratamiento decrecientan en la escala de Kinsey de 4.4 a 1.7 en promedio y suponen que éste es un criterio de éxito aceptable; James y col. (1977), establecen como criterio de éxito que los pacientes decrecientan dos puntos en la escala de Kinsey y se mantengan así por cierto periodo, etc. Este defecto generalizado provoca, primero, que sea muy difícil establecer comparaciones entre los diferentes tratamientos, y segundo, que en ocasiones los resultados aparezcan inflados por la misma ambigüedad de criterios.

Se observó, así mismo, que la mayoría de los autores llevaron a cabo seguimientos, (Gold y Neufeld, 1965; MacCulloch, y col. 1965; Herman y col. 1974b; MacConaghy, 1975; Freeman, y Meyer, 1975; James, 1978.), con promedio de duración de año y medio aproximadamente, lo que constituye un tiempo muy corto para poder valorar los efectos a largo plazo de cualquier terapia. Además estos seguimientos no fueron realizados en forma adecuada, ya que fueron hechos mediante llamadas telefónicas, entrevistas, o inclusive en algunas no se menciona cómo

se hicieron.

La relación que guardan estos aspectos (éxito de tratamiento y seguimiento), es muy importante, ya que hay que tener presente que el establecimiento de criterios de éxito terapéutico implica la evaluación del comportamiento antes y después del tratamiento en su medio ambiente natural, proceso que no fue realizado por ninguno de los autores citados.

Por el contrario, todos los "criterios de éxito", implícitos y explícitos en estos artículos, están basados en lo que se conoce como criterios experimentales para manifestar que el comportamiento homosexual se había modificado, y dependiendo de la magnitud del cambio se consideró como "ligera mejora", "muy mejorado", "mejorado", etc.

Con la posible excepción de Maletzky y George, (1973), que reportan observaciones de colaterales, ningún investigador utilizó criterios de éxito basados en un enfoque terapéutico por el contrario, se basaron en el experimental. Y como resultado de esta deficiente implementación de los criterios terapéuticos surgen los seguimientos deficientes, que hechos mediante entrevistas o llamadas telefónicas, se desligan del sistema de medición y evaluación que se hace a las conductas elegidas como relevantes a modificar; esto es, el comportamiento homosexual se mide, analiza y evalúa en base a ciertos instrumentos (pletismógrafo, test, etc.) y criterios (experimental o estadístico) y durante el surgimiento se cambian las formas de medición (entrevistas, telefonemas, etc.)

y se asume que no existe mayor problema al hacerlo y los resultados obtenidos, en las medidas de seguimiento, se interpretan como relevantes.

Este problema se evita si se adoptan criterios terapéuticos adecuados que permitan tener conductas específicas a comparar (método de comparación social) o elementos de reporte externo (evaluación subjetiva), que puedan ser extensivos en el tiempo y que sean acordes con los elementos de medición de la terapia.

Finalmente, como caso especial habría que mencionar a Masters y Johnson (1979), que apodan el problema de la homosexualidad reportando un porcentaje aproximado de 35% de fracaso, con seguimientos de 1 a 5 años, lo que los lleva a concluir que "si bien las estadísticas registradas... son interesantes, no permiten sacar conclusiones seguras..."

En pocas palabras, es difícil valorar los aportes de estos autores, debido principalmente a que no se conoce con detalle cual fué el tratamiento terapéutico empleado (que será reportado posteriormente, según los autores) y la nueva orientación que se dá al tratamiento de la conducta homosexual, por lo que en este caso tampoco es posible afirmar la efectividad del tratamiento, siendo importante considerar esta aproximación como una tentativa válida e importante.

En conclusión, ninguno de los trabajos citados en esta revisión pueden demostrar que el tratamiento o tratamientos

surgidos sean efectivos, ya que tienen importantes deficiencias metodológicas como las aquí expuestas; estas son: la adopción de criterios inadecuados (experimentales en lugar de terapéutica), con lo cual el terapeuta se preocupa más por que las medidas tomadas en el laboratorio cambien, que por observar si el cambio se trasfiere al medio ambiente natural. Además otra consecuencia de esta adopción de criterio es que al no existir criterios terapéuticos claros, se cae en el de fasamiento de medidas, esto es, las medidas usadas durante el tratamiento difieren de las empleadas en el seguimiento y, basándose en estas últimas, se afirma que los efectos terapéuticos persisten o se han revertido, lo cual constituye un error metodológico que puede conducir a conclusiones erróneas, ya que por un lado los instrumentos que se han usado en los seguimientos no han probado aún su validez, y por otro, al su poner que al cambiar los instrumentos de evaluación (de una condición a otra no implica mayor problema) puede ser peligroso, cuando no se ha probado la pertinencia de ese cambio, agrega la dificultad de poder hacer comparaciones adecuadas entre ambas condiciones (de laboratorio y situación natural).

Un problema más estriba en que los terapeutas no se preocupan o se preocupan muy poco por evaluar el comportamiento en el medio ambiente natural del paciente, con lo cual se deja de lado el criterio terapéutico y por ende los beneficios y ventajas que éste puede aportar; aunado a esto los tiempos de seguimiento que son muy variables y sobre todo cortos, y

en muchas ocasiones ni siquiera se llevan a cabo, con lo cual dichos seguimientos contribuyen muy poco o en nada en la valoración real de los tratamientos terapéuticos.

A conclusiones similares llega Bancroft (1977) con respecto a la efectividad de los tratamientos, aún cuando sus críticas se basan en un análisis un tanto diferente al antes expuesto mientras que Marks (1976), también coincide con muchos de estos puntos de vista cuando afirma que las investigaciones controladas son la excepción más que la regla y que son necesarios trabajos más sistemáticos con seguimientos adecuados de un año o más.

Considerando el anterior análisis, es pertinente señalar la convivencia de tomar en cuenta la diferencia sustancial que existe entre los criterios experimental y terapéutico, y aún más, el apegarse en la medida de lo posible a los loneamientos que cada uno de éstos guarda, para de esta forma, poder valorar con mayor precisión los avances reales de los tratamientos, tanto a nivel de gabinete como a nivel aplicado, para así poder sugerir líneas de acción más precisas.

CONCLUSION

Conclusión

Por último, ya que se ha analizado tanto la validez de los instrumentos así como los logros terapéuticos, es oportuno abordar el problema de la posible validez del constructo "conducta homosexual".

Partiendo del hecho de que la validez de constructo no es otra cosa que el validar la teoría que subyace al constructo, para este caso concreto se puede decir que los supuestos teóricos subyacentes al comportamiento homosexual no han podido ser cabalmente validados.

Esta afirmación descansa en el análisis previo, esto es, si se retoman los conceptos expuestos en principio se recordará que la validación de constructo es muy importante más no puede ocurrir independientemente del instrumento que mide, de ahí que se tenga que:

a) La mayoría de los instrumentos (e.g. red de repertorios, MMPI, etc.) no pudieron probar su validez y tan solo algunos tuvieron un grado aceptable de confiabilidad, (e.g. Bentler) y únicamente el pletismógrafo, como medida de preferencia sexual, tuvo validez. Esto último es de gran importancia ya que el uso más frecuente del pletismógrafo ha sido en la valoración del cambio durante la terapia, campo en el cual dicha pletismografía tiene serias limitantes de validez (aún cuando su confiabilidad es un hecho) esta validez está cuestionada en base al concepto de fraccionamiento de respuesta. Tema que guía el siguiente inciso.

b) Siendo la excitación sexual la conducta sobre la cual se han enfocado la inmensa mayoría de los tratamientos de cambio, y el pletismógrafo el instrumento básico con el que se han reportado dichos cambios de excitación; es posible decir que, si bien en un principio se aceptó como un hecho que dicho instrumento de medición tenía validez, el tema de fraccionamiento de respuesta vino a ser determinante para poder establecer con mayor precisión la valía real del uso de esta herramienta, ya que las pruebas acumuladas y expuestas han demostrado lo dudoso que resulta la erección peneana como indicador apropiado del cambio en la excitación sexual, esto es, no se ha logrado determinar con éxito que dicha erección peneana sea representativa de la excitación sexual, (dicho de otra forma, no se ha conseguido la validez de contenido) y por consiguiente los datos arrojados por el pletismógrafo, como instrumento que mide la erección, y supuestamente cambios en la excitación sexual, no ha podido ser validado apropiadamente, es decir, las hipótesis derivadas del uso del pletismógrafo (cambios en la excitación) no han podido ser validados cabalmente.

En este sentido, la pura validez del pletismógrafo para detectar preferencias sexuales, no es suficiente para decir que se tienen herramientas de medición apropiadas para medir el constructo conducta homosexual.

Aún cuando el pletismógrafo es útil en la identifica-

ción de preferencias sexuales parece no ser suficiente como una medida de diagnóstico, en el sentido de que da poca luz para orientar las posibles estrategias de intervención terapéutica.

c) Los instrumentos de evaluación usados durante los seguimientos han sido particularmente difíciles de analizar, por el momento no es posible hablar acerca de la validez de estos instrumentos y sólo se puede mencionar que los datos arrojados por dichos instrumentos poseen cierta estabilidad.

d) Esta carencia de instrumentos de medición que contengan niveles satisfactorios de confiabilidad y validez, conllevan un doble problema; primero, la información que arrojan estos instrumentos aporta muy poca información sobre la conducta en cuestión, segundo, aparte de que en poco o en nada contribuyen a sugerir las formas de intervención terapéutica, tampoco proporcionan nuevas líneas de investigación en la terapia del comportamiento homosexual.

Por otra parte, con respecto a los estudios terapéuticos, y en relación a la validez de constructo, se puede decir que:

a) Ningún investigador ha podido proponer un tratamiento que demuestre efectividad considerable; esto quedó de manifiesto al encontrarse coincidencia entre los análisis de Bancroft (1977), y Marks (1976), y el realizado en este trabajo, lo cual es de suma importancia si se piensa que, al in-

tentarse modificar la conducta homosexual se han elegido ciertas variables que se "supone" pertenecen a este constructo y que si estas variables no se han modificado, (o se han modificado, y el comportamiento homosexual persiste) por tanto la validez de constructo no se ha logrado. La prueba más palpable de esta falta de validez de constructo queda manifestada, primero, en el bajo nivel de éxito alcanzado por la mayoría de los investigadores; segundo, las deficiencias metodológicas no permiten hacer comparaciones entre los diferentes estudios para llegar a conclusiones seguras que permitan una correcta valoración, no solo de las terapias, sino también de las variables (validez de constructo) y su representatividad (validez de contenido).

b) Las deficiencias metodológicas encontradas, en la gran mayoría de los reportes, imposibilitan evaluar con seguridad las técnicas terapéuticas empleadas; un ejemplo de esto son los seguimientos que con todas sus restricciones (que van desde la carencia de sistematicidad, falta de criterios adecuados, utilización de instrumentos de medición con baja confiabilidad, etc.), no permiten valorar apropiadamente la permanencia de la respuesta ni la posible transferencia del entrenamiento, factores que son de suma trascendencia si se piensa en la aportación que éstos harían en caso de ser valorados con propiedad, ya que suponiendo que realmente algún tratamiento resultara efectivo y fuera demostrada

la permanencia de respuesta y la transferencia del entrenamiento, es entonces que la posible validez de constructo podría ser mayor, aún cuando existe la objeción de que el hecho de que un tratamiento sea efectivo no garantiza que las variables utilizadas para la descripción de ese problema sean adecuadas.

IZT. 1000310

Visto desde otra perspectiva, el constructo conducta homosexual podrá ser validado solo cuando la teoría subyacente sea confirmada mediante procedimientos empíricos que demuestren que las variables elegidas pertenecen realmente al dominio contemplado y, de la misma forma, los instrumentos de medición serán validados en la medida que detecten y confirmen los cambios e hipótesis derivadas de su uso. Similarmente, la validez de contenido tendrá que demostrarse de manera empírica.

Así basándose en el análisis anterior, es oportuno proponer, primero, mayor rigor metodológico que permita la valoración más precisa de las técnicas y tratamientos usados en la modificación de conducta; segundo, poner mayor atención en la confiabilidad y validez de los instrumentos usados durante los tratamientos y de no existir instrumentos con estas características, crear herramientas especiales para medir lo que se quiere o pretende; tercero, propiciar investigación no solo a nivel de laboratorio sino también fuera de este, tal propósito tiene como fin el ampliar el campo de la tera-

pia de la conducta; cuarto, pensar que mucho del comportamiento homosexual, al cual se enfrenta el terapeuta de la conducta, es muy complejo y regularmente no está en función de un solo factor sino de un conjunto muy amplio de factores; por ejemplo, Herman, Barlow y Agras, (1974, a y b), Tanner, (1974), James, (1978), han puesto de manifiesto que no solo es necesario modificar la excitación sexual desviada sino también el implementar habilidades heterosociales en los pacientes.

Este último punto tiene diferentes implicaciones; por un lado sugiere el abordar e investigar nuevos tipos de variables, estas son las de tipo social. Tal vez mucho de esta inclinación se deba a dos factores, primero, el bajo nivel de éxito alcanzado con los tratamientos enfocados a modificar la excitación sexual; segundo, las evidencias se inclinan a lo aprendido de la sexualidad en contraste con lo heredado, (Masters y Johnson, 1979; Gagnon, 1977; Bell y Weinberg, 1979; Giraldo, 1981).

Sin embargo, el ampliar el rango de variables a estudiar (al campo del aprendizaje social) implica abordar nuevos problemas. En principio, no hay generalidades en cuanto a cuáles son las variables sociales relevantes que determinen el comportamiento homosexual, ya que los patrones establecidos de masculinidad y feminidad cambian constantemente de una cultura a otra, e inclusive dentro de la misma cultura en diferen-

tes momentos. Las investigaciones sobre patrones característicos entre homosexuales y heterosexuales (e.g. postura, preferencia de actividades, vestido, diversiones, desarrollo familiar, etc.) han arrojado, de igual forma, poca luz en cuanto a cuáles son las variables sobresalientes en la homosexualidad, (Baneroff, 1977; Tripp, 1978; Giraldo, 1981; Gagnon, 1977; Bell y Weinberg, 1979).

Sin embargo no se puede afirmar que se haya fracasado en la búsqueda de las variables relevantes en la conducta homosexual, ya que la investigación en este campo aún es joven, por lo cual sería importante intentar determinar de manera empírica cuáles serían las variables sociales relevantes; investigar la influencia de la interacción de las variables de excitación sexual y conductas sociosexuales.

Posiblemente la conducta homosexual, como mucho otros comportamientos complejos, esté en función de más de una variable, esto es, suponer que la sola modificación de la erección peneana es suficiente para modificar la conducta homosexual es un error (pero suponer que la modificación de la erección no es importante puede también serlo); hay que considerar también a las variables de tipo social como un factor importante en el desarrollo y mantenimiento de la conducta homosexual.

Por el momento, sería importante prestar mayor atención a los entrenamientos, ya conocidos, para incrementar las con-

ductas sociosexuales (e.g. entrenamiento asertivo) y evaluar sistemáticamente la interacción con las variables de excitación sexual, tanto desde la perspectiva metodológica, de transferencia, generalización, mantenimiento, significancia clínica, etc., y reevaluar la validez de contenido y de constructo de estas interacciones.

Bibliografía

- Abel, G.G. Assessment of sexual deviation. En Hersen, H. and Bellack, A. Behavioral assessment a practical handbook. New York: Pergamon Press, 1976, pag. 437-456.
- Ballesteros, F.R. y Carrobles, I.J.A. Evaluación conductual: Metodología y aplicaciones. Madrid: Pirámide S.A. 1981.
- Bancroft, J. The application of psychophysiological measures to the assessment and modification of sexual behavior. Behavior Research and Therapy, 1971, 9, 119-130.
- Bancroft, J. Desviaciones de la conducta sexual. Barcelona: Fontanella, 1977.
- Barlow, D.H. Assessment of sexual behavior. En Cimínero, A. R. Calhoun, K.S. and Adams, H.E. Handbook of behavioral assessment. New York: Wiley, A. Intercence Publication, 1977, 461-505.
- Barlow, D.H. y Abel, G.G. Desviación sexual en modificación de conducta. En Craighead, W.E. Kazdin, A.E. y Mahoney, M.J. Principios, técnicas y aplicaciones. Barcelona: Omega, S.A. Casanova, 1981, 365-385.
- Bell, A.P. y Weinberg, M.S. Homosexualidades: Informe Kinsey. Madrid: Debate Colección Universitaria, 1979.
- Coleman, J.C. Sexual "deviation" abnormal psychology an modern life. Chicago: John Wile and Songs, 1975.

- Freeman, W. and Meyer, R.G. A behavioral alteration of sexual preferences in the human male. *Behavior Therapy*, 1975, 6, 206-212.
- Gagnon, J. *Sexualidad y conducta social*. México: Pax-México, Librerías Carlos Césrman S.A. 1980.
- Giraldo, N.O. *Explorando las sexualidades humanas: Aspectos psicosociales*. México: Trillas, 1981.
- Gold, S. and Neufeld, I.L. A learning approach to the treatment of homosexuality. *Behavior Research and Therapy*, 1965, 2, 201-204.
- Haynes, S.H. *Principles of behavioral assessment*. New York, Gardner Press, Inc. 1978.
- Herman, S.H. Barlow, D.H. and Agras, S. An experimental analysis of classical conditioning as a method of increasing heterosexual arousal in homosexuals. *Behavior Therapy*, 1974, 5, 33-47, (a).
- Herman, S.H. Barlow, D.H. and Agras, S. An experimental analysis of exposure to "explicit" heterosexual stimuli as an effective variable in changing arousal patterns of homosexuals. *Behavior Research and Therapy*, 1974, 12, 335-345, (b).
- Hollenbeck, R.A. Problems of reliability in observational research. En Sackett, P.G. *Observing behavior*. Baltimore: University Park Press, 1978, 79-99, Vol. II.

- James, S. Treatment of homosexuality II. Superiority of desensitization/arousal as compared with anticipatory avoidance conditioning: Results of a controlled trial. Behavior Therapy, 1978, 9, 28-36.
- James, S. Orwin, A. and Turner, R.K. Treatment of homosexuality I. Analysis of failure following a trial of anticipatory avoidance conditioning and the development of an alternative treatment system. Behavior Therapy, 1977, 8, 840-848.
- Kazdin, A.E. Artifact, bias, and complexity of assessment: The A B C of reliability. Journal of Applied Behavior Analysis, 1977, 10, 141-150.
- Kazdin, A.E. Modificación de la conducta y sus aplicaciones prácticas. México: El Manual Moderno S.A. 1978. (a).
- Kerlinger, H.F. Investigación del comportamiento: Técnicas y metodología. México: Interamericana, 1975.
- Kinsey, A.C. Pomeroy, W.B. and Martin, C.E. Sexual behavior in human male. Filadelfia: Saunders, W.B. 1948.
- Lathrop, P. Introduction to psychological research: Logic, design, analysis. New York: Harper and Row, 1969.
- Lehrman, N. Las técnicas sexuales de Masters y Johnson. Barcelona: G.E.D.I.S.A. 1979.
- MacConagy, N. Aversive and positive conditioning treatment of homosexuality. Behavior Research and Therapy. 1975, 13, 309-319.

- MacCulloch, M.J. Feldman, M.P. and Pinshoff, J.M. Application of anticipatory avoidance learning to the treatment of homosexuality II: Avoidance response latencies and pulse rate changes. Behavior Research and Therapy, 1965, 3, 21-43.
- Maletzky, B.M. "Boosters" sessions in aversion therapy: The permanency therapy of treatment. Behavior Therapy, 1977, 8, 460-463.
- Maletzky, B.M. and George, F.S. Case histories and shorter communications, the treatment of homosexuality by "assisted" covert sensitization. Behavior Research and Therapy, 1973, 14, 655-657.
- Marks, M.I. Management of sexual disorders. En Leitenberg, H. Handbook of behavior modification and behavior therapy. Englewood Cliff.: Prentice -Hall, 1976.
- Masters, W. y Johnson, E.V. Respuesta sexual humana. Buenos-Aires Argentina: Intermédica. 1978.
- Masters, W. and Johnson, E.V. Homosexuality in perspective. Boston: Little, Brown and Company. 1979.
- Nunnally, C.J. Psychometric theory. U.S.A.: Book Company. 1976.
- Plaza y James, S.A. Diccionario enciclopédico básico. Barcelona: Autor. 1974.
- Risley, T.R. Behavior modification: An experimental therapeutic endeavor. In L.A. Hamerlynk, P.D. Davison and L.E. Acker (Eds), Behavior modification and ideal health service pp 103- 127 Colgary, Alberta Canada: U.C.P., 1970.

- Selecciones del Reader's Digest. El gran libro de la salud.
México: Autor. 1971.
- Solomon, P. y Patch, D.V. Manual de psiquiatría. México.
México: Manual Moderno S.A. 1976.
- Solyom, L. and Miller, A. A differential conditioning as the
initial phase of the behavior therapy of homosexuality.
Behavior Research and Therapy, 1965, 3, 147-160.
- Standford, A.D. Tustin, R.D. and Priest, N.P. Increasing he-
terosexual arousal in two adult male homosexuals using
a differential reinforcement procedure. Behavior Therapy.
1975, 6, 689-693.
- Stanley, R.G. and Wineze, P.G. Orgasmic reconditioning: A con-
trolled study of its effects upon the sexual arousal and
behavior of adult male homosexuals. Behavior Therapy,
1976, 7, 155-166.
- Stullman, Enciclopedia familiar de la medicina y la salud.
New York: Autor, 1964, Vol. I.
- Tanner, A.B. Shock in intensity and fear of shock in the mo-
dification of homosexual behavior in males by avoidance
learning. Behavior Research and Therapy. 1973, II, 213-
218.
- Tanner, A.B. A comparison of automated aversive conditioning
and a waiting list control in the modification of homo-
sexual behavior in males. Behavior Therapy, 1974, 5,
29-32.

Tripp, A.C. La cuestion homosexual. Madrid: E.D.A.F. 1978.

Ullman, L.P. and Krasner, L.A. A psychological approach to abnormal behavior. Englewood Califf. Prentice-Hall, 1975.

Walen, A. Hauserman, R. and Lavin, M. Clinical guide to behavior therapy. New York. Prentice-Hall, 1977.