

- CHANO, V. & BAGUENA, M.J. (1983): Un cuestionario de Locus de control (LUCAM). *Análisis y Modificación de Conducta*, 9(20), 5-40.
- ER, J.B. (1966): Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80, 609.
- WSKI, C.J. & BLACKWELL, M.W. (1987): The relationship of locus of control to anxiety and coping among student teachers. *College Student Journal*, 187-189.
- OE, A. (1985): Theoretical bases for task selection in cardiovascular psychophysiology. En A. Steptoe y col. (eds.), *Clinical and methodological issues in cardiovascular psychophysiology*. Berlin-Heidelberg: Springer Verlag.
- DOORNEN, L.J.P. (1988): *Physiological stress reactivity: Its relationship to aviorial style, mood, sex, and aerobic fitness*. Doctoral Dissertation. Vrije universiteit te Amsterdam.
- SLER, D. (1984): *Escala de inteligencia de Wechsler para adultos*. Manual adaptado por Yela y colaboradores, Madrid: TEA ediciones.

## ELABORACIÓN DE UN SISTEMA DE EVALUACIÓN Y DIAGNÓSTICO CONDUCTUAL DEL RETRASO EN EL DESARROLLO

J. Gil Roales-Nieto (\*)  
A. Polaino-Lorente (\*\*)

(\*) UNIVERSIDAD DE GRANADA

(\*\*) UNIVERSIDAD COMPLUTENSE DE MADRID

### RESUMEN

El trabajo de investigación representa la elaboración de un Sistema de Evaluación y Diagnóstico Conductual (SEDC) para el Retraso en el Desarrollo, desde la perspectiva de la evaluación conductual. Para la confección de las escalas de desarrollo que lo componen se eligieron 120 sujetos con edades entre los 6 meses y los 6 años; evaluándose directamente su nivel de desarrollo en todas las áreas conductuales. Las escalas así resultantes forman la base del SEDC, que ha sido estructurado en forma que permite (1) la valoración o medición de los repertorios del sujeto, (2) el diagnóstico conductual del mismo, (3) la especificación de los objetivos a programar y secuencia de los mismos, (4) la evaluación de la efectividad de la intervención, y (5) la posibilidad de convertirse en instrumento de prevención del retraso en el desarrollo.

### SUMMARY

A Behavioral Assessment and Diagnosis System (BADS) for developmentally retarded children is presented as a result of an extensive study from a behavioral standpoint. 120 children, aged 6 months to six years old were evaluated with regard

to five different general repertoires. The educational use of the BADS permits (1) the assessment of the actual subject's repertory; (2) the behavioral diagnosis of the subject; (3) to specify the type of behavioral goals and the level in which training has to begin; (4) the evaluation of the intervention effectiveness; and (5) the possibility to become a instrument to prevent developmental retardation.

El trabajo de investigación que recoge este artículo, representa un intento de extender los principios y técnicas de la evaluación conductual (Cone y Hawkins, 1977; Haynes, 1978; Hersen y Bellack, 1981; Nelson y Hayes, 1979), a uno de los campos aplicados más tradicionales en el estudio de la medida en psicología y educación: el área del retraso en el desarrollo o deficiencia mental. Asimismo, también representa un intento de solucionar una carencia práctica largo tiempo detectada (Gil Roales-Nieto, 1983).

La evaluación y diagnóstico de los sujetos retrasados realizada durante décadas exclusivamente, o casi, en base a criterios paramétricos, ha sido puesta en duda en cuanto a su utilidad aplicada - clínica, educativa y social - (p.ej. Bigelow, 1977; Bijou y Grimm, 1975; Bijou y Paterson, 1971; Haywood, 1977; Haywood et al., 1975; Kazdin y Straw, 1976).

Los intentos por superar las limitaciones y consecuencias de un "diagnóstico" del retraso escuetamente basado en una etiqueta médica y un cociente de inteligencia, son casi tan antiguos como la posición misma que pretenden modificar. Sucesivamente han ido apareciendo alternativas que, con mayor o menor fortuna, han intentado en ciertos casos completar -y en otros sustituir-, el diagnóstico médico-normativo del retraso. Por citar sólo algunos de los ejemplos más significativos, mencionaríamos los trabajos de Doll (1936, 1947), Mercer (1978), y Gunzburg (1969), así como la posición general establecida por la American Association on Mental Deficiency en su Manual de 1959.

Todos ellos representaron avances interesantes y relativamente fructíferos, pero al mismo tiempo compartían defectos comunes con el punto de vista normativo que les impidieron convertirse entonces en las alternativas globales que la práctica psicológica y educativa del retraso mental exigía. Una limitación común es que continuaban manteniendo separadas las funciones diagnóstica y terapéutica-rehabilitadora.

A pesar de que la evaluación y diagnóstico del retraso mental o retraso en el desarrollo (en lo sucesivo rd), atrajo y continúa atrayendo, la atención de un considerable número de investigadores (ver, por ejemplo, las cifras que citan en su revisión Matson y Breunig, 1982), en la actualidad carecemos de un método de evaluación y diagnóstico que sea de aceptación y uso generalizado. Por otra parte, los que han tenido una mayor difusión son, todavía hoy, duramente contestados (Bender 1983a, 1983b; Coulter y Morrow, 1977; Haywood, 1977).

En lo referente a la práctica clínico-educativa con sujetos retrasados, la evaluación resulta todavía más caótica. Bender (1983a), establecía que los evaluadores suelen seleccionar sus instrumentos en función de ciertos criterios como (a) la disponibilidad en ese momento de uno u otro procedimiento de evaluación, (b) haber sido o no entrenado en el uso de un procedimiento en particular, (c) la recomendación de otros colegas sobre la eficacia y facilidad de su empleo, (d) el hecho de no conocer otros procedimientos alternativos al utilizado, y (e) el grado de facilidad o dificultad en su interpretación y puntuación, además de la economía de su aplicación. Cuestiones como la validez de contenido, la fiabilidad y la utilidad diagnóstica y pronóstica de cada instrumento, parecen ser sistemáticamente ignoradas por quienes los emplean.

La búsqueda de sistemas alternativos viene siendo una constante en el campo que nos ocupa. Kysela, Anderson y Marfo (1981), señalaron algunos de los factores que, a su juicio, motivaron dicha búsqueda a lo largo de las dos últimas décadas. Para los autores citados, una razón básica fue el claro y desdichado efecto que las categorías tradicionales de diagnóstico tienen sobre la percepción de las personas así diagnosticadas (proceso de estigmatización social). En segundo lugar, la aparición de una amplia literatura experimental sobre el desarrollo infantil y la adquisición del lenguaje motivó un replanteamiento a nivel de evaluación, medida y control. Y, por último, la emergencia de métodos de evaluación por referencia a un criterio, en lugar de a una norma (véase Bijou, 1975) como había sido lo habitual hasta ese momento, supuso un cambio radical en la conceptualización de los instrumentos de evaluación.

Nuestro trabajo se enmarca en los criterios tradicionales de investigación diagnóstica enriquecidos con las aportaciones derivadas de la corriente denominada Evaluación Conductual. Las características específicas de esta corriente aplicada al caso concreto del rd han sido analizadas en otro lugar (Gil Roales-Nieto, 1984), y en síntesis implican (a) la consideración de la conducta retrasada como variable dependiente en función de otras variables independientes que, en numerosos casos, pueden ser modificadas mediante la oportuna intervención, (b) la adopción de un procedimiento orientado preferentemente según un criterio en lugar de una norma, y (c) la consideración de los determinantes psicopatológicos y orgánicos de la conducta retrasada, ya que estos determinan las condiciones límites de la persona retrasada. En ningún caso, sin embargo, estos determinantes deberán ser juzgados como motivos a priori por cuya virtud se renunciara a la planificación - o a su intento al menos - de las condiciones ambientales, familiares y de cualquier otro tipo que puedan llegar a establecerse como tratamiento del rd.

Así, el objetivo prioritario dentro de esta alternativa conductual al diagnóstico tradicional, no será sino la provisión de los datos necesarios que directamente encaminan a la intervención. La evaluación y el diagnóstico deben, por tanto, servir

al tratamiento; de lo contrario no se acierta a entender cuál pueda ser la eficacia funcional del proceso evaluador.

Uno de los principales problemas en dicha aproximación es la relativa carencia de instrumentos precisos y estandarizados que faciliten nuestro objetivo principal: la provisión de los fundamentos para el desarrollo, ejecución y contrastación de programas de intervención conductual en el sujeto con rd (Rojahn y Schroeder, 1983). Ni las numerosas "checklists" existentes, ni los diseños curriculares más complejos, cubren totalmente las necesidades específicas de la práctica clínico-educativa en su aspecto diagnóstico y en lo que se refiere al contexto socio-cultural en el que se genera.

De aquí que el objetivo fundamental del presente estudio haya consistido en diseñar un procedimiento que abarcase satisfactoriamente los aspectos de evaluación, diagnóstico conductual y seguimiento o evaluación continua, de manera que estuviese en consonancia con los requisitos considerados indispensables para el trabajo evaluador en rd (para una exposición y análisis de los mismos véase Gil Roales-Nieto, 1984). En concreto, se pretendía elaborar un procedimiento que (a) abarcara todas y cada una de la áreas o repertorios comportamentales que resultan necesarios para proveer una estimación, lo más completa posible, del nivel comportamental del sujeto con rd, y (b) cumpliera, entre otros, con los siguientes requisitos:

1. Ser un procedimiento fiable.
2. Ser un procedimiento con validez de contenido.
3. Proporcionar objetivos secuenciados de tratamiento, social y clínicamente válidos.
4. Servir como sistema de valoración y medición.
5. Servir como sistema de diagnóstico conductual.
6. Servir como sistema de evaluación continua o seguimiento.

## METODO

### Sujetos

Se eligieron un total de 120 sujetos de ambos sexos y edades comprendidas entre los 6 meses y los seis años, de entre la población infantil de la ciudad de Granada. En los criterios de selección se tuvieron en cuenta la edad, el sexo, el nivel socioeconómico familiar (medido a través de una encuesta sociopsicológica elaborada para tal fin), y la ausencia de enfermedades graves, defectos físicos y alteraciones conductuales que pudieran representar una merma considerable en el desarrollo. Los sujetos se distribuyeron en seis grupos en función de la edad, equilibrando intragrupos el resto de aspectos mencionados.

### Situaciones de evaluación

Las evaluaciones se hicieron en una de las tres formas siguientes: (a) evaluación por observación directa en situación estructurada, (b) entrevistas a padres y profesores, y (c) evaluación colectiva de productos conductuales permanentes. Para el primer y el tercer tipo de evaluaciones, se contó con habitaciones o aulas reservadas a tal fin en cada uno de los centros, durante el tiempo que duró el estudio. Las condiciones exigidas durante la evaluación incluyeron ausencia de ruidos, buena iluminación, ausencia de objetos no implicados en la evaluación, etcétera.

La especial distribución de las personas presentes durante la evaluación por observación directa en situación estructurada, siguió el esquema que aparece en la figura 1. Sólo en el caso de los sujetos de menor edad, dicha distribución se varió permitiendo que alguno de los padres, cuidadores habituales del niño u otra persona familiar para él, estuvieran presentes.

Figura 1. Situación típica de evaluación.

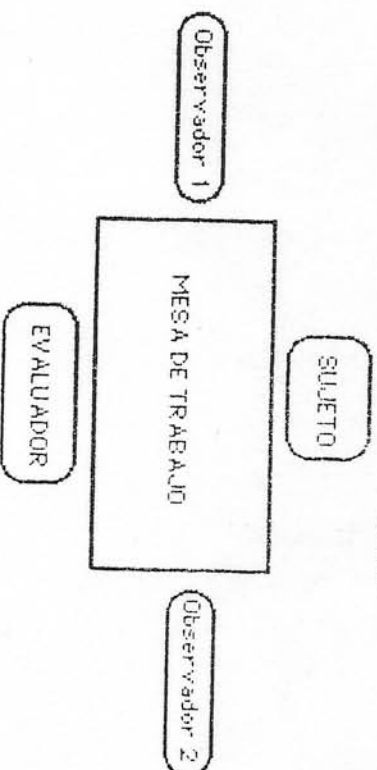


FIGURA 1. Distribución de las personas presentes durante las evaluaciones por observación directa en situación estructurada.

### Procedimiento

La obtención de datos se realizó mediante uno de los tres tipos de evaluaciones mencionadas anteriormente. La evaluación por observación directa en situación estructurada, implicó la presentación de los Ss<sup>d</sup> de control por el evaluador, correspondientes a la respuesta que se pretendía evaluar, y el registro por parte de

Los observadores de la respuesta emitida por el sujeto. Los instrumentos de evaluación y registro empleados estaban normalizados para cada una de las escalas evaluadas.

En este caso, los criterios de respuesta fueron los siguientes:

1. Si el sujeto emita la respuesta correcta en el primer ensayo, se daba por válido el ítem y se pasaba a evaluar el siguiente, excepto en el caso de ítems de conducta verbal comprensiva y/o conceptual, en cuyo caso se exigía un criterio de tres ensayos correctos por cada ítem.
2. En el caso de que el sujeto fallara el primer ensayo, se realizaban un segundo y un tercer ensayo. Tras tres ensayos consecutivos incorrectos se consideraba el ítem como no superado.

El código de registro consistió en tres categorías: correcto (CO), incorrecto (I) y ausencia de respuesta (N). Se estipuló una latencia máxima de respuesta de tres segundos. Cuando se producían fallos en la precisión - desacuerdo entre evaluador y observador - se repetía la evaluación en posteriores sesiones, sin que ello se informase a los observadores.

En cuanto a la interacción sujeto-evaluador, éste procuró hacer siempre la evaluación una actividad altamente reforzante para el niño. La ejecución correcta se reforzó con elogios, fichas, actividades o juegos. Al final de cada sesión las fichas podían cambiarse por diversos reforzadores, dependiendo todo ello de la subescala evaluada y de la edad del niño. Ciertas tareas se mostraron como intrínsecamente reforzantes para la mayor parte de los sujetos (p.e., cortar, pintar, actividades motoras, etc.).

No obstante, una vez que los sujetos y los evaluadores se habituaban a la interacción, desaparecían problemas como negativas a la cooperación, emisión de distractivas menores, ausencia de atención, etc.; y, en general, los niños acudían a las sesiones de evaluación de muy buen grado, resultando para ellos una actividad extraescolar altamente motivante.

La evaluación por entrevista se limitó a la obtención de algunos datos relativos al área de autonomía funcional. No obstante, en estos casos se eligieron algunos ítems de entre los así evaluados y se comprobó la fiabilidad de la información a través del contraste con una evaluación directa de los mismos.

### **Evaluadores, observadores y jueces.**

Los observadores fueron los encargados de registrar los datos. Se emplearon siempre dos observadores que registraban la conducta simultáneamente y en forma independiente. Los observadores fueron elegidos de entre los participantes en un seminario sobre Análisis Conductual Aplicado realizado por el primer autor en la Universidad de Granada. En la información que recibieron se hizo mención

a su participación como observadores y evaluadores en un estudio cuyo objetivo era conocer los patrones de rendimiento o ejecución de los niños sobre una extensa área de comportamientos.

Los observadores fueron entrenados según un método específico elaborado para tal fin (puede verse en Gil Roales-Nieto, 1984). Podemos resumir diciendo que recibieron entrenamiento teórico en el manejo del código de objetividad que iba a ser utilizado, y entrenamiento práctico en observación de sesiones clínicas con sujetos retrasados en su desarrollo, y en sesiones de evaluación similares a las propias de la investigación.

Durante todo el proceso de entrenamiento se midió la precisión y la confiabilidad de los observadores, no dando por finalizada una fase del entrenamiento hasta la obtención de los adecuados índices de acuerdo (superiores a .80).

Además del entrenamiento anterior, los observadores recibieron formación teórica y práctica sobre la forma en que debían conducirse en las sesiones de evaluación.

Además se contó con los servicios de ocho evaluadores quienes actuaron en forma rotatoria con las diversas parejas de observadores. Cuatro de ellos eran psicólogos profesionales con la suficiente experiencia en evaluación y actuación clínica con niños como para no necesitar entrenamiento especial alguno, exceptuando las recomendaciones y guías de actuación propias del presente estudio.

Para los cuatro evaluadores restantes se precisó de un entrenamiento. Recibieron la formación completa de los observadores y también instruidos en administración de contingencias, utilización del material de evaluación y manejo de los manuales de evaluación. Por último, recibieron formación sobre cómo manejar la situación de evaluación, controlar los problemas que surgieran, facilitar la tarea de los observadores, etc.

Los cuatro evaluadores no profesionales actuaron primero como observadores (aproximadamente unos 20 sesiones de evaluación) con el objetivo de familiarizarse con las situaciones.

Los observadores actuaron en ocasiones como jueces en la calificación de productos conductuales permanentes (evaluaciones realizadas en prescripción, predibujo, etc.). El código utilizado fue el mismo que para la observación directa. En todos los casos se proporcionaba a los observadores que actuaban como jueces, los criterios que debían aplicar.

### **Selección de los ítems a evaluar**

El universo de ítems se obtuvo de varias fuentes disponibles, bien de naturaleza teórico-experimental (manuales de psicología del desarrollo, estudios descriptivos de desarrollo...) bien de naturaleza práctica (instrumentos de evaluación ya

existentes, escalas de conducta, "tests" ...). Se consultaron un total de ocho fuentes teóricas principales y numerosas secundarias; fueron treinta y dos los instrumentos de evaluación y diagnóstico consultados (cuestionarios, escalas, etc. Los datos relativos a las fuentes secundarias pueden verse en Roales-Nieto, 1984, págs 65 a 67).

Con la información obtenida de todas las fuentes se elaboró un conjunto o universo de ítems que, en principio, llegó a contar con 2463 ítems de todas las áreas comportamentales.

Otra fuente de información consultada fue la opinión de expertos. Fueron consultados tres expertos con amplia experiencia en diagnóstico y tratamiento de niños retrasados, así como varios expertos universitarios en desarrollo humano. A todos se les pidió que facilitarían una relación de los objetivos de entrenamiento que según su criterio deberían emplearse para niños retrasados cuyo nivel de desarrollo no superase el nivel normal de seis años. Se obtuvieron así un total de 721 ítems formulados como objetivos de entrenamiento y que no estaban comprendidos en el conjunto anterior. El proceso de selección se inició, por tanto, con un universo de 3184 ítems.

A fin de facilitar la tarea de selección, todos los ítems fueron agrupados en cinco grandes áreas o repertorios comportamentales tal y como aparece en el cuadro 1.

El proceso de selección de ítems se realizó con el objetivo de eliminar, reformar o añadir aquellos ítems que fuera necesario. Los criterios utilizados en esta operación fueron los siguientes:

- (a) fueron eliminados aquellos ítems que,
  - se solaparan con otros en cuanto a su contenido.
  - no representaran conductas demostrablemente entrenables o que pudieran sugerir programas de entrenamiento (aún y cuando tuvieran que ser redefinidas), o bien que representarían hipotéticos rasgos, estados o características generales del comportamiento, o sólo fueran descriptivos.
- (b) se reformaron aquellos ítems que,
  - no pudieran convertirse directamente en objetivos de entrenamiento.
  - estuvieran formulados con un enunciado negativo.
  - contuvieran más de un comportamiento a evaluar.
  - puntuaran entre dos y cuatro en el código de objetividad.
- (c) se añadieron aquellas conductas-conductas-objetivo o ítems que, representando conductas consideradas claves (en el sentido de previas y básicas para alguna de las demás respuestas adaptativas), sin embargo, no estuvieran entre los seleccionados en el universo primario de ítems.

Dos investigadores universitarios en psicología infantil y tres psicólogos clínicos con amplia experiencia en evaluación y tratamiento de niños retrasados, sirvieron como jueces de los criterios arriba mencionados. En general, cada ítem no se eliminó, modificó o incluyó a menos que cuatro de los cinco jueces coincidieron en ello. Los juicios se hicieron en forma independiente.

CUADRO 1. Áreas y subáreas del SEDC.

ÁREA DE CONDUCTA MOTORA	
Motora Gruesa Subáreas	Motora Fina Subáreas
Control postural de la cabeza. Control postural de la horizontalidad. Gateo. Control postural de la verticalidad y locomoción.	Presión y Manipulación I. Presión y Manipulación II. Habilidades de cortar. Prescritura y predibujó I. Prescritura y predibujó II. Prescritura y predibujó III.
Sedestación. Coordinación motora de extremidades. Habilidades de fuerza y tracción.	
ÁREA DE REPERTORIOS PREVIOS	
Subáreas	
Desarrollo sensorial. Conducta de atención. Conducta de Imitación.	
ÁREA DE CONDUCTA VERBAL Y CONCEPTUAL	
Subáreas	
Imitación vocal. Identificación de prendas de vestir. Identificación de partes del cuerpo. Identificación de objetos comunes I. Identificación de objetos comunes II. Identificación de objetos comunes III. Identificación de acciones. Situación y orientación espacial. Situación y orientación temporal. Conceptos numéricos. Discriminación de formas. Discriminación de tamaños. Discriminación de grosores. Discriminación de conceptos contrarios. Conceptos múltiples. Discriminación de colores. Pronombres personales. Pronombres posesivos. Repertorio verbal.	
ÁREA DE AUTONOMÍA FUNCIONAL	
Subáreas	
Habilidades de alimentación Habilidades de vestido. Habilidades de aseo. Autoayuda y autoocuidado. Control de esfínteres.	
ÁREA DE CONDUCTA INTERPERSONAL Y SOCIAL	
Subáreas	
Habilidades de interacción social y cooperación. Habilidades ciudadanas y comunitarias. Habilidades domésticas. Manejo de dinero.	

De los 3184 ítems primitivos, quedaron 1719. De ellos 1518 se consideraron materia de evaluación y pasaron a formar parte de las escalas de evaluación exploratoria, siendo agrupados en áreas, subáreas y escalas, según se menciona más adelante.

#### Selección de áreas y subáreas de evaluación.

Se agruparon los ítems en las siguientes áreas o repertorios conductuales:

1. Área de conductas o habilidades motoras, que a su vez se subdividió en Conducta Motora Gruesa y Conducta Motora Fina.
2. Área de conductas o repertorios básicos.
3. Área de conducta verbal y conceptual.
4. Área de habilidades de autonomía funcional.
5. Área de conducta interpersonal y social.

A su vez, estas cinco grandes áreas se subdividieron en subáreas que representaban repertorios más específicos, siguiendo un criterio de independencia clínica.

La división resultante en subáreas aparece en los cuadros 2, 3, 4, 5, y 6 para cada una de las áreas (1).

Cuadro 2. División del área de conducta motora resultante.

ÁREA DE CONDUCTA MOTORA	
SUBÁREA MOTORA GRUESA Control postural de la cabeza Control postural de la horizontalidad Gallo Control postural de la verticalidad y locomoción Sedestación Coordinación motora de extremidades Habilidades de fuerza y tracción	SUBÁREA MOTORA FINA Prensión y manipulación I Prensión y manipulación II Habilidades de cortar Preescritura y dibujo I Preescritura y dibujo II Preescritura y dibujo III

Cuadro 3. División del área de repertorios previos resultante.

ÁREA DE REPERTORIOS PREVIOS
Desarrollo sensorial Conducta de atención Conducta de imitación

(1) Las definiciones de las áreas y subáreas pueden solicitarse al primer autor. No se incluyen aquí por razones de espacio.

Cuadro 4. División del área de conducta verbal y conceptual resultante.

ÁREA DE CONDUCTA VERBAL Y CONCEPTUAL
Imitación vocal Identificación de prendas de vestir Identificación de partes del cuerpo Identificación de objetos I Identificación de objetos II Identificación de objetos III Identificación de acciones Situación y orientación espacial Situación y orientación temporal Conceptos numéricos Discriminación de formas Discriminación de tamaños Discriminación de grososres Discriminación de conceptos contrarios Conceptos múltiples Discriminación de colores Pronombre personales Pronombres posesivos Repertorio verbal - conversación-

Cuadro 5. División del área de Autonomía Funcional resultante

ÁREA DE AUTONOMÍA FUNCIONAL
Habilidades de alimentación Habilidades de vestido Habilidades de aseo Autoguarda y autocuidado Control de esfínteres

Cuadro 6. División del área de Conducta Interpersonal y Social resultante

AREA DE CONDUCTA INTERPERSONAL Y SOCIAL
Habilidades de interacción social y cooperación
Habilidades ciudadanas y comunitarias
Habilidades domésticas
Manejo de dinero

## RESULTADOS

### Confiabilidad y Precisión

Para obtener la confiabilidad de los diferentes aspectos del proceso de evaluación se utilizaron un índice clásico de en evaluación conductual, el porcentaje de acuerdo (Pa), y otro menos utilizado pero que presenta ciertas ventajas que lo hacen recomendable, el índice kappa (K) de Cohen (1960).

Teniendo en cuenta que se evaluaron un total de 1518 ítems en cada sujeto, y que en buena parte de estos ítems se realizaron tres ensayos con cada sujeto, las dimensiones de la matriz de datos resultante que contuviera todos los porcentajes de acuerdo obtenidos, hacen inviable su operatividad. Por ello adoptamos la decisión de ofrecer un resumen de los índices obtenidos.

Así, la confiabilidad de los observadores fue obtenida, por ensayo, es decir, considerando los acuerdos y desacuerdos de los dos observadores a lo largo de todos los ensayos de cada sesión (vale decir de la evaluación de todos los ítems de una subescala) respecto de cada uno de los sujetos evaluados. De entre todas las sesiones, se eligió al azar un porcentaje que incluyera sujetos de todas las edades y evaluaciones en todas las subescalas. En definitiva, se obtuvieron los índices de confiabilidad (Kappa y Pa), sobre 188 mediciones de confiabilidad, y ambos ofrecen resultados más que satisfactorios. Sólo un 3,7% de los Pa resultaron inferiores a .90, y ninguno inferior a .80 (límite mínimo comúnmente aceptado); asimismo, sólo 3 índices kappa (el 1,6% del total) resultaron inferiores a .60, y sus valores críticos en el 85,6% son significativos a un nivel de .05; en el 75% de los casos lo son al .01 y en el 59,5% de los casos lo son al .001.

La confiabilidad interjueces, para aquellas escalas que implicaban el registro de productos permanentes, ofrece también datos de elevada fiabilidad. Todos los índices kappa son mayores de .60 y significativos, a la vez que los Pa resultaron superiores a .80 en todos los casos.

TABLA I. Coeficiente de consistencia interna (K-20) de las escalas.

<b>SUBAREAS</b>	
<b>MOTORA GRUESA</b>	
Gateo .....	.9124
Sedestación .....	.9816
Coordinación motora .....	.9851
Habilidades de fuerza y tracción .....	.9547
<b>MOTORA FINA</b>	
Prensión y manipulación I .....	.9898
Prensión y manipulación II .....	.9888
Prescritura y predibujo I .....	.9740
Prescritura y predibujo II .....	.9947
Prescritura y predibujo III .....	.9903
<b>REPERTORIOS PREVIOS</b>	
Imitación y seguimiento de instrucciones .....	.9789
<b>AUTONOMÍA FUNCIONAL</b>	
Habilidades de vestido .....	.9726
Habilidades de aseo .....	.9700
Habilidades de autoayuda .....	.9770
Habilidades de autocuidado .....	.7937
Habilidades de control de esfínteres .....	.9238
<b>CONDUCTA INTERPERSONAL Y SOCIAL</b>	
Habilidades domésticas .....	.9635
Manejo de dinero .....	.9072
<b>CONDUCTA VERBAL Y CONCEPTUAL</b>	
Imitación vocal .....	.9921
Identificación de objetos I .....	.9947
Identificación de objetos II .....	.9931
Identificación de objetos III .....	.9857
Identificación de prendas de vestir .....	.9890
Identificación de partes del cuerpo .....	.9875
Identificación de acciones .....	.9948
Situación y orientación espacial .....	.9741
Situación y orientación temporal .....	.9566
Conceptos numéricos .....	.9915
Discriminación de formas .....	.9274
Discriminación de tamaños .....	.9578
Discriminación de grososres .....	.9580
Discriminación de conceptos contrarios .....	.9757
Conceptos múltiples .....	.9868
Discriminación de colores .....	.9920
Pronombres personales .....	.9089
Pronombres y adjetivos posesivos .....	.9891

Por último, los datos obtenidos en la comprobación de la precisión de los observadores, garantizan que los observadores están realmente registrando las conductas que se pretendía medir y no otras. Todos los índices de precisión calculados para cada observador superan en el caso del Palos valores de .90 (rango de .90 a 1). En el caso de los índices kappa el rango de .77 a 1, por lo que todos los índices resultaron muy superiores a los valores usualmente aceptados como fiables.

### Fiabilidad de las escalas

Para obtener la fiabilidad de las escalas hemos optado (por diversas razones entre las que cabe destacar la relativa ausencia de índices más adecuados a las características del problema) por el método de hallar su consistencia interna mediante la clásica fórmula 20 de Kuder-Richardson, 1937; Brown, 1970).

En cuanto a la fiabilidad de las escalas evaluadas, 34 de ellas (de un total de 35) han obtenido índices de consistencia interna superiores a .90 (rango .90 a .99), y sólo una, la subárea de Autocuidado, obtuvo un coeficiente menor, en concreto de .70. En la tabla 1 se exponen los índices obtenidos para cada subescala.

### Obtención de las subescalas: grupos de edad y secuencias de ítems

En primer lugar se obtuvieron curvas de ejecución que representan el porcentaje de ítems de cada subárea ejecutada correctamente por cada uno de los sujetos. En segundo lugar se realizó un análisis de cluster (BMDF-2M) para cada una de las escalas (subáreas), lo que proporcionó la orientación necesaria sobre la formación de los diferentes grupos de edad. En tercer lugar, se estructuraron tablas de porcentajes de sujetos que ejecutaron correctamente cada ítem, en base a los grupos de edad obtenidos en el punto anterior. Con esto se obtuvieron las secuencias de ítems para cada escala.

De esta forma, con los grupos de edad, los ítems correspondientes a los mismos y la secuencia de ítems, se formaron cada una de las diferentes subescalas que dieron lugar a las escalas definitivas del Sistema.

Los resultados para todas las áreas representarían un volumen de datos que superaría con creces las disponibilidades de espacio del presente artículo. Expondremos, por tanto, únicamente los resultados correspondientes a una de las subáreas (escalas), en concreto la correspondiente a la evaluación de "Conceptos Múltiples", pertenecientes al área de Conducta Verbal y Conceptual. Dicha escala servirá como ejemplo ilustrativo del proceso seguido en todas las demás.

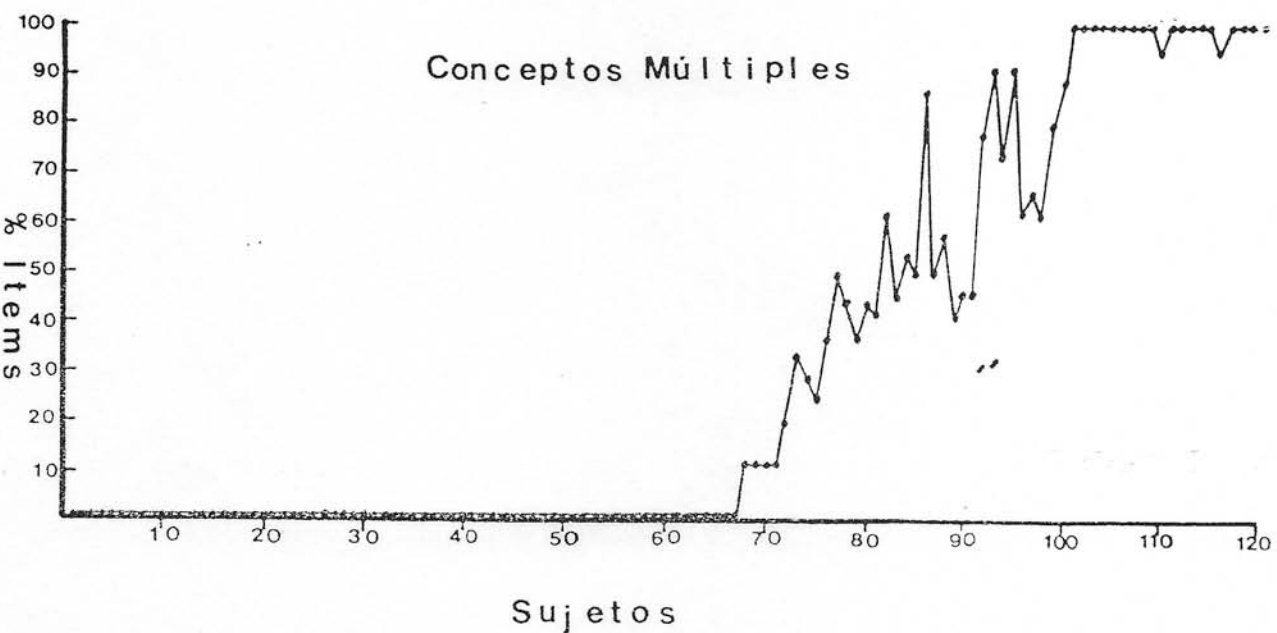
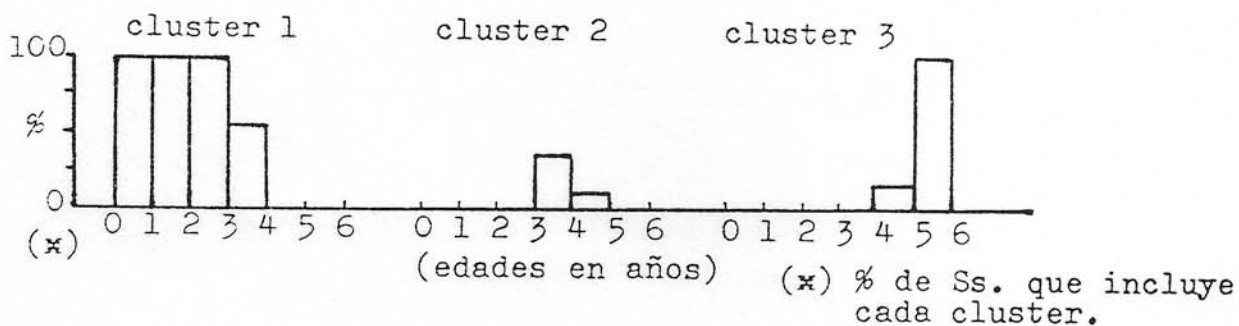




Tabla de Clusters

Cluster nº	nº de sujetos	sujetos
1	71	1 al 71
2	9	76 al 80, 89 y 90
3	23	86, 93, 95, 101 al 120

Histogramas de distribución de los sujetos



La figura 2 muestra la curva general de ejecución para todos los sujetos en todos los ítems de la subárea de Conceptos Múltiples. El parámetro utilizado es el porcentaje de ítems que verifica cada sujeto. La curva, por tanto, muestra la progresión de la ejecución a través de la edad en dicha subárea, y nos aporta una primera orientación general sobre la edad aproximada de aparición de cada conducta-objetivo, a la vez que la edad aproximada en que ocurren los efectos "suelo" y "techo".

Sobre estos datos se realizó el análisis de clusters referido anteriormente. Se eligió el punto medio de rango de amalgamiento, como criterio principal para la decisión sobre los clusters estimados como núcleos de agrupamiento de los sujetos. Se utilizó, además, un segundo criterio de agrupamiento, que consistió en la constatación de los patrones de agrupamiento que se producían hasta formas clusters independientes; esto es, agrupaciones de sujetos que se unen entre sí y sólo se unen a otros clusters en bloque, y no de forma individual.

Los clusters, y el número de sujetos que agrupa cada uno de ellos, correspondientes al ejemplo que tratamos aparecen en la Figura 3, en la que también se recogen los histogramas de distribución de los sujetos en cada cluster.

Sobre estos datos, los grupos de edad que se formaron en esta subárea fueron los siguientes:

1. Hasta 3 años y 6 meses. Representado por el cluster 1 y que recoge de los sujetos 1 al 71 (a.i.).
2. De 3 años y 6 meses a 4 años y 6 meses. Representado por el cluster 2, y que recoge de los sujetos 72 al 94 (a.i.).
3. De 4 años y 6 meses en adelante. Representado por el cluster 3 y que recoge de los sujetos 95 al 120 (a.i.).

La secuencia de ítems para cada uno de los grupos de edad anteriores se configuró de la forma siguiente: un grupo de edad no recoge ningún ítem si ninguno de los que forman la subárea resultó superado por, al menos, el 55% de los sujetos (criterio utilizado para la inclusión de un ítem en un determinado grupo de edad). Cada grupo de edad y su correspondiente secuencia de ítems conformaron lo que denominamos como Subescala; formándose como Escala con todas las subescalas resultantes en cada subárea de evaluación.

Los ítems que no fueron superados por el 55% de los sujetos en alguno de los grupos de edad se eliminaron por entender que pertenecen a ejecuciones correspondientes a edades superiores al rango de edad elegido en el presente estudio.

En la Tabla 2 se recogen las subescalas resultantes, en su forma definitiva, en relación con la subárea de Conceptos Múltiples que estamos tomando como ejemplo.

TABLA 2. Escala de la subárea de Conceptos Múltiples.

En esta subárea se evalúa la combinación entre color más tamaño y entre color más tamaño más forma.

Toda la evaluación se hace en conducta verbal comprensiva. Por último, la latencia de respuesta en todos los ítems es como máximo de 3 segundos.

## CRITERIO DE RESPUESTA

En todos los casos es comprensiva, es decir, una respuesta motora apropiada a la estimulación presentada.

## RELACIONES DE CONTROL

(A) Color más tamaño

- estimulación visual: en todos los ensayos incluirá la combinación evaluada más otras 4 diferentes que no deben tenerla misma forma ni el mismo color entresí.

- estimulación auditiva: "señala el... (nombre de la combinación evaluada)".

(B) Color más tamaño más forma

- estimulación visual: igual que en (A). Los cuatro estímulos control no deben ser iguales en su forma, color ni tamaño.

- estimulación auditiva: "señala el... (combinación evaluada)".

Subescala 1. Rango de edad: 3 años y 6 meses a 4 años y 6 meses

ITEM Nº	COMBINACIONES EVALUADAS (empleando "bloques lógicos")	RELACION DE CONTROL
VC/CM 1	rojo pequeño	A
VC/CM 2	rojo grande	A
VC/CM 3	azul pequeño	A
VC/CM 4	azul grande	A
VC/CM 5	amarillo grande	A
VC/CM 6	amarillo pequeño	A
VC/CM 7	cuadrado pequeño amarillo	B
VC/CM 8	cuadrado grande rojo	B
VC/CM 9	cuadrado pequeño azul	B
VC/CM 10	cuadrado grande azul	B
VC/CM 11	cuadrado grande rojo	B
VC/CM 12	cuadrado pequeño rojo	B
VC/CM 13	triángulo grande amarillo	B
VC/CM 14	triángulo grande azul	B
VC/CM 15	triángulo pequeño amarillo	B
VC/CM 16	círculo grande amarillo	B
VC/CM 17	triángulo pequeño azul	B
VC/CM 18	círculo pequeño amarillo	B
VC/CM 19	cuadrado grande amarillo	B
VC/CM 20	triángulo pequeño rojo	B
VC/CM 21	círculo grande azul	B
VC/CM 22	círculo pequeño rojo	B
VC/CM 23	triángulo grande rojo	B
VC/CM 24	círculo pequeño azul	B

## DISCUSIÓN Y CONCLUSIONES

Se estableció como objetivo del presente estudio, la elaboración de un Sistema de Evaluación y Diagnóstico Conductual para el retraso en el desarrollo. Dicho Sistema tendría que cumplir una serie de requisitos, tales como ser fiable y estar elaborado confiablemente, ser válido y útil para la prescripción y programación de objetivos de intervención, y servir como sistema de evaluación continua. Además, debería cumplir también la condición más general de evaluación continua. Además, repertorios conductuales, necesaria para proporcionar a abarcar todas las áreas o completa posible del repertorio del sujeto evaluado.

Respecto al objetivo principal, los datos indican que efectivamente se ha conseguido elaborar un sistema conductual de evaluación y diagnóstico para el retraso en el desarrollo que cumple los requisitos estipulados. Nuestra afirmación se fundamenta en el análisis de todas y cada una de las sucesivas clapas que se han cubierto, desde la inicial de obtención del universo de ítems o conductas-objetivo, hasta el final de realización de las escalas definitivas. El Sistema, en efecto, abarca todas y cada una de las áreas conductuales y de información general sobre el sujeto a evaluar que son necesarias para proporcionar una estimación completa de su repertorio y de las variables con el relacionadas. Entre la información aportada por los instrumentos auxiliares del Sistema (generales, de contingencias y de conductas-problema), y la conseguida mediante la utilización de las diferentes subescalas, se hace posible obtener una estimación global y específica de los problemas que cada sujeto evaluado presenta.

A la vez, los datos obtenidos sobre fiabilidad nos permiten afirmar que los instrumentos del Sistema son fiables, los observadores trabajaron precisa y confiablemente y que en igual forma lo hicieron cuando actuaron como jueces. En igual forma, podemos concluir que nuestro Sistema cumple con el requisito de poseer validez de contenido. Las conductas-objetivo o ítems que lo forman han superado hasta cuatro diferentes filtros, analíticos unos basados en juicios expertos otros.

Otro de los requisitos del Sistema era la facilitación de objetivos de tratamiento secuenciados y social y clínicamente válidos. La conclusión en este punto vendría proporcionada por el análisis global de las escalas resultantes. En efecto, las escalas proporcionan objetivos de intervención ya que los ítems que las componen están garantizados, a través de un proceso de validez de contenido, que representan conductas-objetivo que un sujeto debe poseer a la edad determinada, para su funcionamiento diario en la comunidad donde reside; han sido elaborados en forma objetiva, especificando claramente las relaciones de control (esto es, qué debe hacer el evaluador y en qué situaciones), los criterios de respuesta (esto es, qué debe hacer el evaluado), y los aspectos paramétricos y de ejecución que deben acontecer para que un sujeto sea diagnosticado como necesitado de intervención en una conducta-objetivo determinada.

Los objetivos de tratamiento han sido secuenciados en función de los resultados obtenidos en el análisis de los sujetos experimentales, de forma que los ítems están ordenados de menor a mayor dificultad de ejecución, y cada uno de ellos representa una conducta-objetivo previa a la que precede y necesitada de la que le antecede (entre respuesta de la misma clase o categoría). El educador o terapeuta puede, por tanto, conocer la secuencia de ítems o conductas-objetivo que debe entrenar a partir de la propia evaluación, por lo que el Sistema cumple uno de los aspectos fundamentales de la función diagnóstica.

Por último, los objetivos de tratamientos secuenciados resultan social y clínicamente válidos por dos razones fundamentales. Una la ya mencionada validez de contenido que presentan, y otra porque son objetivos que presentan los niveles de ejecución socialmente reconocidos como normales.

Mediante la aplicación del Sistema podemos, por tanto, medir los repertorios del sujeto y valorar sus déficits y excesos a la vez que el nivel de su conducta adaptativa, obteniendo sus perfiles de desarrollo. El Sistema funciona así como una extensa y detallada línea de base que permite conocer el punto exacto de partida contra el cual hemos de contrastar futuras medidas.

El Sistema puede cumplir el requisito de servir como un instrumento de diagnóstico conductual. Y lo hace en la medida que permite que la valoración y medidas anteriores no signifiquen una mera descripción de habilidades del sujeto sino algo que directamente posibilita:

- prescribir el tratamiento o intervención necesario, esto es (a) programar qué conductas-objetivo ha de aprender el sujeto, qué conductas-objetivo habrá que eliminar y cuáles necesitan ser modificadas en uno u otro sentido, (b) programar la secuencia en que esto va a ser llevado a cabo;

- determinar cómo y cuánto el sujeto difiere en su comportamiento de los patrones de ejecución considerados socialmente normales, lo que puede conseguirse mediante la obtención del Perfil Comportamental del sujeto.

Otras funciones del diagnóstico conductual quedan garantizadas también. Por ejemplo, la clasificación del sujeto en un grupo afín (por necesidades administrativas o de planificación institucional) es posible mediante el Perfil Comportamental resultante de la evaluación. Por otra parte, el propio perfil, el resultado de la evaluación de conductas-problema, y los datos de la evaluación de contingencias y psicología, combinados entre sí permiten la emisión de un pronóstico a corto, medio y largo plazo sobre el curso del retraso y el posible impacto de la intervención.

El Sistema cumple también la condición de servir como instrumento de evaluación continua o seguimiento, ya que los ítems o conducta-objetivo en muchos casos incluyen un mini-análisis de tareas y en otros casos representan eslabones de una cadena más compleja, lo que permite captar sensiblemente los avances del sujeto aunque estos sean pequeños. Por otra parte, las mismas características que les

otorgaban su utilidad como instrumento de valoración y medida son las que permiten ser un eficaz instrumento de evaluación continua.

Bender (1983a) afirmaba que "se ha reconocido sólo recientemente por los evaluadores y aquellos encargados de la responsabilidad de evaluar a los niños discapacitados, la necesidad de instrumentos especializados, entrenamiento especializado para su administración y del conocimiento técnico de todos los aspectos conductuales y componentes que comprenden un apropiado y equitativo sistema de evaluación" (pág. 22). El Sistema elaborado es especializado, como tal requiere un entrenamiento especializado para su administración y exige el conocimiento de los aspectos conductuales y de los componentes implicados en todo proceso de evaluación.

En resumen, las aplicaciones y utilidad que el Sistema aporta a la práctica clínico-educativa del retraso en el desarrollo, implican (1) la valoración o medición de los repertorios del sujeto, (2) el diagnóstico conductual del mismo, (3) la facilitación de los objetivos a programar y secuencia de los mismos, (4) la evaluación de la ejecución del profesional o terapeuta, (5) la evaluación de la efectividad de la intervención, y (6) la posibilidad de servir como instrumento de prevención.

Algunas cuestiones, la mayoría de las cuales se han derivado de la propia realización de este estudio, quedan todavía por estudiar. Por ejemplo, queda por establecer la validez de tratamiento del Sistema. Por otra parte, sería muy conveniente la elaboración de un mini-sistema, más específico, que permitiera su aplicación en programas de intervención precoz, y un sistema simplificado para padres, cuya cooperación como co-terapeutas es imprescindible.

## BIBLIOGRAFÍA

- BENDER, M. (1983a): Psychological Assessment of Young Children, 0-3 years. Ponencia presentada en el *Curso Internacional para la Prevención de la Deficiencia Mental y Estimulación Precoz* (Santiago de Compostela).
- BENDER, M. (1983b): Psychological Tests and Methods for Handicapped Children. Ponencia presentada en el *Curso Internacional para la Prevención de la Deficiencia Mental y Estimulación Precoz* (Santiseigo de Compostela).
- BIGELOW, G. (1977): The Behavioral Approach to Retardation. en T. Thompson y J. Grabowski (Eds.) *Behavior Modification of the Mentally Retarded*, 2ª ed., pp. 17-45 (Nueva York: Oxford Press).

- BIJOU, S.W. y GRIMM, J.A. (1975): Behavioral Diagnosis and Assessment in teaching young Handicapped Children. En T.Thompson Y J. Dockens III, *Proceedings of the National Symposium on Behavior Modification*, pp. 161-180. New York. Academic Press.
- BIJOU, S.W. y PETERSON, R.F. (1971): The Psychological Assessment of Children: A Functional Analysis. En P. McReynolds (Ed.), *Advances in Psychological Assessment, vol II*. Palo Alto, CA.: Science and Behavior Books.
- BROWN, F.G. (1970): *Principles of Educational and Psychological Testing*. Hindsdale, Ill.: Dryden Press.
- COHEN, J.A. (1960): A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37-46.
- CONE, J.D. y HAWKINS, R.P. (1977): *Behavioral Assessment: New Directions in Clinical Psychology*. New York: Bruner and Mazel.
- COULTER, W.A. y MORROW, H.W. (1978) (eds.), *Adaptive Behavior*. Nueva York: Grune and Stratton.
- DOLL, E.A. (1936): *The Vineland Social Maturity Scale: Revised Condensed Manual of Directions*. Vineland. Nueva Jersey: The Training School.
- DOLL, E.A. (1965): *Social Maturity Scale*. Circle Pines. Minn.: American Guidance Service, Revisión de 1965.
- GIL ROALES-NIETO, J. (1983): Retardo en el Desarrollo (Deficiencia Mental): Necesidad de un Modelo alternativo de Diagnóstico y Evaluación. *Avances en Psicología Clínica Latinoamericana*, 2, 69-78.
- GIL ROALES-NIETO, J. (1984): *Una aproximación comportamental al Diagnóstico del Retardo en el Desarrollo* (Tesis Doctoral., Facultad de Psicología, Universidad Complutense de Madrid).
- GUNZBURG, H.C. (1969): *Manual para el P.A.C.*. Fundación Centro de Enseñanza Especial, Madrid.
- HAYNES, S.N. (1978): *Principles of Behavioral Assessment*. Nueva York: Gardner Press.
- HAYWOOD, H.C. (1977): Alternatives to Normative Assessment. En P. Miltler (Eds.) *Research to Practice in Mental Retardation, V. II*. pp. 11-18. Baltimore.
- HAYWOOD, H.C.; FILLER, J.W. Jr; SHIFMAN, M.A. y CHATELANAT, G. (1975): Behavioral Assessment in Mental Retardation. En P. McReynolds (Eds.) *Advances in Psychological Assessment, vol. 3*. San Francisco, CA.: Jossey Bass.
- HERSEN, M. y Bellack, a.s. (1981): *Behavioral Assessment: A Practical Handbook*. 2ª ed. Nueva York: Pergamon Press.
- KAZDIN, A.E. y STRAW, H.K. (1976): Assessment of Behavior of the Mentally Retarded. En H. Hersen y A.S. Bellack. *Behavioral Assessment: A Practical Handbook*. Nueva York: Pergamon Press.
- KUDER, G.F. y RICHARDSON, M.W. (1937): The Theory of Estimation of

- Tests Reliability. *Psychometrika*, 2, 151-160.
- KYSELA, G.M.; ANDERSON, D. y MARFO, K. (1981): Issues in Mental Retardation in Canada: Transitions. *Applied Research in Mental Retardation*, 2, 145-164.
- MATSON, J.L. y BREUNING, S.E. (1982): A Review and Analysis of Applied Research in Mental Retardation: 1975-1980. *Applied Research in Mental Retardation*, 3 (2), 185-189.
- MERCER, J.R. y LEWIS, J.F. (1978): *System of multicultural/pluralistic Assessment*. Nueva York: Psychological Corporation.
- NELSON, R.O. y HAYES, S.C. (1979): Some Current Dimensions of Behavioral Assessment. *Behavioral Assessment*, 1, 1-16.
- ROJAHN, J. y SCHROEDER, S.R. (1983): Behavioral Assessment. En J.L. Matson y J.A. Mulick. *Handbook of Mental Retardation*, pp. 227-243. Nueva York: Pergamon Press.