

Evaluación de patrones de aprendizaje como cambio

R. López Feal

Universidad de Barcelona

INTRODUCCION

Desde finales del siglo pasado hasta la primera mitad del siglo actual las investigaciones sobre aprendizaje han sido patrimonio casi exclusivo de los laboratorios de psicología experimental y han estado centradas prioritariamente en los métodos de condicionamiento clásico y de condicionamiento operante. Teniendo en cuenta sus rigurosas condiciones de control y la simplicidad de su paradigma E-R asociado, en estos tipos de aprendizaje no se plantean problemas fundamentales de medición ni a nivel teórico ni a nivel operacional. Tampoco se plantean problemas relevantes de medición entre los teóricos del aprendizaje cognitivo de orientación conductista. Sin embargo, estos problemas surgen desde el momento en que entra en crisis el paradigma clásico Estimulo-Respuesta o sus variantes. En esta crisis influyen dos factores determinantes:

- a) Uno que tiene su origen en la teoría cognitiva del aprendizaje y que supone, como señala Simon (1980), la sustitución del paradigma Estimulo-Respuesta por el paradigma Señal-Respuesta. Este cambio lleva, por una parte, a la ampliación de los estímulos en forma de energía física a estímulos en forma de símbolos y, por otra, a la ampliación de las investigaciones sobre aprendizaje a formas y situaciones más complejas, centradas fundamentalmente en el ámbito del aprendizaje humano.
- b) El otro factor determinante de la crisis del paradigma clásico surge desde el momento en que la medición del aprendizaje no sólo se tiene en cuenta la estimación de la influencia de la intervención de los estímulos o señales sino también la de la influencia que pueden tener en el aprendizaje características diferenciales entre individuos o grupos, tanto a nivel genético como ambiental. Han sido numerosas las publicaciones sobre este tema durante los últimos años. Entre ellas destacamos, desde la problemática de la medición, el «reading» editado por Dillon (1983) que recoge trabajos de los principales expertos en el campo de la investigación sobre diferencias individuales en las estructuras y funciones cognitivas.

Esta ampliación del marco conceptual del aprendizaje obliga a tener en cuenta no sólo datos conductuales sino también datos mediacionales producidos por procesos de pensamiento, motivación y personalidad y por antecedentes remotos más allá de los *inputs* inmediatos. Al mismo tiempo, el hecho de que cada vez tenga mayor acepta-

ción la existencia de distintos patrones de aprendizaje que tienen su origen en características diferenciales de tipo genético y ambiental está contribuyendo a que la teoría del aprendizaje deje de ser patrimonio exclusivo de la Psicología Experimental y a que sus problemas teóricos y metodológicos sean compartidos por la Psicología Diferencial. Con ello se está dando un paso más hacia la unificación de las dos metodologías tradicionales de la psicología expuestas formalmente por Cronbach, ya en el año 1957.

Antes de concluir estas reflexiones previas al desarrollo de los puntos que vamos a tratar en este artículo, queremos plantear la necesidad de distinguir entre el concepto de aprendizaje como adquisición de esquemas o patrones verbales, motóricos, conceptuales o de otro tipo y el uso o ejecución de esos esquemas. En el primer caso, el aprendizaje, como ciencia pura, está interesado por los principios generales que rigen estos esquemas o patrones. En el segundo caso el interés prioritario está centrado en la utilización de estos patrones una vez adquiridos y en los factores que producen distintos tipos de cambio en dichos patrones como consecuencia fundamentalmente de la intervención en contextos de laboratorio, educacionales, laborales, familiares, sociales o clínicos.

Hechas estas aclaraciones, vamos a entrar en el desarrollo de la temática planteada que estará centrada, fundamentalmente, en la vertiente aplicada del aprendizaje, dado que es aquí donde tienen un papel más relevante las técnicas de medición del aprendizaje como cambio.

Como referencia para tratar la medición y estimación del aprendizaje desde el contexto de cambio vamos a tomar los siguientes puntos: En primer lugar plantaremos el marco conceptual del aprendizaje como cambio y luego expondremos los distintos tipos de diseños, los instrumentos de medición y alguno de los métodos de tratamiento de datos sensibles a la estimación del cambio.

MARCO CONCEPTUAL DEL APRENDIZAJE COMO CAMBIO

Para el estudio del aprendizaje o de cualquier otro proceso de cambio hay que introducir dentro del mismo sistema conceptual lo que los físicos han denominado inercia temporal y decaimiento entrópico (Dwyer, 1983, págs. 333-350); es decir, hay que tener en cuenta simultáneamente un componente o patrón que puede permanecer estable y un componente de cambio producido por una intervención externa, por un proceso de maduración o por interacción de ambos factores. Así, de acuerdo con el componente de inercia temporal debemos asumir que los patrones de aprendizaje manifiestan alguna clase de tendencia a permanecer estables sobre el tiempo y de acuerdo con el decaimiento entrópico manifiestan simultáneamente una tendencia al cambio, dando lugar a un equilibrio inestable entre fuerzas que tratan de mantener el funcionamiento del sistema de acuerdo con los patrones iniciales y fuerzas que tratan de cambiar estos patrones.

Estos dos componentes deben tenerse en cuenta en cualquier estudio sobre aprendizaje o sobre cualquier otro tipo de cambio, aunque en Psicología no siempre se ha considerado así, como lo demuestra el hecho de que cada uno de estos componentes ha sido estudiado indepen-

dientemente por la teoría clásica de los tests, que se ha centrado en la obtención de estimadores de estados permanentes y por la Psicología Experimental, más preocupada por el estudio y estimación de cambios producidos por intervención o manipulación externa. Otros conceptos, también procedentes del campo de la física, que hay que introducir en el estudio del cambio y del aprendizaje como cambio, son, de acuerdo con Golembiewski y Billingsley (1980), por una parte, el cambio de grado dentro de un mismo estado, como, por ejemplo, el cambio de temperatura del agua en estado líquido, y, por otra parte, el cambio de estado, como sería el caso cuando el agua pasa del estado líquido al gaseoso por un aumento adecuado de temperatura.

La necesidad de integrar dentro de un mismo sistema los conceptos de inercia temporal, cambio, cambio de grado y cambio de estado han llevado a Golembiewski, Billingsley y Yeager (1976) y a Zmud y Armenakis (1978) a proponer una tipología de tipos de cambio que pretende cumplir con un fin similar al planteado por Campbell y Stanley (1966) para la organización de hipótesis rivales dentro de categorías de invalidez interna y externa.

Los autores anteriormente citados distinguen en su tipología tres clases de cambio a los que denominan, respectivamente, cambio Alpha, cambio Gamma y cambio Beta, definiendo cada uno de los tipos del siguiente modo:

- a) El cambio Alpha comprende una variación en el nivel de algún estado existencial, dado un instrumento de medida constantemente calibrado y relacionado con el dominio conceptual. Este tipo de cambio se correspondería con el cambio de grado dentro de un mismo estado existencial producido por maduración, manipulación, intervención o por influencias del entorno no controlado. Señala Schmitt (1982) que el análisis usual del cambio Alpha comprende la comparación de las medias de respuestas pretest y postest, después de una intervención.
- b) El cambio Gamma comprende una redefinición o reconceptualización de algún dominio, lo que supone un cambio substantivo en la estructura de referencia dentro de la que los fenómenos son percibidos y clasificados. Este tipo de cambio ocurre cuando un sujeto o un grupo cambian cualitativamente sobre el tiempo su conocimiento de un concepto, sus rasgos de personalidad, su estructura cognitiva, sus hábitos de comportamiento o cualquier otra característica de su sistema psicológico. Así, por ejemplo, conceptos asociados a procesos cognitivos o hábitos de comportamiento pueden significar cosas completamente distintas en distintos momentos evolutivos o antes y después de una intervención o manipulación que pretende incrementar el conocimiento de un concepto, modificar actitudes o cambiar hábitos de comportamiento. En consecuencia, mientras que el cambio Alpha representaría sólo cambio de magnitud dentro de un mismo estado, el cambio Gamma implicaría cambio de estado. Es decir, el cambio Alpha no exige una redefinición del concepto original y el cambio Gamma exige una nueva redefinición de dicho concepto original. Zmud y Armenakis (1978) ponen como ejemplo de cambio Gamma el concepto abstracto de «liderazgo» que puede significar cosas completamente distintas para los sujetos en un tiempo T1 y T2,

especialmente si a través de un proceso de intervención se incluyen sesiones que pretenden incrementar el conocimiento de este concepto. Así, el cambio Gamma constituye la finalidad de muchas intervenciones planificadas.

- c) El cambio Beta comprende una variación en el nivel de algún estado existencial caracterizado por el hecho de que alguno de los intervalos del continuo de medida asociado a un dominio conceptual ha sido recalibrado. Es decir, este tipo de cambio ocurre cuando los sujetos recalibran la escala de medida sobre dos o más administraciones del mismo instrumento de medida, aumentando o reduciendo los intervalos entre puntos sobre una escala. Zmud y Armenakis (1978) ponen como ejemplo de cambio Beta el de la percepción de cierta conducta con el valor «2» de una determinada escala en el tiempo T y el de la percepción de la misma conducta con valor «3» en el tiempo T. Así, aunque cambio Beta implica cierto cambio cualitativo, al contrario de lo que ocurre con el cambio Alpha, dicho cambio no implica cambio estructural completo, que es lo que ocurre con el cambio Gamma.

Como síntesis de esta breve exposición sobre la conceptualización del cambio, recogemos la propuesta de Golembiewski y Billingsley (1980) en el sentido de que el estudio planificado del cambio, y del aprendizaje como cambio, debe estar guiado por el siguiente paradigma primario: primero conceptualizar, luego investigar y después reconceptualizar, si éste es el caso.

TIPOS DE DISEÑOS ASOCIADOS A LA MEDICION DEL APRENDIZAJE COMO CAMBIO

Aunque los estudios sobre aprendizaje se enmarcan en el contexto más amplio de los estudios que tratan la estimación del cambio, poseen la característica diferencial de que tienen como objetivo prioritario la estimación del cambio producido por una intervención externa planificada y controlada.

Durante muchos años las investigaciones sobre aprendizaje han estado vinculadas a los laboratorios de psicología experimental. Sin embargo, desde la publicación del artículo de Cronbach (1957), se inicia una vía de unificación de la metodología experimental y no-experimental que influye tanto en la ampliación de los contenidos del campo del aprendizaje objeto de estudio como en la ampliación de los diseños de recogida de datos y de los métodos de tratamiento de dichos datos.

Entre las numerosas publicaciones que más han contribuido en los últimos años a la potenciación de esta vía unificadora iniciada por Cronbach y a la consecución de taxonomías integradoras de diseños de observación y de métodos asociados para la estimación de cambio sobre el tiempo destacamos las de Campbell (1963), Campbell y Stanley (1966), Kenny (1975) y Cook y Campbell (1979). Dwyer (1983) integra estas taxonomías en dos tipos: uno que agrupa los diseños de caso único y el otro los diseños longitudinales. El primer tipo comprende aquellos diseños que recogen observaciones sobre un caso único o sujeto único en tantos puntos en el tiempo como sea posible, partiendo de un mínimo de dos momentos temporales. En cuanto a los diseños de investigación longitudinal, los define como diseños de observación

de una muestra de casos o sujetos en dos o más puntos o momentos temporales y, siguiendo la taxonomía de Cook y Campbell (1979), divide este tipo de diseños en tres clases bien diferenciadas. La primera clase está constituida por el tradicional diseño de experimento verdadero o diseño aleatorizado que se caracteriza por la selección al azar de dos grupos equivalentes de sujetos, uno denominado grupo control, no sometido a tratamiento, intervención o manipulación planificada, y el otro denominado grupo experimental, sometido a un tratamiento, intervención o manipulación, una vez seleccionado. La segunda clase de diseños de este tipo está constituida por el diseño de grupo de control no equivalente, denominado también diseño cuasi-experimental longitudinal con medidas pre y postratamiento. En esta clase de diseños también se da un grupo de control y un grupo experimental sometido a tratamiento. Sin embargo, al contrario del experimento verdadero, la asignación de los sujetos a cada uno de los grupos no se hace al azar sino en función de diferencias psicológicas, demográficas, sociológicas o de otro tipo, establecidas antes del tratamiento. Entre los criterios diferenciales más frecuentemente utilizados para la asignación de sujetos a cada uno de los grupos se encuentra el de establecer un punto de corte a través de un test pretratamiento. Por último, la tercera clase de este tipo de diseños está constituida por los denominados diseños longitudinales pasivos o diseños correlacionales en terminología tradicional. Estos diseños se diferencian de los incluidos en las otras dos clases anteriores en que no se da ningún tipo de intervención experimental entre las mediciones obtenidas en distintos momentos temporales. Además, dichos diseños pueden utilizarse con un solo grupo de sujetos o con más de un grupo, aunque en este último caso no hay grupo experimental sino dos o más grupos diferenciados en alguna característica psicológica, sociológica o de otro tipo.

Una vez expuestas las características generales de los diseños propuestos hasta la fecha para la medición del cambio, vamos a plantear brevemente las posibilidades de aplicación de cada uno de estos diseños en el contexto de la medición del aprendizaje como cambio producido por intervención planificada.

Los diseños de caso único se utilizan en situaciones manipulativas de laboratorio, sobre todo con animales, en situaciones escolares de enseñanza individualizada, en enseñanza compensatoria y en situaciones de intervención terapéutica planificada. La línea base, que se toma como punto de partida para la intervención en este tipo de diseños, es difícil de objetivar, sobre todo en situaciones donde no hay control de laboratorio. También son de difícil control fuera de laboratorio los procesos de intervención. Por otra parte, como veremos en el punto correspondiente a instrumentos de medición del aprendizaje, estos diseños, cuando se utilizan en el contexto del aprendizaje humano, exigen instrumentos de medida pretratamiento y postratamiento distintos de los tradicionales de norma de grupo.

Dentro de los diseños longitudinales, los diseños de experimento verdadero son los más rigurosos, ya que permiten tanto un control inicial por aleatorización de los grupos como un control de los procesos de tratamiento o intervención, sobre todo en situaciones de laboratorio. Por otra parte, al contrario de los otros diseños longitudinales, tienen potencialmente la posibilidad de establecer empíricamente relaciones causa-efecto. En consecuencia, con esta clase de diseños se pueden controlar las múltiples fuentes de invalidez interna y externa enu-

meradas por Campbell y Stanley (1966), que amenazan la objetivización de los efectos del tratamiento, finalidad prioritaria de esta clase de diseños.

En cuanto a los diseños longitudinales de grupo de control no equivalente, tienen la ventaja de que, al contrario de los experimentos verdaderos, se abren a la investigación diferencial, al seleccionar el grupo experimental y el grupo de control a partir de una o más características diferenciales de distintos tipos, existentes o impuestas, entre los dos grupos. La tarea principal en esta clase de diseños es separar el cambio debido a efectos de un tratamiento del cambio debido a efectos de las otras variables diferenciales y de otras variables no controladas. Tampoco podemos olvidar en esta clase de diseños, como señala Dwyer (1983) y Kenny (1979), la incertidumbre acerca de la influencia de «terceras variables» o variables espúreas que pueden estar correlacionadas con la manipulación y tener una relación causal, desde fuera del sistema formalizado, con la variable dependiente. En consecuencia, en estos diseños no aleatorizados las múltiples amenazas de invalidez interna y externa sólo se pueden eliminar controlándolas sucesivamente. Para ello, el contexto de laboratorio es más propicio que las situaciones de campo, aunque éstas son las más frecuentes en esta clase de diseños.

Por último, los diseños longitudinales de observación pasiva tienen varias ventajas sobre los diseños pasivos transversales entre las que distintos autores destacan la de poder establecer relaciones causa-efecto unidireccionales, la de poder probar las influencias espúreas en algunos modelos y la de no interrumpir el sistema que está siendo estudiado. Sin embargo, son muchas las desventajas de estos diseños en relación con las otras dos clases de diseños longitudinales anteriormente valoradas y muchas las fuentes de invalidez interna y externa difíciles de controlar al no disponer de grupos aleatorizados ni de la posibilidad de una manipulación planificada y controlada experimentalmente. Por todas estas razones, las influencias causales desde diseños longitudinales pasivos requieren ciertos supuestos acerca de la manera en que el sistema de relaciones cambia sobre el tiempo. Estos supuestos deben ser argumentados desde una teoría sustantiva o bien explorados en áreas de investigación específica. En uno y otro caso las bases para aislar y estimar relaciones causa-efecto tienen menos posibilidades de control que las que aportan los experimentos aleatorizados.

Como resumen acerca del importante papel que juega el diseño en la medición y estimación del aprendizaje, recogemos algunas de las observaciones que hacen distintos autores sobre el tema. Éstas son:

1. Cualquiera que sea el tipo o clase de diseño, se debe partir de algún tipo de control inicial o de alguna forma de estacionariedad. Este control inicial puede ser la línea base, la aleatorización o la objetivización rigurosa de las diferencias iniciales entre el grupo control y el grupo experimental, según el diseño que se va a utilizar.
2. Deben hacerse esfuerzos especiales para asegurarse de que las mediciones sean similares en situaciones pretest y posttest y para los sujetos tratados y no tratados, en el caso de diseños experimentales verdaderos y en los de grupo control no equivalente.
3. Para comparar mediciones pretest y posttest los tests deberían

ser similares estructuralmente, aunque no necesariamente paralelos.

4. En estudios sobre cualquier tipo de aprendizaje, y especialmente de aprendizaje humano, hay que asumir definitivamente el principio de que el cambio es multivariante por naturaleza, con el fin de superar el esquema tradicional de relación bivalente asociado con demasiada frecuencia a numerosas investigaciones sobre el aprendizaje, sobre todo en situaciones de laboratorio.
5. Por último, hay que abrirse a la consideración de que el estudio del cambio, y del aprendizaje como cambio producido por intervención externa, no debe limitarse sólo a estimar el aumento de magnitud o complejidad de las conductas, conocimientos o patrones cognitivos, ya que, como señalan Cronbach y Furby (1970), algunos procesos desaparecen, otros permanecen y varios son reemplazados por procesos cualitativamente diferentes.

En definitiva, a través de estas observaciones hemos pretendido poner de relieve que, además de los problemas generales y específicos asociados a diseños de aprendizaje, hay que tener en cuenta, como paso previo, la conceptualización de estados y cambios, así como la conceptualización de distintos tipos de cambios que pueden producirse como consecuencia de una intervención o de otras múltiples causas que hay que controlar integrándolas en el sistema formalizado o aislándolas, si la integración no es posible.

INSTRUMENTOS DE MEDICION DEL APRENDIZAJE COMO CAMBIO

Tanto a nivel de diseños de recogida de datos como a nivel de métodos de tratamiento de datos, se está en vías de superar la dicotomía mantenida durante muchos años entre datos experimentales y datos no experimentales gracias a las aproximaciones iniciadas por Campbell y Stanley (1966), en el campo de los diseños, y por Cohen (1968) en el de los métodos de análisis multivariantes, al proponer el modelo general de la regresión múltiple como sistema genérico de análisis de datos. Un paso más hacia este proceso de convergencia y armonización entre los dos métodos tradicionales de la Psicología científica exige una revisión sistemática del papel que pueden jugar distintos instrumentos de medición en psicología construidos a partir de diferentes diseños. Esta revisión tiene especial interés para la medición en situaciones de estado y de cambio producido por intervenciones de distintos tipos.

Vamos a plantear a continuación una serie de puntualizaciones sobre el papel que pueden tener diferentes tipos de tests en la evaluación del aprendizaje humano en situaciones iniciales de selección de sujetos o grupos y en situación de evaluación pre y postratamiento. Los tipos de tests en los que nos centraremos en esta exposición serán los tests de norma de grupo (López Feal, 1986), los tests referidos al criterio (Hambleton y Novick, 1973; Berck, 1982) y los tests adaptativos (Urry, 1977; Weiss, 1983).

Para una mejor comprensión de este papel, resumiremos las características que diferencian a los tests de norma de grupo de los tests referidos al criterio o tests edumétricos, como los denomina Carver (1974). Estas son:

1. Los tests de norma de grupo, o de diferencias individuales, clasifican diferencialmente a los sujetos de un grupo normativo en una característica psicológica o de otro tipo a partir de su desviación con respecto a una puntuación estadística estándar, obtenida desde una medida de tendencia central, que suele ser la media aritmética del grupo. Por su parte los tests referidos al criterio evalúan a cada sujeto no en relación con otros sujetos sino en relación con el nivel que haya alcanzado en un dominio teórico o contenido previamente escalado.
2. Las puntuaciones directas de cada sujeto no significan nada en los tests de norma de grupo si no son comparadas con las del grupo normativo. Por el contrario, en los tests referidos al criterio las puntuaciones directas de un sujeto tienen un significado referido al dominio o contenido del test, siempre que este dominio haya sido delimitado, estructurado y escalado adecuadamente.
3. Los tests tradicionales de norma de grupo, por su propio diseño de construcción, son insensibles a la medición de logros o cambios, no asumiendo ninguna condición de tratamiento, ya que su finalidad se centra en conceptualizar diferencias individuales entre sujetos en aptitudes generales, capacidades o rasgos. Así, el patrón de cambio en este tipo de test sería el grupo y no el contenido. Por su parte, los tests referidos al criterio están centrados en la medición de procesos evolutivos o de aprendizaje en lugar de la medición de estados. Este hecho repercute en el tipo de ítems a seleccionar al construir un test: selección de ítems sensibles al cambio por maduración o intervención en los tests referidos al criterio y selección de ítems estables en el caso de tests de norma de grupo.
4. En tests tradicionales construidos de acuerdo con los diseños de norma de grupo, la fiabilidad es sinónimo de estabilidad o consistencia de las medidas, por lo que un test asociado a dicho diseño es psicométricamente fiable si establece las mismas discriminaciones entre individuos en dos o más ocasiones. Por el contrario, en tests referidos al criterio la fiabilidad no se centra en la obtención de la máxima consistencia de diferencias individuales sino en la maximización de las diferencias entre medidas de test pretratamiento o precambio y postratamiento o postcambio.
5. La validez tradicional de tests de norma de grupo se obtiene por comparación de diferencias individuales en un test con diferencias individuales en otra variable referida al mismo grupo. Por su parte, en tests referidos al criterio la validez viene determinada por su sensibilidad al cambio producido por maduración o por intervención. Así, por ejemplo, un test tendría la máxima validez en diseños de test referidos al criterio cuando un sujeto o distintos sujetos no responden correctamente a ningún ítem del test en una situación pretratamiento o precambio y responden correctamente todos los ítems en una situación postratamiento o postcambio.

Por todas estas características diferenciales radicalizadas por Carver, entre los diseños de construcción de test de norma de grupo y test referidos al criterio o contenido, ambos tipos de tests deben tener distintas aplicaciones, aunque con demasiada frecuencia se ha confundido

su uso no sólo en psicología aplicada sino también en el contexto de la investigación básica.

Veremos qué papel puede tener cada uno de estos tipos de instrumentos de medida en el contexto de los distintos diseños para el estudio del cambio, centrándonos especialmente en los diseños de cambio producido por intervención, pero, previamente, haremos referencia a un tercer tipo de diseños de test de reciente aparición, denominados tests adaptativos computerizados. Como características más relevantes que los diferencian de los otros tipos anteriormente esbozados, resaltamos las siguientes:

1. Son tests interactivos e individualizadores que permiten establecer un perfil o línea base de cada sujeto y una estimación de sus logros en sus recursos o capacidades después de un proceso de cambio producido por maduración y/o intervención planificada. Por ello, el nivel de dificultad de cada uno de los items administrados a cada sujeto se adapta durante el proceso de aplicación a sus recursos a su nivel de capacidad o a su nivel de desarrollo, según el caso.
2. Parten de un banco de items abierto, en lugar de estar constituidos por un número de items fijado e inalterable como ocurre con los otros tipos de tests.
3. En lugar de aplicar los mismos items del banco y el mismo número a todos los sujetos, a cada sujeto se le aplican aquellos que se adaptan mejor a las características que pretenden medirse, siendo seleccionados dichos items en cada caso en función del proceso de interacción del sujeto con los items del banco de items respectivo. Por ello, el proceso de aplicación a cada sujeto concluye cuando se ha estabilizado un determinado tipo de respuesta.
4. Aunque este tipo de tests puede ser de papel y lápiz o puede aplicarse a través de entrevistas, en los últimos años se está imponiendo el sistema de aplicación por ordenador. Por esta razón, las publicaciones más recientes (Vale, 1981; Weiss, 1983) utilizan la denominación «Tests Adaptativos Computerizados», que exigen la creación de bancos abiertos de items y el disponer de una red de microordenadores.
5. Por último, los tests adaptativos computerizados suelen utilizar los modelos de Rasgo Latente o Curva Característica de Items (Bejar, 1983) para la selección de items que constituirán un determinado banco, aunque Cliff (1977) ha propuesto la alternativa de utilizar modelos de Base Ordinal para la construcción de este tipo de tests. También es frecuente la utilización de estas dos clases de modelos en la construcción de tests referidos al criterio.

Antes de terminar con esta breve presentación de estos dos diseños de tests alternativos a los de Norma de Grupo, queremos llamar la atención sobre las posibilidades potenciales de estos nuevos diseños de construcción de instrumentos de medida en la investigación básica y en la psicología aplicada y sobre la necesidad de integrarlos en la formación básica del psicólogo, a pesar de que la utilización masiva e indiscriminada de los tests de norma de grupo durante muchos años constituye un freno a dicha integración en nuestras universidades.

En cuanto al papel que pueden jugar estos tres tipos de tests en el contexto de los diseños alternativos para el estudio del cambio producido por manipulación o intervención planificada, vamos a exponerlo a continuación.

Comenzaremos por los tests de norma de grupo. Este tipo de tests, por su carácter de escalas diferenciales y por su propio diseño de construcción, que los hace, en su forma tradicional, insensibles al cambio por intervención externa, no pueden ser utilizados como instrumentos de medida pretratamiento, ni en los diseños temporales de sujeto único, ni en ninguna de las clases de diseños longitudinales expuestos en el punto anterior. Sin embargo, se pueden utilizar como criterio para establecer la línea base en diseños de sujeto único, para seleccionar la población que va a ser aleatorizada en los diseños de experimento verdadero, para diferenciar el grupo de control y el experimental en los diseños longitudinales de grupo de control no equivalente, a partir de una puntuación de corte entre uno y otro grupo, y para identificar un grupo de sujetos o diferenciar varios grupos en diseños longitudinales de observación pasiva. Por su parte, los tests referidos al criterio, por su carácter de instrumentos de medida de dominios o contenidos y por su diseño de selección de ítems sensibles al cambio por maduración o intervención, pueden ser utilizados como tests pretratamiento y postratamiento, tanto en diseños temporales de sujeto único, como en cualquiera de las clases de diseños longitudinales citados anteriormente. En cuanto a los tests adaptativos computerizados, dado su carácter interactivo e individualizador, tienen grandes posibilidades potenciales como instrumentos de medida pretratamiento y postratamiento, sobre todo en diseños temporales de sujeto único sometido a intervención o tratamiento planificado. También pueden ser utilizados en cualquiera de las clases de diseños longitudinales citados anteriormente, tanto para la creación de uno o varios grupos de sujetos con patrones de respuesta similares o diferenciados, como para la obtención de medidas pretratamiento y postratamiento. En particular resaltamos la utilidad potencial de estos diseños de test como soporte para el análisis de datos de enseñanza individualizada asistida por ordenador.

Antes de terminar con este apartado centrado en la presentación esquemática de instrumentos de medida implicados en la medición del aprendizaje como cambio, queremos poner de relieve que su definitiva integración armonizada en los procesos de investigación del cambio por intervención planificada exige una previa conceptualización y reconceptualización de los contenidos que van a ser medidos en un contexto de cambio. También exige una mayor aproximación de la situación de test a los problemas reales con los que debe enfrentarse el sujeto, lo que puede ser facilitado con la creación de Hardware periférico de tipo visual y auditivo asociado a diseños de tests adaptativos computerizados.

METODOS DE TRATAMIENTO DE DATOS SENSIBLES A LA ESTIMACION DEL CAMBIO PRODUCIDO POR INTERVENCION

La extensa literatura aparecida durante los últimos años sobre tratamiento de medidas de cambio en general, y de medidas de cambio producidas por intervención, ofrece numerosos métodos estadísticos

para estimar diferencias entre situaciones precambio o pretratamiento y postcambio o postratamiento asociados a los distintos tipos y clases de diseños expuestos en el apartado anterior.

Entre las publicaciones más relevantes sobre el tema, destacamos el texto publicado por Cook y Campbell en 1979, donde, además de tratarse la problemática teórico-metodológica asociada a la taxonomía integradora de los diseños que presentan, exponen distintos métodos estadísticos adecuados a distintos diseños. Entre estos métodos destacamos distintas variedades de Análisis de Varianza (ANOVA), de Análisis de Covarianza (ANCOVA) y de Análisis de Series temporales, entre los que merecen especial mención los modelos Autorregresivos Integrados de Medias Móviles (ARIMA). Además, hacen una presentación de los Modelos Causales asociados a Estructuras de Covarianza, aunque todavía los vinculan al contexto originario de los diseños de observación pasiva.

Otros autores, como Cascio y Kurtones (1977), proponen técnicas particulares para estimar la influencia de la intervención a partir de la comparación pretratamiento, postratamiento, después de señalar que ninguna técnica es suficiente para abarcar todos los fines posibles que pueden proponerse en la investigación del cambio. Por su parte, otros como Corder-Bolz (1978) evalúan, a través de un estudio Monte Carlo, seis métodos tradicionales utilizados en el estudio del cambio y no llegan a conclusiones definitivas sobre cuál de ellos puede ser más adecuado. Teniendo en cuenta el enfoque de este artículo, no vamos a dedicar este apartado a la descripción de cada uno de estos múltiples métodos parciales propuestos para el estudio del cambio, sino que, retomando las distintas clases de cambio conceptualizados por Golembiewski y otros (1976), plantaremos las posibilidades que tiene el modelo integrador LISREL elaborado por JØreskog y SØrbom (1983) como modelo asociado a estructuras de varianza-covarianza.

Para una breve introducción a la historia, estructura formal y metodología de aplicación de este modelo, remitimos al artículo publicado por López Feal (1982), por lo que aquí nos centraremos en aquellos puntos no tratados en dicho artículo y más relacionados con el tema de la estimación del cambio. En este sentido, queremos resaltar que este modelo, elaborado conjuntamente por estadísticos, econométricos, psicómetras y sociólogos, además de integrar distintas submatrices de relaciones de covarianza, puede tener, entre otras, las siguientes utilidades:

1. Sus modelos de Análisis Factorial Confirmatorio son útiles en la construcción de instrumentos de medida para un solo momento temporal o para varios momentos. Además, en uno y otro caso pueden operar con una sola población o con varias poblaciones simultáneamente. En consecuencia, pueden utilizarse para construir instrumentos de medida pretratamiento o precambio y postratamiento o postcambio, tanto en diseños de un solo grupo como en diseños de grupo de control no equivalente y en experimentos verdaderos.
2. Su modelo de Ecuaciones Estructurales Lineales puede estimar relaciones causa-efecto entre variables exógenas o independientes y variables endógenas o dependientes y de estas últimas entre sí. Estas relaciones causa-efecto también están abiertas a procesos temporales y a la utilización de una sola población o de varias poblaciones simultáneas.

3. Cuando se utiliza el sistema LISREL completo, formado por los dos modelos de medida y el modelo de ecuación estructural, se pueden generar variables latentes o factores comunes que representan variaciones y covariaciones entre las mediciones observables o indicadores empíricos.
4. Este modelo, a través de su programa de computadora asociado, puede llevar a cabo el análisis simultáneo de estructuras de media y covarianza.
5. Por último, dada su flexibilidad, estos modelos permiten especificar *a priori* distintos modelos alternativos asociados a los mismos datos. Los argumentos para estas especificaciones *a priori* pueden proceder de una teoría hipotetizada, referida al modelo causal, el modelo de medida o a ambos, de consideraciones experimentales conocidas, de resultados de estudios previos basados en datos extensivos o de otras bases.

Antes de terminar con esta breve presentación del modelo LISREL, vamos a presentar las tres ecuaciones que lo integran. Estas son:

1. Ecuación del modelo de medida para las variables endógenas o dependientes:

$$x = \Lambda_x \xi + \delta \quad [1]$$

2. Ecuación del modelo de medida para las variables exógenas o independientes:

$$y = \Lambda_y + \epsilon \quad [2]$$

- 3) Ecuación del modelo estructural que representa los efectos de las variables exógenas o independientes sobre las variables endógenas o dependientes y de éstas entre sí:

$$\eta = \beta \eta + \gamma \xi + \zeta \quad [3]$$

Los términos asociados a estas ecuaciones tienen el siguiente significado:

- x e y representan respectivamente, los vectores asociados a las variables observables exógenas o independientes y endógenas o dependientes.
- Λ_x y Λ_y son matrices que representan respectivamente coeficientes Lambda o pesos de las variables observables X e Y sobre las variables latentes.
- δ y ϵ son vectores de componentes de unicidad asociados respectivamente a variables observables X e Y (incluyen un componente de especificidad y un componente de error de medida al azar para cada variable).
- ξ y η representan vectores de variables latentes asociados, respectivamente, a las variables observables X e Y .
- β y Γ son matrices de coeficientes. Los coeficientes β representan efectos causales directos de las variables sobre otras variables y los coeficientes Γ representan efectos causales directos entre las variables y las variables.
- ζ es un vector de residuales o términos de perturbación que representa la varianza no explicada por el sistema causal formalizado.

Este sistema integrado de ecuaciones puede utilizarse como sistema complejo o bien cada ecuación puede actuar independientemente. En uno y otro caso es posible operar con datos desde una sola población y con datos simultáneos desde varias poblaciones. Cuando se opera simultáneamente con varias poblaciones la forma particular de estas ecuaciones puede especificarse igual para todos los grupos en todos los parámetros o en algunos, o puede tener una especificación distinta en todos los parámetros para los diferentes grupos.

En el caso frecuente de que se pretenda un análisis simultáneo de estructuras de medias y covarianzas, se estima la matriz de momento desde cero en lugar de la matriz de covarianzas de población asociada a las ecuaciones [1], [2] y [3], por lo que en el segundo miembro de cada una de estas ecuaciones hay que añadir, respectivamente, los vectores v_x , γ_y y α , que representan términos constantes de ordenadas en el origen.

Pasamos seguidamente a conectar las distintas clases de cambio, conceptualizados por Golembiewski y otros (1976, 1980), con el modelo LISREL. Para ello tomaremos como referencia el artículo recientemente publicado por Schmitt (1982), quien expone un ejemplo aplicado de estimación y prueba de cambios Alpha, Gamma y Beta a través del modelo de medida del sistema LISREL, utilizando un test pretratamiento-postratamiento. Conviene aclarar que, aunque el ejemplo de Schmitt está referido al modelo de medida, la estimación y prueba de estas clases de cambio puede hacerse extensible al modelo de ecuación estructural o modelo causal con variables latentes previamente generadas por modelos de medida y también al modelo de ecuación estructural con relaciones causa-efecto entre variables observables, es decir, sin modelo previo de medida.

Cualquiera que sea el caso, para la estructuración de cambio Alpha entre una situación pretratamiento y una situación postratamiento evaluadas, habría que utilizar un análisis simultáneo de estructuras de media y covarianza y luego probar a través de medidas de ajuste la significación de la diferencia entre el valor de Alpha en las situaciones pretratamiento y postratamiento. Los parámetros Alpha estarían representados en las ecuaciones del modelo LISREL por v_x en el modelo de medida asociado a las variables X, por v_y en el modelo de medida asociado a las variables Y, y por α en el modelo de ecuación estructural.

Para la estimación del cambio Gamma se comparan patrones estructurales o funcionales conceptualizados y especificados de una determinada manera en la situación pretratamiento y reconceptualizados de otra manera en la situación postratamiento y luego se estima, a través de una prueba de ajuste global, el grado en que se cumple la hipótesis especificada o bien se contrastan patrones alternativos de cambio y se prueba cuál de ellos ajusta mejor los datos.

Cargas factoriales o parámetros β y γ claramente diferenciados entre medidas pretratamiento y medidas postratamiento también son indicadores de cambios en los patrones estructurales o funcionales, siempre que estas diferencias se constaten entre todos los parámetros estimados.

Por último, la identificación de cambio Beta viene dada por cambios parciales ocurridos en los patrones estructurales o funcionales. En consecuencia, cambios significativos en algunas de las cargas factoriales o en algunos de los parámetros β o γ pueden ser indicadores de cambios Beta ocurridos después de un tratamiento o intervención.

En el caso de que, en lugar de utilizar un solo grupo pretratamiento-postratamiento, se utilice un grupo experimental y un grupo control no varía la interpretación de estas distintas clases de cambio a través del modelo LISREL, aunque sí el tipo de hipótesis, ya que en estos diseños se hipotetizan cambios producidos por otras causas en el grupo de control.

Antes de terminar con este punto queremos recordar la importancia fundamental que tiene la utilización de instrumentos de medida metodológicamente adecuados a la medición del cambio y teóricamente adaptados a las conceptualizaciones y reconceptualizaciones subyacentes a los objetivos que se pretenden en cada investigación. Sin estos requisitos, ni el modelo LISREL, ni cualquier otro modelo sofisticado de tratamiento de los datos puede garantizar que los cambios Alpha, Gamma y Beta, o cualquier otro tipo de cambio que se proponga, están realmente representados a través de los parámetros estimados por el modelo. También consideramos importante resaltar la necesidad que tienen la psicología de disponer de instrumentos de medida libres de influencias de muestras particulares de sujetos, de contenidos particulares y de influencias temporales y, al mismo tiempo, sensibles al cambio. En este sentido, una de las vías más prometedoras está representada por los diseños de tests de Rasgo Latente o Curva Característica de Item. Muchas han sido las aportaciones en esta dirección, entre las que destacamos, en lo que concierne al estudio del cambio, las de Fischer, quien en una reciente publicación aparecida en 1983 propone una serie de diseños asociados a modelos de Rasgo Latente para la medición del cambio a partir de observaciones cualitativas.

Referencias

- BEJAR, I. I. «Introduction to Item Response models and their assumptions», en R. K. Hambleton (Ed.): *Applications of Item Response Theory*. Vancouver, B. C.: Educational Research Institute of British, 1983, 1-23.
- BEREITER, C. «Some persisting dilemmas in the measurement of change», en C. W. Harris (Ed.): *Problems in measuring Change*. Madison: The University of Wisconsin Press, 1963, 3-20.
- BERK, R. A. (Ed.). *Criterion-Referenced Measurement*. Baltimore: The John Hopkins University Press, 1982.
- CAMPBELL, D. T. «From description to experimentation: Interpreting trends as Quasi-experiments», en C. W. Harris (Ed.): *Problems in measuring Change*. Madison: The University of Wisconsin Press, 1963, 212-242.
- CAMPBELL, D. T., y STANLEY, J. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally, 1966. (Hay traducción castellana en Buenos Aires, Amorrortu Editores, 1973.)
- CARVER, R. P. «Two dimensions of tests: Psychometric and Edumetric». *American Psychologist*, 1974, 29, 512-518.
- CASCIO, W. F., y KURTINES, W. M. «A practical method for identifying significant change scores». *Educational and Psychological Measurement*, 1977, 37, 889-896.
- CLIF, N. «A theory of consistency of ordering generalizable to Tailored Testing». *Psychometrika*, 1977, 42, 375-399.
- COHEN, J. «Multiple regression as a general data-analytic system». *Psychological Bulletin*, 1968, 70, 426-443.
- COOK, T. D., y CAMPBELL, D. T. *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. Boston: Houghton Nifflin Company, 1979.
- CORDER-BOLZ, C. R. «The evaluation of change: New evidence». *Educational and Psychological Measurement*, 1978, 38, 959-976.
- CRONBACH, L. J., y FURBY, L. «How we should measure "change" or should we?» *Psychological Bulletin*, 1970, 74, 68-80.
- DILLON, R. F. (Ed.). *Individual Differences in Cognition: (vol. I)*. Nueva York: Academic Press, 1983.

- DWYER, J. H. *Statistical models for the Social and Behavioral Sciences*. Nueva York: Oxford University Press, 1983.
- FISCHER, G. H. «Some latent trait models for measuring change in qualitative observations», en D. J. Weiss (Ed.): *New Horizons in testing. Latent Trait theory and Computerized Adaptive Testing*. Nueva York, Academic Press, 1983, 309-329.
- GOLEMBIEWSKI, R. T.; BILLINGSLEY, K., y YEAGER, S. «Measuring change and persistence in human affairs». *Journal of Applied Behavior Science*, 1976, 12, 133-157.
- GOLEMBIEWSKI, R. T., y BILLINGSLEY, K. «Measuring change in OD Panel Designs: A response to Critics». *Academy of Management Review*, 1980, 5, 97-103.
- HAMBLETON, R. K., y NOVICK, M. R. «Toward an integration of theory and method for criterion referenced tests». *Journal of Educational Measurement*, 1973, 10, 159-170.
- HOLTZMAN, W. H., «Statistical models for the study of change in the single case», «En C. W. Harris (Ed.): *Problems in measuring Change*. Madison: The University of Wisconsin Press, 1963, 199-211.
- JNRESKOG, K.G. y SNRBOM, D. LISREL. *Analysis of Linear Structural Relationships by the method of Maximum Likelihood. User's Guide*. Versiones V y VI. Chicago: International Educational Services, 1983.
- KENNY, D. A. «A Quasy-Experimental Approach to Assesing treatment effects in the non equivalent Control Group Designs». *Psychological Bulletin*, 1975, 82, 345-362.
- KENNY, D. A. *Correlation and Causality*, Nueva York: John Wiley, 1979.
- LÓPEZ FEAL, R. «Los modelos causales asociados a estructuras de covarianza: Una vía para la unificación de los "dos métodos de la psicología científica"». *Anuario de Psicología*, 1982, 27, 5-21.
- LÓPEZ FEAL, R. *Construcción de instrumentos de medida en ciencias conductuales y sociales*. Barcelona: Alamex, 1986.
- MCCLEARY, R., y HAY, R. A. *Applied Time Series analysis for the Social Sciences*. Beverly Hills: Sage, 1980.
- SCHMITT, N. «The use of analysis of covariance structures to asses Beta and Gamma change». *Multivariate Behavioral Research*, 1982, 17, 343-358.
- SIMON, H. A. «The Behavioral and Social Sciences». *Science*, 1980, 209, 72-78.
- URRY, V. W. «Tailored Testing: A succesful applications of latent trait theory». *Journal of Educational Measurement*, 1977, 14, 473-492.
- VALE, C. V. «Design and implementation of a microcomputer-based adaptative testing system». *Behavior Research Methods and Instrumentation*, 1981, 13, 399-406.
- WEISS, D. J. *New Horizons in testing. Latent Trait Test theory and Computerized Adaptive Testing*. Nueva York: Academic Press, 1983.
- ZMUD, R. W., y ARMENAKIS, A. A. «Understanding the measurement of change». *Academy of management*, 1978, 3, 661-669.

Resumen

Como resumen de este artículo queremos manifestar que no son frecuentes las publicaciones que pretenden una armonización entre conceptualizaciones-reconceptualizaciones de contenidos de cambio, diseños de recogida de datos y métodos de tratamiento de datos. Menos frecuentes todavía son las publicaciones orientadas a conectar tipos y clases de diseños de recogida de datos con instrumentos de medida, a causa del becho histórico de que, mientras que los diseños de recogida de datos han sido dominio casi exclusivo de los experimentalistas, los psicómetras se han centrado en los instrumentos de medida.

También queremos manifestar que el fin de este artículo es el de contribuir a esta armonización desde el contexto del método científico como marco de referencia de los distintos componentes que entran en juego.

Summary

It is not usual to find publications aiming a harmonization between conceptualization-reconceptualization of change contents, data collecting designs, and data processing methods.

Publications intending to relate types and classes of data collecting designs and measurement instruments are even less frequent, because of the historical fact that while data collecting designs have been experimentalists' almost exclusive domain, psychometricians have focused in measurement instruments. Also we wish to manifest that the purpose of this article is to contribute to that harmonization, taking the scientific method as a reference framework for the different components that come into action.