

Traducción del artículo original de McEachin, Smith & Lovaas publicado American Journal On
Mental Retardation, 1993, Vol 97, No 4, 359-372

**"Long-Term Outcome for Children With Autism Who Received Early
Intensive Behavioral Treatment"**

John J. McEachin, Tristram Smith, and O.Ivar Lovaas

Traducción al castellano

ÍNDICE DE CONTENIDOS

Resultado a largo plazo para niños con autismo que recibieron un tratamiento conductual
intensivo precoz

John J. McEachin, Tristram Smith y O. Ivar Lovaas

COMENTARIOS SOBRE McEACHIN, SMITH Y LOVAAS

La asignación casi-arbitraria puede ser tan convincente como la asignación arbitraria

Donald M. Baer

Efectos satisfactorios en espera de replicación independiente

R. M. Foxx

Replicación y ampliación del tratamiento conductual del trastorno autista

Alan E. Kazdin

El resultado del tratamiento es esperanzador

Gary B. Mesibov

Condición normal en relación con la condición de alto funcionamiento en los niños con
autismo

Peter Mundy

Comentarios sobre Replicación y evaluación del resultado

Tristram Smith, John J. McEachin y O. Ivar Lovaas

"Long-Term Outcome for Children With Autism Who Received Early Intensive Behavioral Treatment"

John J. McEachin, Tristram Smith, and O.Ivar Lovaas

Resultado a largo plazo para niños con autismo que recibieron un tratamiento conductual intensivo precoz¹

John J. McEachin, Tristram Smith y O. Ivar Lovaas
Universidad de California, Los Ángeles

Después de una intervención conductual muy intensa, un grupo experimental de 19 niños con autismo en edad preescolar consiguieron ubicaciones escolares menos restrictivas y CI más elevados que otro grupo de 19 niños similares de 7 años (Lovaas, 1987). El presente estudio hizo un seguimiento de tal resultado mediante la evaluación de los sujetos a una edad media de 11,5 años. Los resultados mostraron que el grupo experimental preservaba sus logros por encima del grupo control. Los 9 sujetos experimentales que habían logrado mayores resultados a la edad de 7 años recibieron evaluaciones particularmente extensas, indicando que 8 de ellos eran indistinguibles de los niños normales en las pruebas de inteligencia y la conducta adaptativa. Así, el tratamiento conductual puede producir unos logros importantes de larga duración para los niños con autismo.

El *autismo infantil* es un trastorno marcado por una alteración grave en el funcionamiento intelectual, social y emocional. Su aparición tiene lugar en la infancia, y el pronóstico parece ser extremadamente pobre (Lotter, 1978). Por ejemplo, en el estudio de seguimiento prospectivo más largo y con un diseño metodológico sólido, Rutter (1970) encontró que sólo 1 de cada 64 sujetos con autismo (menos del 2%) podía considerarse libre de problemas clínicamente significativos en la edad adulta, como se evidenciaba por el mantenimiento de un puesto de trabajo, vida independiente y llevando una vida social activa y apropiada a su edad. Los sujetos restantes presentaban diversas disfunciones, como marcadas rarezas en su comportamiento, aislamiento social y amplia psicopatología. La mayoría de sujetos requería unas condiciones de vida supervisadas.

Los profesionales han utilizado una gran variedad de intervenciones en su esfuerzo por ayudar a los niños con autismo. Durante muchos años, ninguna prueba científica demostró que alguna de estas intervenciones mejorara el pronóstico a largo plazo de estos niños (DeMyer et al., 1981). Sin embargo, a partir de los años sesenta, una de tales

¹ Este estudio fue financiado por la beca n°. MH-11440 del Instituto Nacional de Salud Mental. El estudio se basó en una tesis doctoral entregada a la Universidad de California, Los Ángeles, departamento de Psicología, como requisito parcial del curso de doctorado. Los autores desean expresar su más sincero agradecimiento a los diversos estudiantes de la UCLA que actuaron como terapeutas y ayudaron a realizar este estudio. Rendimos un agradecimiento especial a Bruce Baker y Duane Buhrmester, que ayudaron en el diseño del presente estudio. Las peticiones para volver a publicar este artículo, copias de la Escala de valoración clínica o información adicional sobre el presente estudio deben dirigirse a O. Ivar Lovaas, 405 Hilgard Ave., UCLA, Department of Psychology, Los Angeles, CA 90024-1563.

intervenciones, el tratamiento conductual, se ha mostrado prometedora. Se ha demostrado que el tratamiento conductual aumenta las conductas adaptativas como el lenguaje y las habilidades sociales, mientras disminuye las conductas perjudiciales como la agresión (DeMyer, Hingtgen y Jackson, 1981; Newsom y Rincover, 1989; Rutter, 1985). Además, el tratamiento conductual ha sufrido continuas modificaciones y mejoras como resultado de esfuerzos de investigación continuados en diversos lugares (Lovaas y Smith, 1988).

Algunas pruebas recientes han indicado que el tratamiento conductual se ha desarrollado hasta el punto en que puede producir mejoras substanciales en el funcionamiento global de los niños con autismo (Simeonsson, Olley y Rosenthal, 1987). Lovaas (1987) dedicó aproximadamente unas 40 horas semanales de tratamiento conductual individualizado por un período de 2 años o más a un grupo experimental de 19 niños con autismo que tenían menos de 4 años. Tal intervención también incluía la formación de los padres y la ubicación general en entornos preescolares normales. Cuando se volvieron a evaluar a la edad media de 7 años, los sujetos del grupo experimental habían ganado una media de 20 puntos CI y habían logrado mayores avances en el rendimiento educativo. Nueve de los 19 sujetos completaron el primer curso de primaria en clases generales (educación no especial) sin ayudas, y tenían CI que aumentaron hasta el nivel medio. Por el contrario, a los dos grupos control, con un total de 40 niños, también diagnosticados como autistas y comparables al grupo experimental al inicio del tratamiento, no les fue tan bien. Tan sólo un sujeto de los grupos control (2,5%) alcanzó niveles normales de funcionamiento intelectual y educativo. Estos datos sugieren que el tratamiento conductual es efectivo. Sin embargo, la durabilidad de los logros del tratamiento es incierta. En otro estudio anterior, Lovaas, Koegel, Simmons y Long (1973) hallaron que los niños con autismo experimentaban una regresión al finalizar el tratamiento. Otros estudios han mostrado que los niños con autismo pueden presentar mayores dificultades cuando entran en la adolescencia (Kanner, 1971; Waterhouse y Fein, 1984). Asimismo, como se observó en el primer seguimiento (Lovaas, 1987), “algunos déficits residuales pueden permanecer en el grupo de funcionamiento normal, que no pueden detectarse por parte de maestros y padres y sólo pueden aislarse en evaluaciones psicológicas más cercanas, especialmente cuando los niños crecen” (p. 8). Esta posibilidad nos indica la necesidad de una evaluación más detallada y unos seguimientos continuados del grupo a lo largo del tiempo.

La presente investigación consistía en dos partes: En la primera parte, examinamos si varios años después de la evaluación a la edad de 7 años, el grupo experimental del estudio de Lovaas (1987) había mantenido sus logros de tratamiento. Los sujetos del grupo experimental y uno de los grupos control realizaron pruebas estandarizadas de funcionamiento intelectual y adaptativo. Después se contrastaron los grupos entre sí, y su actuación presente se comparó con su actuación en las evaluaciones anteriores. La segunda parte de la investigación se centró en aquellos sujetos que habían alcanzado el mejor resultado al final del primer curso en el estudio de Lovaas (1987; es decir, los 9 sujetos que fueron clasificados como funcionamiento normal de los 19 sujetos del grupo experimental). Examinamos hasta qué punto dichos sujetos con mejores resultados podían considerarse libres de sintomatología autista. Se construyó una batería de pruebas para evaluar una variedad de déficits posibles: por ejemplo, pautas de pensamiento idiosincrásico, amaneramientos e intereses; falta de relaciones cercanas con la familia y los amigos;

dificultad en la convivencia con las personas; debilidad relativa en ciertas áreas del funcionamiento cognitivo, como el razonamiento abstracto; no trabajar hasta el nivel de su habilidad en la escuela; monotonía en el afecto; ausencia o peculiaridad en el sentido del humor. Los posibles puntos fuertes a identificar incluían un funcionamiento intelectual normal, buenas relaciones con los miembros de la familia, habilidad de funcionar independientemente, uso apropiado del tiempo libre y socialización paritaria adecuada. Se tomaron numerosas precauciones metodológicas para garantizar la objetividad de la exploración de seguimiento.

Método

Sujetos y antecedentes

Se han descrito las características de los sujetos y su tratamiento en otro documento (Lovaas, 1987), por lo que en el presente artículo tan sólo ofreceremos un resumen. El estudio de tratamiento inicial contenía 38 niños que, en el momento del inicio, eran muy jóvenes (menores de 40 meses en caso de mutismo, menores de 46 meses en caso de ecolalia) y habían recibido un diagnóstico de autismo por parte de un psicólogo o psiquiatra profesional sin implicación en el estudio. Se dividieron estos 38 sujetos en un grupo experimental y un grupo control. La asignación a los grupos se realizó en base a la disponibilidad de personal. Al principio de cada trimestre académico, se formaron los equipos de tratamiento. El director clínico y los miembros del personal determinaban entonces si existía alguna plaza para el tratamiento intensivo. De ser así, la siguiente derivación que llegaba entraba en el grupo experimental; de otro modo, el sujeto se incluía en el grupo control. El grupo experimental consistía en 19 niños que recibieron 40 o más horas semanales de tratamiento conductual individualizado durante 2 o más años. El grupo control estaba formado por 19 niños que recibieron una intervención mucho menos intensiva (10 horas semanales o menos de tratamiento conductual individualizado, además de una variedad de tratamientos proporcionados por agencias comunitarias, como la formación de padres o las clases de educación especial). El estudio inicial también incluía un segundo grupo control, formado por 21 niños con autismo que recibieron seguimiento a lo largo del tiempo por parte de una agencia cercana, pero que nunca fueron derivados para este estudio. Sin embargo, estos 21 sujetos no estaban disponibles para el presente estudio. En las medidas estandarizadas de inteligencia, el segundo grupo control no difería ni del grupo experimental ni del primer grupo control al inicio del tratamiento, y tampoco difería del primer grupo control en la segunda evaluación de los sujetos a la edad de 7 años. Estos resultados sugieren que, medidos con pruebas estandarizadas, a) los niños con autismo que nos fueron derivados para tratamiento eran comparables a los niños con autismo descritos en otros estudios y b) el tratamiento mínimo proporcionado al primer grupo control no alteró su funcionamiento intelectual.

El análisis estadístico de un amplio conjunto de medidas pretratamiento confirmó que el grupo experimental y el grupo control eran comparables al inicio y presentaban grandes igualdades en variables tan importantes como el CI y la gravedad de la alteración. La edad cronológica media (EC) en el diagnóstico para los sujetos del grupo experimental era de 32

meses. Su CI medio era 53 (gama de 30 a 82; todos los CI son puntuaciones desviadas). La EC media de los sujetos del grupo control era de 35 meses; su CI medio era 46 (gama de 30 a 80). La mayoría de sujetos era mudo, todos presentaban grandes deficiencias en el lenguaje receptivo, ninguno jugaba con iguales o mostraba un juego con juguetes apropiado a su edad, todos eran emocionalmente reservados, la mayoría presentaba rabietas graves, y todos mostraban conductas ritualistas y estereotipadas (autoestimuladoras) extensas. Así, parecían ser una muestra representativa de niños con autismo (Lovaas, Smith y McEachin, 1989). Lovaas (1987) presentó un informe más completo de los datos de ingreso.

Los niños del grupo experimental y el grupo control recibieron sus respectivos tratamientos por parte de estudiantes terapeutas formados que trabajaban en la casa del niño. Los padres también trabajaban con sus hijos y recibieron una extensa formación y supervisión sobre técnicas de tratamiento apropiadas. Siempre que fue posible, los niños se integraron en centros de preescolar generales. El tratamiento se centraba principalmente en el desarrollo del lenguaje, la mejora de la conducta social y la promoción del juego cooperativo con iguales juntamente con un juego con juguetes independiente y apropiado. Al mismo tiempo, se destinaron esfuerzos considerables a la disminución de rituales excesivos, rabietas y conducta agresiva. (Para una descripción más detallada del programa de intervención, véase el manual de tratamiento [Lovaas et al., 1980] y los videos instructivos que complementan el manual [Lovaas y Leaf, 1981].)

En el período del presente seguimiento (1984-1985), la EC media de los niños del grupo experimental era de 13 años (gama = de 9 a 19 años). Todos los niños que habían alcanzado el funcionamiento normal a los 7 años habían terminado el tratamiento en ese momento. (El *funcionamiento normal* se definió operativamente como una puntuación dentro del campo normal en las pruebas estandarizadas de inteligencia y la finalización satisfactoria del primer curso en una clase general, de educación no especial, sin ayudas.) Por otra parte, algunos de los niños que no habían logrado el funcionamiento normal a los 7 años habían continuado el tratamiento a petición de los padres. El período de tiempo transcurrido desde que los sujetos del grupo experimental habían terminado el tratamiento variaba de 0 a 12 años (media = 5), mientras que en el caso de los niños con funcionamiento normal era de 3 a 9 años (media = 5).

La edad media de los sujetos en el grupo control era de 10 años (gama de 6 a 14). El período de tiempo transcurrido desde que finalizaron el tratamiento variaba de 0 a 9 años (media = 3). Así, los sujetos del grupo experimental tenían tendencia a ser mayores y a tener un período después del tratamiento más largo que en el caso de los sujetos del grupo control. La diferencia en la edad tenía lugar debido a que las primeras derivaciones para el estudio se asignaron al grupo experimental, ya que tales derivaciones eran escasas al principio (7 en los primeros 3,5 años) y los terapeutas podían tratarlas todas. (Como hemos mencionado anteriormente, se asignaba a los sujetos al grupo experimental si los terapeutas podían tratarlos; de otro modo, eran asignados al grupo control.)

Se realizaron análisis estadísticos para comprobar si la tendencia de asignar las primeras derivaciones al grupo experimental producía parcialidad. Por ejemplo, puede ser que las primeras derivaciones tuvieran un funcionamiento mayor al inicio del tratamiento o un mejor pronóstico que las derivaciones siguientes. Si fuera así, entonces el procedimiento para la asignación de sujetos habría favorecido al grupo experimental. Para evaluar dicha

posibilidad, correlacionamos el orden de derivación con el CI al inicio y el CI en el primer seguimiento (a los 7 años de edad). Se calcularon correlaciones de Pearson entre ambos grupos y en cada grupo. Tales análisis indicaron que el orden en que los sujetos fueron derivados no estaba asociado con el CI al inicio o el CI final. En consecuencia, aunque la tendencia de incluir las primeras derivaciones en el grupo experimental podía crear parcialidad, los datos indican que esto era poco probable.

Procedimiento

El procedimiento de evaluación consistía en la confirmación de la ubicación escolar y la administración de tres pruebas estandarizadas. La información sobre la ubicación escolar se obtuvo de los padres de los sujetos, que las clasificaron como clases generales o de educación especial (por ejemplo, una clase para niños con autismo o retraso mental, retrasos en el lenguaje, discapacidades múltiples, o discapacidades en el aprendizaje). Las tres pruebas estandarizadas eran las siguientes:

1. *Test de inteligencia*. La Escala Wechsler de inteligencia para niños-Revisada (Wechsler, 1974) se administró cuando los sujetos eran capaces de proporcionar respuestas verbales, lo que implicó a los 9 sujetos con mejores resultados del grupo experimental además de 8 de los 10 sujetos restantes del grupo experimental y 6 de los 19 sujetos del grupo control. Para los sujetos que no eran capaces de proporcionar una respuesta verbal, se administraron la Escala de actuación internacional de Leiter (Leiter, 1959) y la Prueba de vocabulario visual de Peabody-Revisada (Dunn, 1981). Todas estas pruebas habían sido utilizadas ampliamente para la evaluación del funcionamiento intelectual de los niños con autismo (Short y Marcus, 1986).

2. *Las Escalas Vineland de conducta adaptativa* (Sparrow, Balla y Cicchetti, 1984). La Vineland es una entrevista estructurada que se administra a los padres para evaluar hasta qué punto sus hijos presentan las conductas que se requieren para afrontar efectivamente el entorno diario.

3. *El Inventario de personalidad para niños* (Wirt, Lachar, Klinedinst y Seat, 1977). Esta medida es un cuestionario de 600 ítems de verdadero/falso que rellenan los padres y que evalúa hasta qué punto sus hijos presentan diversas formas de alteración psicológica (por ejemplo, ansiedad, depresión, hiperactividad y conducta psicótica).

Se pretendía con las tres pruebas una evaluación global del funcionamiento intelectual, social y emocional. Todas las pruebas habían sido estandarizadas en poblaciones medias. De ahí que proporcionaran una base objetiva para la comparación de los sujetos con niños sin discapacidades a lo largo de las diversas áreas de evaluación.

Se obtuvieron datos de todos los sujetos, a excepción de una chica del grupo control, de la que se sabía estaba institucionalizada y con un funcionamiento pobre. Los 9 sujetos con mejores resultados (los que habían sido clasificados como funcionamiento normal a la edad de 7 años) fueron objeto de evaluaciones especialmente amplias, como se señala más adelante. De los 28 sujetos restantes, 17 fueron evaluados por miembros del personal en nuestro programa de tratamiento y 11 recibieron evaluaciones de agencias exteriores como escuelas o clínicas psicológicas. (En algunos casos, las agencias exteriores no administraron todas las medidas incluidas en la batería.)

Evaluación de los sujetos con mejores resultados. Para garantizar la objetividad en la evaluación de los sujetos con mejores resultados, programamos la administración y puntuación a ciego de todas las pruebas para los sujetos del siguiente modo. Un psicólogo no implicado en el estudio reclutaba estudiantes de segundo ciclo de psicología clínica para administrar las pruebas. Los examinadores no tenían conocimiento del historial de los niños, y el psicólogo sólo les comunicó que la prueba formaba parte de un estudio de investigación sobre la evaluación de niños. El psicólogo les notificó que la naturaleza del estudio tan sólo requería la obtención de algunos datos informativos estándar sobre antecedentes: edad, ubicación escolar y curso, y el nombre y número de teléfono de los padres. Para aumentar la heterogeneidad de la muestra y controlar cualquier parcialidad por parte del examinador, cada uno también examinaba uno o más sujetos con la misma edad que los sujetos del grupo experimental y sin historial de alteración conductual. Se asignó a los examinadores aleatoriamente un número aproximadamente igual de sujetos para las pruebas en el grupo experimental y en el grupo de comparación. Dos sujetos del grupo experimental no vivían en el área local. Por tanto, para cada uno de ellos, el psicólogo reclutó a un examinador del pueblo del sujeto, así como a un sujeto control con la misma edad, y los datos se recogieron del modo descrito. Además, el examinador del niño rellenaba una escala de valuación clínica siguiendo una entrevista estructurada que cubría una lista de temas estándar, incluyendo amistades, relaciones familiares y actividades escolares y comunitarias. Se diseñó la entrevista para obtener contenidos y para muestrear estilo interpersonal. La escala de valuación consistía en 22 ítems, cada uno con una puntuación de 0 (mejor condición clínica) a 3 (desviación marcada). Se diseñaron los ítems para incluir posibles áreas de dificultad para los niños con autismo de inteligencia media (por ejemplo, conducta compulsiva o ritualista, empatía e interés hacia los demás, sentido del humor), como también áreas de dificultad potencial para la población general de niños (por ejemplo, estado de ánimo depresivo, ansiedad, hiperactividad). (Puede obtenerse la escala al completo y una copia de las instrucciones para la entrevista clínica mediante carta al tercer autor.)

Resultados

El grupo experimental en comparación con el grupo control

En esta primera sección, examinamos los efectos globales del tratamiento mediante la comparación entre los datos de seguimiento de los 19 sujetos que recibieron el tratamiento intensivo (experimental) y los datos de aquellos que recibieron el tratamiento mínimo (control). Se obtuvieron los datos de todos los sujetos sobre su ubicación escolar y de todos los sujetos a excepción de uno sobre su CI. En la Vineland, se obtuvieron puntuaciones para 18 de los 19 sujetos del grupo experimental, y para 15 de los 19 sujetos del grupo control. La disponibilidad más baja correspondió al Inventario de personalidad infantil, con puntuaciones para 15 sujetos del grupo experimental y 12 sujetos del grupo control. Los sujetos del grupo control con puntuaciones en el Inventario de personalidad infantil no parecían diferir de los sujetos que no tenían tales puntuaciones, tras comparar con la prueba de la *t* las diferencias en CI al inicio, CI a los 7 años, o CI en el presente estudio.

Como hemos mencionado anteriormente, 17 de los 29 sujetos que no estaban en el grupo con mejores resultados fueron evaluados por los miembros del personal del proyecto, 11 por agencias externas y 1 no recibió evaluación. Para comprobar si los miembros del personal del proyecto eran imparciales en sus evaluaciones o en su selección de los sujetos a evaluar, utilizamos la prueba de la t para comparar los sujetos que ellos evaluaron con los sujetos evaluados por agencias externas en CI al inicio, CI a los 7 años, y CI en el presente estudio. No se hallaron diferencias significativas entre los sujetos evaluados por los miembros del personal del proyecto y los sujetos evaluados por agencias externas.

Ubicación escolar. En el grupo experimental, 1 de los 9 sujetos del grupo con mejores resultados que había asistido a una clase general a los 7 años (J.L.) estaba ahora en una clase de educación especial. Sin embargo, 1 de los otros 10 sujetos había pasado de una clase de educación especial a una clase general e iba a un instituto universitario en la época del presente seguimiento. El resto de sujetos del grupo experimental no habían alterado su clasificación. En general, pues, la proporción de sujetos del grupo experimental en clases generales no había cambiado con respecto a la evaluación a los 7 años de edad (9 de 19, o 47%). En el grupo control, ninguno de los 19 sujetos se hallaba en una clase general, como se había dado en la evaluación a los 7 años. La diferencia en la ubicación escolar entre el grupo experimental y el grupo control presentaba significación estadística, $\chi^2(1, N = 38) = 19,05, p < .05$.

Funcionamiento intelectual. Las puntuaciones en las pruebas para el grupo experimental y el grupo control en funcionamiento intelectual, conductas adaptativas y maladaptativas, y funcionamiento de personalidad se resumen en la Tabla 1. Como puede observarse en la tabla, el grupo experimental en el seguimiento tenía un CI medio significativamente superior al del grupo control. Esta diferencia era significativa, $t(35) = 2,97, p < .01$. Once sujetos (17%) del grupo experimental obtuvieron CI a escala global de 80 como mínimo; tan sólo 3 sujetos (17%) del grupo control alcanzaron tales cifras. Las puntuaciones eran similares a las obtenidas por el grupo experimental y el grupo control a los 7 años (CI medio de 83 y 52, respectivamente), indicando que el grupo experimental había mantenido los logros en funcionamiento intelectual entre la edad de 7 años y el período de la presente evaluación.

Tabla 1
Puntuaciones medias y DE por grupo y medida en el seguimiento

Medida	Grupo			
	Experimental		Control	
	Media	DE	Media	DE
CI	84,5	32,4	54,9	29,1
Vineland ^a				
Comunicación	5,1	28,4	51,9	26,7
Habilidades de vida diaria	73,1	26,9	45,9	25,4
Socialización	75,5	26,8	49,7	19,9
Compuesto de conducta adaptativa	71,6	26,8	45,7	21,3
Conducta maladaptativa	10,6	8,2	17,1	7,2
Escalas del IPI ^b				
Elevación media	61,8	10,2	64,8	8,1

Traducción del artículo original de McEachin, Smith & Lovaas publicado American Journal On
Mental Retardation, 1993, Vol 97, No 4, 359-372

Escalas > 70	4,0	3,9	6,2	2,8
--------------	-----	-----	-----	-----

^a Escala Vineland de conducta adaptativa. ^o Inventario de personalidad infantil.

Conducta adaptativa y maladaptativa. En la Vineland, la puntuación media global o del compuesto fue de 72 en el grupo experimental y 48 en el grupo control. (La puntuación media para la población general en esta prueba es de 100, con una desviación estándar [DE] de 15.) En las tres subescalas —comunicación, vida diaria y socialización— cada puntuación fue paralela a la puntuación del compuesto. La interacción entre los grupos y las subescalas no fue significativa, indicando que a lo largo de las tres subescalas el grupo experimental puntuaba consistentemente más alto que el grupo control. Como puede observarse en la Tabla 1, la conducta maladaptativa era significativamente superior en el grupo control, $t(31) = 2,39, p < .05$. La puntuación media para el grupo control se hallaba en el campo clínicamente significativo, mientras que no ocurría lo mismo con la del grupo experimental. (Las puntuaciones de 13 y superiores se consideran indicativas de niveles clínicamente significativos de conducta maladaptativa a las edades de 6 a 9 años; 12 o superiores, a las edades de 12 a 13 años; y 10 o superiores, a los 14 años o mayores.) Así, los resultados indican que el grupo experimental presentaba más conductas adaptativas y menos conductas maladaptativas que el grupo control.

Funcionamiento de la personalidad. Las puntuaciones del grupo experimental y el grupo control no diferían en la elevación global de la escala, con puntuaciones t medias de 62 y 65, respectivamente. (En esta prueba, la puntuación t media para la población general es aproximadamente 50 [DE = 10].) Las puntuaciones t por encima de 60 se consideran indicativas de desviación posible o leve, mientras que las puntuaciones t superiores a 70 sugieren un problema clínico significativo, es decir, un problema que requiere atención profesional. Se halló una interacción significativa entre los grupos y las escalas individuales en esta prueba, $F(15, 390) = 2,36, p < .01$. Los resultados de la prueba de Tukey indicaron que la diferencia más fiable entre los grupos tenía lugar en la escala de psicosis, en que los sujetos del grupo experimental tenían una media de 78 y los sujetos del grupo control tenían una media de 104, $F(1, 26) = 8,53, p < .01$. Siete sujetos del grupo experimental puntuaron dentro del campo clínicamente preferente (inferior a 70), mientras que ningún sujeto del grupo control consiguió una puntuación tan baja. Tan sólo otra escala presentó una diferencia significativa, preocupaciones somáticas, $F(1, 26) = 4,60, p < .05$. Los sujetos del grupo control tendían a mostrar un nivel inferior a la media de quejas somáticas (media de 45 en comparación a la media de 54 en el grupo experimental).

Grupo con mejores resultados en comparación con el grupo no clínico de comparación

Una prueba de t indicó que no había diferencias significativas en edad entre el grupo con mejores resultados y el grupo de comparación de niños sin historial de alteración conductual clínicamente significativa. Los sujetos del grupo con mejores resultados tenían una edad media de 12,42 años (gama de 10,0 a 16,25) en comparación con 12,92 años (gama de 9,0 a 15,17) para el grupo no clínico de comparación. Las puntuaciones en la WISC-R y la escala de valuación clínica se obtuvieron para todos los sujetos; 1 sujeto del grupo experimental y 2 sujetos del grupo no clínico de comparación no tenían puntuaciones

de la Vineland, y 2 sujetos del grupo experimental y 1 sujeto del grupo no clínico de comparación no tenían puntuaciones del Inventario de personalidad infantil. Tanto la Vineland como el Inventario de personalidad infantil fueron rellenados por los padres. En los casos en que no se obtuvieron tales puntuaciones, los padres habían rehusado participar. En las medidas que proporcionaban puntuaciones estandarizadas, se midió el funcionamiento de los sujetos con mejores resultados con mayor precisión al comparar el grupo con mejores resultados con las normas de la prueba. Por tanto, este análisis es de gran interés. Los datos para el grupo no clínico de comparación son útiles principalmente para confirmar que los procedimientos de evaluación eran válidos y para proporcionar un grupo de contraste para la medida sin normas, la Escala de valoración clínica. Para el grupo no clínico de comparación, bastará con resumir los resultados de la siguiente manera: En la WISC-R, este grupo obtuvo CI medios de 116 verbal, 118 actuación, y 119 escala global. En la Vineland, el grupo obtuvo puntuaciones estándar medias de 102 comunicación, 100 habilidades de vida diaria, 102 socialización, y 101 compuesto. La puntuación media en la escala del Inventario de personalidad infantil fue de 49. Así, el grupo no clínico de comparación presentó un funcionamiento medio o por encima de la media en todas las áreas evaluadas.

La siguiente sección se centra en el funcionamiento del grupo con mejores resultados respecto a CI, conducta adaptativa y maladaptativa, y medidas de personalidad, y contrasta los sujetos con mejores resultados con los sujetos del grupo de comparación en la Escala de valoración clínica.

Funcionamiento intelectual. La Tabla 2 presenta los datos de CI para cada sujeto del grupo con mejores resultados y las puntuaciones medias para el grupo. La tabla muestra que, en conjunto, los 9 sujetos con mejores resultados obtuvieron buenos resultados en la WISC-R. Sus CI les situaron en la parte alta del campo normal, unos dos tercios de una *DE* por encima de la media. Sus CI a escala global iban de 99 a 136.

Las puntuaciones de los sujetos se distribuían uniformemente a lo largo de una gama entre 80 a 125 en CI verbal, y de 88 a 138 en CI actuación. Los sujetos presentaron una media de 3 puntos superior en CI actuación que en CI verbal. Dos de ellos (J.L. y A.G.) presentaban una diferencia de 20 puntos como mínimo entre sus CI actuación y verbal.

En cada subprueba de la WISC-R, la media para la población general es de 10 (*DE* = 3). Puede observarse en la Tabla 2 que los sujetos con mejores resultados puntuaron más alto en similitudes, diseño de bloque, y ensamblaje de objeto. Puntuaron menos en distribución de dibujo y aritmética. Por tanto, los sujetos puntuaron consistentemente en la media o por encima de la media.

Tabla 2
Puntuaciones en la WISC-R de los sujetos con mejores resultados

Sujeto	Verbal					Actuación					CI WISC-R		
	Infor	Simil	Aritm	Vocab	Compr	FinDi	DisDi	DisBl	EnsOb	Cod	CIV	CIA	Global
R.S.	12	12	13	9	11	10	9	13	12	11	106	106	106
M.C.	17	19	11	14	10	12	16	19	19	11	125	138	136
M.M.	14	13	10	14	11	12	11	11	11	8	114	102	109
L.B.	12	16	11	13	15	7	12	17	17	19	119	131	128
J.L.	6	9	7	4	8	18	11	16	14	7	80	123	100
D.E.	9	17	8	10	15	13	9	12	9	17	98	114	105

**Traducción del artículo original de McEachin, Smith & Lovaas publicado American Journal On
Mental Retardation, 1993, Vol 97, No 4, 359-372**

A.G.	7	14	12	11	13	9	4	8	11	10	108	88	99
B.W.	12	11	10	10	9	7	10	9	11	10	102	95	99
B.R.	11	14	11	13	16	12	10	12	11	10	118	106	114
Media	11.1	13.9	10.3	10.9	12	11.1	10.2	13	12.8	11.4	108	111	111

Nota. Infor = Información, Simil = Similitudes, Aritm = Aritmética, Vocab = Vocabulario, Compr = Comprensión, FinDi = Finalización de dibujo, DisDi = Distribución de dibujo, DisBl = Diseño de bloque, EnsOb = Ensamblaje de objeto, Cod = Codificación, CIV = CI verbal, CIA = CI actuación, Global = CI a escala global.

Conducta adaptativa y maladaptativa. La Tabla 3 presenta los datos para el grupo con mejores resultados en las Escalas Vineland de conducta adaptativa. Puede observarse que el grupo con mejores resultados obtuvo puntuaciones cercanas a la media en la Escala del compuesto y en las subescalas de comunicación, vida diaria y socialización. Sin embargo, la Tabla 3 muestra que algunos de los sujetos con mejores resultados tenían puntuaciones marginales, incluyendo a J.L., B.W. y M.M. Aún así, todos los sujetos del grupo con mejores resultados obtuvieron puntuaciones del compuesto dentro del campo normal. Como puede observarse en la Tabla 3, en la Escala de conducta maladaptativa (partes I y II) la puntuación media para el grupo con mejores resultados indica que, en conjunto, estos sujetos no presentaban niveles clínicamente significativos de conducta maladaptativa. Tres de ellos puntuaron en el campo clínicamente significativo en comparación con un sujeto del grupo no clínico de comparación, que tenía una media de 7,7 en esta escala.

Tabla 3
Puntuaciones en la Escala Vineland de conducta adaptativa para los sujetos con mejores resultados

Sujeto	Conducta adaptativa				Conducta maladaptativa
	Com	HVD	Soc	Comp	
R.S.	83	98	102	92	6
M.C.	119	93	86	98	16
M.M.	119	79	114	105	2
L.B.	107	108	112	108	4
J.L.	77	103	94	88	13
D.E.	93	81	82	80	15
A.G.	101	97	99	98	5
B.W.	83	74	105	83	9
B.R.	--	--	--	--	--
Media	96	92	99	94	8.8

Nota. Com = Comunicación, HVD = Habilidades de vida diaria, Soc = Socialización, Comp = Compuesto de conducta adaptativa.

Funcionamiento de la personalidad. Los resultados del Inventario de personalidad infantil se resumen en la Tabla 4. Los sujetos con mejores resultados obtuvieron perfiles válidos en el Inventario de personalidad infantil, medidos en tres escalas de validación (Mentira, Frecuencia, y Defensividad). Como puede observarse en la Tabla, los sujetos puntuaron en el campo normal en todas las escalas. Presentaban tendencia a puntuar más alto en Investigación intelectual, Psicosis y Frecuencia. La Investigación intelectual evalúa el desarrollo intelectual lento, y la Psicosis y Frecuencia evalúan conductas inusuales o

extrañas. Tan sólo la Investigación intelectual estaba por encima del campo normal, y esta escala se ve afectada por el historial anterior de los sujetos. Por ejemplo, la escala contiene sentencias como “Mi hijo habló por primera vez antes de los dos años de edad”, que sería falso para los sujetos con mejores resultados sin considerar su nivel actual de funcionamiento.

Como se indica en la Tabla 4, 4 sujetos con mejores resultados tenían una sola escala elevada por encima del campo clínicamente significativo y un quinto (J.L.) presentaba nueve escalas elevadas, incluyendo las puntuaciones más altas en el grupo con mejores resultados en Investigación intelectual, Psicosis y Frecuencia. Así, este sujeto parecía ser el responsable de gran parte de la elevación de las puntuaciones en tales escalas. En comparación, había 3 sujetos en el grupo no clínico de comparación con al menos una escala elevada.

Escala de valuación clínica. En esta escala, 8 de los sujetos con mejores resultados puntuaron entre 0 y 10, y el noveno (J.L.) puntuó 42. La media fue de 8,8, con una desviación estándar de 12,9. Todos los sujetos del grupo no clínico de comparación puntuaron entre 0 y 5 (media = 1,7, $DE = 2,1$). Al ser desiguales las DE , usamos una estadística no paramétrica, un U test de Mann-Whitney, que reveló una diferencia significativa entre los grupos, $U = 19$, $p < .05$. Así, los sujetos con mejores resultados presentaban mayor desviación que los sujetos del grupo de comparación, pero la mayor parte de la desviación parecía proceder de un solo sujeto, J.L.

Tabla 4
Puntuaciones t del Inventario de personalidad infantil para los sujetos con mejores resultados

Sujeto	Punt. t		M	F	Df	Aj	Rn	I-I	Ds	Sm	Dp	Fm	Dl	Ab	An	Ps	Hp	Sc
	Media	<70																
R.S.	56	1	49	54	43	61	53	75	49	44	69	47	46	69	60	65	46	64
M.C.	52	1	48	63	37	43	39	54	38	64	55	54	46	65	51	75	40	55
M.M.	49	0	42	54	43	50	42	64	46	58	48	55	46	47	53	46	54	36
L.B.	51	1	60	50	49	49	37	70	39	55	49	48	51	45	60	51	49	51
J.L.	70	9	42	84	37	85	77	94	65	78	86	65	61	69	78	76	52	72
D.E.	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
A.G.	51	0	38	45	49	57	48	39	53	51	49	69	40	55	55	55	49	63
B.W.	54	1	45	63	50	59	64	48	55	47	44	57	90	44	45	46	62	44
B.R.	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Media	55	2	46	56	44	58	51	64	49	57	57	56	54	56	57	59	50	55

Nota. Media = elevación media a lo largo de las escalas, M = escala mentira, F = frecuencia, Df = defensividad, Aj = ajuste, Rn = rendimiento, I-I = investigación intelectual, Ds = desarrollo, Sm = preocupación somática, Dp = depresión, Fm = relaciones familiares, Dl = delincuencia, Ab = abandono, An = ansiedad, Ps = psicosis, Hp = hiperactividad, Sc = habilidades sociales.

Discusión

El presente estudio constituye un seguimiento posterior y más extenso de dos grupos de jóvenes con autismo que habían sido previamente estudiados por Lovaas (1987): a) un grupo experimental ($n = 19$) que había recibido un tratamiento conductual muy intensivo, y b) un grupo control ($n = 19$) que había recibido un tratamiento conductual mínimo. En este

estudio facilitamos datos sobre estos niños a una edad media de 13 años para los sujetos del grupo experimental y de 10 años para los del grupo control. Los datos se obtuvieron a partir de una batería de evaluación completa.

Los principales resultados de la batería de pruebas son los siguientes: En primer lugar, los sujetos del grupo experimental habían mantenido su nivel de funcionamiento intelectual entre su evaluación previa a los 7 años y la presente evaluación a una edad media de 13 años, según la medida de pruebas de inteligencia estandarizadas. Su CI medio era unos 30 puntos superior al de los sujetos del grupo control. En segundo lugar, los sujetos del grupo experimental también presentaban niveles significativamente más elevados de funcionamiento que los sujetos del grupo control en medidas de conducta adaptativa y personalidad. En tercer lugar, en una evaluación especialmente rigurosa de los 9 sujetos del grupo experimental que habían sido clasificados como con mejores resultados (funcionamiento normal) en el estudio previo (Lovaas, 1987), los resultados de las pruebas indicaron consistentemente que los sujetos presentaban una inteligencia media y niveles medios de funcionamiento adaptativo. Se halló alguna desviación de la media en la prueba de personalidad y en las valuaciones clínicas. Sin embargo, parece que la desviación procedía de las puntuaciones extremas de un sujeto, J.L. (véanse Tablas 2, 3 y 4). Este sujeto también había sido cambiado de una clase de educación no especial a una clase para niños con retraso en el lenguaje, y obtuvo puntuaciones relativamente bajas (alrededor de 80) en la sección Verbal de la prueba de inteligencia y la sección Comunicación de la medida de conducta adaptativa. Así, parecía que ya no presentaba un funcionamiento normal. Sin embargo, los 8 sujetos restantes que habían sido clasificados previamente como con funcionamiento normal exhibieron un CI medio, con una actuación intelectual distribuida con homogeneidad a lo largo de las subpruebas, se mantuvieron en clases generales, no presentaban signos de alteración emocional, y presentaban un desarrollo adecuado de habilidades adaptativas y sociales dentro del campo normal. Además, las impresiones clínicas subjetivas de examinadores a ciego no los distinguieron de los niños sin historial de alteración conductual. Puede concluirse que estos 8 sujetos (42% del grupo experimental) habían obtenido logros mayores y más duraderos y pueden describirse como "con funcionamiento normal". Por el contrario, ningún sujeto del grupo control alcanzó un resultado tan favorable, lo que es consistente con el pobre pronóstico para los niños con autismo observado por otros investigadores (Freeman, Ritvo, Needleman y Yokota, 1985). Para evaluar este resultado, debemos considerar si nuestra metodología era rigurosa o no. La idoneidad de nuestra metodología es esencial, ya que el resultado del presente estudio representa una mejora importante sobre los resultados obtenidos en estudios experimentales previos sobre el tratamiento de los niños con autismo (Rutter, 1985). Los únicos datos de resultados comparables proceden de estudios de caso incontrolados (por ejemplo, Bettelheim, 1967), y subsiguientes investigaciones han indicado que dichos estudios de caso habían sobrestimado en gran medida los resultados obtenidos con el tratamiento que se proponía. De manera similar, los informes de mayores logros en otras poblaciones, como grandes incrementos de CI en niños de procedencia pobre, también se han basado en pruebas altamente cuestionables (Kamin, 1974; Spitz, 1986). Tales informes poseen el potencial para causar graves daños al desorientar a consumidores y profesionales.

Una descripción detallada de todas las garantías metodológicas que deberían erigirse en un estudio de tratamiento está fuera del objetivo del presente artículo (véase Kazdin, 1980; Kendall y Norton-Ford, 1982; Spitz, 1986). Sin embargo, señalamos que incluimos un gran número de garantías metodológicas tanto en el estudio original (Lovaas, 1987) como en la presente investigación:

1. El grupo experimental y el grupo control recibieron baterías de evaluación similares al inicio y presentaron varias similitudes en múltiples e importantes variables. Además, el número de sujetos del grupo control que se predijo hubieran alcanzado el funcionamiento normal de haber recibido el tratamiento intensivo, fue aproximadamente el mismo que el número de sujetos del grupo experimental que sí alcanzaron el funcionamiento normal con el tratamiento intensivo (Lovaas y Smith, 1988). Así, el procedimiento de asignación de sujetos produjo grupos comparables antes del tratamiento, lo que implica una fuerte indicación de que el funcionamiento superior del grupo experimental después del tratamiento era resultado del mismo tratamiento, y no un procedimiento parcial para asignar los sujetos al grupo experimental.
2. Todos los sujetos permanecieron en los grupos a los que fueron asignados al inicio. Tan sólo 2 sujetos abandonaron y no fueron remplazados. Por lo tanto, se preservó la composición original de los grupos en esencia.
3. Todos los sujetos fueron diagnosticados independientemente como autistas por médicos o terapeutas profesionales, y existió un alto nivel de acuerdo en el diagnóstico entre los terapeutas independientes. Lo cual proporcionó la prueba de que los sujetos cumplían los criterios para el diagnóstico de autismo.
4. Antes del tratamiento, los sujetos parecían comparables a los diagnosticados de autismo en otros estudios. La prueba de ello procede del segundo grupo control que se incorporó al estudio inicial del tratamiento. Este grupo fue evaluado por otro equipo investigador (independiente del nuestro), presentó CI similares al inicio basándose en las mismas medidas de inteligencia que utilizamos nosotros, y sin embargo mostró datos en los resultados similares a los de otros estudios. Una prueba adicional puede obtenerse de la similitud de nuestros datos al inicio con los datos presentados por otros investigadores (Lovaas et al., 1989). Por ejemplo, aunque Schopler y sus colaboradores (Schopler, Short y Mesibov, 1989) sugirieron que nuestra muestra tenía un CI medio superior al de otras muestras de niños con autismo, sus propios datos no parecen diferir de los nuestros (Lord y Schopler, 1989). Así, es evidente que nuestros sujetos eran un grupo típico de niños con autismo en edad preescolar, y no un grupo selecto de niños con autismo de elevado nivel de los que cabría esperar un funcionamiento normal con un tratamiento mínimo o nulo.
5. El primer grupo control, que recibió unas 10 horas semanales de tratamiento conductual individualizado, no se diferenciaba después del tratamiento del segundo grupo control, que no recibió tratamiento alguno por nuestra parte. Ambos grupos obtuvieron unos resultados substancialmente menos favorables que el grupo experimental. Al ser todos los grupos similares antes del tratamiento, este resultado confirma que nuestros sujetos tenían problemas que respondían solamente al tratamiento intensivo, y no problemas como ser poco obedientes o refrenarse (escondiendo un funcionamiento intelectual medio subyacente que respondería a intervenciones a escala menor).

6. Las familias de los sujetos iban de condición socioeconómica elevada a baja, y en conjunto no se diferenciaban de la población general (Lovaas, 1987). Así, aunque nuestro tratamiento precisaba una amplia participación de la familia, aparentemente un grupo diverso de familias fue capaz de satisfacer tal requisito.

7. El tratamiento se ha descrito detalladamente (Lovaas et al., 1980; Lovaas y Leaf, 1981), y se ha demostrado experimentalmente la efectividad de muchos componentes del tratamiento en un número importante de investigaciones a lo largo de los últimos 30 años (véase Newsom y Rincover, 1989). Por tanto, nuestro tratamiento puede ser replicado, un punto que se discute más detalladamente a continuación.

8. Los resultados del presente seguimiento, que se extendió a varios años después de la finalización del tratamiento en la mayoría de sujetos, son un signo esperanzador de que los logros del tratamiento se han mantenido durante un amplio período de tiempo.

9. Se administró un amplio conjunto de medidas, evitando una dependencia extrema en las pruebas de inteligencia, que presentan limitaciones si se usan aisladamente (por ejemplo, parcialidad al aprender de la prueba, seleccionar una prueba que puede ofrecer resultados más favorables, dejar de evaluar otros aspectos del funcionamiento como la competencia social o el rendimiento escolar) (Spitz, 1986; Zigler y Trickett, 1978).

10. La utilización en el seguimiento de un grupo de comparación normal, pruebas estandarizadas y valuación a ciego permitió una evaluación objetiva, detallada y cuantificable de la efectividad del tratamiento. Se efectuó una evaluación especialmente rigurosa en aquellos sujetos que presentaban mayores mejoras.

En conjunto, estas garantías proporcionan una base sólida considerable para afirmar que los resultados favorables de los sujetos del grupo experimental pueden atribuirse al tratamiento que recibieron, y no a factores externos como una mejora producida sin tratamiento, procedimientos parciales para seleccionar a los sujetos o asignarlos a los grupos, o baterías de evaluación reducidas o inadecuadas.

A pesar de las numerosas precauciones que hemos tomado, pueden surgir ciertos temores por lo que respecta a la validez de los resultados. Quizás el más importante sea que la asignación al grupo experimental o al grupo control se llevó a cabo en base a la disponibilidad de terapeutas en vez de un procedimiento más arbitrario como la alternación de derivaciones (asignando la primera derivación al grupo experimental, la segunda al grupo control, la tercera al grupo experimental, y sucesivamente). Sin embargo, parece poco probable que la asignación fuera parcial a la luz de los datos de pretratamiento que hemos presentado sobre la similitud entre los grupos experimental y control. Por otra parte, aún desconocemos si existe una variable pretratamiento que prevea los resultados, que no estuviera entre las 19 variables que escogimos pero que podría haber discriminado entre los grupos. En una publicación anterior (Lovaas et al., 1989), respondimos detalladamente a esta preocupación por la asignación de los sujetos, como a otros posibles problemas asociados al estudio original. Otras cuestiones adicionales pueden plantearse a partir de la presente investigación de seguimiento:

1. El grupo experimental tenía más edad que el grupo control en el momento de la evaluación de seguimiento. Hemos explicado este resultado anteriormente y apuntado que los análisis de datos indican que era poco probable que esta diferencia de edad reflejara una parcialidad en la asignación de sujetos.

2. Las evaluaciones de seguimiento de 17 de los sujetos con funcionamiento más bajo en el presente estudio se llevaron a cabo por parte de miembros del personal de nuestro proyecto, que podrían haber modificado los resultados de las pruebas. Sin embargo, como hemos mencionado anteriormente, una revisión no encontró prueba alguna de dicha parcialidad.

3. La Escala de valuación clínica, basada en una entrevista con los sujetos que habían sido clasificados de funcionamiento normal en el estudio original, no posee normas o datos sobre fiabilidad y validez. Sin embargo, consideramos la entrevista como una simple revisión sobre si los examinadores habían detectado signos residuales de autismo u otros problemas de la conducta que hubieran pasado desapercibidos de algún modo en las otras tres medidas (bien estandarizadas) del estudio y sus 30 subescalas. No contemplamos la entrevista como un instrumento que por sí mismo aporte resultados concluyentes. No existe actualmente ninguna otra entrevista que se adaptara a nuestro propósito. En próximas investigaciones, queremos usar una entrevista que Michael Rutter y sus colaboradores están desarrollando con el propósito de detectar signos residuales de autismo en individuos con una inteligencia media.

4. Como en la mayoría de estudios de seguimiento a largo plazo, no disponemos de ciertos datos. Pero no hay prueba alguna de que tales datos hubieran cambiado los resultados globales.

5. En nuestro análisis del grupo con mejores resultados, apuntamos que las medias del grupo se desviaban de lo "normal" en una subescala del Inventario de personalidad infantil y en la Escala de valuación clínica. Entonces atribuimos dicha desviación a las puntuaciones extremas de un sujeto, y no a problemas generales dentro del grupo. Reconocemos que las medias grupales raramente se interpretan de tal modo. Sin embargo, como han señalado estadistas y metodologistas (por ejemplo, Barlow y Hersen, 1984), muchas veces las medias grupales representan la actuación de algunos o ningún sujeto dentro del grupo. Nuestro caso fue una de esas ocasiones, como demuestran claramente los datos de los sujetos individuales (Tablas 2, 3 y 4). La desviación se hallaba casi exclusivamente en un solo individuo, sin estar distribuida uniformemente entre todos los sujetos, y de ahí nuestra presentación de los resultados.

El vacío más importante que debe llenar la investigación en estos momentos es la replicación por parte de investigadores independientes que usan metodologías rigurosas. Dados los instrumentos de evaluación objetiva que utilizamos y la detallada descripción que hemos proporcionado del tratamiento (Lovaas et al., 1980), dicha replicación debe ser posible. Sin embargo, el tratamiento es complejo y, para una replicación correcta del mismo, un investigador debe poseer: a) una base sólida en la investigación de la teoría del aprendizaje; b) conocimientos detallados del manual de tratamiento que utilizamos; c) unas prácticas supervisadas de unos 6 meses mínimos en el trabajo individualizado con clientes que presentan retrasos en el desarrollo, con énfasis en la discriminación del aprendizaje y la construcción de lenguaje complejo; y d) el compromiso de dedicar 40 horas semanales de tratamiento individualizado a cada cliente, 50 semanas al año, durante un mínimo de 2 años. Nuestros sujetos con mejores resultados precisaron de un mínimo de 2 años de tratamiento intensivo para alcanzar niveles medios de funcionamiento (otra indicación de que dichos sujetos tenían discapacidades generalizadas y no eran simplemente desobedientes).

Un segundo vacío a llenar se refiere a la mayoría de niños que no se beneficiaron hasta el punto de alcanzar el funcionamiento normal con el tratamiento intensivo. Quizás un inicio más temprano en el tratamiento era todo lo que se precisaba para obtener mejores resultados en muchos de estos niños. Más pesimísticamente, quizás tales niños requieren intervenciones nuevas y diferencias que aún no se han descubierto e implantado. En cualquier caso, resulta esencial desarrollar servicios más apropiados para estos niños.

Finalmente, una área bastante especulativa pero al mismo tiempo prometedora para la investigación es determinar hasta qué punto la intervención precoz altera las estructuras neurológicas en los niños con autismo. El autismo es casi a ciencia cierta el resultado de déficits en tales estructuras neurológicas (Rutter y Schopler, 1987). Sin embargo, estudios en laboratorio con animales han demostrado que las alteraciones en la estructura neurológica son posibles como resultado de los cambios en el entorno en los primeros años de vida (Sirevaag y Greenough, 1988), y puede creerse que tales alteraciones también son posibles en los niños. Por ejemplo, los niños menores de 3 años presentan sobreproducción de neuronas, dendritas, axones y sinapsis. Huttenlocher (1984) argumentó que, con la estimulación adecuada del entorno, tal sobreproducción podría permitir a las criaturas y los preescolares compensar las anomalías neurológicas de modo mucho más completo que lo que se da en niños mayores. Hay que ir con cautela al generalizar estos resultados con niños medios a la intervención precoz para niños con autismo, especialmente porque la naturaleza exacta de las anomalías neurológicas de los niños con autismo es aún poco clara (por ejemplo, Rutter y Schopler, 1987). Sin embargo, los resultados sugieren que la intervención intensiva precoz podría compensar las anomalías neurológicas en dichos niños. Si se hallaran pruebas de tal compensación, podría explicarse mejor porqué el tratamiento de nuestro estudio fue efectivo. Más generalmente, podría contribuir a una comprensión de las relaciones entre el cerebro y la conducta en los niños.

Bibliografía

[...]

COMENTARIOS SOBRE McEACHIN, SMITH Y LOVAAS

La asignación casi-arbitraria puede ser tan convincente como la asignación arbitraria

Donald M. Baer

Universidad de Kansas

El significado de los resultados de McEachin, Smith y Lovaas (1993) depende de los resultados del trabajo anterior de Lovaas (1987), que mostró que se habían conseguido futuros normalizados a corto plazo para la mitad de los sujetos. Esto es un triunfo para la ciencia conductista y su aplicación clínica, dada un área de problema de gran importancia caracterizada por fracasos anteriores concurrentes. El presente estudio (McEachin et al.) demuestra la durabilidad del triunfo.

No estamos acostumbrados a los triunfos en la psicoterapia. Estamos acostumbrados a proclamaciones completamente indefendibles de triunfo en la psicoterapia, normalmente basadas en un diseño de investigación y una medición pésimos. Algunos de nosotros estamos acostumbrados a proclamaciones similarmente indefendibles de que algo tan pretendidamente simple como la terapia conductual no puede tener éxito en un problema tan complejo como el autismo. Así, este estudio y su predecesor deberían someterse a una aplicación crítica de las constricciones estándar de la metodología de la investigación.

En mi opinión, ambos estudios satisfacen las estructuras metodológicas relevantes perfectamente: Están caracterizados por un diseño experimental estándar, simple, sólido y adecuado; una asignación funcionalmente aleatoria de los sujetos en grupos (debatida intensamente más adelante); un examen detenido de los resultados de tal asignación anterior a la intervención; un amplio conjunto de medidas de resultados; especial atención sobre pruebas a ciego siempre que sea posible; análisis de datos detallado e intensivo; examen minucioso y amplio de parcialidades potenciales; e interpretación de datos conservadora. Aparentemente, los éxitos obtenidos deben atribuirse a la terapia aplicada.

Una posible limitación, aunque improbable, a dicha conclusión la constituye la asignación de sujetos a los grupos experimental y control: Si un terapeuta sensible hubiera consistentemente asignado sólo los casos mejores en potencia al grupo experimental y sólo los casos peores en potencia al grupo control, esto podría haber maquillado y mejorado el tratamiento del grupo experimental. Pero no podría haber hecho que un tratamiento inefectivo pareciera efectivo: Recordemos que Rutter (1985) halló que sólo 1 de cada 64 niños mejora sin tratamiento; la asignación sistemática de este niño al grupo experimental no produciría *estos* resultados.

Idealmente, los sujetos deberían asignarse aleatoriamente al grupo experimental y al grupo control. En el estudio de Lovaas (1987), esto no fue posible: el tratamiento del grupo experimental requería recursos masivos de personal formado, y los investigadores apenas disponían de recursos adecuados; en consecuencia, los niños derivados al estudio, normalmente de uno en uno, podían asignarse al tratamiento experimental sólo si se disponía de los recursos necesarios; y después del primer año, no siempre estuvieron disponibles. Por el contrario, los niños siempre podían asignarse al programa menos exigente del grupo control. Así, si el grupo experimental no estaba disponible, se asignaba

el niño al grupo control, lo que provocó un reparto casi equitativo de los niños derivados entre los dos grupos.

Sin embargo, la cuestión crucial es si tales procedimientos obtuvieron un reparto generalmente aleatorio. Yo pienso que sí: ante tal tesitura, la condición del niño como caso mejor en potencia o caso peor en potencia, aún cuando fuera perceptible para el terapeuta, no podría haber afectado la disponibilidad de dichos recursos en el momento en que el niño estaba disponible para la asignación y, por tanto, a mi juicio, la asignación fue fundamentalmente aleatoria. (La cuestión seguirá siendo un tema de debate, ya que la norma que podría sacarla del reino del debate -utilización de una asignación verdaderamente aleatoria- no se cumple. Además, es muy típico de muchos de nosotros en la comunidad científica retirar las normas e imponer nuestro juicio en su lugar cuando no nos gustan los resultados del seguimiento de las normas del diseño competente. Quizás algunos lectores que desecharán estos resultados porque la asignación de los sujetos no fue prístinamente aleatoria de hecho están desechando los procedimientos que produjeron tales resultados y la teoría que produjo tales procedimientos; quizás hubieran hecho exactamente lo mismo si hubiera sido posible asignar los sujetos mediante una tabla de números aleatoria. De ser así, los investigadores bien pueden usar procedimientos de asignación casi aleatorios cuando los realmente aleatorios no son posibles: tienen poco que perder.)

Es la asignación en sí misma lo que estamos analizando. Si la asignación fue aleatoria funcionalmente en el primer estudio, entonces el primer estudio era sólido y convincente, y también lo es el segundo por cuanto *no* ha habido desgaste en el subgrupo relevante de los 9 sujetos del grupo experimental que sacaron un gran beneficio del tratamiento que recibieron en el primer estudio (o en los controles). No hace falta que nos hagamos las preguntas estándar sobre el desgaste potencialmente parcial. Si se puede dudar de que la asignación fue aleatoria funcionalmente en el primer estudio y, más aún, si puede sospecharse que presentaba una parcialidad sistemática en favor de los sujetos del grupo experimental, entonces el segundo estudio también presenta parcialidad.

Por mi parte, no veo casi posibilidad de tal parcialidad. Podría escribir que no veo posibilidad alguna de tal parcialidad; podría omitir el *casi* si supiera con más detalle cómo los investigadores-terapeutas llegaron al juicio de que había recursos suficientes para permitir la asignación de cada nueva derivación al grupo experimental. Si tal decisión fue fácil, una cuestión obvia del número de personas, sujetos, horas, distancias y tiempos, de modo que cualquiera que mirase los datos del momento podría haber llegado al mismo juicio de disponible o no disponible, y si pudiera ver que el juicio fue mecánico, fiable e inevitable, entonces perdería cualquier sospecha de parcialidad. Si, por el contrario, pareciera más un asunto de opinión del terapeuta que derivaba al niño sobre si había recursos adecuados para permitir que *este* niño fuera asignado al grupo experimental en ese momento, entonces vería una pequeña posibilidad de que un terapeuta sensible podría haber realizado sistemáticamente asignaciones parciales de los mejores sujetos hacia el grupo experimental. Pienso que es una posibilidad remota; nunca he presenciado tal grado de sensibilidad clínica. Además, lo que tal terapeuta mágico percibiera como “mejor” sobre las sensibilidades de los sujetos hacia el tratamiento futuro sería algo que escaparía completamente a la psicometría aplicada a cada niño: Los sujetos del grupo experimental y del grupo control presentaban medias sorprendentemente similares en la mayoría de

**Traducción del artículo original de McEachin, Smith & Lovaas publicado American Journal On
Mental Retardation, 1993, Vol 97, No 4, 359-372**

variables, y en aquellas en que se diferenciaban no me parecen predictores del éxito del tratamiento; incluso, el hecho de que los sujetos del grupo experimental fueran mayores parece señalar lo contrario.

Así, siguiendo las normas, creo que las conclusiones que presenta el estudio son fiables, correctas y excepcionalmente importantes.

Bibliografía

[...]

Efectos satisfactorios en espera de replicación independiente

R. M. Foxx

Universidad del Estado de Pennsylvania, en Harrisburg

Durante casi 30 años, Lovaas y sus coinvestigadores han derribado barreras; corrido riesgos; educado e inspirado a profesionales, personas profanas y padres; literalmente creado el campo del tratamiento conductual de personas con autismo; mostrado cuidado y limitación al presentar sus resultados; desarrollado una reputación merecida por su rigor científico en la evaluación de los resultados de sus investigaciones; proporcionado seguimientos a largo plazo; y provocado controversias. En los años sesenta, sus artículos generaban excitación y controversia porque se oponían abiertamente a las teorías y tradiciones dominantes. Sin embargo, se replicaron sus resultados por parte de otros y, después, satisficieron la prueba científica del tiempo.

Otra vez se hallan ante esta prueba. Hasta que sus espectaculares efectos no sean replicados por otros, los artículos de McEachin, Smith y Lovaas (1993) y de Lovaas (1987) ocuparán una especie de limbo científico, y continuará la polarización y la controversia asociadas a ellos. Algunos hallaron que los efectos globales son convincentes, verosímiles y creíbles, no sólo porque dichos artículos ofrecen pruebas sólidas, sino también porque Lovaas y sus colaboradores han superado la prueba de la replicación con anterioridad. Tales individuos les dirán que conocen el valor de hacer predicciones en base a la conducta anterior. Otros que siguen caminos más eclécticos o proceden de Missouri nunca aceptarán los resultados hasta que una replicación independiente haya proporcionado un apoyo irrefutable. Yo mismo me hallo en algún lugar en el medio. Me gustaría creer pero debo guardarme el juicio hasta que se haya proporcionado una verificación independiente.

Una fuente mayor de mis dudas se refiere a un problema metodológico. Tanto en el estudio de McEachin et al. (1993) como en el de Lovaas (1987), existe un problema de proporción suficiente para preocupar a cualquiera que evalúe sus provocadores resultados. Se refiere a cómo tuvo lugar la asignación de los sujetos a los grupos. En pocas palabras, la asignación de sujetos a los grupos experimental y control no fue aleatoria. Los autores reconocen tal dificultad y ofrecen un número de explicaciones verosímiles por las que este fallo metodológico no debiera restar valor a los efectos globales presentados. También proporcionan multitud de pruebas que testifican en contra de la parcialidad de los resultados. Sin embargo, otros influyentes investigadores del autismo no se han convencido a la vista de tales explicaciones y argumentos (Schopler, Short y Mesibov, 1989).

Muchos estarán de acuerdo en que la coordinación de equipos de estudiantes/terapeutas para que proporcionen 40 horas semanales de terapia era una tarea desalentadora que sin duda se vio influida por las fases del año académico. Si la asignación de grupos se llevó a cabo solamente en base a la disponibilidad de los estudiantes/terapeutas y nada más, entonces quizás no debiera preocuparnos tanto el tema de la parcialidad de la selección. Sin embargo, este tema no desaparecerá y continuará persiguiendo este proyecto hasta que sea tratado por un intento sistemático y sólido de replicación. Si tal esfuerzo revela efectos similares con sujetos asignados aleatoriamente, entonces el tema quedará zanjado.

Por lo tanto, es muy importante que Lovaas y sus colaboradores hagan todo lo posible para facilitar la replicación de sus efectos por parte de un grupo investigador independiente. Tal facilitación no sólo debería incluir el compartir los aspectos prácticos de la metodología y cualquier información pertinente relativa a la selección y asignación de los sujetos, sino también proporcionar información de las experiencias clínicas de los autores con sus sujetos. De hecho, podemos pensar que puede ser bastante complicado replicar estos efectos completamente sin haber imitado a fondo la considerable sensibilidad clínica e inteligencia de Lovaas.

Así, animaría a los autores a escribir otro artículo en que compartieran sus pensamientos especulativos en lo relativo a los factores y/o variables que creen que han contribuido al “funcionamiento normal” actual de los 8 sujetos. Por ejemplo, ¿fueron importantes el apoyo de los padres o sus conocimientos conductuales? ¿Era más probable que dichos sujetos tuvieran padres de matrimonios estables o quizás sólo uno de los padres trabajaba? ¿Fueron capaces en algún momento los estudiantes/terapeutas de realizar predicciones de éxito una vez se habían implicado con el niño, la familia y el entorno familiar? Seguramente los autores deben tener una abundancia de sensaciones “instintivas” o de intuiciones clínicas que podrían compartir.

Todos los buenos investigadores clínicos desarrollan un sexto sentido sobre sus sujetos, como conocer que lo que no hay que hacer es a veces tan importante o más que conocer lo que hay que hacer. Creo que esto es particularmente cierto en el grupo de UCLA. Espero que compartirán sus percepciones en un próximo artículo para que otros dispongan de todas las bases necesarias para replicar sus extraordinarios resultados.

Los artículos que presentan resultados excitantes son un poco como comer unos deliciosos entremeses en un restaurante. Presentan una variedad de pensamientos jugosos, se devoran rápidamente, y le dejan a uno hambriento de más y mejores delicias. Sin embargo, no importa cuán deliciosas sean estas ofrendas iniciales, un comensal experto no valora el servicio del restaurante o felicita al cocinero hasta que se ha presentado, probado y consumido el último plato. Si todo ha ido bien, entonces tenemos confianza y recomendamos a futuros comensales la calidad de la comida y la fiabilidad del servicio.

En la ciencia clínica, este proceso evaluativo es completo y el valor real de un artículo se juzga después de que sus efectos de tratamiento o sus resultados han sido replicados por investigadores independientes. Sólo entonces podemos sacar el coñac y los puros, saborear realmente los resultados, y premiar a los investigadores con el equivalente científico a las cuatro estrellas Michelin.

Si el artículo original de Lovaas (1987) era el aperitivo, entonces su continuación en McEachin et al. (1993) representa la sopa. Como su predecesor, tienta nuestros paladares científicos/clínicos, mantiene nuestra hambre, y aumenta nuestro deseo de continuar con la comida. Aún así, a veces incluso los aperitivos o las sopas más interesantes pueden amenazar la promesa de la comida, pues contienen demasiado de esto y poco de aquello. Sin embargo, estos tempranos reveses desagradables pueden superarse con la consistente excelencia de los platos siguientes.

La importancia potencial de los dos artículos requiere que tenga lugar una replicación independiente a través de un estudio multicentros financiado federalmente. De este modo,

**Traducción del artículo original de McEachin, Smith & Lovaas publicado American Journal On
Mental Retardation, 1993, Vol 97, No 4, 359-372**

sabremos si nuestras papilas gustativas científicas y clínicas estaban siendo solamente estimuladas o bien preparadas para una experiencia realmente memorable y duradera.

Bibliografía

[...]

Replicación y ampliación del tratamiento conductual del trastorno autista

Alan E. Kazdin

Universidad de Yale

La importancia del trastorno autista es difícil de sobrestimar como una disfunción clínica para los niños y sus familias. La aparición temprana, alteración generalizada y consecuencias a largo plazo en conjunto señalan los retos que el trastorno refleja. La teoría e investigación continúan avanzando; también ha progresado la investigación en el tratamiento. Aún se requieren intervenciones que puedan mejorar las habilidades adaptativas, reducir síntomas y aumentar la calidad de vida de los niños y sus familias.

Durante unas 3 décadas, Lovaas ha dirigido investigaciones sobre programas para desarrollar y entender el tratamiento de los niños con trastorno autista. En el presente artículo, McEachin, Smith y Lovaas (1993) presentan un estudio de seguimiento de la investigación original (Lovaas, 1987), diseñado para evaluar un tratamiento intensivo para los niños autistas. En el estudio original, Lovaas evaluó un tratamiento multifacético que incluía tratamiento intensivo (40 horas de atención individualizada durante 2 o más años), diversos agentes terapéuticos (por ejemplo, estudiantes, padres), atención continuada después de la fase de tratamiento intensivo (tratamiento reducido después de los 2 años si era preciso), e intervención en múltiples entornos (casa, escuela, clínica). Los resultados principales, resumidos en McEachin et al., indicaron logros marcados en diversos campos, incluyendo mejoras en el funcionamiento intelectual y ubicación en clases generales.

Contextos para la evaluación

El estudio original (Lovaas, 1987) y publicaciones afines han sido criticados desde diversos frentes. El progreso científico depende a menudo de tal criticismo; los enérgicos retos a la teoría subyacente, métodos y prácticas estimulan el debate y, más importante aún, la investigación adicional. Desde esta dialéctica científica, esperamos alcanzar logros. Una evaluación crítica del presente estudio es importante. Al mismo tiempo, el estudio original y el presente informe pueden observarse desde la perspectiva de diferentes contextos, incluyendo el campo de investigación más amplio sobre las intervenciones para niños, la investigación sobre tratamientos para el trastorno autista, y el programa de investigación del que partió este estudio en concreto.

Consideremos como una de estas perspectivas el contexto de la investigación de intervenciones psicoterapéuticas para niños y adolescentes. Existen en la actualidad centenares de estudios de resultados controlados sobre tratamientos para niños y adolescentes (véase Kazdin, Bass, Ayers y Rodgers, 1990; Kratochwill y Morris, 1991; Mash y Barkley, 1989). Los estudios contemporáneos del tratamiento para niños y adolescentes comparten un número de características. En el estudio de tratamiento habitual, el foco de atención se centra en el joven con problemas relativamente menores (por ejemplo, conductas destructivas, molestas); se pide a los sujetos que asistan al tratamiento, no son derivados; las intervenciones se llevan a cabo en las escuelas; el tratamiento es relativamente breve (de 8 a 10 sesiones); en la mayoría de casos, no hay una evaluación de

seguimiento de los efectos del tratamiento; cuando tiene lugar el seguimiento, la duración media es de 5 meses después de la finalización del tratamiento (véase Kazdin et al., 1990). En cada una de las anteriores características, los estudios de 1993 y 1987 de Lovaas y sus colaboradores son notables. Además, destacan otros puntos, como una prueba de una forma intensiva de tratamiento y la comparación de variaciones alternativas de tratamiento (versiones más intensivas en comparación con versiones menos intensivas). La evaluación de seguimiento que McEachin et al. refleja incluye esfuerzos para reducir o eliminar parcialidades por parte de los evaluadores y para establecer en el máximo grado posible que las condiciones del tratamiento y el control (tratamiento menos intensivo) eran muy probablemente equivalentes antes de la intervención. Asimismo, tienen gran importancia el uso de medidas estandarizadas como criterios primarios de resultados (por ejemplo, CI, conducta adaptativa y síntomas de disfunción) y una medida de impacto social (ubicación escolar).

Evaluación y replicación

Hay muchas cuestiones que pueden plantearse en relación con el estudio y en relación con la dirección necesaria. Una preocupación que se ha escuchado es que los jóvenes no eran “realmente” autistas o quizás no presentaban las suficientes alteraciones como en el caso de la mayoría de jóvenes que son derivados para tratamiento. Como una amenaza a la generalidad de los resultados, a menudo es importante plantear cuestiones sobre la representatividad de una muestra; si tienen base, las dudas pueden calmar el entusiasmo y las pretensiones de los resultados. Sin embargo, no rechazan necesariamente la importancia de la demostración. A no ser que exista un tratamiento estándar, establecido previamente y efectivo para los niños con trastorno autista del tipo incluido en este estudio, las supuestas características especiales que estos jóvenes presentan pueden ser dudosas. La importancia de una demostración en el tratamiento del trastorno autista no precisa que todos los casos sean tratados efectivamente o que incluso un gran segmento de un grupo representativo sea tratado efectivamente. La cuestión de interés es si cualquier niño con una alteración significativa puede ser tratado de modo que muestre cambios fiables e importantes.

Desde el punto de vista de la generalidad de los resultados, podemos dirigir un foco importante sobre el estudio mismo. Especialmente, queremos saber dentro de la muestra cómo y por qué los resultados del tratamiento difieren para desarrollar hipótesis sobre los mecanismos, factores y características de los pacientes (niño, padre, familia) que pueden moderar los resultados. Es esencial intentar entender las diferencias entre los sujetos con mejores y peores respuestas dentro del grupo de tratamiento. Un esfuerzo examinador en McEachin et al. (1993) apunta a posibles direcciones. El corto número de sujetos con una determinada condición y el número limitado de predictores imposibilita el uso óptimo de los análisis familiares (por ejemplo, regresión múltiple y logística). Por ello, en este punto, se requiere más investigación futura.

Existe la necesidad urgente de replicar el estudio original de Lovaas (1987). Los efectos marcados y duraderos del tratamiento para jóvenes con autismo, aún siendo un pequeño porcentaje dentro de la condición de tratamiento, deben identificarse. McEachin et al. es una investigación inicial, no en el programa de estudios desde el que ha evolucionado, sino

desde el punto de vista de la intervención global para niños jóvenes evaluados en pruebas clínicas controladas con un seguimiento a largo plazo. Con la promesa de un tratamiento que puede ofrecer un cambio marcado, la replicación asume una prioridad esencial. ¿Pueden reproducirse fiablemente los resultados de este estudio? Es esencial saberlo debido a la promesa que los resultados conllevan.

La replicación de la investigación también es importante para tratar puntos metodológicos y posteriores análisis. En el estudio de replicación, sería importante incluir diversas condiciones.

1. *Los casos deben asignarse aleatoriamente:* El estudio previo indicaba que los grupos eran en gran medida equivalentes antes de la intervención. Sin embargo, es inherente a los estudios con muestras pequeñas que pueda haber desigualdad entre los grupos e incapacidad de detectar las diferencias preintervención mediante las comparaciones estadísticas habituales (Kazdin, 1992).

2. *Se debería usar una muestra de casos más amplia para permitir la evaluación de las personas dentro del grupo de tratamiento que puedan variar en su grado de respuesta al tratamiento:* Estudios como el presente no son fáciles de llevar a cabo, y por eso, si se inician, debemos maximizar la producción de información y conocimiento. En un estudio de replicación, no sólo queremos evaluar los resultados producidos por el tratamiento (efectos principales), sino también disponer de pruebas sólidas para ver posibles interacciones entre paciente y tratamiento y predicciones sobre los factores que producen tales efectos.

3. *Deberían emplearse instrumentos estándar de diagnóstico para definir la alteración, incluyendo toda la gama de trastornos diagnosticables.*

4. *Debería usarse una amplia batería de evaluación para evaluar el alcance de la alteración y el funcionamiento que muestran los niños y jóvenes con autismo.*

En conclusión, el estudio de Lovaas (1987) ha sugerido lo que puede conseguirse con niños con un diagnóstico de trastorno autista. Se han proporcionado resultados prometedores y se han planteado cuestiones importantes. Entre tales cuestiones, por ejemplo, nos preguntamos sobre las razones o los mecanismos del cambio terapéutico, la base para el grado de respuesta diferencial al tratamiento, y las características de los jóvenes en relación a su disfunción que moderan el tratamiento. Dados los resultados prometedores y las cuestiones que plantean, resulta difícil esperar más de un solo estudio de intervención. El siguiente paso es la replicación, no sólo para evaluar la reproductibilidad, sino para adelantar cuestiones que mejorarán nuestra comprensión y el impacto del tratamiento.

Bibliografía

[...]

El resultado del tratamiento es esperanzador

Gary B. Mesibov

Universidad de Carolina del Norte en Chapel Hill

McEachin, Smith y Lovaas (1993) es un seguimiento de Lovaas (1987), en el que se observó que un grupo experimental de 19 niños con autismo en edad preescolar, tras haber seguido un programa de intervención conductual intensiva, había logrado ubicaciones escolares menos restrictivas y CI más elevados que en el caso de los sujetos de los grupos control.

No es mi objetivo cuestionar los resultados globales. Durante muchos años, los enfoques conductistas se han mostrado eficientes en su trato con los estudiantes que tienen autismo, y esta intervención intensiva debería producir cambios positivos y duraderos. Lo que está en cuestión es la magnitud de los cambios y lo que representan.

Debemos tener cuidado al interpretar estos resultados, ya que los sujetos no fueron distribuidos en grupos aleatoriamente. Aunque McEachin et al. (1993) intentaron demostrar por todos los medios que los grupos experimental y control eran similares, la única manera segura de evitar la posibilidad de parcialidad en los grupos es la asignación aleatoria de los sujetos.

Se plantean cuestiones sobre la representatividad y la comparabilidad de la muestra porque: a) se utilizaron diferentes edades de corte para los niños con mutismo y ecolalia, y los autores no especifican cómo se eligieron tales edades; b) el grupo control presentaba menos clientes con alto funcionamiento que lo esperado en un grupo de tal tamaño: normalmente un 20-30% de personas con autismo presentan un alto funcionamiento, sin tener en cuenta los servicios que reciben; y c) se utilizaron diferentes protocolos de pruebas para los clientes de los diferentes grupos.

La otra cuestión respecto al estudio de McEachin et al. (1993) es que los lectores pueden llegar a la conclusión de que los niños se han curado. Ésta ha sido la desafortunada consecuencia de otras presentaciones y estudios publicados por dichos autores. Aunque sus resultados son impresionantes, no llegan a demostrar el funcionamiento normal.

Las principales medidas utilizadas en el presente estudio eran restricción en la ubicación escolar, CI y conducta adaptativa. Aunque los logros en tales áreas son significativos e importantes, no indican un funcionamiento normal o incluso cercano a la normalidad. Hay muchas personas autistas con funcionamiento alto que presentan CI y conducta adaptativa casi normal y participan en programas de la escuela pública general, pero continúan siendo discapacitados graves. Pensar de otro modo significa minimizar la gravedad de los aspectos social, cognitivo y comunicativo del autismo.

Muchas de las habilidades que se requieren para el funcionamiento normal no han sido medidas por McEachin et al. (1993). No informan sobre las interacciones sociales, amistades, habilidades conceptuales y la comunicación social de los estudiantes, unas habilidades que diferencian a los niños con autismo de sus iguales sin discapacidad. Son aspectos importantes del síndrome autista y merecen mayor atención en próximos estudios. En resumen, el estudio confirma que las intervenciones conductuales son efectivas a largo plazo. Su grado de efectividad, aún lo desconecemos. Los padres de niños con autismo que

**Traducción del artículo original de McEachin, Smith & Lovaas publicado American Journal On
Mental Retardation, 1993, Vol 97, No 4, 359-372**

llevan sufriendo largo tiempo deben sentirse alentados por lo que argumentan McEachin et al. (1993), pero también deben recordar que aún queda mucho camino por recorrer.

Bibliografía

[...]

Condición normal en relación con la condición de alto funcionamiento en los niños con autismo

Peter Mundy
Universidad de Miami

Entender la naturaleza del autismo y cómo tratar efectivamente a los niños con este trastorno pernicioso presenta un reto considerable para los investigadores y los terapeutas. Por tanto, informes de resultados de intervención positivos, particularmente resultados a largo plazo como los que proporciona el grupo de Lovaas, representan contribuciones valiosas y excepcionales a este campo. Sin embargo, debido a su valor y excepcionalidad, cada uno de tales informes debe ser examinado y quizás criticado para sondear los límites de la información que proporcionan: de ahí estos comentarios sobre el reciente estudio de McEachin, Smith y Lovaas (1993).

McEachin et al. (1993) presentaron datos sobre los resultados a largo plazo de una intervención conductual intensiva que se inició con niños autistas de 3 años ($N = 19$) y finalizó cuando alcanzaron la edad de 6-7 años. Sus datos sugieren que un subgrupo de 9 niños del grupo (de intervención) experimental mostraron marcados logros intelectuales, que se mantuvieron de los 9 a los 19 años. Además, los autores sugieren que 8 de estos 9 niños del grupo con resultados óptimos no presentaron alteración emocional y demostraron habilidades sociales adaptativas dentro del campo "normal".

El informe inicial de resultados de este programa de investigación único y seminal (Lovaas, 1987) encontró críticas metodológicas considerables y dudas en lo relativo a los efectos de la intervención sobre el CI (Schopler, Short y Mesibov, 1989). Sin embargo, la resolución del grupo de Lovaas respecto a los marcados efectos de la intervención sobre el CI en un subgrupo de niños parece cada día más válida. Algunas investigaciones independientes actuales indican que existe una considerable plasticidad en la actuación cognitiva ante las pruebas por parte de los niños con autismo (Lord y Schopler, 1989) y que, si la intervención tiene lugar en el período preescolar, *algunos* niños responden con enormes cambios en su actuación en las medidas de CI (Harris, Handleman, Gordon, Kristoff y Fuentes, 1991).

Mi preocupación, pues, no es la validez de los efectos sobre la actuación en pruebas de inteligencia, aunque este punto pueda provocar un merecido debate. Al contrario, me centro en la afirmación de McEachin et al. (1993) de que su grupo con mejores resultados puede describirse de haber alcanzado un "funcionamiento normal" en las esferas del funcionamiento cognitivo, emocional y adaptativo. Para tratar este tema, su artículo debe considerarse dentro del contexto de la investigación reciente en individuos autistas con alto funcionamiento. La cuestión a considerar es, ¿Fue el resultado de la intervención una remisión completa de los síntomas autistas en un subgrupo (es decir, "funcionamiento normal") o, por el contrario, un subgrupo que presentaba síntomas que típicamente presentan los niños autistas con alto funcionamiento?

Schopler et al. (1989) se cuestionan la idoneidad de las medidas de resultados utilizadas en el informe anterior de este programa de intervención (Lovaas, 1987). Así, puede ser también importante examinar la idoneidad y las interpretaciones de las medidas utilizadas para evaluar los resultados en el estudio de McEachin et al. (1993). Las medidas que se

utilizan en este estudio, o en cualquier otro estudio, quizás no proporcionan una imagen completa de los resultados cognitivos, socioemocionales y adaptativos de los niños en este programa de intervención. McEachin et al. emplearon cuatro medidas para evaluar los resultados cognitivos, emocionales y adaptativos con el objetivo de determinar si “los sujetos con mejores resultados podían verse libres de sintomatología autista” (MS p. 5). Las medidas de los resultados incluían la Escala de inteligencia para niños Wechsler-revisada - WISC-R (con el grupo con mejores resultados), la Escala Vineland de conducta adaptativa, el Inventario de personalidad infantil, y la Escala de valuación clínica elaborada por McEachin et al. Debido a que no se presentan suficientes datos de fiabilidad y validez de la última escala para permitir una interpretación adecuada, no vamos a considerarla aquí.

La elección de una evaluación intelectual como la WISC-R para evaluar el resultado fue claramente razonable. Sin embargo, quizás no es posible concluir sólo a partir de los datos de la prueba de inteligencia o de la ubicación escolar si los individuos con autismo e inteligencia media presentan cognición típica o atípica. Los niños y adultos autistas con alto funcionamiento pueden presentar una actuación adecuada en las medidas estandarizadas de inteligencia y completar el bachillerato o asistir a escuelas universitarias, pero al mismo tiempo presentar una actuación cognitiva pobre en tareas que requieren flexibilidad en el proceso de pensamiento o proceso de pensamiento operativo (Ozonoff, Pennington y Rogers, 1992; Rumsey, Rapoport y Sceery, 1985; Szatmari, Bartolucci, Bremner, Bond y Rich, 1989; Yirmiya, Sigman, Kasari y Mundy, 1992). De hecho, McEachin et al. (1993) apuntaron que las observaciones de los padres en el Inventario de personalidad infantil sugerían que 4 de los 7 sujetos con mejores resultados tenían puntuaciones extremadamente elevadas en la Escala de investigación intelectual (70, 75 y 94) o una puntuación moderadamente elevada (65). Esta escala es sensible a pruebas de alteración cognitiva (Lachar, Kline y Boersma, 1986), y estos datos pueden indicar la presencia de anomalías cognitivas específicas o en el proceso de pensamiento en el 57% de los niños con mejores resultados evaluados con esta medida. En cualquier caso, la literatura actual sugiere que puede ser necesario ir más allá de la actuación en las pruebas de inteligencia para evaluar adecuadamente los resultados cognitivos en los niños con autismo que presentan un CI dentro del campo normal.

Investigaciones recientes con niños autistas con alto funcionamiento también sugieren que el uso del Inventario de personalidad infantil y la Vineland puede haber producido una imagen limitada del resultado emocional, social y adaptativo del grupo con mejores resultados.

El estado emocional y social de los individuos autistas con alto funcionamiento puede verse comprometido de maneras específicas aunque no fácilmente observables. Las personas autistas con alto funcionamiento pueden mostrar habilidades adaptativas y resultados sociales relativamente buenos, pero seguir discapacitados por procesos de pensamiento extraños, pensamientos y preocupaciones obsesivos, estado de ánimo depresivo, etc. (Dykens, Volkmar y Glick, 1991; Rumsey et al., 1985; Szatmari et al., 1989). Es decir, los individuos autistas con alto funcionamiento pueden presentar tendencia a experimentar síntomas interiorizados, que no exteriorizados, de alteración emocional y social. Sin embargo, las medidas de información de los padres sobre el estado emocional y conductual de los hijos, como el Inventario de personalidad infantil o la Vineland, se muestran mejores

en el aislamiento de las formas exteriorizadas de alteración como la oposicionalidad, y no en formas interiorizadas de alteración como las preocupaciones obsesivas (por ejemplo, Edelbrock, Costello, Dulcan, Conover y Kalas, 1986). Lovaas (1987) reconoció tal posibilidad al decir que “pueden permanecer déficits en el grupo de funcionamiento normal que los maestros o los padres no puedan detectar” (p. 8). Sin embargo, la elección de las medidas de resultados en McEachin et al. no refleja tal reconocimiento. Así, una imagen más completa de los resultados de este programa de investigación deberá esperar al uso de medidas multi-informadores, incluyendo una entrevista de diagnóstico fiable y válido de los niños.

A pesar de la cuestión sobre la sensibilidad de la evaluación, también es importante examinar la interpretación del estudio de McEachin et al. (1993), especialmente en lo relativo al Inventario de personalidad infantil. Sugieren que estos datos, con la excepción de las puntuaciones de un niño, indicaban que los sujetos del grupo con mejores resultados no presentaban alteración emocional. Sin embargo, sin incluir a este niño (J.L.), que tenía elevaciones múltiples en las escalas del Inventario de personalidad infantil, cuento otros 2 niños en la Tabla 4 con puntuaciones dentro del campo clínico (70+) en el Inventario de personalidad infantil. Un niño tenía una elevación en la escala Psicosis (M.C.) y, aunque parezca mentira, el otro presentaba elevación en la escala Delincuencia (B.W.). Además, otro niño presentaba elevaciones límite de escala de 69 en las dos escalas Depresión y Abandono (R.S.). Así, además del sujeto J.L., los datos de la información de los padres sobre el Inventario de personalidad infantil identificaron 4 niños que parecían presentar una alteración emocional o cognitiva notable. Dada la falta potencial de sensibilidad de tal medida mencionada anteriormente, y el hecho de que sólo 7 niños fueron evaluados con el Inventario de personalidad infantil, parece difícil interpretar estos datos como prueba de la consecución del funcionamiento normal en el grupo con mejores resultados.

En resumen, la interpretación del conjunto de datos de McEachin et al. (1993) debería haberse realizado más cuidadosamente y verse limitada por los límites de las medidas de los resultados. Las medidas de los resultados utilizadas pueden no haber tenido la suficiente especificidad o amplitud para tratar la cuestión sobre si el grupo con mejores resultados presentó una remisión de síntomas autistas o presentó síntomas típicos de los niños autistas con alto funcionamiento. Hago constar, sin embargo, que esta crítica se basa en los datos recientes sobre los individuos autistas con alto funcionamiento que quizás no estaban disponibles cuando McEachin et al. iniciaron su estudio de resultados. Además, esta crítica no se ofrece en un intento de desacreditar los efectos positivos en los resultados que se han conseguido en este programa de intervención. Sino que se ofrece para sugerir que los límites de la interpretación de este conjunto de datos deben reconocerse y tratarse en próximas investigaciones con la misma muestra. McEachin et al. empezaron el camino en esta dirección al apuntar que su próxima investigación debía incluir un instrumento de entrevista con padres que fuera sensible a los síntomas de los niños autistas con alto funcionamiento. Este primer paso sería bien recibido, pero muchos otros pasos son posibles y deben considerarse.

Bibliografía

[...]

Comentarios sobre Replicación y evaluación del resultado

Tristram Smith, John J. McEachin y O. Ivar Lovaas
Universidad de California, Los Ángeles

El debate puede ayudar o herir, dependiendo de si agudiza el pensamiento sobre un tema o si simplemente crea un ambiente hostil que imposibilita la investigación futura. Desafortunadamente, en el tema del tratamiento de niños con autismo, el debate destructivo ha sido lo más frecuente (Lovaas, 1989), y, por este motivo, los presentes comentarios son especialmente bien recibidos. A nuestro parecer, aunque los comentaristas expresan diversas reacciones a los resultados que hemos presentado, todos ellos establecen un equilibrio entre el rechazo arbitrario y la aceptación sin sentido crítico de nuestros resultados, y están de acuerdo en muchos puntos principales: En primer lugar, parece que nuestros sujetos con mejores resultados consiguieron logros notables. En segundo lugar, hemos presentado una razón verosímil al atribuir estos logros a nuestro tratamiento en vez de a factores externos como la recuperación espontánea. En tercer lugar, este resultado requiere replicación, ya que un solo estudio no puede proporcionar pruebas concluyentes sobre la eficacia de un tratamiento. Finalmente, las replicaciones no deberían simplemente ser un refrito de lo que ya se ha hecho. Al contrario, deberían basarse en trabajos anteriores para mejorar el diseño, los procedimientos de evaluación y plantearse nuevas cuestiones. Tan sólo podemos añadir que es importante que las próximas investigaciones revelen modos de mejorar el tratamiento. La mayoría de nuestros sujetos no se hallaban en el grupo con mejores resultados, y la identificación de los servicios más apropiados para estos sujetos y niños como ellos representa un enorme reto.

El acuerdo en los comentarios sobre estos temas es impresionante. Sin embargo, puede crear la impresión que la realización de replicaciones y la evaluación de resultados son más sencillas de lo que en realidad son. A continuación, intentaremos provocar más deliberación sobre estos temas y contestar las críticas que aparecen en los comentarios.

Replicación

Empezamos a colaborar en los esfuerzos de replicación en 1988. El progreso ha sido lento, pero ya hay tres proyectos de replicación a punto de empezar. Los principales investigadores de cada proyecto tienen el objetivo de incorporar las importantes precisiones metodológicas que Kazdin (1993) apunta. A pesar de estos prometedores desarrollos, nuestra experiencia confirma ampliamente la observación de Kazdin que pocos investigadores se interesan por el desarrollo del tipo de investigación sobre tratamiento que hemos realizado.

Un impedimento importante es que las dificultades logísticas que conlleva la replicación de nuestro tratamiento y evaluación puedan parecer enormes. Foxx (1993) observó astutamente que en nuestro tratamiento se esconde mucho más de lo que puede expresarse en la sección de Método y nos sugirió que escribamos un artículo sobre los matices que hemos observado. Un artículo de tal naturaleza sería útil, como lo son el manual de tratamiento y los videos que ya hemos publicado (Lovaas et al., 1980; Lovaas y Leaf,

1981). En definitiva, sin embargo, sólo se aprende un poco a través de la lectura y la observación. Para proporcionar un tratamiento competente, se debe recibir una formación extensiva, práctica y supervisada. Lo que es también cierto en el caso de la evaluación, ya que los procedimientos estándar de administración de pruebas normalmente no funcionan con los niños que tienen autismo (Freeman, 1976). Un requisito adicional es el dominio de los principios teóricos subyacentes en el tratamiento y la evaluación. Dicho dominio es necesario para dar significado a las experiencias formativas, aplicarlas adecuadamente después de finalizar la formación y mantenerse al corriente de los nuevos desarrollos. Así, los investigadores que lleven a cabo un proyecto de replicación deben realizar una extensa preparación. Estimamos que esta preparación requiere como mínimo de 6 a 12 meses.

Naturalmente, esta preparación es tan sólo el principio. Una vez empezado el camino, el proyecto de replicación puede durar unos 5 años y ocupar la mayor parte del tiempo del investigador hasta su finalización. No es probable que investigadores mayores, más establecidos, quieran dedicar tanto esfuerzo a la replicación de un programa de tratamiento existente. Los jóvenes investigadores, a los que a veces sus universidades juzgan por la rapidez en que publican, pueden apartarse de este tipo de proyecto, porque requiere mucho tiempo. Presiones como éstas han llevado a un exceso de estudios de tratamiento a pequeña escala (y una escasez de estudios a gran escala), en detrimento de los clientes, que necesitan servicios validados científicamente, a largo plazo y globales.

Evaluación del resultado

Cuando iniciamos el Proyecto de autismo juvenil en 1970, se consideraba el autismo como una condición crónica, y se veía el tratamiento conductual como una intervención superficial que como mucho proporcionaba un aligeramiento de los síntomas. Los datos de los resultados de los clientes que tratamos en los años sesenta (Lovaas, Koegel, Simmons y Long, 1973) apoyaban esta percepción, ya que los clientes presentaron una mejora limitada durante el tratamiento y a menudo recaían después. Afortunadamente, investigadores de muchos lugares identificaron maneras de mejorar el tratamiento. Incorporamos tales mejoras en nuestro nuevo proyecto, al tiempo que reteníamos aspectos del tratamiento que habían sido satisfactorios. Así, el tratamiento conductual se desarrolló de modo acumulativo. En consecuencia, esperamos obtener mejores resultados en el Proyecto de autismo juvenil que en el estudio de Lovaas et al. (1973). Sin embargo, no anticipamos que tantos de nuestros clientes alcanzarían el funcionamiento normal, y no diseñamos una batería de evaluación para nuestra primera evaluación de seguimiento (Lovaas, 1987) con tal resultado en mente. Nuestro objetivo era mostrar que los clientes podían conseguir logros significativos con el tratamiento mejorado y que tales logros les permitirían funcionar en entornos menos restrictivos.

Incluso cuando diseñamos el seguimiento que se presenta en este artículo, no predijimos lo que encontraríamos. Por una parte, esperábamos descubrir que nuestros sujetos con mejores resultados continuaban funcionando con normalidad. Por otra parte, sabíamos que era posible que hubiesen desarrollado problemas. Crecer es difícil incluso para los niños sin historial psiquiátrico. Nuestros clientes podían haber recaído, desarrollado efectos secundarios imprevistos a partir del tratamiento, dejado de mostrar síntomas de autismo,

pero al mismo tiempo con riesgo de desarrollar otros problemas diagnosticables, etc. Así, necesitábamos instrumentos de evaluación “de uso múltiple” adecuados para examinar un amplio abanico de problemas.

Algunos de los comentaristas de esta edición sugieren que debíamos haber incluido más pruebas para los signos residuales de autismo. Muchas de las pruebas defendidas en los comentarios son diseñadas para evaluar una teoría concreta sobre el “déficit central” de autismo (por ejemplo, flexibilidad alterada en los procesos de pensamiento, ausencia de una teoría de la mente). Reconocemos el valor potencial de añadir tales pruebas nuevas, y lo haremos en próximos seguimientos. Sin embargo, somos de la opinión de que nuestra batería de evaluación ofrece grandes oportunidades de detectar signos de autismo (véase la siguiente sección), y afirmamos que los comentaristas tienen demasiada fe en las pruebas de teorías específicas. Los individuos con autismo difieren de las otras poblaciones de incontables maneras. Los investigadores no se ponen de acuerdo sobre cuál de tales diferencias es el “déficit central”, y sus visiones han cambiado a lo largo del tiempo. De hecho, desde 1965, uno o más investigadores han propuesto un nuevo déficit central cada año, y algunos han declarado haber descubierto el déficit central en más de una ocasión. Durante este período, ha habido cuatro tendencias mayores como mínimo (compárese con Yirmiya y Sigman, 1992): Hasta los años sesenta, investigadores como Bettelheim y Kanner ponían el énfasis en la dificultad de los niños con autismo en conseguir lazos emocionales con sus padres. A mediados de los sesenta, esta visión recibió fuertes críticas, e investigadores como Rimland y Goldfarb cambiaron el centro de atención hacia los déficits en la percepción y la estimulación sensorial. En los años setenta, esto dejó paso a teorías sobre las alteraciones cognitivas y del lenguaje. Teorías muy influyentes fueron establecidas por Rutter, DeMyer y Wing, entre otros. En los años ochenta, algunos investigadores continuaron esta línea, pero la mayor parte de la atención se volvió hacia las alteraciones sociales, sociocognitivas o socioemocionales. Mundy y sus colaboradores contribuyeron significativamente al aumento del interés en dichas variables.

Puede que los investigadores lleguen a identificar el déficit central del autismo. Sin embargo, tenemos nuestras dudas de que pueda pasar. Los individuos con autismo difieren de otras poblaciones de tantas maneras que es muy fácil producir nuevas teorías sobre cuál puede ser el déficit central del autismo, pero es muy poco probable que ninguna de estas teorías demuestre ser la verdadera. Un problema incluso más fundamental es que a lo mejor no existe un déficit central. Los investigadores que buscan el déficit central del autismo toman tres premisas que pueden ser válidas o falsas: En primer lugar, el autismo es más que una simple etiqueta otorgada a ciertos niños: es una entidad real. En segundo lugar, existe una forma de autismo, no varias. En tercer lugar, el autismo tiene un déficit central, a diferencia de otras enfermedades como la parálisis cerebral que no lo tienen (Rutter, 1978; Rutter y Schopler, 1987). Todas estas premisas son perfectamente razonables, tan razonables que muchos investigadores las dan por sentado. Sin embargo, ninguna de ellas ha sido verificada hasta el momento, a pesar de los 50 años que llevan intentándolo. Por tanto, pueden acabar siendo un “callejón sin salida”. En vez de ir por este callejón, nosotros y otros conductivistas hemos intentado desarrollar maneras alternativas de tratar los problemas de los niños diagnosticados con autismo (Lovaas y Smith, 1989). Al evaluar los resultados, nos hemos basado en baterías de evaluación que son: a) lo bastante amplias para

analizar una variedad de áreas importantes del funcionamiento (en vez de centrarse en un problema central), b) lo bastante globales para detectar problemas que podemos no haber previsto (en vez de probar una hipótesis específica), y c) lo bastante duraderas para soportar la prueba del tiempo incluso si las teorías del autismo cambian o si el concepto del autismo se rechaza completamente.

Respuesta a las críticas

Varios comentaristas citan la crítica de Schopler, Short y Mesibov (1989) a nuestro primer informe sobre el Proyecto del autismo juvenil (Lovaas, 1987). Mundy (1993) se refiere al argumento de Schopler et al. de que las pruebas de inteligencia son medidas engañosas del resultado. Schopler et al. afirmaban que tales pruebas, cuando se administran a niños con autismo, pueden evaluar la conducta alterada y no el funcionamiento intelectual. Este argumento se ve ampliamente rebatido por la extensa investigación que se ha llevado a cabo sobre la utilización de tales pruebas y otras medidas de habilidad con niños que tienen autismo (Rutter, 1983). Kazdin (1993) aludía a la aseveración de Schopler et al. de que nuestra muestra no era un grupo representativo de niños con autismo. Concretamente, Schopler et al. comentaban que nuestra muestra presentaba un CI al inicio superior a lo que mostraban la mayoría de muestras de niños con autismo y que los niños autistas con un CI elevado pueden mejorar independientemente del tratamiento. Como señalamos en la discusión (McEachin, Smith y Lovaas, 1993), parece que ambas afirmaciones se contradicen con los propios datos de los autores. Por ejemplo, una muestra estudiada por Lord y Schopler (1989) tenía un CI medio casi idéntico al de nuestra muestra, y los sujetos con mayor CI tendían a presentar disminución del CI a lo largo del tiempo. (Para comentarios adicionales sobre la crítica de Schopler et al., véase Lovaas, Smith y McEachin, 1989).

La críticas sobre nuestro último informe (McEachin, 1993) se centran en nuestro procedimiento de asignación de sujetos y la batería de evaluación. Agradecemos la apreciación de Baer (1993) sobre la asignación de sujetos y no tenemos nada que añadir. Sin embargo, queremos tratar el tema de la evaluación.

El principal problema, como señala detalladamente Mundy (1993), es que los sujetos con mejores resultados de nuestro estudio pueden presentar el tipo de déficits cognitivos y socioemocionales que exhiben los individuos autistas con "alto funcionamiento". Creemos que este resultado sería sorprendente, ya que nuestros sujetos tenían historias vitales muy diferentes de la gran mayoría de tales individuos. Proporcionamos literalmente miles de horas de tratamiento, utilizando procedimientos validados experimentalmente, con el objetivo de ayudar a nuestros sujetos a hacer amigos, desarrollar intereses apropiados a su edad, adaptarse y aprender en clases generales, etc.

Para evaluar posibles similitudes entre nuestros sujetos con mejores resultados e individuos autistas con alto funcionamiento, Mundy (1993) recomienda el uso de medidas adicionales para evaluar áreas como la flexibilidad del proceso de pensamiento y las operaciones de teoría mental. Sin embargo, en los estudios que cita, los sujetos llamados "de alto funcionamiento" pueden fácilmente distinguirse de nuestro grupo con mejores resultados y no "normal". Sus diferencias pueden detectarse sin recurrir a medidas especializadas que

tienen sus propias limitaciones (véase la sección anterior). Por ejemplo, en el estudio de Mundy (Yirmiya, Sigman, Kasari y Mundy, 1992), 7 de los 18 sujetos se describieron como que presentaban “grandes dificultades en las relaciones sociales y la vida profesional” (p. 152); el resto sufría dificultades incluso más generalizadas. En el estudio de Dykens, Volkmar y Glick (1991), 9 de los 11 sujetos vivían en una casa de grupo, piso supervisado o otra instalación residencial. Al contrario que estos sujetos, nuestros sujetos con mejores resultados vivían en sus casas, y todos excepto uno asistían a clases generales. Los sujetos de Dykens et al. también tenían un CI de actuación medio de 76, en comparación a una media de 111 en nuestros sujetos con mejores resultados. En el estudio de Ozonoff, Pennington y Rogers (1991), los sujetos autistas con alto funcionamiento tenían un CI verbal medio prorrateado de 83 en la WISC-R, excluyendo la subescala de comprensión. Basándonos en otro estudio (Yirmiya y Sigman, 1992), asumimos que las puntuaciones en esta subescala deberían haber sido mucho menores. Nuestros sujetos eran muy diferentes, con un CI verbal medio de 108, incluyendo la subescala de comprensión, en la que todos actuaron en el campo normal o por encima. Los sujetos de Ozonoff et al. también tenían una puntuación media en las escalas de valoración del autismo infantil de 34, indicando una conducta obviamente autista. Otro ejemplo de como diferían nuestros sujetos con mejores resultados de los individuos autistas con alto funcionamiento aparece en Venter, Lord y Schopler (1992). Los sujetos de este estudio obtuvieron una puntuación media de la Vineland compuesta de 42 en comparación a 94 de nuestros sujetos con mejores resultados. En todos estos estudios, aunque los sujetos tenían un funcionamiento superior que la mayoría de individuos con autismo, presentaban déficits que eran visibles en las medidas estándar de los resultados. Citando a una de las fuentes de Mundy, “Todos [los sujetos] serían considerados ciertamente peculiares por los legos” (Rumsey, Rapoport y Sceery, 1985, p. 468).

No nos parece razonable utilizar medidas especializadas cuando muchas medidas clínicas estándar pueden decirnos lo mismo. Por ahora, nadie ha demostrado ningún déficit cognitivo o socioemocional importante en un grupo de sujetos con autismo que asisten a clases generales y tienen un CI verbal por encima de 90, puntuaciones en la subescala de comprensión por encima de 90, y puntuaciones en la Vineland compuesta por encima de 80 (cosa que describe a 8 de los 9 sujetos con mejores resultados de nuestro seguimiento). Por el contrario, en los raros casos en que los individuos con un historial de autismo han cumplido estos criterios, se les ha descrito de recuperados (Szatmari, Bartolucci, Brunner, Bond y Rich, 1989). Añadiendo el hecho de que los sujetos presentaban perfiles corrientes del Inventario de personalidad infantil y las entrevistas clínicas (véase la discusión posterior), consideramos que es adecuado considerar su funcionamiento socioemocional como esencialmente normal.

Mundy (1993) y Mesibov (1993) aportaron otras críticas específicas a nuestro estudio:

1. Mundy sugiere que los sujetos con mejores resultados podían haber presentado déficits “específicos aunque no detectables” (p. 382) que nuestra batería de evaluación no pudo identificar. En concreto, especula que estos sujetos podían tener formas interiorizadas de alteración como pensamientos extraños o ansiedad excesiva, y cita dos estudios que muestran como los padres tienden a minimizar tales problemas. Sin embargo, el cuestionario que usamos para los padres, el Inventario de personalidad infantil, contiene

procedimientos de derivación empírica que compensan este tipo y otros de parcialidad en las respuestas. Además, en nuestra entrevista clínica, se instruyó a terapeutas expertos para que preguntaran a los sujetos directamente sobre problemas interiorizados y para observar signos externos de la existencia de tales problemas. Es posible que problemas significativos escaparan a nuestra atención a pesar de tales garantías, pero pensamos que es poco probable.

2. Mundy pone gran énfasis en la elevación en la escala de investigación intelectual del Inventario de personalidad infantil. En nuestro artículo (McEachin et al., 1993), atribuimos la elevación a los ítems sobre el desarrollo temprano. La escala contiene dos de estos ítems: "Una vez mi hijo tuvo problemas al hablar" y "Mi hijo habló por primera vez cuando tenía dos años". Estos ítems -con respuesta verdadero y falso, respectivamente- añadían de 10 a 12 puntos a las puntuaciones de J.L., L.B., R.S. y M.M. (los sujetos cuyas puntuaciones estaban elevadas) y el mismo número de puntos a las puntuaciones de los restantes sujetos con mejores resultados. Por eso, no nos preocupan las puntuaciones en la escala de Investigación intelectual. Para situar las puntuaciones en contexto, deberíamos mencionar que, como indica su nombre, esta escala tan sólo es un instrumento de investigación (Wirt, Lachar, Klinedinst y Seat, 1984) y nadie hasta Mundy ha sugerido que debería preceder a la WISC-R cuando se dispone de la posterior (ni Lachar, Kline y Boersma, cuyo capítulo de 1986 cita Mundy en conexión a este tema).

3. Mundy señala que algunos de los sujetos con mejores resultados presentan puntuaciones elevadas en el Inventario de personalidad infantil. Es importante destacar que un número similar de sujetos en nuestra muestra no clínica también presentaba puntuaciones elevadas, como reflejamos en nuestra sección de *Resultados* (McEachin et al., 1993). Además, varias de las elevaciones obtenidas por los sujetos con mejores resultados pueden haber sido falsas, como hemos explicado anteriormente.

4. Mundy rechaza nuestra entrevista clínica porque le faltan datos de fiabilidad y validez. Estamos de acuerdo en que dicha información es importante, pero pensamos que, aún sin ella, los resultados de la entrevista son notables. Es interesante que terapeutas formados, independientes de nosotros y sin conocimiento de los historiales de los sujetos, hallaran pocas dificultades de adaptación en 8 de los 9 sujetos con mejores resultados.

5. Mesibov argumenta que es inapropiado otorgar a nuestros sujetos con mejores resultados la etiqueta de "funcionamiento normal", porque no llegamos a medir "interacciones sociales, amistades, habilidades conceptuales y comunicaciones sociales" (p. 380). De hecho, todas estas variables fueron medidas en la entrevista clínica y el Inventario de personalidad infantil, todas a excepción de las habilidades conceptuales fueron medidas por la Vineland, y las habilidades conceptuales fueron medidas por la WISC-R.

6. Mesibov observa que no explicamos por qué utilizamos diferentes edades de corte para los niños con mutismo y los niños con ecolalia. Anteriormente, él y sus colaboradores pidieron tal explicación (Schopler et al., 1989) y nosotros se la ofrecimos (Lovaas et al., 1989).

7. Mesibov argumenta que nuestro grupo control presentaba más alteraciones que los grupos típicos de niños con autismo y especifica que el 20-30% deberían presentar alto funcionamiento. Si el *alto funcionamiento* significa tener un CI superior a 70 (la definición más común), entonces 26% de nuestros sujetos del grupo control (5 de 19) entran en tal

definición. Este resultado es consistente con lo que podría predecirse a partir de las medias y desviaciones estándar del CI que aparecen en la Tabla 1 de McEachin et al. (1993). Para establecer tal predicción pueden usarse procedimientos estadísticos bien establecidos (convirtiendo un CI de 70 en una puntuación *Z* y consultando una tabla para el porcentaje esperado de puntuaciones en este nivel o por encima).

8. Mesibov comenta correctamente que clientes diferentes recibieron diferentes baterías de evaluación. Como señalamos en nuestra sección de *Método*, las baterías de evaluación se individualizaron de dos modos: En primer lugar, usamos un procedimiento estándar para seleccionar la prueba de inteligencia que mejor se adaptaba al nivel de funcionamiento de cada sujeto. En segundo lugar, empleamos garantías metodológicas especialmente estrictas en nuestras evaluaciones de los sujetos con mejores resultados. Mesibov sugiere que tales procedimientos debilitan nuestro estudio, pero creemos que tuvieron el efecto contrario, ya que aumentaron la idoneidad y el rigor de la evaluación. Es una práctica generalizada el seleccionar las pruebas de inteligencia individualmente para acomodar las grandes diferencias que hay entre los niños con autismo, y Mesibov ha sido uno de los principales defensores de dicha práctica. El uso de garantías metodológicas adicionales en la evaluación de los sujetos con mejores resultados era asimismo importante, ya que minimizaba la posibilidad de obtener resultados parciales.

Aunque no estamos de acuerdo con las críticas específicas que se han planteado, coincidimos en el punto general de que deberíamos esforzarnos en mejorar nuestra metodología. Nuestro seguimiento del 1987 era mejor que el seguimiento del 1973, y el presente seguimiento es aún mejor, aunque ciertamente no es perfecto. Esperamos que los esfuerzos de replicación, cuando finalmente tengan lugar, constituyan un nuevo avance.