

Durante los últimos años la investigación cognitiva comparada ha sido estimulada por diversos intereses. Las comparaciones transculturales y entre las clases sociales, han provocado el deseo de comprender los procesos mentales de los individuos en las sociedades industrializadas mediante el cotejo y el contraste con miembros de otros grupos. Este lúcido libro de dos importantes especialistas se basa en la afirmación de que las características que separan a los adultos de una sociedad tecnológica respecto de diversos grupos de comparación (características como la adquisición de un lenguaje, el desarrollo de pulgares opuestos para aferrar herramientas, o la invención de un sistema de escritura) pueden estudiarse como factores causales potenciales de las capacidades cognitivas del hombre actual. Una obra de lectura imprescindible para el estudioso de la conducta contemporánea.

Paidós

Psicologías del siglo XX

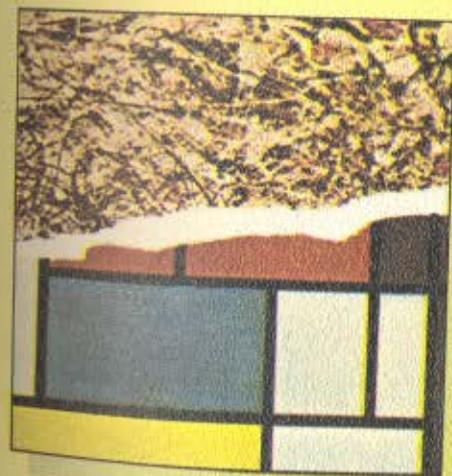
COGNICIÓN Y PENSAMIENTO

*Cómo pensamos.
Estudios comparados*

*Michael Cole
Barbara Means*

Paidós

Psicologías del siglo XX



Título original
Comparative Studies of how People Think
An Introduction
Harvard University Press,
Cambridge, Mass. and London, England
Copyright © 1981 by the President and Fellows of
Harvard College
ISBN 0 - 674 - 15260 - 3

Traducción
Stella Abreu

Diseño de tapa:
Gustavo Macri

1a. edición, 1986

© Copyright de todas las ediciones en castellano by
Editorial Paidós, S.A.I.C.F.
Defensa 599 - Buenos Aires

Impreso en la Argentina
Printed in Argentina

Queda hecho el depósito que previene la ley 11.723
La reproducción total o parcial de este libro, en cualquier forma que sea,
idéntica o modificada, escrita a máquina por el sistema "multigraph", mi-
meógrafo, impreso, por fotocopia, fotoduplicación, etc., no autorizada
por los editores, viola derechos reservados. Cualquier utilización debe ser
previamente solicitada.

ISBN 950 - 12 - 3563 - 7

INDICE

PREFACIO	9
I. ANTIGUOS INTERESES Y NUEVAS EXIGENCIAS	11
II. LA LOGICA NORMATIVA DEL DISEÑO EXPERIMENTAL	22
III. ¿QUE OCURRE CUANDO TODO LO DEMAS NO ES IGUAL?	43
Cómo obtener grupos comparables (45). La equivalencia de tratamiento para diferentes grupos (53). Limitaciones de la simple descripción de diferencias entre grupos (69). Relación entre las diferencias de los grupos y sus causas básicas (71).	
IV. COMPARACION DE PRUEBAS Y DE GRUPOS	78
Comparación de las pautas de desempeño (83). Enfoque del grupo por interacción de pruebas (87). Estudios de entrenamiento (98). Advertencias (109).	
V. ENFOQUES BASADOS EN MODELOS	112
Modelos de información cualitativa de procesos (114). Modelos de simulación por computadora (122). Modelos matemáticos (129). Medición funcional (134). Ventajas de las estrategias de investigación basadas en modelos (143). Advertencias (148).	

cualquier diferencia de desempeño entre los grupos. La asignación al azar es necesaria para asegurar la equivalencia de los dos grupos como término medio, no sólo en función de las características de los sujetos, sino también de las variables circunstanciales relacionadas con la particular ocasión en que un sujeto es sometido a prueba. Estas garantías no existen en grupos determinados por factores naturales y no por asignación al azar.

En este lugar, para la metodología de investigación convencional es beneficioso que al emprender un estudio, los investigadores posean considerables conocimientos acerca de sus sujetos, tanto por los resultados de previas investigaciones como por la información más implícita que emana de antecedentes culturales y ambientales compartidos. Contando con esos conocimientos, los investigadores que inician un estudio de la memorización pueden comenzar las pruebas con un razonable grado de seguridad de que los sujetos —todos ellos de nivel universitario— interpretarán debidamente las consignas dadas, mostrarán un aceptable nivel de memorización para una lista de cierta longitud y considerarán representativas de las categorías pertinentes a las palabras seleccionadas según normas habituales de solución. Este punto es crucial para toda la lógica del modelo de agrupamiento, que parte del supuesto de que el experimentador puede especificar por anticipado las categorías (o asociaciones de categorías) que el sujeto utiliza para la organización. El modelo tiene sentido sólo si las categorías establecidas por los investigadores son aquellas comúnmente accesibles a los sujetos y utilizadas por éstos. Supuestos de este tipo, basados en estudios previos y en el conocimiento de los sujetos, pierden fuerza en la investigación comparada.

Las dos últimas diferencias entre investigación comparada e investigación experimental indican que en aquélla la tarea es complicada en algunos aspectos muy fundamentales. Los factores que salvaguardan la validez intrínseca de un experimento convencional no funcionan en los estudios comparados. De ahí que en la investigación comparada es necesario desarrollar métodos más complejos y cautelosos, y poner cuidado en no extraer conclusiones espurias acerca de las diferencias entre grupos, como se verá en el capítulo siguiente.

III

¿QUE OCURRE CUANDO TODO LO DEMAS NO ES IGUAL?

En la investigación cognitiva no comparada, se presentan diferentes versiones de una tarea (por ejemplo recordación libre) a grupos equivalentes o formados por sujetos asignados al azar, y los investigadores tratan de ver de qué modo las variaciones que introdujeron afectan el desempeño. Aunque es común decir que éstos son estudios sobre *procesos* psicológicos, subrayamos antes que estudios de este tipo ofrecen una teoría de la prueba o, en otras palabras, permiten conocer las variables pertinentes que producen diferencias precisables en el desempeño. En el extenso ejemplo presentado en el capítulo anterior, los desempeños considerados se referían a agrupamiento por categoría y a cantidad de palabras memorizadas.

¿Pero cuál es la lógica de un plan de investigación cuando los grupos no son equivalentes, cuando de hecho no queremos que sean equivalentes porque nos interesa observar la manera en que diferentes grupos memorizan palabras o realizan alguna otra prueba cognitiva? En lenguaje corriente, nuestro interés reside en las diferencias entre los grupos respecto del modo en que las personas piensan. Al emprender una investigación, nos puede mover el interés por probar un aspecto teórico de una prueba psicológica de uso muy difundido. Pero la mayoría de las veces nos interesan las diferencias psicológicas entre grupos, diferencias en lo que se puede llamar aptitudes, motivos o comprensión. Cualquiera que sea la motivación para iniciar un experimento sobre grupos manifiestamente diferentes en aspectos relacionados con la prueba, existe una consecuencia ineludible: estaremos violando la lógica que permite extraer conclusiones válidas del tipo de estudios considerados en el capítulo anterior.

Al no poder asignar al azar los sujetos a los diferentes tratamientos, cuando el "tratamiento" tiene que ver con edad, cultura o especie, no podemos garantizar que los grupos sean equivalentes en todos los aspectos excepto el que estamos investigan-

do. Por el contrario, los grupos a comparar difieren en general de manera muy marcada en cuanto a un sinnúmero de características relacionadas con grandes diferencias en experiencias previas, rasgos psicológicos o una combinación de ambos. Los grupos de edades diferentes, por ejemplo, difieren no sólo en la edad cronológica, sino también en el nivel y el tipo de escolaridad, en conocimientos específicos, enfrentamiento con ciertos tipos de situaciones de prueba, destreza física y desarrollo neurológico. Los grupos de culturas diferentes difieren en lenguaje, pautas de crianza, tipo de experiencias educacionales, valor asignado a diversas cualidades personales y cognitivas y salud física. "Edad" y "cultura" no son más que rótulos globales que abarcan convenientemente una serie de diferencias imposibles de especificar por completo.

Se viola así de dos maneras la condición de que "todo lo demás sea igual" que rige en el método experimental. 1) Los grupos a comparar diferirán por lo general no en una sino en una gran cantidad de dimensiones. Nos resultará difícil por lo tanto probar que una particular diferencia entre ellos es la causa de las diferencias observadas en el desempeño. 2) No podremos garantizar que el tratamiento sea equivalente para los diferentes grupos en todos los pasos de la tarea experimental. Si desde el punto de vista psicológico los grupos reciben tratamientos diferentes, se enturbiará aun más la interpretación de las diferencias observadas entre ellos. Es preciso analizar estos dos problemas antes de considerar los enfoques generales de la investigación comparada¹.

¹ En ocasiones, en lugar de comparar directamente grupos diferentes, el investigador puede poner a prueba métodos y materiales desarrollados con un grupo, en otro grupo especial (por ejemplo, administrando a niños retardados una serie de pruebas de Piaget). En esos casos, el interés específico puede ser estudiar el funcionamiento del grupo especial en sí mismo, sin compararlo necesariamente con el grupo no especial. Si ése es realmente el único propósito que se persigue, no se evalúan las diferencias entre los grupos y por lo tanto no surgen los problemas teóricos

COMO OBTENER GRUPOS COMPARABLES

El método experimental fue concebido como una técnica para aislar y probar individualmente las variables, con el propósito de demostrar relaciones causa-efecto. La imposibilidad de aislar y manejar cualquiera de las variables individuales de un sujeto en un estudio comparado viola una de las condiciones fundamentales de este método e impide atribuir un efecto experimental a un determinado aspecto de las diferencias entre los grupos. Por ejemplo, es posible que un grupo de esquizofrénicos muestre un rendimiento diferente del de un grupo de sujetos normales en una prueba de recuerdo del reconocimiento, no porque la esquizofrenia en sí y de por sí afecte este tipo de memorización, sino porque los esquizofrénicos 1) están menos educados, 2) se distraen con más facilidad, 3) se interesan menos en lograr un buen desempeño o 4) de algún modo están

inherentes a atribuir las diferencias a alguna de las variables de los sujetos. Sin embargo, siempre es tentador cotejar los resultados de la investigación con el grupo especial con los resultados de otra investigación con un grupo de sujetos diferentes. Algunas de las generalizaciones más cuestionables en investigación comparada se originaron en el tipo de estudios que confunden secuencia con consecuencia, y que comparan grupos especiales con grupos no especiales, sometidos a prueba en diferentes investigaciones por diferentes experimentadores y en diferentes condiciones.

Independientemente de que el propósito sea extraer conclusiones comparativas o simplemente estudiar un grupo especial en sí, la equivalencia de tratamiento planteará problemas importantes. Si las consignas, los materiales y los procedimientos tomados de la psicología cognitiva no comparada no significan lo mismo para el grupo especial que para el grupo con el cual fueron desarrollados, ya no estaremos probando lo que nos proponíamos probar. En consecuencia, los resultados que obtengamos serán a menudo tergiversantes u origen de errores muy graves, aunque no se establezcan comparaciones entre los grupos.

afectados por experiencias inherentes a la vida en una institución psiquiátrica. Cuando se observa una diferencia al comparar grupos como los mencionados, el hecho de que difieran en tantos aspectos hace difícil individualizar la variable responsable.

El riesgo que significa atribuir automáticamente las diferencias entre grupos a variables como las comúnmente usadas para definir la pertenencia a un grupo (por ejemplo edad, raza o sexo), queda demostrado por el trabajo que Zigler y colaboradores (1966) realizaron con niños con retardo mental. Anteriores investigaciones que comparaban el desempeño de niños retardados y normales de la misma edad mental, habían llevado a la conclusión generalmente aceptada de que la conducta de los niños retardados era más "rígida". Esta conclusión se basaba en comprobaciones experimentales, como por ejemplo la mayor persistencia de los sujetos retardados, en comparación con los normales, cuando se les pedía realizar repetidamente una misma tarea simple. Según Zigler, esta persistencia, que muchos habían interpretado como una rigidez mental casi psicológica inherente al retardo, es por el contrario un mecanismo originado en las diferencias de experiencias sociales previas entre ambos tipos de niños. Debido a su relativa carencia de interacciones sociales previas, los niños retardados recluidos en instituciones tienden a valorizar más los aspectos sociales de la interacción con un experimentador adulto y por lo tanto a persistir en una tarea hastiante para que la interacción continúe.

Al examinar las historias sociales previas de los retardados internados, Zigler comprobó que los niños más privados de experiencias sociales positivas eran los más persistentes en la tarea experimental. Investigaciones ulteriores demostraron que niños internados con coeficiente intelectual normal presentaban igual grado de persistencia que los niños retardados internados, en tanto que en los retardados no internados no se observaba un nivel superior de persistencia que en los niños con coeficiente intelectual normal. Así, una diferencia en el desempeño de una tarea repetitiva por parte de dos grupos, anteriormente atribuida a diferencias de nivel intelectual, resulta estar más relacionada con diferencias de experiencias sociales previas. Cuando tratamos de "manejar" el nivel intelectual al comparar individuos

normales y retardados, estamos manejando al mismo tiempo una gran cantidad de otros factores (como por ejemplo la historia social). En vista de este hecho, lógicamente carece de fundamento todo intento de atribuir una diferencia entre los dos grupos a cualquiera de las variables de los sujetos. Además, la nueva luz que las comprobaciones de Zigler arrojaron sobre la naturaleza de la variable del sujeto que causa el diferente grado de persistencia (la experiencia social y no el nivel intelectual), trae consigo una nueva interpretación del constructo cognitivo evaluado por la tarea experimental (nivel de motivación y no rigidez).

Se ha hecho muy común la falacia de atribuir diferencias en el desempeño de grupos a una única diferencia entre esos grupos, ignorando otros posibles factores causales. Este tipo de interpretación se observa a menudo en estudios del desarrollo y transculturales y en comparaciones de grupos con diferentes psicopatologías y de diferentes razas o clases sociales. El mismo error está presente en muchos de los recientes estudios de la "evolución de vida", que comparan el funcionamiento intelectual de sujetos ancianos y jóvenes. Por ejemplo, se ensayó someter a prueba a un grupo de sujetos universitarios o de mediana edad, en comparación con un grupo de ancianos (convenientemente seleccionados de instituciones geriátricas), para después atribuir las diferencias de desempeño al deterioro mental (otra vez supuestamente fisiológico) inherente a la vejez. El problema es que, por supuesto, es muy posible que los dos grupos difieran no sólo en edad sino también respecto de las actuales condiciones de vida, el nivel educacional alcanzado, la salud física y las motivaciones. Un estudio de Mistler-Lachman (1977) ilustra la necesidad de examinar las correlaciones de las diferencias entre grupos. Mistler-Lachman estudió la memoria a diferentes edades, utilizando un modelo experimental llamado liberación de la inhibición proactiva. Cuando estudiantes universitarios realizan repetidas veces una prueba de recordación con elementos muy similares (por ejemplo cifras de tres dígitos como 723, 614, 892), sus desempeños empeoran a medida que la repiten: pero si se administra la prueba con otro tipo de elementos (por ejemplo series de tres letras como BQD, PLX, MUK), sus desempeños vuelven a ser buenos, presumiblemente porque se ha "liberado"

la inhibición generada por las cifras de tres dígitos. Mistler-Lachman sometió a la misma prueba a ancianos seleccionados de entre los más competentes de una institución geriátrica, con el requisito adicional de que no debían sufrir ninguna deficiencia mental manifiesta, y comprobó que la modificación de los estímulos no los beneficiaba tanto como a los estudiantes. Este tipo de diferencia entre estudiantes universitarios y sujetos ancianos ha sido interpretada a menudo como prueba de que con la vejez comienzan las deficiencias de aprendizaje. Mistler-Lachman, en cambio, consideró la posibilidad de que la edad no fuera la única responsable de la diferencia y que en ésta intervinieran factores relacionados con la vida en una institución geriátrica. Por lo tanto, incorporó al estudio un tercer grupo de ancianos que vivían integrados a la comunidad. Al comparar a este último grupo con el de estudiantes universitarios, no observó diferencia en el grado de mejoramiento de la memoria después de pasar de un tipo de elementos al otro. El desempeño de los ancianos integrados a la comunidad demostró que no se puede atribuir a la edad en sí misma la diferencia entre estudiantes universitarios y ancianos recluidos en instituciones geriátricas, respecto de la capacidad de beneficiarse con un cambio en el tipo de elementos.

La primera estrategia en la que se piensa para descartar otras explicaciones de las diferencias entre grupos, consiste en constituir grupos de comparación de manera tal que sean equivalentes en función de características que no están siendo estudiadas (por ejemplo el tipo de residencia) pero que también podrían producir el efecto pronosticado. Si esta estrategia da resultado, como en el ejemplo de Mistler-Lachman, se hace posible evitar una conclusión errónea. Pero el problema con la equivalencia es que nunca se puede estar seguro de haberla logrado en grado suficiente. Se podrían seleccionar grupos de ancianos y de adultos jóvenes para un estudio comparado, de manera que fueran equivalentes en cuanto a tiempo de escolaridad, nivel económico y ambiente de vida². Pero estos recaudos no solucionan necesariamente el problema de evitar variables significativas no contro-

² Este tipo de equiparación es muy difícil de lograr en la práctica.

ladas. Siempre habrá variables potencialmente importantes respecto de las cuales los grupos no pueden ser igualados. Por ejemplo, en una comparación entre ancianos y adultos jóvenes, sería imposible equiparar grupos etarios en un estudio transversal, en función del período histórico en que los sujetos crecieron. Podría haber efectos resultantes de haber sido criados y educados en Norteamérica durante la Primera Guerra Mundial, a diferencia de en la década de 1950, y por lo tanto sería imposible controlar estos efectos generacionales en un estudio comparativo de grupos etarios. Dejando de lado estas variables siempre desorientadoras, podemos lograr la equivalencia en la mayoría de las variables que se sabe o se sospecha que están asociadas con la variable en estudio, pero nunca podremos tener la certeza de haber igualado los grupos respecto de todas las características significativas de los sujetos. Así, puede resultar crucial una variable respecto de la cual los grupos de comparación no han sido equiparados. Este no es sólo un problema teórico, sino un obstáculo con el que la investigación comparada tropieza repetidamente.

El estudio transcultural realizado por Berry (1968, 1971) demuestra que ni los esfuerzos más diligentes bastan para equiparar a los sujetos en función de variables potencialmente tergi-versantes. Berry emprendió la investigación movido por anteriores observaciones acerca de los efectos de la cultura en la percepción visual. Los investigadores que estudiaban estos efectos habían encontrado diferencias entre grupos europeos y no europeos, respecto de la susceptibilidad a la ilusión óptica de Muller-Lyer (véase figura 1). Los europeos se dejaban engañar más a menudo por la ilusión; es decir, tendían a percibir la línea con "puntas de flecha" como más larga que la otra (Rivers, 1901, 1905; Segall, Campbell y Herskovitz, 1966). Una interpretación muy aceptada de esta diferencia cultural giraba alrededor de supuestas diferencias de experiencias perceptuales. Los europeos viven en ambientes "escuadrados", caracterizados por líneas y ángulos rectos e intersecciones rectangulares. De ahí que se planteó la hipótesis de que estas experiencias llevaban a los europeos a ver las líneas que se intersectan en las figuras de Muller-Lyer como ángulos rectos que se extienden hacia afuera en el

espacio. Los ambientes en que viven muchos de los grupos no europeos incluidos en estos estudios no ofrecen experiencia comparable respecto de la percepción de figuras rectangulares desde diferentes ángulos; por lo tanto se supuso que la relativa falta de este tipo de experiencia era la causa de la menor susceptibilidad a la ilusión. Sin embargo, personas de culturas europeas y no europeas difieren en gran cantidad de aspectos. Con posterioridad, la hipótesis de la "rectangularidad" fue sometida a pruebas más rigurosas, comparando personas pertenecientes al mismo grupo cultural pero que vivían en ambientes de rectangularidad variable: por ejemplo, africanos que vivían en ambientes rurales tradicionales y miembros de la misma tribu habitantes de ciudades africanas con arquitectura de estilo europeo. Al comparar estos grupos, en muchos casos los investigadores no encontraron diferencias en la susceptibilidad a la ilusión (Jahoda, 1966).

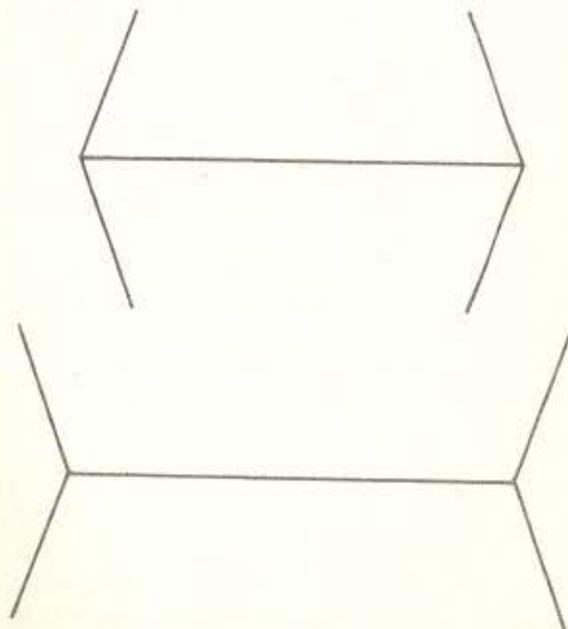


Fig. 1. La ilusión de Muller-Lyer.

Berry señaló como posible causa de la discrepancia de las observaciones, el hecho de no haber controlado las diferencias del desarrollo perceptual. La susceptibilidad a la ilusión de Muller-Lyer disminuye con la edad, quizá como resultado del mayor desarrollo de la percepción. Aunque la susceptibilidad de los europeos y los no europeos occidentalizados tendería a aumentar por la influencia de los ambientes más rectangulares, cualquier ventaja que pudieran tener respecto del desarrollo perceptual en comparación con no europeos miembros de sociedades tradicionales, actuaría en el sentido opuesto, es decir disminuyendo su susceptibilidad a la ilusión. Berry trató de discriminar estos factores comparando el desempeño en la prueba de Muller-Lyer de dos subgrupos pertenecientes a la cultura esquimal, que diferían en cuanto a la rectangularidad ambiental pero eran equivalentes respecto de los puntajes en un test de desarrollo de la percepción (así como en edad y sexo). Sobre la base de estos controles, Berry encontró que la rectangularidad de los ambientes ejercía un efecto significativo dentro de la cultura esquimal. Sin embargo, se puso de manifiesto después que no había sido suficientemente cuidadoso al equiparar los grupos en función de características potencialmente importantes de los sujetos, aparte de la rectangularidad.

Investigaciones posteriores revelaron que además de cultura, ambiente, edad, sexo o desarrollo perceptual, existe otra característica importante. Pollack y colaboradores (Pollack y Silvar, 1967; Pollack, 1970) demostraron que la densidad de pigmentación del ojo está en relación inversa con la susceptibilidad a la ilusión de Muller-Lyer. Como los individuos de piel oscura tienen pigmentación ocular más densa, este factor ofrecía una explicación alternativa de las diferencias entre europeos y no europeos en cuanto a la reacción a la ilusión³. Además, aun los grupos esquimales tan cuidadosamente equiparados por Berry diferían en el grado de pigmentación cutánea: los habitantes de ambientes más rectangulares tenían piel algo más clara. Estas com-

³ Respecto de los diferentes grados de susceptibilidad a la ilusión en función del desarrollo, es de destacar que la densidad de la pigmentación ocular aumenta también con la edad.

probaciones hicieron que Berry realizara nuevos estudios transculturales sobre el mismo tema. Examinó el desempeño en la prueba de Muller-Lyer de un grupo transicional urbano y un grupo tradicional rural en cada una de cinco culturas. Los diez grupos fueron evaluados en función de la pigmentación cutánea y la rectangularidad ambiental. Al estudiar estas variables como pronosticadores de la susceptibilidad, Berry encontró que la pigmentación cutánea era la más estrechamente relacionada con el desempeño. Una vez eliminado estadísticamente el efecto de la rectangularidad, se encontró una fuerte relación entre pigmentación y desempeño en la prueba de Muller-Lyer (r de Kendall = 0, 70); en cambio, eliminado el efecto de la pigmentación, se encontró sólo una muy moderada relación entre rectangularidad y susceptibilidad a la ilusión (r de Kendall = 0, 26). Lo que importa destacar es que resultaron insuficientes los empeñosos esfuerzos iniciales de Berry por ofrecer una prueba válida de la hipótesis de la rectangularidad, equiparando grupos respecto de características reconocidas como posibles causas de las diferencias entre grupos. Una variable que nadie había siquiera considerado resultó ser crítica, y a la luz de su importancia la investigación previa es ambigua. Siempre puede haber (y generalmente *hay*) factores desconocidos que varían entre los diferentes grupos y que afectan los fenómenos observados. De ahí que la equiparación sencillamente no puede garantizar la equivalencia de los grupos como lo hace la asignación al azar.

Además, la equiparación puede producir sus propios problemas. Generalmente es un proceso costoso, de modo que si equiparamos variables no vinculadas con el fenómeno en estudio, habremos malgastado tiempo y esfuerzo. Por otra parte, la equiparación sistemática respecto de ciertas características puede causar la "desequiparación" sistemática respecto de otras. Por ejemplo, al comparar entre adultos jóvenes de un país europeo y norteamericanos, si equiparamos los grupos respecto de la cantidad de años de educación formal, es posible que los europeos provengan de un grupo relativamente más selecto y privilegiado que los norteamericanos, debido a que en los países europeos el acceso a la educación superior es más restringido.

Otro tipo de equiparación consiste en formar grupos equivalentes en función de una capacidad que se supone estrechamente relacionada con la conducta en estudio. (El coeficiente intelectual es el candidato elegido en muchos encuadres aplicados.) Es común que se aplique este criterio cuando se busca comparar la efectividad que determinada indicación o intervención tiene para dos grupos. En general este procedimiento no es recomendable porque a menudo produce efectos falsos derivados de la regresión estadística. (En el Apéndice se examinan estos efectos y el modo en que pueden llevar a conclusiones erróneas sobre las diferencias entre grupos.)

LA EQUIVALENCIA DE TRATAMIENTO PARA DIFERENTES GRUPOS

El otro problema importante en investigación comparada es que se debe ofrecer a los grupos condiciones de tratamiento equivalentes. El simple hecho de seguir el mismo procedimiento o de usar iguales consignas y materiales con los dos grupos no constituye automáticamente un tratamiento equivalente. El problema se manifiesta en aspectos generales tanto como específicos. En general, hay razones para creer que estudiantes universitarios, niños preescolares, adolescentes retardados, individuos esquizofrénicos e indígenas australianos no comprenden igualmente bien ni están igualmente motivados para la mayoría de los procedimientos experimentales corrientes. Lo mismo se puede decir, de manera más específica, cuando se evalúa la equivalencia de un determinado conjunto de materiales o procedimientos. Cada uno de los aspectos del método de un experimentador —los materiales, las consignas y el procedimiento— puede hacer que el tratamiento no sea equivalente para los diversos grupos de comparación.

Materiales

En una investigación comparada entre especies o entre grupos que hablan idiomas diferentes, resulta claro que los materiales

no son igualmente conocidos ni tienen el mismo significado para los dos grupos en estudio. (Por ejemplo, si nativos de Liberia fueran sometidos a una prueba de recordación, la lista debería incluir "café" y no "patatas".) No obstante, es común que se pase por alto esta diferencia al comparar grupos de diferente edad, educación y origen étnico dentro de una cultura. Además, no basta que los materiales sean *más o menos* conocidos y significativos para todos los grupos estudiados, sino que tienen que ser *iguales* en cuanto a conocimiento, significado y cualquier otro estímulo que afecte el desempeño. En muchos casos no se tiene en cuenta la necesidad de igual conocimiento (por supuesto, un requisito difícil de satisfacer), y los investigadores se limitan a demostrar cierto grado de conocimiento por parte del grupo potencialmente menos hábil; por ejemplo, mostrando en un estudio del desarrollo que los sujetos más jóvenes son capaces de nombrar los objetos utilizados en la prueba. Este procedimiento, sin embargo, no basta para garantizar un tratamiento equivalente.

Nuestra preocupación por el grado de conocimiento y significación no es simple retórica. Una serie de estudios sobre recordación ilustran la influencia fundamental que puede ejercer sobre los respectivos desempeños la significación diferente que el material tiene para los grupos. Típicamente, en una tarea de recordación libre, los sujetos de más edad evocan mayor cantidad de palabras que los más jóvenes. Si bien esta comprobación ha sido interpretada como prueba de que los grupos etarios difieren en los procesos que movilizan para ejecutar la tarea propuesta, esta interpretación no puede ser aceptada sin retaceos si se tiene en cuenta que un determinado conjunto de palabras rara vez es tan conocido o significativo para los niños pequeños como para los niños mayores o los adultos. Si factores como la significación o el conocimiento de las palabras influyen en el nivel de recordación que exhiben los estudiantes adultos, cabe suponer que afectan también las diferencias relacionadas con el desarrollo de la capacidad de recordación. Chi (1978) confirmó la importancia de los factores de familiarización en un estudio

realizado con un grupo de niños que jugaban muy bien al ajedrez y un grupo de investigadores recién graduados que eran menos expertos en ese juego. En la recordación de dígitos los adultos aventajaron a los niños, como es habitual, pero éstos eclipsaron a los adultos cuando los elementos a evocar fueron posiciones de ajedrez. Todos los adultos estaban más o menos familiarizados con los "elementos" a recordar (todos sabían jugar ajedrez), pero lo estaban en menor grado que los niños. (En esta situación se encuentran los niños ante la mayoría de las pruebas de recordación.)

En otro estudio, Richman, Nida y Pittman (1976) buscaron comprobar si los niveles de memorización eran equivalentes para niños de diferentes edades, cuando los elementos a recordar eran igualmente conocidos para los mayores y los más pequeños. Elaboraron escalas de significación para los diferentes grupos etarios, para lo cual formaron un conjunto de palabras de tres letras e hicieron que los niños nombraran todos los vocablos que se les ocurrieran a partir de cada una de las palabras de tres letras dadas. Se tomó como grado de significación el promedio de la cantidad de vocablos inducidos por cada palabra de partida. Como era de esperar, la significación media de cada palabra aumentó con la edad. Cuando sobre la base de estas escalas, se formaron listas de palabras de modo que la lista administrada a los más pequeños fuera para ellos tan significativa como lo era la lista de los mayores para éstos, no se observaron diferencias de desempeño entre grupos con diferencias de edad de hasta cuatro años. Esta ausencia de diferencias contrastaba con los resultados obtenidos cuando se administraba la misma lista a los dos grupos; en este caso, los niños mayores, para quienes las palabras eran más significativas, mostraban un nivel superior de memorización.

Las diferencias en la calidad estimada de los materiales de una tarea para diversos grupos pueden afectar aspectos no sólo *cuantitativos* del desempeño (como en los estudios que acabamos de describir), sino también *cualitativos*. Una investigación sobre asociación de palabras realizada por Stolz y Tiffany (1972) ilustra esta situación. En una prueba de asociación de palabras,

se administra al sujeto una serie de palabras y se le pide que responda a cada una con el primer vocablo que le venga a la mente. Se ha demostrado que los grupos etarios difieren en el tipo de vocablos asociados que mencionan. Es típico que los adultos asocien sinónimos, antónimos o supraordinales de la palabra clave y que correspondan a la misma parte del discurso (por ejemplo oscuridad-luz), mientras que los niños tienden a asociar una parte diferente del discurso, un vocablo que suena parecido a la palabra clave o que podría seguirla en una oración (oscuridad-noche). A menudo se citan estos resultados para reafirmar la hipótesis de que niños y adultos difieren fundamentalmente en el modo en que organizan la información. Sin embargo, el trabajo de Stolz y Tiffany demuestra que comparar las asociaciones de palabras entre grupos etarios puede inducir a errores, en vista de que los adultos están mucho más familiarizados que los niños con las palabras clave. Estos investigadores formaron dos listas para una prueba de asociación de palabras: la primera comprendía adjetivos utilizados a menudo en idioma inglés, y la segunda sinónimos de esos adjetivos de uso menos frecuente. Por ejemplo, la primera lista incluía las palabras "muchos", "terco" y "prolijo", y la segunda "innumerables", "recalcitrante" y "melindroso". Se probaron las respuestas de un grupo de estudiantes universitarios a una de las dos series de palabras y después, mediante una prueba de vocabulario, se evaluó su grado de familiarización con las palabras menos frecuentes. El tipo de respuestas varió de acuerdo con la lista que se les hubo administrado. Los habituales tipos de asociaciones "maduras" predominaron con las palabras frecuentes y familiares, mientras que las asociaciones "infantiles" surgieron típicamente en respuesta a las palabras menos comunes y conocidas. Esto demuestra claramente la importancia del contenido de una tarea experimental. Si el grado de conocimiento y frecuencia de uso de las palabras de una prueba afecta la "madurez" de las asociaciones hechas por estudiantes universitarios, pocas son las conclusiones que se pueden extraer de las diferencias entre asociaciones de palabras hechas por grupos de diferentes edades, cuando los sujetos más jóvenes y los de más edad no están igualmente familiarizados con las palabras seleccionadas. Como un mismo conjunto de

palabras no resulta igualmente conocido para ambos grupos, es difícil lograr estímulos equivalentes para hacer una comparación imparcial de las tendencias observadas en las asociaciones de palabras.

Por supuesto, el problema de la no equivalencia de los materiales no se limita a los estudios relacionados con elementos verbales. Una investigación de Cole (1976) sobre el aprendizaje de conceptos por parte de niños ilustra los efectos que pueden tener aspectos teóricamente insignificantes de los materiales no verbales. En una prueba convencional de aprendizaje de conceptos, se realiza una serie de ensayos, en cada uno de los cuales el sujeto ve dos estímulos que varían en diferentes dimensiones (por ejemplo tamaño, forma y color). Uno de los estímulos de cada par es "correcto" por cuanto representa el particular valor de la dimensión seleccionado por el experimentador (por ejemplo, "negro"). Después de aprender a elegir regularmente el estímulo correcto en cada ensayo, es frecuente que el sujeto sea sometido a una prueba de transferencia o "cambio", en la que el estímulo correcto está caracterizado por un valor diferente de la misma dimensión (por ejemplo "blanco") o bien por un valor de una dimensión diferente (por ejemplo "grande", valor de tamaño y no de color). Pruebas como la descrita han sido utilizadas con mucha frecuencia al investigar el aprendizaje tanto en animales como en niños. El desempeño de los niños en diversos tipos de pruebas de cambio ha sido evaluado e interpretado como indicador de aprendizaje "conceptual" (reflejado en respuestas basadas en la dimensión y el valor correctos), similar al de los adultos, o simplemente del aprendizaje de asociar determinados elementos (por ejemplo el cuadrado blanco grande) con la recompensa (véase Medin, 1975, para el estudio de conceptualizaciones más complejas acerca de los procesos que intervienen en este tipo de aprendizaje). Se afirma clásicamente que los preescolares aprenden tareas como la descrita de manera no conceptual, mientras que los niños mayores solucionan los problemas conceptualmente (Kendler y Kendler, 1962; Tighe y Tighe, 1972). En un estudio sobre el aprendizaje conceptual, en el que sometió a prueba a niños de tres y cinco años, Cole utilizó dos series de estímulos diferentes. Uno comprendía bloques geomé-

tricos similares a los usados en la mayoría de las investigaciones sobre el tema. El otro comprendía muñecos de tamaño, sexo y color variados. Inesperadamente, el desempeño de los niños de tres años ante los problemas de cambio dependió del tipo de material que se les suministraba. Cuando la prueba se hizo con bloques geométricos, el número de niños que mostraron aprender conceptualmente fue menor para los de tres años que para los de cinco. Por el contrario, cuando la prueba se hizo con muñecos, aumentó la cantidad de niños de tres años que demostraron hacer aprendizaje conceptual, al punto que desapareció la diferencia respecto del otro grupo. Por lo tanto, que el hecho de observar o no, diferencias de edad en una prueba de uso común pueda servir de indicador de un cambio fundamental en el desarrollo de los procesos de aprendizaje, parece depender del particular tipo de materiales empleados.

La no equivalencia de los materiales de una prueba afecta también las conclusiones respecto de las diferencias de desarrollo cognitivo según clases sociales y razas. Simmons (1979) observó que lo que denomina *prominencia cultural* de los materiales de prueba ejerce un efecto sobre la complejidad de las estrategias de clasificación de los niños. Las figuras culturalmente prominentes son aquellas que muestran objetos, personas o actividades muy conocidos y valorados por un particular subgrupo cultural. En trabajos anteriores (Sigel, Anderson y Shapiro, 1966; Sigel y McBane, 1967; Sigel y Olmstead, 1970) se había demostrado que al clasificar las figuras del Test de estilos conceptuales de Sigel (Sigel Conceptual Styles Test-SCST), los niños de clases bajas utilizaban los agrupamientos descriptivos y por categorías menos que los de clase media; en cambio, recurrían más a menudo al agrupamiento por relaciones. Sigel sugirió (1970) que esos diferentes estilos de clasificación reflejaban diferencias de capacidad de representación según las clases sociales, resultantes de que los niños de hogares de clase baja tienen menos "experiencias de distanciamiento" (actividades que promueven la diferenciación y la abstracción). Al no contar con pruebas directas para demostrar estas diferencias en las actividades desarrolladas en el hogar, Sigel utilizó los resultados del test para apoyar su hipótesis. Simmons (1979) consideró posible que las diferencias

en la prominencia cultural de los materiales de prueba para niños de clases baja y media ofrecieran otra explicación de los diferentes estilos de clasificación de figuras según las clases sociales, y planteó la hipótesis de que esas diferencias resultarían mínimas si se igualara la prominencia cultural de las figuras para los subgrupos culturales estudiados.

Para obtener los estímulos apropiados, Simmons entrevistó a niños negros y blancos de clases media y baja, indicándoles que hicieran una escala de sus preferencias y grado de participación en: actividades académico-culturales; juegos y deportes, y ocupaciones de tipo empleados-profesionales, obreros y animador-atleta. Las actividades académico-culturales y las ocupaciones de tipo empleados-profesionales predominaron entre los niños blancos y de clase media, mientras que los juegos y deportes y las ocupaciones de tipo obrero y animador-atleta fueron más frecuentes entre los niños negros y de clase baja. Las pruebas piloto realizadas indicaron que las figuras del SCST eran, en general, culturalmente prominentes para los niños blancos y de clase media.

Simmons elaboró entonces un test de 18 elementos compuesto de tres subconjuntos: uno con seis elementos culturalmente prominentes para niños blancos y de clase media (por ejemplo una aeronave comercial); otro con seis elementos culturalmente prominentes para niños negros y de clase baja (un boxeador profesional), y un tercero con seis elementos neutros, es decir sin prominencia diferente según los grupos (un anciano tocando el violín). Para cada elemento, se presentaban tres figuras, de las cuales los niños elegían dos que según ellos "debían estar juntas" y daban las razones de su elección. Se estimuló a los niños a que hicieran más de una combinación para cada elemento y no se fijó límite de tiempo. Tanto el test de Simmons como el SCST (para el que se utilizó el mismo procedimiento) fueron administrados a todos los sujetos, varones de quinto y sexto grados, con cuatro combinaciones de raza (negra y blanca) y de clase social (media y baja, según educación y ocupación de los padres). Se premió a los niños por cada par de figuras que formaran sobre la base de la pertenencia a una misma categoría conceptual (por ejemplo, "los dos son juegos").

Simmons llegó a las mismas comprobaciones que Sigel, en cuanto a la diferencia relacionada con las clases sociales, en el SCST, pero no en su propio test. Por el contrario, el test de Simmons mostró una importante interacción de clases sociales y razas; en lo concerniente al uso de categorías para organizar las clasificaciones, los niños negros de clase baja superaron a los negros de clase media, y los niños blancos de clase media a los blancos de clase baja. Los niños blancos en general se basaron más en las categorías en el SCST, mientras que los niños negros lo hicieron más a menudo en el test de Simmons. Al analizar los subconjuntos de elementos en este último test, se observó que los sujetos agrupaban más a menudo por categorías para clasificar elementos más prominentes para su subgrupo cultural. En el caso de los elementos "neutros", el desempeño de los cuatro grupos fue similar.

Las comprobaciones de Simmons sustentan su hipótesis de que los resultados anteriores que indicaban diferentes estilos de clasificación de figuras según las clases sociales, pueden haber dependido de las relaciones sujeto-estímulo, más que de los estilos conceptuales diferentes según los estratos sociales. Sugieren también que utilizar estímulos igualmente conocidos no garantiza que éstos sean equivalentes para determinados tipos de comparaciones, sino que es necesario asimismo considerar las preferencias de los sujetos.

Consignas

Las consignas son otro aspecto importante de la prueba, para el cual igualdad no significa equivalencia. Por supuesto, no es de esperar que una consigna como "seleccionar el cuerpo geométrico que completa correctamente la matriz", transmita el mismo significado a un estudiante universitario y a un niño de cuatro años. Aunque el ejemplo es exagerado, ilustra en esencia la situación que de manera más atenuada aparece reiteradamente en la bibliografía. Es frecuente que los psicólogos que estudian el desarrollo utilicen las mismas consignas con niños de diferentes edades; simplemente suponen que esas consignas cumplirán el mismo propósito respecto de todos los sujetos. Un estudio

realizado por Abramyan (1977) ilustra la poca solidez de semejante suposición. El estudio se basaba en una prueba de aprendizaje conceptual algo diferente de la usada por la mayoría de los investigadores norteamericanos. Los estímulos eran diapositivas que mostraban una figura geométrica de un color sobre un fondo de color contrastante (por ejemplo un círculo verde sobre un fondo amarillo). Se indicaba a los niños que respondieran al fondo de la diapositiva, es decir que apretaran un botón cuando vieran que la diapositiva tenía fondo amarillo, pero no cuando el fondo fuera blanco. Los niños realizaban muchos ensayos con las mismas dos diapositivas, al tiempo que se reforzaban las indicaciones ("Mira, aquí es amarillo, aprieta"). Después efectuaban ensayos con un nuevo par de diapositivas, en las que la figura geométrica que antes había estado sobre el fondo "positivo" aparecía ahora sobre el fondo "negativo" (por ejemplo, la diapositiva negativa podía mostrar el círculo verde sobre el fondo blanco). Este cambio de estímulos reveló que los niños de tres años no se guiaban por las consignas para ejecutar la prueba. En lugar de basar sus respuestas en el color del fondo, las basaban en la figura que aparecía en la diapositiva. Al mostrarles las nuevas diapositivas, respondían incorrectamente a la que mostraba la figura geométrica que antes había aparecido contra el fondo positivo. Los de seis años, en cambio, respondían de acuerdo con las consignas, es decir tomando en cuenta el fondo de las diapositivas. En una etapa ulterior, Abramyan demostró que las consignas de la prueba podían ser modificadas de manera de hacerlas igualmente efectivas para los niños más pequeños. Se reemplazaron las figuras geométricas por siluetas de aviones y los fondos fueron descriptos como condiciones meteorológicas; así, la explicación dada a los niños consistió en que debían permitir volar al avión (apretando el botón) con un día soleado (fondo amarillo), pero no con tormenta (fondo gris). Cuando merced a este recurso las consignas se hicieron más significativas y prominentes, los niños de tres años respondieron de acuerdo con el fondo de las diapositivas, incluso cuando se modificaba el color de los aviones.

Las consignas pueden no resultar equivalentes para dos grupos, porque difieren no sólo en el modo en que son entendidas

por los sujetos, sino además en el grado en que los motivan para esforzarse. Goodnow (1976) destacó la preponderancia de este problema en las investigaciones transculturales. Los niños de culturas no occidentales tienden a mejorar su desempeño en una tarea experimental en mayor medida que los occidentales, cuando se les permite un segundo intento, diciéndoles que "esta vez traten de hacerlo bien". Al comentar este efecto, dice Goodnow: "Es como si entre los niños bien socializados como los de nuestra sociedad, frases como 'ésta es una prueba' o 'tengo varios juegos para que juegues' fueran señal suficiente para alertarlos sobre la necesidad de desplegar cierto grado de esfuerzo. Se debe hacer 'lo mejor posible' en tales ocasiones o, idealmente, en todo momento" (pág. 180).

Procedimiento

Así como muchas veces los diversos grupos de comparación no están igualmente familiarizados con los materiales de un experimentador, lo mismo ocurre con las actividades que los sujetos deben realizar. Los investigadores suelen indicar a grupos diferentes que desarrollen determinada actividad que es más conocida para un grupo que para otro, y después se basan en las diferencias de desempeño para extraer conclusiones acerca de procesos o capacidades cognitivas. El estudio transcultural sobre reproducción de modelos que Robert Serpell realizó con escolares ingleses y de Zambia, demuestra la falacia de este tipo de procedimiento. Trabajos anteriores habían mostrado que sujetos africanos y occidentales exhibían desempeños similares cuando se les indicaba diferenciar o responder verbalmente a estímulos de orientación diferente (Serpell, 1971a., 1971b). Sin embargo, cuando la prueba consistía en copiar un *modelo* utilizado como estímulo, se observaron diferencias sorprendentes entre niños de Zambia y occidentales de igual nivel educacional (Serpell y Deregowski, 1972; Deregowski, 1972). Aunque otras investigaciones transculturales habían puesto de manifiesto las mismas dife-

rencias, estos resultados contradictorios intrigaron a Serpell, que consideró la posibilidad de que se debieran únicamente al particular medio de reproducción utilizado en las escuelas de tipo occidental, en este caso papel y lápiz. Razonó que la capacidad de los niños de reproducir modelos podría estar relacionada con su experiencia en el manejo del medio elegido. En consecuencia, administró tres pruebas de reproducción diferentes a grupos de escolares ingleses y de Zambia. En la primera, la reproducción se hacía modelando arcilla, material elegido por la experiencia que los niños de ambas culturas tienen en su manejo. Para la segunda prueba se eligió la reproducción con alambre, teniendo en cuenta la vasta experiencia de los niños de Zambia en la confección de figuras de alambre. La tercera prueba, copiar figuras con papel y lápiz, fue elegida por ser aquella en la que los niños ingleses tenían más práctica. Las figuras reproducidas por los niños incluían un cuadrado con diagonales, una estrella de cinco puntas y el bosquejo de una persona. Los resultados mostraron que los niños de Zambia superaron a los ingleses en la reproducción de modelos con alambre, mientras que los ingleses hacían mejores copias con papel y lápiz. Los dos grupos se desempeñaron igualmente bien con el modelado en arcilla. Usando varias versiones de la misma prueba, Serpell logró aclarar parte de la confusión causada por la aparente incapacidad de los sujetos de aplicar procesos que parecían utilizar en una situación experimental (individualización de diferencias) a otra situación (reproducción) en la que, a juicio del investigador intervenían los mismos procesos.

Aspectos aparentemente insignificantes de un procedimiento hacen a menudo que el tratamiento deje de ser equivalente. Como los sujetos de diferentes grupos intervienen en un experimento con grados muy dispares de experiencia, conocimiento y comprensión, pueden ser afectados en diferente medida por algún detalle del procedimiento experimental teóricamente no significativo para la aptitud o capacidad cognitivas supuestamente en estudio. Algunos de los estudios sobre conservación de cantidades ilustran el modo en que determinados detalles del procedimiento pueden afectar el desempeño de ciertos sujetos. En una prueba convencional de conservación numérica, se le mues-

tran al niño dos filas paralelas compuestas por igual cantidad de discos y se le pide que estime su relación numérica en dos oportunidades. La primera vez, con los discos alineados correspondiéndose uno a uno (produciendo filas de igual longitud), el niño debe decir si las dos filas contienen igual cantidad de discos; la segunda vez debe hacer la misma estimación, pero después de que los discos de una de las filas han sido distanciados o acercados, de manera que las dos filas ya no tienen la misma longitud. La respuesta típica de los niños más pequeños, después de haber reacondicionado una de las filas, es que las dos filas ya no contienen igual cantidad de discos; los niños mayores, en cambio, continúan estimando que las filas no han variado en número. Rose y Blank (1974) sostuvieron que es posible que ciertos detalles del procedimiento influyan en la aparente incapacidad de los niños pequeños de comprender que el reacondicionamiento de los elementos no modifica la cantidad de éstos. Como en circunstancias cotidianas, a un niño no se le suele formular dos veces la misma pregunta acerca de una cantidad, a menos que se haya producido algún cambio importante, el procedimiento corriente lleva a pensar que conviene dar una respuesta diferente. Rose y Blank demostraron que si se modifica la prueba y se le pide abiertamente al niño que compare las filas una única vez, después de haber reacondicionado los discos, mejoran mucho los resultados obtenidos por los más pequeños en cuanto a responder conservando la relación numérica.

McGarrigle y Donaldson (1974-1975) desarrollaron una idea similar. Basaron su argumento en el concepto general de que el niño pequeño que aún no ha logrado un total dominio del lenguaje depende, en cuanto a la provisión de un contexto que le sirva para reconocer significados a partir de la palabra, de los aspectos no lingüísticos de la interacción social. En la prueba de conservación numérica descrita antes, se le pide al niño formular la segunda estimación sobre relación numérica, inmediatamente después que el experimentador ha reacondicionado una de las filas de discos. Normalmente, una pregunta sobre cantidades sería considerada fuera de lugar en el contexto de este cambio, mientras que sería muy pertinente una pregunta sobre longitud. Si para interpretar lo que se le pregunta, el niño pequeño

se basa en las acciones del experimentador, más que en sus palabras, llega a la conclusión de que se le está formulando la pregunta más lógica, es decir sobre longitud, y al responder en consecuencia fracasa en la prueba convencional. McGarrigle y Donaldson corroboraron esta idea investigando lo que ocurría si el reacondicionamiento de la fila parecía ser accidental y no una acción intencional por parte del experimentador. Notaron que esta maniobra ejercía un fuerte efecto: mientras que sólo trece de los ochenta niños pequeños comprendidos en el estudio conservaron la relación numérica en la versión corriente de la prueba, en la que el cambio es manifiestamente intencional, cincuenta lo hicieron después de un cambio "accidental". Por consiguiente, la posibilidad de obtener resultados diferentes en cuanto a la conservación numérica en niños de diferentes edades (y por lo tanto de extraer conclusiones sobre diferencias cognitivas asociadas con la edad) puede depender de detalles teóricamente insignificantes del procedimiento de prueba, como por ejemplo solicitar una estimación numérica una o dos veces o hacer que el reacondicionamiento de la fila aparezca como accidental o como intencional.

Blank y Rose (1975) demostraron que las diferencias de metodología *dentro de un mismo estudio* pueden hacer que se interpreten mal los patrones de resultados. Les preocupaba la falta de correspondencia observada en anteriores estudios sobre transferencia intermodal de discriminación de formas en los niños. En una prueba de transferencia intermodal, se le presentan al niño dos formas, en primer término con una modalidad (por ejemplo visual), hasta que aprende a seleccionar regularmente la forma designada como correcta. Después se le presentan las mismas dos formas con otra modalidad (por ejemplo táctil, haciendo que el niño palpe las formas, pero sin verlas), para comprobar si elige la misma forma que había aprendido a seleccionar con la modalidad anterior (demostrando el efecto de transferencia). En un estudio previo, Blank, Altman y Bridger (1968) observaron que niños preescolares transferían una discriminación aprendida con la modalidad visual a la modalidad táctil, pero no una discriminación aprendida con el tacto a la modalidad visual. Los resultados sugerían que los preescolares carecen de un

medio eficaz para codificar la información táctil. No obstante, Blank y colaboradores sospecharon que la asimetría observada en los niños entre las transferencias intermodales visual-a-táctil y táctil-a-visual se podía deber en parte a los diferentes métodos de prueba utilizados en cada una de las modalidades. En la prueba visual, el niño debía levantar de una plataforma el objeto-estímulo que consideraba correcto y tomaba el premio que se encontraba debajo si la elección había sido acertada. En la versión táctil, los objetos-estímulo eran puestos en las manos del niño, que debía entregar al examinador el que consideraba correcto, recibiendo un premio si acertaba en la elección. Era posible que uno de estos procedimientos en sí mismo, independientemente de la modalidad, resultara más dificultoso para los niños.

Para probar esta posibilidad, Blank y Rose midieron en primer término el tiempo que necesitaban los preescolares para aprender a discriminar la misma forma en cada modalidad (visual y táctil) y con cada modo de presentación (en la plataforma y sosteniendo con las manos). Observaron que cuando los niños sostenían los objetos (que en la modalidad visual estaban dentro de una caja de plástico transparente para que no palparan las formas), aprendían con más rapidez tanto en la modalidad visual como en la táctil. En un segundo estudio, se probó en los niños la transferencia intermodal de discriminación de formas, sólo que esta vez se instituyó para ambas modalidades como modo de presentación el de "sostener con las manos" que, de acuerdo con el estudio anterior, facilitaba el aprendizaje. El desempeño de los sujetos resultó ser igualmente bueno para la transferencia visual-a-táctil y para la táctil-a-visual, una vez eliminado el efecto tergiversante de las diferentes formas de presentación de los objetos-estímulo.

Las interacciones de las características de la prueba

Para complicar aun más las cosas, las características del procedimiento, los materiales y las instrucciones de una prueba suelen interactuar entre sí y con la experiencia y los conocimientos previos de un grupo acerca de las "reglas del juego experimental", como consecuencia de lo cual un grupo puede interpretar

la prueba de una manera, mientras que el otro lo hace de un modo muy diferente. Blank puso de manifiesto este tipo de efectos en el contexto de una prueba de aprendizaje de discriminación (1975). Esta prueba consiste en mostrar al niño varias veces un par de estímulos, insistiendo en que elija uno, independientemente de la posición de éste en el par. Por ejemplo, se designa como correcto un círculo rojo, que se puede acompañar de un triángulo rojo. Los niños pequeños aprenden con facilidad a hacer la elección correcta pero, curiosamente, parecen ser incapaces de fundamentar su elección. Cuando se les pregunta "¿Por qué elegiste esa figura?", no contestan o bien dan una respuesta evidentemente fuera de lugar, como por ejemplo "Porque quería caramelos", en lugar de dar una contestación apropiada ("Porque es un círculo"), como hacen los niños mayores. Blank señaló que lo que realmente pregunta el experimentador con su formulación "por qué" es "¿Qué características del objeto te llevaron a identificarlo como correcto?" Este significado entra en conflicto con las preguntas "por qué" a las que el niño está habituado en su experiencia cotidiana y que por lo general se refieren a las motivaciones que tuvo para realizar determinada acción. Ocurre que el niño pequeño carece de los conocimientos que en cambio poseen los niños mayores y los adultos acerca de la naturaleza de una situación de prueba, que les permiten saber que es improbable que el experimentador les pregunte sobre el motivo que los decidió a elegir el estímulo correcto.

Para probar su hipótesis, Blank administró una prueba de aprendizaje de discriminación a un grupo de preescolares y después preguntó a cada uno, usando una de varias formulaciones diferentes, en qué habían fundado la respuesta. A la mitad de los niños se les preguntó cómo sabían en qué objeto-estímulo estaban escondidas las golosinas o bien por qué eligieron ese particular objeto. A los restantes se les preguntó "¿Cuál era el que tenía las golosinas?" Dentro de cada grupo, la mitad de los niños tenían los objetos frente a ellos cuando se les formuló la pregunta, mientras que para la otra mitad los objetos estaban fuera de la vista. Los niños interrogados con los objetos frente a ellos se limitaron a señalar el correcto, independientemente del tipo de formulación de la pregunta. Con los objetos fuera de la

vista, casi todos los niños interrogados con la formulación "cuál" dieron una contestación apropiada; casi nunca lo hicieron, en cambio, los niños interrogados con las formulaciones "cómo" y "por qué". El estudio demostró así que la posible interacción entre los procedimientos y las consignas de una prueba puede afectar el desempeño de ciertos grupos. Las respuestas "maduras" surgieron sólo con los estímulos fuera de la vista y preguntando a los niños cuál era el elemento correcto y no por qué eligieron determinado objeto. En efecto, el procedimiento es parte de las consignas, ya que el sujeto se basa en las acciones del experimentador tanto como en sus palabras, cuando busca señales que le aclaren en qué consiste la prueba⁴. Además, los conocimientos de los sujetos acerca de la naturaleza de un experimento determinan la manera en que interpretan las exigencias de la prueba. Los grupos formados en función de edad, cultura y aptitudes difieren en sus conocimientos sobre las respuestas y las conductas que son apropiadas en un experimento. Puesto que esas diferencias existen, un mismo conjunto de consignas y procedimientos puede no ser entendido del mismo modo por diferentes grupos.

De ahí que rara vez podamos decir con certeza que determinado tratamiento experimental es realmente equivalente para dos grupos de comparación. Como lo demostraron los ejemplos presentados, tratamientos nominalmente idénticos son a menudo psicológicamente no equivalentes. Variar los procedimientos o las consignas para diferentes grupos con el fin de hacerlos psicológicamente equivalentes (por ejemplo, dando a los niños más pequeños consignas más detalladas y menos problemas a solucionar que a los mayores), no basta para saber si esos esfuerzos cumplieron realmente su propósito. No hay una pauta directa para evaluar la equivalencia de tratamiento para dos grupos de

⁴ Hall, Salvi, Seggev y Caldwell (1970) hicieron otra buena demostración de este tipo de efecto, usando una prueba de seudolectura, que según la interpretación de algunos investigadores, indicaba que los niños preescolares carecían de "disposición neurológica" para el tipo de "síntesis cognitiva" que exige la lectura.

comparación. Tomar la equivalencia de desempeño como pauta para demostrar la equivalencia de tratamiento sólo probaría que es imposible encontrar legítimas diferencias entre grupos. Este dilema refleja la debilidad fundamental de los diseños experimentales en los que se comparan diferentes grupos en función del desempeño de una única prueba cognitiva.

LIMITACIONES DE LA SIMPLE DESCRIPCION DE DIFERENCIAS ENTRE GRUPOS

Dijimos antes que la naturaleza misma de la investigación comparada hace que no podamos estar seguros de haber dado tratamiento equivalente a los dos grupos de comparación, y aun suponiendo que lo hayamos hecho, no podemos saber con certeza cuál de las muchas diferencias entre los grupos es responsable de cualquier diferencia observada en el desempeño. El diseño de los estudios comparativos simplemente no da cabida a aseveraciones de relación causal, que son la meta fundamental de la investigación experimental. La solución que proponemos para este problema es prescindir de las interpretaciones causales en los estudios comparativos y tratar sus resultados como tratamos una correlación observada entre dos características de un sujeto, es decir como indicación de que dos cualidades tienden a variar paralelamente, sin inferir que una es causa de la otra. (Wood, 1974, adopta este enfoque.) Se podría decir que es útil describir simplemente las diferencias de desempeño de diferentes grupos aunque no sea posible establecer una relación causal entre esas diferencias y cualquiera de las variables estudiadas. "Después de todo" —dirían quienes sostienen este punto de vista— "es útil a los fines de la educación saber que la copia de formas geométricas presenta para los niños de cinco años dificultades que los de ocho parecen haber superado, cualquiera que sea la causa de esta diferencia."

Sin embargo, esta línea de razonamiento adolece de un defecto fundamental. Si no es posible determinar sin ambigüedades cuáles son las diferencias de los procesos cognitivos que produjeron las diferencias de desempeño, es decir cuál de las muchas diferencias preexistentes entre niños de cinco y ocho años

está asociada con las manifiestas diferencias de sus aptitudes para copiar figuras, cualquier descripción de resultados puede llevar a conclusiones erróneas. Tanto el desarrollo teórico como la aplicación práctica de los resultados de la investigación se perjudicarán seriamente si las diferencias del funcionamiento cognitivo son atribuidas a la variable equivocada. Los ejemplos de los diferentes grados de susceptibilidad a las ilusiones ópticas en diversos grupos ilustran las limitadas aplicaciones y la pérdida de esfuerzos que puede producir el hecho de no reconocer cabalmente las causas de las diferencias cognitivas. La susceptibilidad a la ilusión de Muller-Lyer fue estudiada en más de dos mil personas pertenecientes a más de catorce grupos culturales dispersos en todo el mundo, antes de que las comprobaciones de Pollack, Berry y otros demostraran que la densidad de la pigmentación ocular podía influir en los resultados.

Desde un punto de vista pragmático, la simple descripción de las diferencias entre grupos en este caso es insuficiente, porque arroja escasa luz sobre el tipo de medidas que deberían ser tomadas para mejorar el desempeño cognitivo de un grupo. Se ha dicho que los africanos de zonas al sur del Sahara encuentran particularmente difíciles, en comparación con sujetos europeos o norteamericanos, las pruebas que exigen manejar relaciones espaciales o individualizar contornos; la ilusión de Muller-Lyer no es más que un ejemplo de ese tipo de pruebas. Una de las aplicaciones prácticas de la capacidad evaluada mediante estas pruebas es reconocer las diferentes alturas o profundidades indicadas en un mapa con diferentes matices de un color. Desde un punto de vista práctico, se podría ayudar a los estudiantes africanos, entrenándolos especialmente para discriminar entre láminas de diversos matices de un color y para individualizar contornos indicados con variaciones de tonos, para mejorar su capacidad de reconocer profundidades en un mapa. Aunque algo prematuro, una intervención como la mencionada podría parecer razonable en vista de las descripciones de las aparentes dificultades de los estudiantes africanos para realizar ese tipo de tareas. Sin embargo, dos trabajos de Jahoda (1971, 1975) mostraron que estas intervenciones no tienen fundamentos válidos. Este investigador estudió la percepción cromática y su relación con la

percepción espacial en estudiantes universitarios africanos y escoceses. Demostró que la mayor densidad de la pigmentación ocular que hacía a los africanos relativamente poco susceptibles a la ilusión de Muller-Lyer, producía también menor sensibilidad a la luz azul. En su primer estudio sobre este problema, Jahoda observó que al reconocer contornos en un mapa, los africanos tenían más dificultades con las zonas de agua, indicadas en tonos de azul, que con zonas montañosas, representadas en tonos de castaño rojizo; para los estudiantes escoceses, en cambio, ambos problemas eran de dificultad equivalente. En la práctica, por lo tanto, para mejorar la capacidad de los estudiantes africanos de reconocer los contornos en un mapa, el primer paso sería evidentemente representar las zonas de agua con tonos de un color más fácilmente reconocible por personas con densa pigmentación de la retina. No obstante, en su estudio de seguimiento, Jahoda amplió el tiempo asignado a los sujetos para dar sus respuestas y después relacionó sus desempeños con una prueba de percepción espacial. Con esta modificación de las condiciones, el cambio de color de las líneas de contorno no produjo variaciones en el desempeño, y no se observó relación entre la sensibilidad a las líneas de contorno azules y el rendimiento en el test de percepción espacial. La moraleja de esta historia, en lo que atañe a nuestros fines, es que para que los esfuerzos tendientes a reducir el aparente déficit cognitivo de un determinado grupo en comparación con otro resulten efectivos, se deben basar en algo más que la descripción de las diferencias observadas en los respectivos desempeños. Necesitamos conocer las circunstancias exactas en las que se manifiesta el déficit, los procesos que intervienen en la prueba en estudio y las diferentes maneras en que las personas utilizan esos procesos en circunstancias diversas.

RELACION ENTRE LAS DIFERENCIAS DE LOS GRUPOS Y SUS CAUSAS BASICAS

En el capítulo II subrayamos que la principal dificultad con que tropieza el investigador en psicología cognitiva es que las ac-

tividades que son el objeto de estudio, es decir los procesos cognitivos que tratamos de caracterizar, generalmente no son accesibles a la observación directa. En ningún caso es posible medir los procesos (o estructuras, o capacidades) cognitivos aisladamente. De hecho, ni siquiera son definibles aisladamente. Tanto la definición como la medición de las actividades cognitivas se basan en el estudio de la conducta en pruebas claramente definidas, que ofrecen un buen margen de certeza respecto de los diversos estímulos a los que está reaccionando el sujeto y de las respuestas que se pueden esperar de él.

La necesidad de tener en cuenta los efectos de factores específicos de una prueba al evaluar un tipo supuestamente general de actividad cognitiva (memorización, ejercitación, asociación), sumada a las ideas a menudo no demasiado claras acerca de cuáles son esos factores, es una de las principales fuentes de complejidad en el diseño de experimentos destinados a aislar los efectos de cada variable en psicología cognitiva no comparada. Se requiere una serie de estudios para elegir entre varias hipótesis y es posible que la hipótesis en cuestión, salvo que sea razonablemente específica, sólo pueda ser probada en el contexto de una serie de experimentos vinculados por una teoría.

En los estudios comparativos del tipo de los que comenzamos a analizar en este capítulo, extraer conclusiones es extremadamente difícil. No contamos con ninguna de las garantías de igualdad de los elementos que no están en estudio y que en investigación no comparada simplifican la labor. No podemos presuponer que todos los sujetos entienden las consignas del mismo modo; no podemos presuponer que los estímulos dentro del experimento propiamente dicho son equivalentes para todos ellos (es decir igualmente conocidos, igualmente fáciles de discriminar o igualmente asociados unos con otros). Y lo más importante, no podemos asignar los sujetos al azar a las condiciones experimentales en estudio y dar por sentado después que estas no equivalencias afectan a todos los grupos por igual.

En pocas palabras, estamos en dificultades. La tarea se simplificaría si estuviéramos dispuestos a no ir más allá de la simple demostración de una diferencia entre los grupos en una prueba específica. Pero largas enumeraciones de semejanzas y diferen-

cias entre un grupo en estudio y algún grupo control no proporcionan bases suficientes para reconocer las causas de las diferencias o para contribuir a la elaboración de programas de tratamiento, en los muchos casos en que ofrecer tratamiento es la finalidad última de la investigación.

No hay, en nuestra opinión, un modo fácil de eludir el problema; esto es parte de la lógica del trabajo. Pero existen diferentes estrategias que pueden ser utilizadas para disminuir el peligro de extraer conclusiones no válidas. Antes de considerarlas en detalle en los capítulos IV y V, ejemplificaremos cómo las dificultades a las que nos referimos pueden desmerecer una investigación, para lo cual describiremos un estudio comparativo ya publicado. Para conservar el anonimato, pasaremos por alto ciertos detalles, ya que abundan las críticas desaprensivas a las debilidades de la investigación en este campo y lo que nos interesa es ilustrar los tipos *generales* de error que genera la investigación cognitiva comparada. De todos modos, los fundamentos, la metodología y los resultados descriptos a continuación son verídicos y fueron extraídos del informe publicado.

El estudio en cuestión se basaba en la hipótesis general de que el agrupamiento en la recordación libre es una pauta para medir los procesos organizativos que están en la base no sólo de la recordación verbal sino de otras capacidades, como la comprensión de la lectura. En este caso en particular, se planteó la hipótesis de que la ausencia de procesos organizativos era característica de individuos con retardo mental y constituía una posible causa de sus dificultades para entender lo que leían. Para probar esta hipótesis, se tomó una muestra de adolescentes con retardo mental y se los sometió a un test uniformado de lectura. Sobre la base de los puntajes en este test, se dividió la muestra en dos grupos, uno con puntaje superior y el otro con puntaje inferior a la media del total de la muestra en comprensión de lectura. Ambos grupos hicieron cuatro intentos de memorización de una única lista, compuesta de cinco elementos de cada una de cuatro categorías. Como la finalidad última de los investigadores era aumentar la capacidad de comprensión de la lectura entrenando a los sujetos retardados en la organización de material verbal, la lista se presentó de dos modos. Para la mitad de los

sujetos de cada grupo, se ubicaron al azar los elementos de las cuatro categorías en los cuatro intentos; para la otra mitad, en los dos primeros intentos las palabras fueron agrupadas (se ordenaron consecutivamente los elementos de cada categoría), y en los dos últimos las palabras se ubicaron al azar. Los principales resultados fueron que los sujetos con puntaje por encima del promedio agrupaban más que aquellos con puntaje por debajo del promedio, y que los sujetos a los que se administró la lista ordenada por categorías en los primeros dos intentos mostraron más capacidad de agrupamiento que aquellos con presentación al azar en los últimos dos intentos (ubicación al azar) así como en los primeros dos intentos (ordenados). De estos resultados se extrajeron las siguientes conclusiones: la hipótesis sobre la existencia de una relación entre agrupamiento y comprensión de lectura es correcta; es posible incrementar la organización en la memorización y podría ser posible facilitar la comprensión de la lectura en los sujetos retardados, remediando sus deficiencias de organización en la incorporación de material verbal.

Examinemos este estudio en función de los problemas que caracterizan la investigación cognitiva comparada. Aunque se ha intentado relacionar una diferencia entre dos grupos respecto del rendimiento en una prueba (lectura o más precisamente lectura y agrupamiento) con un proceso cognitivo más básico y general (organización de la información verbal), no se ha especificado la naturaleza exacta de este proceso en lo que concierne a la comprensión de la lectura. Además, los planteos son ambiguos por ser materia tan controvertida la naturaleza de los procesos que intervienen en el agrupamiento, así como la hipótesis de que éste es una causa directa de la memorización. Dicho de otro modo, en nuestra opinión no se han analizado las pruebas de memorización y de comprensión, ni el papel que desempeña la organización en cada una de ellas, como tampoco se ha medido directamente la "capacidad de organización" que se considera un factor básico de la diferencia de rendimiento entre los grupos. En ausencia de una medición directa de los procesos organizativos, una relación como la obtenida en este estudio, entre comprensión de lectura y agrupamiento, podría reflejar *cualquier* proceso común que quisiéramos postular como posible

(por ejemplo algún tipo de capacidad verbal, motivación o atención). Como dijimos antes, en investigación cognitiva comparada es insensato utilizar las diferencias de desempeño en una prueba para extraer conclusiones sobre procesos. Si comparáramos diferentes grupos de tratamiento y no diferentes grupos de sujetos, al menos podríamos atribuir la diferencia de rendimiento a determinada variable individual, aunque no lográramos precisar el proceso cognitivo que interviene. Pero en nuestro caso, hasta esa ayuda nos está negada.

La importancia que tiene una diferencia entre la presentación al azar y la presentación ordenada es asimismo difícil de interpretar sin un modelo explícito del rendimiento en la prueba. No se ha especificado qué se está entrenando con la presentación ordenada o, en otras palabras, qué están haciendo de diferente los sujetos con listas de palabras ordenadas y aquellos con palabras ubicadas al azar. La afirmación de que el estudio confirma la posibilidad de incrementar el agrupamiento durante la memorización en los sujetos retardados es especialmente endeble, puesto que el diseño no permite discernir entre agrupamiento por categorías y aprendizaje mecánico. Cuando en los diferentes intentos del estudio, los elementos de cada categoría se presentaron juntos, es posible que hayan sido recordados consecutivamente no porque los sujetos utilizaran las categorías para organizar la memorización, sino simplemente porque los recordaban de acuerdo con el orden en que los fueron incorporando. De ahí que no resulta significativa una diferencia en puntajes de agrupamiento entre la lista ordenada y la de ubicación al azar. Aunque el grupo con presentación ordenada en los primeros dos intentos demostró agrupar más también en el tercero y el cuarto intentos (cuando las palabras de la misma categoría ya no aparecían juntas), los puntajes podrían reflejar simplemente el efecto del aprendizaje mecánico en los intentos anteriores con la misma lista⁵.

⁵ Una diferencia en los niveles de memorización entre los grupos con listas de palabras ordenadas y de palabras ubicadas al azar, habría corroborado la interpretación de que el ordenamiento repercute en la organización de la lista, pero en el estu-

Además de no presentar una interpretación convincente de los resultados en función de los procesos intervinientes (defecto común también en psicología cognitiva no comparada), el estudio presenta el tipo de errores que amenaza la validez intrínseca y que es endémico en investigación comparada. Los investigadores pretendían usar una diferencia de los grupos respecto de la capacidad de lectura, con el fin de probar una hipótesis acerca del papel de los procesos de organización en la lectura y el agrupamiento. Este caso ilustra con claridad el problema básico de obtener grupos comparables, cuando se somete a prueba el efecto de una particular variable del sujeto. Aunque todos los sujetos reunidos para este estudio estaban rotulados como "adolescentes con retardo mental", probablemente existían entre ellos variaciones considerables en cuanto al coeficiente intelectual y a la edad cronológica (los términos "retardo" y "adolescente" son muy abarcativos), así como en función de otros factores pertinentes como tiempo de internación, experiencia previa con tests intelectuales y nivel de motivación. Al dividir la muestra en dos subgrupos sobre la base de los puntajes en lectura, casi con seguridad estos subgrupos diferían también con respecto a las variables recién mencionadas. Lo que en el estudio se describe como relación entre agrupamiento y comprensión de la lectura, bien podría ser una relación del agrupamiento con coeficiente intelectual, edad cronológica, nivel de motivación o cualquier otra característica susceptible de variar entre los dos grupos "de lectura".

También debemos considerar dudosa la supuesta equivalencia de tratamiento para los dos grupos. Aunque la presentación de las palabras que integraban la lista de recordación libre fue visual y auditiva, una diferencia en la capacidad de lectura bien pudo hacer que los lectores menos aventajados tuvieran más dificultades en comprender de entrada las palabras, paso necesario antes de que pudieran aprenderlas. Además, un déficit auditivo-fonológico pudo ser una de las causas que afectó a los sujetos con más problemas para entender la lectura, diferencia que probable-

—
 dio al que nos referimos no se observaron diferencias en los niveles de memorización en función de los dos tipos de presentación.

mente les dificultó la percepción tanto auditiva como visual de las palabras.

Limitar las conclusiones respecto de las diferencias entre sujetos a la descripción del efecto (diciendo por ejemplo que "los sujetos retardados con puntajes relativamente altos en comprensión de lectura agrupan más en la recordación libre que aquellos con puntajes relativamente bajos") no solucionaría el problema. En primer lugar, a la luz de las muchas otras diferencias que podrían caracterizar a los dos grupos de lectura, semejante afirmación podría ser muy desencaminante. En segundo lugar, enunciar simplemente las diferencias de desempeño no es una base adecuada para el propósito que perseguía el estudio, que era extraer conclusiones acerca de las medidas educacionales que deberían ser adoptadas para remediar la situación. Las recomendaciones de este tipo deben estar fundadas en una explicación causal del déficit. Si no es una diferencia en la utilización de los procesos organizativos lo que *causa* la diferencia en la comprensión de la lectura, no tiene sentido tratar de entrenar la capacidad de organización para mejorar la lectura.