

“Psicología Experimental”. Capítulo 16. Pereda.

Diseños aleatorios: diseños factoriales.

Son aquellos por medio de los cuales se estudian, en un mismo experimento, los *efectos de dos o más factores* sobre la conducta de los sujetos; y si dichos factores interactúan entre sí a la hora de influir sobre el fenómeno que se estudia. Este diseño aporta una mayor posibilidad de generalización de las conclusiones, permite una mayor economía en el trabajo y favorece el control en la investigación. Ahora al poder predecir lo que ocurriría cuando actuaran dos o más factores combinadamente sobre la conducta, se aproxima más esta situación a la que se produce en los ambientes naturales de los sujetos.

Factor (A, B, C): es un sistema de clasificación de condiciones experimentales. Es la variable que se manipula para comprobar sus efectos sobre una conducta determinada (*variable criterio: Z, Y, X*). Clasificación de factores:

-En función que impliquen o no concepto de magnitud: establecen categorías pero no ordenar: nivel nominal, *factor cualitativo*; establecen orden entre sus categorías: *factor ordinal*; establecen categorías en función de su magnitud: *factores cuantitativos*, nivel de intervalo y de razón.

-En función del nivel de manipulación: *factores de clasificación*: manipulación natural (cuasi experimental), características que ya posee el sujeto, *factores experimentales*: manipulación artificial, asignación libre y aleatoria de unidades experimentales, control absoluto del experimentador.

-En función del procedimiento seguido para elegir los niveles de los factores: *factores al azar*: niveles elegidos aleatoriamente; *factores fijos*: niveles elegidos por el experimentador siguiendo un criterio arbitrario, o se utilizo todos los posibles niveles que puede adoptar el factor.

Niveles del factor: son los valores que puede adoptar un factor elegidos por el experimentador. Cuando se habla de un diseño factorial se emplea la notación A B, siendo A los p niveles que adopta el primer factor y B los q niveles que adopta el segundo. Ej: 2 x 3 diseño factorial con dos factores; el primero con dos niveles y el segundo con tres. Y con 6 tratamientos.

El número de nivel que adopta un factor es dado por la minuciosidad con que el experimentador desea conocer los efectos de dicho factor y por el tipo de inferencias que desea hacer al terminar el experimento. Cuando emplea dos niveles de un factor trata de establecer de forma grosera la relación que le une con una conducta, cuando emplea más de dos niveles trata de establecer una relación más exacta.

El tratamiento experimental es una determinada combinación de niveles de los factores que se quieren estudiar. El número de tratamientos experimentales de que consta un diseño factorial viene determinado por el número de factores y niveles que se utilizan. En un exp. factorial los sujetos a los que se pasa cada tratamiento son una muestra de alguna población que contiene un número infinito de elementos. [ej. 353-354]

Clasificación de diseños factoriales.

-Según el tipo de situación experimental en que se prueban los efectos de los factores:

Diseños factoriales con medidas independientes: situación tipo I, cada tratamiento es probado en un grupo diferente de sujetos. Ej. 2 x 2 habría cuatro tratamientos y cuatro grupos experimentales.

Diseños factoriales con medidas repetidas: situación tipo II, un solo grupo pasa por todos los tratamientos. Ej. 2 x 2 habría cuatro tratamientos y un solo grupo experimental.

Diseños factoriales con medidas repetidas en algunos factores: situación experimental tipo III, en los mismos habría tantos grupos experimentales como correspondiesen a los factores estudiados con medidas independientes y tantos tratamientos como correspondiesen a las dimensiones del diseño. Ej. 2 x 2 habría cuatro tratamientos y dos grupos experimentales.

-Según el criterio de selección utilizado para elegir los niveles de los factores:

Diseño factorial modelo fijo: se utilizan factores fijos y sistemática, no aleatoria.

Diseño factorial modelo al azar: se utilizan factores al azar y la selección de niveles de forma aleatoria.

Diseño factorial modelo mixto: mezcla de ambos modelos anteriores.

Técnica de análisis de datos de los diseños factoriales.

Se debe utilizar el análisis de varianza, o de covarianza (AVAR), dependiendo del tipo de diseño factorial a emplear. Las fuentes de variación que se van a estudiar a través de AVAR, aplicado a los datos recogidos por medio de un diseño factorial van a ser:

-Varianza inter tratamientos (efectos de los factores y efectos de interacción)

-Varianza error. [Ej. 360]

[Ej de representaciones simbólicas de los diseños factoriales, p. 360 a 678]

Efectos que se pueden observar al emplear un diseño factorial.

-Efectos que producen los factores aisladamente:

Efectos principales: los efectos que han producido un nivel de un factor a la variable criterio, al comparar sus resultados con la media total, es decir, la diferencia entre la media paramétrica de la variable criterio bajo el nivel del factor, y la media paramétrica total.

Efectos diferenciales: hacen referencia a las diferencias existentes entre los efectos principales de dos niveles de un mismo factor. Cuanto mayores sean los efectos diferenciales de un factor, es decir, cuanto mayor sea la variabilidad encontrada entre sus efectos principales, más se alejan los resultados de la hipótesis de que ese factor *no* influye sobre la variable criterio del experimento.

-Efectos producidos por la combinación de factores:

Efectos simples: vienen a expresar los efectos de los niveles de un factor bajo cada nivel de los otros factores.

Efectos de interacción: ocurre cuando el efecto que produce uno de los factores en la V criterio cambia en función de los valores que adopta el otro factor. Las interacciones entre dos V se las denomina interacciones de primer orden; entre tres V, interacciones de segundo orden; las de cuatro V, interacciones de tercer orden.

En la vida real las personas pueden comportarse de manera similar o distinta. Estas últimas hacen necesario el estudio de las interacciones. A través del método estadístico (análisis de varianza) y el método gráfico podremos saber si existe una interacción entre dos variables. [Ej. 373-374]

-Efectos del error experimental: son los efectos producidos por las fuentes de V contaminadoras no controladas por el experimentador, influyen en la V criterio.

[Ej de diseño factorial 375 a 378].

“Diseño de Investigación”. Capítulo 7. León y Montero. [Ej en todo el texto].

En un experimento complejo obtenemos más información que en uno simple. La estrategia del diseño factorial consigue que podamos observar cómo se comporta una variable bajo todas las condiciones de la otra. Si hacemos dos experimentos simples consecutivos, podemos dejar fuera una combinación de niveles crucial para entender el fenómeno en estudio.

Interacción: es el efecto *adicional* a la suma de los efectos individuales de las variables independientes. Se dice que no hay interacción cuando este efecto es nulo, una variable actúa independientemente de las condiciones que se den en la otra. Entonces hay interacción cuando el comportamiento de una variable se modifica según los niveles de otra variable.

-Caso de interacción nula: cuando los patrones de los efectos simples son iguales no existe interacción (rectas paralelas).

-Caso de existencia de interacción: cuando los patrones de los efectos simples no son iguales, existe interacción (las rectas no son paralelas).

-Extensión a más niveles: los dos explicado anteriormente se mantienen para cualquier número de niveles y variables.

-Extensión a más de dos VI: la existencia de interacciones y su representación gráfica empieza a ser complicada.

Diseños factoriales con lupa. [leer p.212 a 217.]

Efectos principales y la interacción.

Nuestro conocimiento se modifica cuando sabemos si hay o no interacción, ya que determina el estudio de los efectos principales: 1) Cuando se comprobó que no existe interacción, se debe estudiar los efectos principales. El hecho de que una variable se comporte de forma similar en presencia de los niveles de la otra es lo que hace que tenga sentido estudiar su influencia de forma conjunta. En esto consiste el estudio de los efectos principales: la influencia de las manipulaciones de una variable independiente sobre una variable dependiente a través de los niveles de otra variable independiente. 2) Cuando se comprobó que hay interacción es más seguro detenerse en su estudio y no sacar conclusiones sobre los efectos principales. Se corre el riesgo de llegar a inferencias absolutamente erróneas.

Interacción ordinal: las líneas están unas por encima de las otras, sin juntarse.

Interacción no ordinal: una línea no mantiene un orden respecto a la otra, se juntan al cruzarse.

Casos especiales de interacción.

Son aquellos en los que, tanto gráfica como numéricamente, aparece interacción y sin embargo no es posible explicar teóricamente. Estas interacciones se han producido por restricciones en el rango de medida de la variable dependiente. Son los casos en los que aparece *efecto techo* o *suelo*.

Efecto techo: en algunas de las variables se producen unos resultados con apariencia de interacción.

Efecto suelo: aparece cuando en uno de los niveles se alcanza el mínimo y no es posible registrar niveles inferiores.

“Diseños de un solo sujeto: reversión y líneas base múltiples”. Capítulo 32. Azaroff.

Diseño de un solo sujeto: la investigación de un solo sujeto genera un principio de la conducta aplicable a un individuo particular. A veces son aplicables a un número de sujetos. Cuando se descubre que los mismos resultados son legítimos (pueden replicarse con muchos sujetos) se dice que constituyen un principio general.

Este diseño reduce al mínimo los efectos de uno de los factores de confusión más poderosos en la investigación de la conducta, *las diferencias del cliente individual*. Permite hacer comparaciones entre la conducta de un individuo bajo una condición y bajo otras condiciones. [Ej. 571]. Es necesario que en el transcurso de todo el programa se mantienen constantes todas las condiciones a excepción de los cambios sistemáticos en la VI para evitar que otras V afecten el resultado. Factores como el tiempo y otras V ambientales no se consideran responsables de los cambios conductuales. El analista conductual se interesa por descubrir factores que afecten a la conducta y que estén sujetos a leyes.

-Diseños de reversión (ABAB).

Relación funcional: un cambio en la VI producirá cambios sistemáticos en la VD. Si la VI es un *reforzador*, la VD incrementa. Si la VI es un *estímulo aversivo*, la VD deberá reducirse. Si la VI es en verdad un reforzador, su *retención* después de la conducta (VD) también deberá tener un efecto sistemático. Cuando los cambios en la VD ocurran como una función de la presentación y retención *contingentes* (posibles) de una VI, hay evidencia de una relación funcional. Los diseños de reversión se basan en esta lógica:

FASE A: se mide el rendimiento en la línea base.

FASE B: se introduce la variable independiente.

FASE A (reversión o sondeo): se elimina la variable independiente. (*Retorno* a las condiciones de línea de base).

FASE B: usualmente la variable independiente se introduce de nuevo.

Variaciones de los diseños de reversión: pueden usarse cuando surjan problemas particulares.

-El reforzador se presenta en todas las fases del estudio, pero no contingente a la respuesta meta durante las condiciones repetidas de línea base. [Ej. 573].

-Solo una fase muy breve de línea base. [Ej. 575].

-Adición de una fase RDO, cuando la conducta dependiente no es reversible, la VI puede aplicarse a todas las conductas a excepción de la VD.

Ventajas: demuestra una relación funcional entre la conducta dependiente y la intervención. Proporciona explicación.

Desventajas: el establecimiento y la medición de las condiciones de línea base lleva tiempo, y estas no contribuyen al cambio conductual real. Se teme que la conducta modificada tal vez no se recobre después del regreso a las condiciones línea base, pero se evidencia que una vez que se ha adquirido una conducta, puede volverse con mayor rapidez. Se cuestiona la *ética* de la implementación de sondeos, con base en que podría ser poco ético el deshacer aunque fuera brevemente los efectos positivos de un programa conductual exitoso, pero estas reversiones son justificables debido a que comprueban la independencia para con la intervención.

-Diseños de línea base múltiple.

A través de conductas: varias respuestas se identifican y se miden en el transcurso del tiempo a fin de proporcionar líneas base en comparación con las cuales puedan evaluarse cambios. Con estas línea base establecidas, el experimentador aplica una variable experimental a una de las conductas, produce un cambio en esta, y tal vez observa muy poco o ningún cambio en las otras líneas base. Pero en vez de revertir el cambio acabado de producir, aplica la variable experimental a una de las otras respuestas que aún no han cambiado. Si se altera en ese momento, aumenta la evidencia de que la variable experimental es en verdad efectiva. La variable podrá aplicarse entonces a otra respuesta más, y así sucesivamente.

Para demostrar el control experimental, se debe estar seguro de que las VD no sean interdependientes ni se encuentren muy interrelacionadas; de no ser así, un cambio en una conducta meta también podría cambiar a las otras. [Ej. 581].

A través de individuos: con este diseño se recopila líneas base de la misma conducta con varias personas. Los efectos de la intervención se comprueban primero con un individuo, mientras se continúan las condiciones de línea base con los otros; después se introduce la intervención con otro. Se quiere demostrar que independientemente del tiempo, sujeto específico, y condiciones ambientales, la conducta de cada individuo cambia de modo sustancial sólo cuando se introduce la intervención. Es mejor utilizar dicho diseño con sujetos de diferentes ubicaciones, para que la conducta de uno de ellos no pueda influir las conductas de los demás. [Ej. 582].

A través de situaciones: en este diseño se recopilan datos de una conducta meta para uno o más sujetos a través de diferentes situaciones o circunstancias. Se quiere demostrar que la conducta cambió sólo cuando se le aplicaba el tratamiento. En las situaciones donde el tratamiento no se aplicó, la conducta cambiaría muy poco o nada.

Se debe asegurar de que las situaciones sean lo distintas, de modo que resulte improbable la ocurrencia de una generalización no programada, es decir, generalizar conductas a otras situaciones.

“Investigación cuantitativa en psicología”. Capítulo 11. Clark.

Muestras y poblaciones.

Estadísticas: como la media, varianza y desviación estándar, describen la muestra que se midió. Cada estadística tiene un equivalente parámetro.

Parámetros: describe la población de la cual proviene la muestra. El equivalente de la media de la muestra es la media de la población; el eq. de la varianza de la muestra es la varianza de la población; el eq. de la desviación estándar de la muestra es la desviación estándar de la población.

Hipótesis: hará una afirmación más general sobre la población de que proviene la muestra.

Elección de una muestra: la muestra debe ser representativa de la población más amplia. Los investigadores deben definir cuál es su unidad de análisis (gente o elementos de la población). Por último se deben colocar restricciones, es decir, factores de limitación a lo que constituye un elemento de población. [Ej. 158].

Muestras aleatorias: cada elemento de población tiene una probabilidad igual, o una probabilidad cuantificable, de ser seleccionado. Con esta, es posible generalizar con cierto grado de exactitud los resultados obtenidos de la muestra a la población.

Muestreo aleatorio simple: una vez elegida la población, se debe determinar el tamaño de la muestra, identificar cada elemento de la población, seleccionar códigos al azar hasta que se identifiquen todos los elementos de la posible muestra.

Pueden ocurrir problemas con la identificación de los elementos de población, que una muestra quede *sesgada*.

También para identificar elementos de una población, debe llevarse a cabo *una encuesta más amplia*. [Ej. 159].

Muestreo sistemático: se debe decidir un tamaño de muestra y que luego se divida el de la población entre el de la muestra; esto da una cifra que se usa como base del muestreo. [Ej. 160].

Muestreo estratificado: requiere la división de la población en subgrupos o estratos mutuamente excluyentes, bien representativos. Dos maneras de conducir el muestreo estratificado: muestreo proporcionado (se aplica si el muestreo del estrato refleja las proporciones de la población); muestreo desproporcionado (si no se requiere que la muestra tenga las proporciones de la población). [Ej. 160].

Muestreo de conglomerados: requiere un muestreo inicial con base en una unidad mayor que el elemento de población. Esto puede hacerse de dos maneras: de una sola etapa (ej se desea encuestar a estudiantes de psicología en Gran Bretaña, identifican todos los lugares donde se imparten cursos sobre la carrera, se seleccionan al azar varios cursos y se encuesta); de varias etapas (ej. se desea encuestar jóvenes de secundaria, se identifican las autoridades educativas de Gran Bretaña y se seleccionan al azar, se identifican escuelas y se seleccionan al azar, se encuentran los alumnos). Se debe tratar de incluir si o si a aquellas personas que no estén disponibles al momento de llevar a cabo el muestreo.

Muestras no aleatorias: no tienen exactitud.

Muestreo accidental: muestreo de las personas que uno encuentra por casualidad. [Ej. 162]

Muestra de cuotas: se establecen cuotas (cantidad) para el número de personas que habrán de incluirse en las submuestras, ej. si desean un número igual de mujeres y varones, una vez que obtengan su cuota para un género se acercarán sólo a miembros del otro. Si las cuotas se establecerán en varias dimensiones, suele usarse el término muestreo dimensional, ej de personas con varias características a la vez, se debe buscar gente muy precisa.

Muestreo propositivo: se utiliza para cuando se desea estudiar una muestra claramente definida. También cuando buscan participantes con características particulares. [Ej. 163]

Muestreo de bola de nieve: usa contactos iniciales para identificar otros posibles participantes. [Ej. 163].

En una **muestra aleatoria** es posible generalizar los resultados de la muestra de la población con cierto grado de exactitud. Esto se obtiene al calcular un *intervalo de confianza*. Nunca podemos estar seguros de que nuestro estimado de un parámetro es exacto, pero sí podemos encontrar un rango de valores que nos de cierto nivel de confianza de que el parámetro debe caer en él. Tal rango se llama intervalo de confianza. Se expresa en porcentaje. Por lo menos tres factores influyen en el tamaño del intervalo de confianza para el mismo grado de confianza: la proporción de la muestra para la que se calcula el intervalo de confianza, el tamaño de la muestra y los tamaños relativos de la muestra y de la población.

Efecto de la proporción en el intervalo de confianza: cuanto más se aleje de 0.5 (50%) la proporción para la que se estima el intervalo de confianza, tanto menor será este.

Efecto del tamaño de la muestra en el intervalo de confianza: cuanto mayor sea la muestra tanto menor será el rango del intervalo de confianza para el mismo nivel de confianza, es decir podemos determinar con mayor exactitud el parámetro de la población.

Efecto del tamaño de la muestra como una proporción de la población: cuanto mayor sea la muestra como proporción de la población, más exacto sería el intervalo de confianza.

Efecto del grado de confianza en el tamaño del intervalo de confianza: cuanto más niveles de confianza desee tener acerca del lugar que se encuentra el parámetro, tanto mayor será el margen de error y el intervalo de confianza. [Ej. 164]

“Técnicas psicométricas. Cuestiones de validez y confiabilidad”. Argibay.

Confiabilidad: toda medición tiende a presentar errores, de manera que el *valor observado* en la medición está formado por el *valor verdadero* y por el *error de medición (errores aleatorios, los cuales no pueden ser controlados ni predichos)*. Por lo tanto un instrumento será más confiable en la medida que *maximice el valor verdadero* (aproxime lo más posible al valor observado) y *reduzca el error*, incrementando así la similitud entre el valor verdadero y el obtenido, y la confiabilidad. [Ej. 17].

La confiabilidad es la consistencia en un conjunto de medidas de un atributo; es la proporción de la variabilidad verdadera respecto de la variabilidad obtenida. En el análisis de confiabilidad debemos considerar tres aspectos:

-**Congruencia o consistencia interna:** consiste en que las distintas partes que componen el instrumento está midiendo lo mismo. Al hablar de confiabilidad del instrumento, estamos diciendo que los reactivos que componen el instrumento son distintas formas del mismo atributo, de manera que estaríamos frente al mismo constructo, observando conductas

diferentes. La congruencia interna del instrumento se establece según la magnitud de los valores que expresan las correlaciones entre sus partes. [Ej. 18]. Los procedimientos para calcularla son dos métodos:

-el primero basado en la división del instrumento en dos mitades: se divide al instrumento en dos mitades, se obtiene la puntuación para cada mitad en forma independiente, y se usan esos puntajes para estimar el coeficiente de correlación (mediante la fórmula de Spearman-Brown) entre ambas mitades.

-el otro es basado en la covarianza de los ítems: se trata a cada ítem como si fuese un test de longitud unidad con una puntuación, coeficiente alpha de Cronbach. El valor de alpha se ve afectado por el número de ítems, entre dos test que tuvieran la misma confiabilidad, el que tuviera menos ítems daría un menor coeficiente alpha. Esto se debe a que al aumentar la longitud del instrumento la varianza aumenta en mayor proporción que la varianza de error. Cronbach propuso una fórmula adicional de consistencia inter-ítem, consiste en una estimación de la correlación media inter-ítem y no se vería afectada por la longitud del instrumento. Puede ser útil su aplicación cuando se desea comparar la confiabilidad de dos instrumentos con número desigual de ítems o de dos factores dentro de un instrumento con un número de reactivos desigual.

Al interpretar el coeficiente de confiabilidad, se debe considerar cuán próximo está el valor máximo (uno), evaluar el margen de error y considerar las distintas circunstancias que involucren el uso del instrumento y tener en cuenta el número de ítems que contenga el instrumento. Es importante que los instrumentos sean confiables, ya que si no lo fuese podría afectar su validez y distorsionar la medición de la variable de interés.

Si un instrumento es errático, es difícil que pueda medir con validez el atributo en cuestión. Sin embargo, un instrumento puede ser confiable y al mismo tiempo no válido.

-**Estabilidad:** lo que se observa es en qué grado se obtienen las mismas medidas al aplicar dos veces el mismo instrumento, mediando entre ambas tomas un tiempo determinado, ya que pueden haber intervenido una serie de factores (modificaciones en la actitud de los sujetos, efectos derivados de la primera aplicación del test que afecten el segundo rendimiento, lapso de tiempo entre ambas tomas, la maduración, la historia, etc.) que sean fuentes de varianza de error, disminuyendo la confiabilidad de las medidas. Uno de los problemas de las técnicas para evaluar la confiabilidad es que pueden confundir las fluctuaciones aleatorias con cambios que se producen realmente en el atributo y estos modificar el puntaje. [Ej. 23]. También puede haber otras fuentes de errores sistemáticos, que alteran los puntajes del test, sin modificar el atributo, ej. factores que modifiquen el contexto de la medida. Los efectos derivados de la historia constituyen un déficit de confiabilidad del instrumento que deberá ser contemplado. Las diferencias de medida (fuentes de variación sistemática) deberían ser controladas dentro de lo posible, para que no sean incluidas como déficit de confiabilidad.

El procedimiento consiste en aplicar la misma prueba en dos momentos diferentes al mismo grupo de sujetos y luego, correlacionar entre sí, los puntajes obtenidos en cada toma. Esta correlación sirve para establecer la confiabilidad del instrumento. Debido a los errores de medición, hay un porcentaje de error aleatorio, incluido en la medición. La correlación debería ser en estos casos alta, ya que se esperaría que estas fuentes de error se vieran minimizadas, y que en la aplicación del instrumento no se incurriera en errores sistemáticos.

-**Equivalencia:** se aplica cuando se quiere determinar la confiabilidad de dos instrumentos que se consideran paralelos, y poder establecer la equivalencia de los mismos que suponen medir el mismo constructo al aplicarlos a los mismos sujetos. Se aplican los dos instrumentos simultáneamente o no, y si ambos son equivalentes tendría que obtenerse un coeficiente de correlación elevado. Otra forma de equivalencia consiste en determinar la confiabilidad entre evaluadores. Se puede aplicar a aquellos instrumentos en los cuales dos o más evaluadores tienen que calificar o puntuar la conducta o rendimiento como parte de la aplicación del instrumento. [Ej. 26].

-**Validez:** tiene que ver con poder determinar si el instrumento está midiendo realmente el atributo que dice medir. La confiabilidad es una cuestión empírica, la validez teórica, ya que la evaluación persigue la explicación. Existen tres tipos:

⇒ de contenido: se evalúa si todos los ítems, de un conjunto de conductas, utilizados para construir el test están dentro del dominio de interés y si constituyen una muestra representativa del universo de conductas.

⇒ de criterio: se persigue relacionar las puntuaciones del instrumento con otras variables llamadas criterio. Es útil para cuando deseamos hacer inferencias a partir de los puntajes que se obtienen en el test respecto de alguna otra variable de interés. [Ej. 27]. Existen dos tipos de validez: *predictiva*: si las puntuaciones del test se utilizan para predecir alguna medida del criterio que se va a realizar a futuro; *concurrente*: si por el contrario relacionamos las puntuaciones del test con alguna medida del criterio tomada en el mismo momento. [Ej. 27]. Ambas tienen el poder de predecir, la diferencia entre estas tiene que ver con el diseño que involucran en cuanto a su dimensión temporal. La predictiva implica un diseño prospectivo, la concurrente un diseño transeccional o transversal.

⇒ de constructo: consiste en tratar de probar que las conductas (no observables y que deben ser operacionalizadas) que registra el test pueden ser consideradas indicadores válidos del constructo al cual refieren. La validación de constructo no se trata de hacer corresponder la puntuación del test con el constructo, ya que un mismo constructo puede tener varios indicadores, es decir, varias operacionalizaciones. Se trata de establecer que las puntuaciones del test constituyen en forma válida una de las manifestaciones del constructo. La validez de constructo, es el principal tipo de validez y la

más difícil de comprobar. Si queremos establecer que las conductas que registra el test, son indicadores válidos del constructo, tendremos que considerar que solo podremos aproximar a ese objetivo. Solo aportar una serie de evidencias. Algunos de los procedimientos para evaluar dicha validez son:

-análisis factorial: se utiliza cuando el test está dividido en factores y sirve para medir la validez de constructo, analiza las intercorrelaciones de un conjunto de datos para establecer determinadas agrupaciones de ítems correlacionados entre sí, las cuales remiten a factores subyacentes que no son observables. No solo se utiliza para evaluar la validez del instrumento, también en su construcción, una vez construido y establecidos los correspondientes factores, la técnica puede aplicarse sobre los datos obtenidos a partir de una muestra de sujetos para establecer si la estructura factorial planteada, puede ser replicada, lo cual nos permitiría hablar de la validez factorial del instrumento. Las estructuras factoriales pueden tender a ser inestables y dependen del tamaño de la muestra.

-diferenciación entre grupos: se aplica el instrumento a dos o más grupos (deben diferir en el atributo que se mide).

-correlaciones con otras medidas del constructo: cuando ya existe otro test validado que mide el mismo constructo, se puede correlacionar el nuevo con aquel o con los tests ya existentes para establecer su validez de constructo.

-matriz multimétodo-multirrasgo: se precisa que existan como mínimo dos métodos diferentes para medir el constructo que se va a validar y se necesitan otros constructos que puedan ser medidos por los mismos métodos. Se miden en los sujetos de la muestra los distintos constructos con método diferentes. Se calculan las correlaciones entre todas las medidas y se forma con ellas una matriz que contendrá los siguientes datos: coeficientes de fiabilidad (correlaciones obtenidas entre medidas del mismo constructo con el mismo método); coeficientes de validez convergente (correlaciones entre medidas del mismo constructo obtenidas con método diferentes); coeficientes de validez divergente (correlaciones de constructo diferentes medidos con igual método y correlaciones de diferentes constructos medidos con distinto método).

“Selección de sistemas de observación”. Capítulo 5. Azaroff.

Medición exacta del comportamiento: se usa para determinar con claridad y objetividad la efectividad de los programas de análisis conductual aplicado y para generar predicciones científicas para el futuro. Este requiere de la selección e implementación de mediciones *objetivas, válidas y confiables*. Proporcionan los datos, o evidencia, sobre cualquier cambio que ocurra en un programa de análisis conductual, sobre estos datos descansa la confiabilidad del análisis de conducta. Objetividad significa que los observadores que registran los datos conductuales no se dejaron influenciar por sus propios sentimientos o interpretaciones. Las medidas válidas son aquellas que miden en efecto de forma directa a la conducta que se supone que miden. Deben ser objetivas. Un instrumento confiable será estándar con independencia de la persona que lo use y de las ocasiones en que se aplique. El objetivo conductual incluye a la conducta meta (comportamiento que habrá de alterar), las condiciones bajo las que el cambio habrá de ocurrir, y los estándares para evaluar la consecución del objetivo, todo está listo para el desarrollo de un sistema efectivo de medición.

Selección de variables dependientes: cuando una conducta meta, se traduce en términos cuantificables, la medida del suceso es la VD. La medida de la conducta meta es la VD clave; pero también puede ser importante medir otras variables, las medidas de las cuales pueden alterarse en función del programa. Una vez que se han seleccionado las VD, se debe escoger un sistema de medición apropiado. [Ej. 82].

Selección de medidas válidas: un sistema de medición válida debe ser apropiado a la variable que intenta medir. Se suele lograr un consenso general (todas las personas participantes estén de acuerdo) sobre lo que se considera una medida válida. [Ej. 82-83].

Selección de medidas confiables: la medición conductual requiere de *consistencia*, es decir se debe demostrar que la conducta bajo observación se mide de la misma manera, bajo el acuerdo entre dos observadores quienes registran simultáneamente. [Ej. 84-85].

Técnicas de registro de la conducta.

-**Mediciones de productos permanentes:** utilizados para comportamientos que dejan evidencia física en la forma de un producto duradero, ej. número de camas hechas. Ventajas de estas mediciones: la validez de las mediciones cuanti/cualitativas, puede demostrarse con facilidad; como los productos permanentes pueden medirse después de un lapso de tiempo, la reactividad (efectos ocasionados por los mismos procedimientos, que pueden distorsionar la validez de los datos) puede reducirse al mínimo. Es más fácil medir los productos permanentes para evaluar la confiabilidad de la medición. [Ej. 86].

-**Medición de los sucesos transitorios:** utilizados para comportamientos los cuales no generan productos permanentes, ej. sonreír. Para medir estas conductas, es necesario registrar observaciones con un observador e instrumentos en vivo. Registro de sucesos: consiste en contar las veces que ocurre una conducta determinada en un intervalo específico. Este registro es apropiado para medir respuestas discretas (aquellas que tienen un inicio y un final claramente definidos). Confiabilidad del registro de sucesos: se coloca a un segundo observador a cierta distancia del primero y se le pide a ambos que registren el comportamiento simultáneamente. Así, podrían compararse los dos totales. Cuando la observación informal sugiere que los observadores perdieron algunas respuestas meta, para asegurar la confiabilidad del registro de sucesos, los periodos de observación deberán dividirse en intervalos.

Registro de duración: usando un reloj, cronómetro etc. Para una mayor precisión pueden compararse las duraciones dentro de los intervalos. [Ej. 87 a 90].

“Registro de intervalo e implementación de sistemas de observación”. Capítulo 6. Azaroff.

Registro de muestra temporal de intervalo (MTI): se utiliza para conductas que no son claramente discretas, es decir, que es difícil decir cuándo comienza y finaliza una respuesta. Cuando se necesita sacar una muestra de varios comportamientos simultáneos se suele usar una *hoja de registro codificado de intervalo*. [Ej. 96-97].

Desventaja: no resulta práctico para el estudio de conductas importantes pero infrecuentes.

Muestreo temporal de intervalo completo: si el sistema requiere que la respuesta se emita en todo el intervalo completo para que su presencia se registre. Se utiliza cuando es importante saber que la conducta no es interrumpida.

Muestreo temporal de intervalo parcial: si se requiere solo de una muestra de la respuesta dentro del intervalo para registrar. Se utiliza para registrar conductas transitorias (gestos o palabras).

Muestreo temporal de intervalo momentáneo: si el sistema requiere que la respuesta ocurra en el momento en que termina el intervalo. Son apropiados para conductas como los movimientos estereotipados. [Ej. 95].

Validez y confiabilidad del registro de intervalo: el MTIC subestima a la medición continua (apropiado para cuando se busca un incremento en la VD). El MTIP sobreestima a la medición continua (adecuado para cuando se busque una reducción de la conducta). El MTI está fuera de lugar (más apto para sacar muestras de conductas de mayor duración). La estimación de una conducta que ocurre con frecuencia pero que no se presta al registro de de suceso, deberá ser posible con intervalos de muestreo temporal breve, ya que mientras más dure el intervalo, más observaciones se perderán.

Por lo general es preferible seleccionar una medición que sea “conservadora” en relación con el resultado buscado. Independientemente del sistema de intervalo utilizado, es necesaria la consistencia en todas las fases de línea base y de tratamiento.

La confiabilidad de la medición en el registro de intervalo se basa en los mismos fundamentos que en otros sistemas de registro. La conducta debe operacionalizarse con precisión y los registradores deben estar entrenados y supervisados. Las verificaciones de confiabilidad deberán hacerse antes de recoger los datos de línea base, y durante cada fase del programa. Un porcentaje alto de acuerdo sugiere que las técnicas de registro son confiables. El acuerdo deficiente entre los observadores suele ser el resultado de la carencia de definiciones operacionales adecuadas.

Implementación de sistemas de observación.

-El personal de registro conductual: los técnicos conductuales deben diseñar e implementar sistemas de registro observacional, supervisar la recolección de datos y diseñar y ejecutar esquemas gráficos para el monitoreo continuo. La observación y registro conduce al cambio conductual y disminuye la carga de trabajo de los técnicos conductuales. El autorregistro es una parte de la administración o programas de autocontrol. Es exacto y confiable, pero no sencillo de implementar. La autoobservación y el autorregistro tienen algunos problemas. Posiblemente la conducta productiva no se mantenga a largo plazo, tal vez no dé lugar a ningún cambio conductual a no ser que vaya acompañado por el refuerzo de la conducta de meta. Por eso es conveniente que los autorregistradores tengan un interés creado por el progreso del programa. Se deberá interrumpir aleatoriamente las actividades usuales para observar y registrar la conducta.

-Sistema de registro automatizado: hay veces que no es práctico ni conveniente tener a un observador presente para registrar la conducta. Los sucesos conductuales transitorios pueden preservarse al grabarlos.

-Capacitación y supervisión del personal de registro conductual: la subjetividad del observador puede confundir la medición. Para ello, deberán tomarse mediciones independientes al menos por dos observadores varias veces.

-Mantenimiento de la confiabilidad observacional: deberán hacerse verificaciones frecuentes de confiabilidad en todo el transcurso de un programa y a los observadores deberá informarles de ello. Un segundo método es el de reforzar la exactitud.

-El registro de los datos conductuales: esto permite al analista conductual estimar el funcionamiento actual de los sujetos en lo que se refiere a sus conductas meta. Los datos se registran y luego se grafican.

-Adaptación: no es apropiado comenzar a recolectar los datos línea de base, que proporcionan un estándar para la medición del cambio futuro, tan pronto como un observador entre un nuevo ambiente, ya que el comportamiento inicial no es típico. Deberá transcurrir el *periodo de adaptación* al nuevo ambiente antes de que se inicie formalmente la fase de línea base.

-Medición de línea base: cuando se presume que la conducta ha retornado a su estado típico, la adaptación es un hecho. La fase de línea base consiste en mediciones repetidas de las VD en el transcurso de varios días. Esta línea base servirá como estándar con el cual se puede comparar el procedimiento de tratamiento.