

## FIABILIDAD TEST-RETEST DE LA ADAPTACIÓN ESPAÑOLA DE LA DIAGNOSTIC INTERVIEW FOR CHILDREN AND ADOLESCENTS (DICA-R)

Lourdes Ezpeleta, Nuria de la Osa, José M<sup>a</sup> Domenech, José Blas Navarro  
y Josep M<sup>a</sup> Losilla

Universitat Autònoma de Barcelona

El artículo examina la fiabilidad test-retest de la adaptación española de la Diagnostic Interview for Children and adolescents-Revised (DICA-R) en una muestra de 110 niños y adolescentes pacientes psiquiátricos externos. Niños y padres fueron entrevistados con la DICA-R en dos ocasiones separadas por un intervalo medio de 11 días. Los resultados del estudio indican que hay estabilidad en la evaluación de los trastornos con la DICA-R, como lo demuestra la ausencia de diferencias significativas entre el test y el retest y el rango de la mayoría de los valores kappa que oscilaron entre buenos y excelentes. Los más estables informando sobre los trastornos fueron los padres y los menos fiables fueron los adolescentes. Los trastornos mostraron cierta atenuación en el retest, aunque las diferencias no fueron significativas. La fiabilidad de los respondientes individuales dependió del tipo de trastorno y de la edad. La entrevista fue más fiable en las categorías más generales que en los trastornos específicos.

*Test-retest reliability of the Spanish adaptation of the Diagnostic Interview of Children and Adolescents.* The paper examines the test-retest reliability of the Diagnostic Interview for Children and Adolescents-Revised (DICA-R). 110 children and adolescent outpatients and their parents were interviewed using the DICA-R on two occasions, separated by a mean interval of 11 days. The results of our study indicate that there is stability in the assessment of the disorders with DICA-R, as shown by the absence of significant differences between test and retest and the majority of kappa values which ranged from good to excellent. The parents were the most stable in reporting disorders and adolescents were the least. Test-retest reliability of DICA-R in the sample was very good. The disorders appeared attenuated in the retest. The reliability of respondent assessment depended on the type of disorder and on age. The interview was more reliable in higher rank of major categories than in specific disorders.

---

Correspondencia: Lourdes Ezpeleta  
Departament de Psicologia de la Salut i de Psicologia  
Social  
Edifici B. Facultat de Psicologia  
08193 Bellaterra (Barcelona)  
E-mail: lourdes.ezpeleta@uab.es

La falta de fiabilidad del diagnóstico psiquiátrico ha impulsado en los últimos años a los investigadores a intentar cambiar este patrón a través de la definición de criterios diagnósticos operacionales, del uso de entrevistas diagnósticas estructuradas basadas

en estos criterios, y derivando los diagnósticos a través de la aplicación de algoritmos computerizados. Apter, Orvaschel, Lases, Moses y Tyano (1989) señalan la necesidad de obtener diagnósticos válidos y fiables no sólo para la investigación, sino también en la práctica clínica.

Cuando un instrumento de medida está en proceso de adaptación o desarrollo, la fiabilidad es un punto importante de estudio porque es necesario un cierto grado de replicabilidad para poder evaluar la validez. La fiabilidad test-retest es una de las formas más rigurosas de medir la fiabilidad porque evalúa la estabilidad de la medida en el tiempo a pesar del cambio en la sintomatología (Robins, 1985). Es sabido que entre la primera y la segunda evaluación no se obtiene la misma información. Es frecuente

observar que las puntuaciones se atenúan en la segunda medida (Andreasen et al., 1981; Lucas, 1992; Lucas, Shaffer, Fisher, Parides y Breton, 1994). Robins (1985) ha señalado entre las razones para estas discrepancias el cambio clínico en el respondiente, la explicación errónea o mala comprensión del propósito del diseño por parte del respondiente, el efecto terapéutico de la primera entrevista o los efectos de la propia entrevista (longitud, fatiga, aburrimiento). Además, el intervalo entre el test y el retest y el grado de inferencia o complejidad de las preguntas también afecta la concordancia entre la información obtenida en los dos momentos (Schwab-Stone et al., 1993).

La Tabla 1 presenta los resultados obtenidos en los estudios de fiabilidad test-retest con las entrevistas diagnósticas más utiliza-

*Tabla 1*  
Fiabilidad Test-Retest de los protocolos diagnósticos estructurados para niños y adolescentes

TRASTORNO	Est. 1	Estudio 2 (Correlación intraclass)						Est. 3 kappa	Est. 4 kappa	Estudio 5		Est. 6 kappa	Estudio 7				Est. 8 kappa	Est. 9 kappa	Est. 10 kappa		
		Niños			Padres					kappa	kappa		N	P	kappa						
		6-9	10-13	14-18	6-9	10-13	14-18								N	A				PN	PA
Déficit aten./hiperact.	.21	.61	.66	.73	.87	.82	.77	.45	.48	.55	.55	1	.43	.24	.78	.86	.43		.47		
T. de Conducta	.47	.58	.73	.77	.88	.84	.86		.55	.87	.55	1	.37	.92	-.01	.87	.71	.63	.56		
Negativismo desaf.		.63	.71	.78	.86	.75	.78	.38		.88	.16	.79	.33	.28	.32	.92			.67		
Depresión Mayor	.36	.30	.53	.81	.83	.79	.75	.50	.63	.72	.77		.00	.45	.77	.31	1	.54	.50		
T. Distímico								.55					.19	.40	.57	.51	.71		.59		
T. del Humor												.90							.63		
Ansiedad Separación	.38	.53	.56	.81	.75	.73	.68	.51	.72		.72		.01	.00	.38	-	.56		.59		
Ansiedad Excesiva	.34	.38	.51	.69	.77	.79	.74	.35					.01	.54	.44	.69	.38		.43		
Fobia		.67	.78	.82	.81	.81	.86												.44		
Obsesivo-Compulsivo			.33	.59	.58	.60	.62														
T. de Ansiedad												.76					.72	.24	.47		
Enuresis								.75													
Psicosis			.45	.72	.69	.49	.75														
Externalizante													.32	.38	.37	.68					
Internalizante													.06	.47	.44	.51					

N, entrevista con niño; P, entrevista con padres; PN, padres de niños; PA, padres de adolescentes.  
*Estudio 1:* Costello et al. (1984). Diagnostic Interview Schedule for Children (DISC). 242 pacientes psiquiátricos de 6 a 18 años. Intervalo de 9 días.  
*Estudio 2:* Edelbrock et al. (1985). DISC. 242 pacientes psiquiátricos internos y externos de 6 a 18 años. Intervalo de 9 días.  
*Estudio 3:* Canino et al. (1987). DISC-C. 91 pacientes clínicos y sujetos de la comunidad de 4 a 16 años. Intervalo medio de 19 días.  
*Estudio 4:* Shaffer et al. (1988). DISC-R. 41 paciente psiquiátrico de 11 a 17 años. Intervalo de 1 a 3 semanas.  
*Estudio 5:* Schwab-Stone et al. (1993). DISC-C. 39 padres y 41 niños pacientes psiquiátricos externos de 11 a 17 años. Intervalo de 1 a 3 semanas.  
*Estudio 6:* Welner et al. (1987). Diagnostic Interview for Children and Adolescents (DICA-C). 27 pacientes internos de 7 a 17 años. Intervalo de 1 a 7 días.  
*Estudio 7:* Boyle et al. (1993). DICA-R. 251 sujetos de la población general de 6 a 16 años. Intervalo de 3 semanas.  
*Estudio 8:* Hodges et al. (1989). Child Assessment Schedule (CAS-C). 32 pacientes psiquiátricos internos de 6 a 12 años. Intervalo de 5 días.  
*Estudio 9:* Chambers et al. (1985). Schedule for Affective Disorders and Schizophrenia (K-SADS). 52 pacientes psiquiátricos externos de 6 a 17 años. Intervalo de 72 horas.  
*Estudio 10:* Roberts et al. (1996). DISC-2.1C. Pacientes psiquiátricos internos y externos de 12 a 17 años. Intervalo de 7 a 14 días.

das en niños. Es difícil comparar los datos porque los informantes, el tipo de muestra y la forma en que se ofrece la información difiere de un estudio a otro. Según estos datos, los mejores resultados son los obtenidos por Shaffer et al. (1988) con la Diagnostic Interview Schedule for Children (DISC), Welner, Reich, Herjanic, Jung y Amado (1987) con la Diagnostic Interview for Children and Adolescents (DICA) y Hodges, Cools y McKnew (1989) con la Child Assessment Schedule (CAS). En estos tres estudios los valores kappa fueron aceptables en todas las categorías estudiadas. Si consideramos las categorías diagnósticas, los peores resultados se logran en negativismo desafiante (ND), depresión mayor (DM), trastorno por ansiedad de separación (TAS) y trastorno por ansiedad excesiva (TAE). Por lo que respecta a los informantes, parece que cuanto más joven es el informante, la fiabilidad es más baja, pero la heterogeneidad de las muestras hace muy difícil sacar conclusiones claras sobre la influencia de esta variable sobre la fiabilidad. El intervalo medio aproximado que ha separado el test del retest en estos estudios es de 8 días.

Nuestro trabajo describe la fiabilidad test-retest de la adaptación española de la DICA-Revised (DICA-R, Reich, Shayka y Taibleson, 1991) en una muestra de niños y adolescentes pacientes psiquiátricos externos.

### Método

#### *Sujetos*

La muestra está compuesta por 110 niños y adolescentes pacientes psiquiátricos externos y sus padres. Los niños (n=59; 54%) tienen edades comprendidas entre 7 y 12 años (35 niños -59%- y 24 niñas -41%) y los adolescentes (n=51; 46%) entre 13 y 17 años de edad (15 chicos -29%- y 36 chicas -71%). Los niños acudían a consulta a cuatro centros de Asistencia Psiquiátrica Infanto-Juven-

il de la red pública. El 98.2% de las familias eran de raza caucásica y el 1.8% restante pertenecían a otras etnias. El nivel socioeconómico según el índice de posición social de Hollingshead (1975) para los niños se distribuyó de la forma siguiente: clase I: 1.9%; clase II: 1.9%; clase III: 11.5%; clase IV: 38.5; clase V: 46.2%. Para los adolescentes la distribución fue: clase I: 2%; clase II: 14%; clase III: 16%; clase IV: 30%; clase V: 38%.

#### *Medidas*

La DICA-R (Reich et al., 1991) fue la primera entrevista diagnóstica estructurada para niños y adolescentes basada en los criterios diagnósticos del DSM-III-R (APA, 1987). Consta de tres versiones que son muy similares en estructura y contenido: una para niños de 6 a 12 años, otra para adolescentes de 13 a 17 años, y una versión para padres de niños de 6 a 17 años que permite recoger información sobre todos los hijos de la familia con edades comprendidas en ese intervalo. El protocolo cubre los trastornos más frecuentes en la infancia y la adolescencia. Las versiones anteriores de la DICA presentaron una alta fiabilidad entre e intra-entrevistadores (Herjanic y Reich, 1982; Welner et al., 1987) y moderado acuerdo entre padres y niños/adolescentes (Herjanic, Herjanic, Brown y Wheatt, 1975; Herjanic y Reich, 1982; Kashani, Orvaschel, Burk y Reid, 1985; Reich, Herjanic, Welner y Gandhi, 1982; Sylvester, Hyde y Reichler, 1987; Welner et al., 1987). La DICA discrimina entre muestras pediátricas y psiquiátricas y correlaciona moderadamente con otras medidas de psicopatología del niño (por ejemplo con el diagnóstico clínico) (Carlson et al., 1987; Herjanic y Campbell, 1977; Sylvester et al., 1987; Welner et al., 1987). Con la DICA-R Boyle et al. (1993) obtuvieron una buena fiabilidad test-retest para los trastornos externalizantes y pobre

para los internalizantes en una muestra de la población general.

En la versión española se ha obtenido un acuerdo entre entrevistadores de bueno a excelente (valores kappa entre .65 y 1) (de la Osa, Ezpeleta, Doménech, Navarro y Losilla, 1996). El acuerdo entre la información del niño/adolescente y sus padres fue bajo. Los niños y adolescentes informaron menos síntomas que sus padres. El grado de concordancia mejoró con la edad del niño (Ezpeleta, de la Osa, Doménech, Navarro y Losilla, 1995). Los diagnósticos derivados de la entrevista concuerdan a un nivel de bajo a moderado con los derivados por el clínico. El mejor acuerdo se obtuvo entre los diagnósticos derivados de la información de los padres y de los niños (Ezpeleta, de la Osa, Doménech, Navarro y Losilla, 1997). Los síndromes derivados de la DICA-R correlacionan significativamente con las dimensiones del Child Behavior Checklist (Achenbach, 1991) teóricamente más relacionadas con ellas. La información de la DICA-R permite discriminar entre grupos de niños normales y clínicos (de la Osa, Ezpeleta, Doménech, Navarro y Losilla, 1996).

#### *Procedimiento*

La traducción de la entrevista del inglés al castellano se hizo en tres pasos: 1) traducción del inglés al español por un psicólogo; 2) revisión de la traducción por un psiquiatra infantil; 3) comparación del contenido de la traducción de cada ítem con la autora de la entrevista para comprobar la equivalencia de los ítems (traducción del castellano al inglés). Las preguntas sobre raza, robos con tarjeta de crédito y la lista de bebidas alcohólicas se adaptaron para los jóvenes españoles.

Existe consenso general sobre la necesidad del entrenamiento en el uso de los protocolos de entrevistas incluso cuando los

entrevistadores tienen experiencia y la entrevista es estructurada. Por ello, en este estudio se ha puesto especial énfasis en el entrenamiento de los entrevistadores (véase de la Osa et al., 1996). Se consideró que los entrevistadores habían terminado satisfactoriamente el entrenamiento cuando la fiabilidad entre-observadores promedio en cinco entrevistas fue igual o superior al valor kappa .80. Durante el estudio, se controló aleatoriamente el acuerdo entre entrevistadores. En cada caso evaluado, estuvieron implicados cuatro entrevistadores, ya que tanto los padres como los niños fueron evaluados simultáneamente por entrevistadores independientes en el test y en el retest. Todos los entrevistadores realizaban la entrevista "a ciegas" ya que no disponían de ninguna información sobre el niño. El intervalo medio entre el test y el retest fue de 11 días. El test se hizo tan cerca como fue posible de la primera visita.

Los diagnósticos se generaron utilizando los algoritmos previamente definidos en la entrevista, excepto para fobia en que se incluyeron las preguntas sobre alteración del funcionamiento. Los diagnósticos se derivaron por separado para los padres y para los niños/adolescentes. La concordancia diagnóstica se calculó con el estadístico kappa (Cohen, 1960), considerando un valor kappa de 0.40 o mayor como fiabilidad buena, y por encima de 0.75 como fiabilidad excelente (Fleiss, 1981). Se realizó la prueba de McNemar (exacta) para estudiar los cambios significativos en el número de sujetos que presentaban un diagnóstico del test al retest. Se utilizó la prueba t para datos apareados para estudiar los cambios en el nivel medio de los síntomas informados.

#### **Resultados**

La duración media de las entrevistas del test fue de 90 minutos, mientras que la longitud promedio del retest fue 81 minutos.

Para el test y el retest, la duración media de la entrevista fue de 68 minutos para los niños, de 85 minutos para los adolescentes, 88 minutos para los padres de los niños y 101 minutos para los padres de los adolescentes.

DICA-R-C

La concordancia diagnóstica entre el test y el retest en la entrevista con el niño generalmente osciló entre buena y excelente (Tabla 2). Sólo en TAS y fobia se obtuvo un acuerdo bajo. No hubo diferencias significativas entre el test y el retest ni en los trastornos considerados individualmente ni en las agrupaciones por categorías diagnósticas. Aunque la diferencia no fue significativa, hubo más diagnósticos en las entrevistas del test que en las del retest. Si se considera el número de síntomas presentes en los trastornos (Tabla 3), aparecieron diferencias significativas en ND y TAE. Hubo también diferencias significativas en el número de síntomas en la categoría general de trastorno por conductas perturbadoras, trastornos por ansiedad, y en el número total de síntomas. En todos los casos la sintomatología se vio atenuada en el retest.

DICA-R-A

Para la DICA-R-A, el acuerdo fue de bueno a excelente con la excepción de cinco categorías (Tabla 2). La única diferencia significativa entre el test y el retest fue en ND y en la categoría general de trastornos por ansiedad, en la que hubo una disminución de casos en el retest. En general, hubo una tendencia a diagnosticar más en el test que en el retest. Los trastornos con menor acuerdo fueron fobia, trastorno de conducta (TC), trastorno por estrés post-traumático (TEPT) y DM (episodio pasado). Cuando se analizó el número de síntomas (Tabla 3), la disminución de la sintomatología fue esta-

dísticamente significativa en ND, trastorno distímico (TD), TAS y TAE. En todas las categorías generales y en el número total de síntomas presentes en la entrevista hubo una disminución en el número de síntomas positivos en el retest.

Tabla 2  
Fiabilidad Test-Retest de la DICA-R-C y DICA R-A

Categorías Diagnósticas		n	Acuerdo			Prevalencia %			
			%	k	Test	Retest	Dif.	p	
Déficit Aten. (presente)	C	57	96	.73	8.8	5.3	3.5	.500	
	A	51	98	.85	5.9	7.8	-2.0	1.0	
Déficit Aten. (pasado)	C	-	-	-	-	-	-	-	
	A	51	96	.81	11.8	11.8	0	1.0	
Negativismo	C	57	93	.56	10.5	7.0	3.5	.625	
	A	53	83	.44	24.5	11.3	13.2	.039	
T. Conducta	C	57	95	-	5.3	0	5.2	.250	
	A	53	81	.27	17.0	13.2	3.8	.754	
Depresión Mayor (presente)	C	56	96	.65	5.4	5.4	0	1.0	
	A	53	87	.67	28.3	26.4	1.9	1.0	
Depresión Mayor (pasado)	C	55	98	.49	1.8	5.5	-3.6	.500	
	A	52	79	.39	26.9	17.3	9.6	.227	
Distimia	C	56	96	-	3.6	0	3.6	.500	
	A	52	90	.65	17.3	15.4	1.9	1.0	
T. Ansiedad Sep.	C	57	75	.35	29.8	19.3	10.5	.180	
	A	53	85	.43	20.8	9.4	11.3	.070	
T. Ansiedad Exc.	C	56	92	.41	12.5	3.6	8.9	.063	
	A	53	81	.52	30.2	22.6	7.6	.344	
Fobia	A	54	93	.31	9.3	1.9	7.4	.125	
	C	51	90	.24	11.8	13.7	-2.0	1.0	
T. Obsesivo-C	C	56	100	1.0	1.8	1.8	0	1.0	
	A	53	100	-	0	0	0	1.0	
T. Estrés post.	C	56	96	-	3.6	0	3.6	.500	
	A	53	94	.38	7.5	1.9	5.7	.250	
Anorexia	C	54	100	1.0	1.9	1.9	0	1.0	
	A	51	94	.74	13.7	11.8	2.0	1.0	
Bulimia	C	55	100	-	0	0	0	1.0	
	A	53	96	.78	7.5	11.3	-3.8	.500	
Enuresis	C	56	80	.54	33.9	28.6	5.4	.549	
	A	53	89	.68	18.9	26.4	-7.6	.219	
Encopresis	C	56	91	.50	8.9	10.7	-1.8	1.0	
	A	53	96	.73	9.4	5.7	3.8	.500	
T. Cond. Perturb.	C	57	88	.52	19.3	10.5	8.8	.125	
	A	53	81	.54	34.0	22.6	11.3	.109	
T. Humor	C	57	95	.70	10.5	8.8	1.8	1.0	
	A	53	79	.58	49.1	43.4	5.7	.549	
T. Ansiedad	C	57	75	.46	40.4	26.3	14.0	.057	
	A	53	77	.55	52.8	37.7	15.1	.039	
T. Cond. Aliment.	C	56	100	1.0	1.8	1.8	0	1.0	
	A	53	92	.75	17.0	20.8	-3.8	.625	
T. Eliminación	C	56	80	.57	37.5	32.1	5.4	.549	
	A	53	87	.68	26.4	32.1	-5.7	.453	

C: DICA-R-C; A: DICA-R-A; en negrita diferencias significativas.

**Tabla 3**  
Comparación del número de síntomas presentes en el Test y en el Retest

Categorías	SUJETOS					PADRES					
		Media de Síntomas					Media de Síntomas				
		n	Test	Ret.	Dif.	p	n	Test	Ret.	Dif.	p
Déficit Aten. (presente)	C	57	3.35	2.72	.63	.093	60	5.35	5.12	.23	.466
	A	51	3.02	3.02	0	1.0	53	2.94	2.72	.23	.401
Déficit Aten. (pasado)	C	-	-	-	-	-	54	4.28	4.05	.22	.448
	A	51	3.45	3.45	.02	.948	52	2.19	2.37	-.17	.385
Negativismo	C	57	1.65	1.16	.49	.023	61	1.79	1.80	-.02	.933
	A	53	2.68	1.70	.98	.000	52	1.94	2.00	-.06	.791
T. Conducta	C	57	.82	.93	-.11	.308	60	.57	.60	-.03	.709
	A	53	1.79	1.55	.25	.079	53	.81	.87	-.06	.497
Depres. M. (presente)	C	56	.73	.82	-.09	.546	61	.57	.69	-.11	.609
	A	53	3.28	2.94	.34	.170	52	1.60	1.85	-.25	.369
Depres. M. (pasado)	C	55	.96	.91	.05	.800	61	1.20	1.54	-.34	.124
	A	52	3.25	2.79	.46	.235	52	2.21	1.92	.29	.364
Distimia	C	56	.13	.13	0	1.0	60	.28	.37	-.08	.512
	A	52	1.19	.77	.42	.042	52	.77	.77	-.17	.141
T. Ans. Sep.	C	57	2.11	1.61	.49	.098	61	1.20	.70	.49	.009
	A	53	1.58	.94	.64	.005	53	.58	.45	.13	.424
T. Ans. Exc.	C	56	1.23	.57	.67	.007	61	1.08	.87	.21	.263
	A	53	2.57	1.68	.89	.003	53	.81	.79	.02	.925
Fobia	C	54	.52	.43	.09	.374	54	.59	.76	-.17	.060
	A	51	.67	.67	0	1.0	53	.30	.42	-.11	.243
Cond. Pert.	C	57	5.82	4.81	1.02	.044	61	7.87	7.82	.05	.918
	A	53	8.30	7.17	1.13	.007	53	5.92	5.79	.13	.730
T. Humor	C	57	1.42	1.42	0	1.0	61	1.80	2.28	-.48	.059
	A	53	5.92	4.87	1.06	.008	53	3.58	3.79	-.21	.496
T. Ansiedad	C	57	4.74	3.23	1.51	.006	61	3.15	2.66	.49	.117
	A	53	6.04	4.23	1.81	.002	53	2.21	1.98	.23	.511
Nº total de síntomas	C	57	14.0	11.3	2.70	.003	61	14.5	14.4	.15	.831
	A	53	25.4	20.9	4.47	.002	53	14.3	13.8	.45	.580

C: DICA-R-C; A: DICA-R-A; en negrita diferencias significativas

**DICA-R-P (Niños)**

Los valores kappa para la mayoría de las categorías diagnósticas alcanzaron un nivel de acuerdo moderado (Tabla 4). Los peores resultados se obtuvieron en DM (presente y pasado). En la sección de fobia se encontraron diferencias significativas entre el número de diagnósticos en el test y retest. Los padres informaron significativamente más problemas fóbicos en el retest y, en general, hubo tendencia a informar más problemas en el retest. Cuando se compara la cantidad de sintomatología (Tabla 3), se obtuvo un número

mero significativamente mayor de síntomas de TAS en el test que en el retest. En los trastornos agrupados en categorías generales no hubo diferencias entre los resultados del test y del retest ni en el número de trastornos identificados ni en el número de síntomas.

**Tabla 4**  
Fiabilidad Test-Retest de la DICA-R-C de niños y adolescentes

Categorías Diagnósticas		n	Acuerdo		Prevalencia %			
			%	κ	Test	Retest	Dif.	p
Déficit Aten. (presente)	C	60	85	.41	23.3	25.0	-1.7	1.0
	A	53	85	.64	11.3	5.7	5.7	.250
Déficit Aten. (pasado)	C	54	91	.51	18.5	18.5	0	1.0
	A	52	91	.88	9.6	7.7	1.9	1.0
Negativismo	C	61	90	.61	14.8	14.8	0	1.0
	A	52	90	.56	9.6	15.4	-5.8	.375
T. Conducta	C	60	98	.66	3.3	1.7	1.7	1.0
	A	53	98	.88	7.5	9.4	-1.9	1.0
Depresión M. (presente)	C	61	88	-.04	3.3	4.9	-1.6	1.0
	A	52	88	.23	7.7	17.3	-9.6	.180
Depresión M. (pasado)	C	61	87	.35	4.9	11.5	-6.6	.219
	A	52	87	.32	21.2	7.7	13.5	.039
Distimia	C	60	95	.49	5.0	1.7	3.3	.500
	A	52	95	.77	13.5	15.4	-1.9	1.0
T. Ansiedad Sep.	C	61	92	.57	14.8	11.5	3.3	.688
	A	53	92	.55	9.4	3.8	5.7	.250
T. Ansiedad Exc.	C	61	93	.62	11.5	13.1	-1.6	1.0
	A	53	93	.55	9.4	3.8	5.7	.250
Fobia	C	54	90	.41	5.6	18.5	-13.0	.016
	A	53	90	.46	9.4	5.7	3.8	.625
T. Obsesivo-C	C	60	99	-	0	1.7	-1.7	1.0
	A	53	100	-	0	0	0	1.0
T. Estrés post.	C	60	99	-	1.7	-	1.7	1.0
	A	53	100	-	0	0	0	1.0
Anorexia	C	54	94	-	0	1.9	-1.9	1.0
	A	52	94	.49	11.5	9.6	1.9	1.0
Bulimia	C	58	100	-	0	0	0	1.0
	A	53	99	.79	3.8	5.7	-1.9	1.0
Enuresis	C	60	90	.73	30.0	35.0	-5.0	.453
	A	53	90	.75	18.9	18.9	0	1.0
Encopresis	C	60	95	.74	16.7	13.3	3.3	.625
	A	53	95	.90	11.3	9.4	1.9	1.0
T. Cond. Perturb.	C	61	77	.49	34.4	34.4	0	1.0
	A	53	96	.89	20.8	20.8	0	1.0
T. Humor	C	61	90	.61	13.1	16.4	-3.3	.688
	A	53	74	.37	30.2	30.2	0	1.0
T. Ansiedad	C	61	77	.41	24.6	27.9	-3.3	.790
	A	53	85	.48	22.6	11.3	11.3	.070
T. Cond. Aliment.	C	61	98	-	0	1.6	-1.6	1.0
	A	53	91	.61	13.2	15.1	-1.9	1.0
T. Eliminación	C	60	92	.83	38.3	43.3	-5.0	.375
	A	53	92	.80	26.4	22.6	3.8	.625

C: DICA-R-C; A: DICA-R-A; en negrita diferencias significativas.

#### DICA-R-P (ADOLESCENTES)

El acuerdo entre la información de los padres de adolescentes proporcionada en test y retest fue de nivel moderado (Tabla 4). El mejor acuerdo se obtuvo en TDAH (pasado), TC, TD, bulimia, enuresis y encopresis. Las categorías de trastorno por conductas perturbadoras y trastornos de la eliminación obtuvieron un acuerdo excelente entre test y retest. El peor acuerdo se encontró en DM, donde hubo diferencias significativas en el episodio pasado (más diagnósticos en el test que en el retest). Para la mayoría de trastornos, se observó una tendencia a la atenuación de la información en el retest y a informar más trastornos en el test que en el retest. No se hallaron diferencias significativas cuando se comparó la sintomatología de las categorías diagnósticas por separado o agrupadas, pero el número de síntomas decreció en el retest (Tabla 3).

#### Discusión

Los resultados de nuestro estudio indican que hay estabilidad en la evaluación de los trastornos con la entrevista DICA-R, tal como lo demuestra la ausencia de diferencias significativas entre el test y el retest y el hecho de que la mayoría de los valores kappa sean buenos o excelentes. Los más estables al informar sobre los trastornos fueron los padres y los menos regulares fueron los adolescentes.

La categoría en la que es más difícil obtener acuerdo cuando se analizan los valores kappa es DM. Esta sección de la entrevista analiza los episodios depresivos presentes y pasados. El episodio presente se centra estrictamente en "las dos últimas semanas", mientras que los episodios pasados han podido ocurrir en cualquier otro momento de la vida del niño. En algunos casos, fue muy difícil codificar la sintomatología reciente. Por ejemplo, los casos en los que el trata-

miento empezó entre el test y el retest pueden mostrar cierta respuesta al tratamiento en el retest. Aunque el intervalo medio entre el test y el retest fue de 11 días, hubo casos en los que hubo un intervalo mayor. En el retest, debido al tratamiento, la sintomatología puede no estar presente "cada día o casi cada día" y "la mayor parte del día". Por otro lado, esta sección tiene una dificultad añadida que es evaluar la presencia o ausencia de los síntomas durante "por lo menos dos semanas". Situar los síntomas dentro de un marco temporal y recordar si se han agrupado o no es, obviamente, una tarea muy difícil para los niños, adolescentes e incluso para los padres y puede hacer que la información que se proporciona varíe. Fallon y Schwab-Stone (1994) han señalado que una de las características de las preguntas que causan que la fiabilidad disminuya es el pedir a los niños que delinee la sintomatología en el tiempo. Otros estudios con entrevistas estructuradas para niños también han subrayado la dificultad para evaluar la depresión (Boyle et al., 1993; Costello, Edelbrock, Dulcan, Kalas y Klaric, 1984; Edelbrock, Costello, Dulcan, Kalas y Conover, 1985).

Otra categoría en la que los valores kappa fueron pobres fue en ansiedad (TAS, fobia y TEPT). Considerando los buenos resultados obtenidos en los estudios de fiabilidad entre entrevistadores (de la Osa et al., 1996), sospechamos que los sujetos dieron información diferente en las dos situaciones. El cambio en las respuestas en el retest *versus* el test puso ser debido, entre otras razones, a haber aprendido a decir "no" y así ir más rápido al saltar las preguntas, a timidez, a pensar que ya se ha suministrado la información en el test, a una reestructuración del auto-concepto sobre la información proporcionada, a una mejor comprensión de las preguntas o a la discusión entre los miembros de la familia de la entrevista inicial. Por ejemplo, en la sección sobre fobia,

la pregunta raíz cuestiona sobre la existencia de un miedo específico (a animales, a la oscuridad, etc.), mientras que las preguntas siguientes interrogan sobre si el niño ha estado más asustado que la mayoría de los niños de su edad y si ha evitado el objeto o situación temido. En el retest, cuando se formulaba la pregunta raíz, los sujetos ya sabían el tipo de miedo que interesaba a los entrevistadores y contestaban directamente (recordando que debía ser más temido de lo que la mayoría de las personas temen ese objeto o situación y esto conducía a evitarlo/a). En adultos, Andreasen et al. (1981) también han encontrado una fiabilidad pobre en la sección de ansiedad de la Schedule for Affective Disorders and Schizophrenia (Endicott y Spitzer 1978). Estos autores explican sus resultados de dos formas: primero, la ansiedad patológica es difícil de distinguir de la tensión normal y, segundo, los síntomas de ansiedad son frecuentemente difíciles de diferenciar de los síntomas de depresión. En comparación con otros trastornos, los trastornos por ansiedad presentan consistentemente una mala fiabilidad entre entrevistadores y test-retest (Boyle et al., 1993; Chambers et al., 1985; Costello et al., 1984; Edelbrock et al., 1985; Hodges et al., 1989; Rey, Plapp y Stewart, 1989). Hodges et al. (1989) señalan que esto podría ser debido a cuestiones nosológicas. En algunas de las secciones de ansiedad (TAS y TAE) y en ND en las entrevistas de niños y adolescentes se informó de un número diferente de síntomas en el test y en el retest, sin cambios en la prevalencia de los trastornos. Posiblemente, el estudio del acuerdo a nivel de síntomas podría ayudar a optimizar estas secciones de la entrevista.

El análisis del acuerdo en categorías generales de trastornos demuestra que la entrevista detecta de manera fiable la presencia o ausencia de cualquiera de los trastornos incluidos en ellas. Diversos autores han encontrado, utilizando diferentes entrevistas,

que es más fácil, obviamente, lograr concordancia en categorías amplias que en categorías más específicas (Canino et al., 1987; Hodges et al., 1989; Welner et al., 1987).

En nuestro estudio, los informantes más fiables fueron los padres, seguidos de los niños y finalmente, los adolescentes. Estamos de acuerdo con otros autores en que la información proporcionada por los padres es más fiable (estable) que la información de niños/adolescentes (Boyle et al., 1993; Edelbrock et al., 1985). Sin embargo, nuestros resultados discrepan de los de Jensen et al. (1993), que encontraron que a medida que los niños se hacen más mayores la información que suministran tiende a ser estable en el tiempo. No obstante, estos autores indican que la consistencia en los informes de los niños estaba mediada por variables diferentes a la edad, como el CI: en los niños con CI más alto se observó menos consistencia. En el estudio de Canino et al. (1987), la edad o el intervalo de tiempo no eran variables influyentes en la fiabilidad de los diagnósticos, pero sí que lo era la gravedad del trastorno: el acuerdo era mejor en los casos graves que en los no casos.

Orvaschel, Puig-Antich, Chambers, Tabrizi y Johnson (1982) indican que hay una tendencia a atenuar la gravedad de la sintomatología independientemente del informante. La única excepción en este sentido en nuestros resultados fue en la información de los padres de los niños. El efecto de atenuación en este caso no se observó ni en los trastornos, ni en las categorías amplias ni en los síntomas. Por el contrario, los padres de los niños informaron menos trastornos en el test que en el retest. Este hecho probablemente es debido a que los padres de los niños más jóvenes entre el test y el retest piensan sobre lo que se les ha preguntado y recuperan más datos de la memoria o de la observación directa de los niños, proporcionando información más relevante para el diagnóstico en la segunda evaluación.



Los hallazgos de este trabajo, que es el primero que estudia fiabilidad test-retest en la DICA-R en pacientes externos, se ven respaldados por el hecho de que: (a) han participado múltiples entrevistadores, muy aleatorizados en el rol entrevistador-observador, que desconocían el diagnóstico del niño (tanto de la entrevista del test como el del clínico); (b) la elección de una muestra de pacientes externos con sintomatología moderada, que es más difícil de detectar; (c) un intervalo test-retest razonable; (d) el uso de una muestra de tamaño considerable para la mayoría de las categorías diagnósticas; y (e) el estudio de un amplio espectro de trastornos.

Finalmente, de nuestros resultados podemos sugerir: (a) la revisión de las secciones de DM y trastornos por ansiedad (TAS, TAE y fobia) y (b) el estudio de la fiabilidad centrado en el análisis de síntomas.

Nuestros resultados nos permiten concluir que la fiabilidad test-retest de la DICA-R en una muestra de pacientes psiquiátricos externos es buena. La entrevista mide la presencia de los trastornos de forma esta-

ble. Aunque no hubo diferencias significativas en la mayoría de las categorías diagnósticas, los trastornos aparecen atenuados en el retest. Los adolescentes son los informadores menos fiables y los padres los más estables. La fiabilidad del respondiente depende del tipo de trastorno y de la edad. La entrevista es más fiable en las categorías amplias trastornos que en las específicas. Son necesarios estudios que se centren en el análisis de ítems para proporcionar datos sobre cómo pueden optimizarse las entrevistas diagnósticas estructuradas para niños y adolescentes.

#### Agradecimientos

Esta investigación se ha llevado a cabo gracias a la ayuda DGICYT #PM91-209 del Ministerio de Educación y Ciencia.

Agradecemos la colaboración del personal y de los niños de los Centros de Asistencia Psiquiátrica Primaria Infanto-Juvenil del Hospital Clínico de Barcelona, del Maresme y del Consorci Hospitalari del Parc Taulí de Sabadell.

#### Referencias

- Achenbach, T.M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Andreasen, N.C., Grove, W.M., Shapiro, R.W., Keller, M.B., Hirschfeld, R.M.A. y McDonald-Scott, P. (1981). Reliability of lifetime diagnosis. *Archives of General Psychiatry*, 38, 400-405.
- American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3d ed.rev.). Washington D.C.: Author.
- Apter, A., Orvaschel, H., Laseg, M., Moses, T. y Tyano, S. (1989). Psychometric properties of the K-SADS-P in an Israeli adolescent inpatient population. *Journal of the American Academy of Child and Adolescent Psychiatry* 28, 61-65.
- Boyle, M.H., Offord, D.R., Racine, Y., Sandford, M., Szatmari, P., Fleming, J.E. y Price-Munn, N. (1993). Evaluation of the Diagnostic Interview for Children and Adolescents for use in general population samples. *Journal of Abnormal Child Psychology*, 21, 6, 663-681.
- Canino, G.J., Bird, H.R., Rubio-Stipec, M.A., Woodbury, M.A., Ribera, J.C., Huertas, S.E. y Sesman, M.J. (1987). Reliability of child diagnosis in a hispanic sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 4, 560-565.
- Carlson, G.A., Kashani, J.H., de Fatima, T.M., Thomas, M., Vaydya, A. y Daniel, A.E., (1987). Comparison of two structured interviews on a psychiatrically hospitalized population of children. *Journal of the American*

- Academy of Child and Adolescent Psychiatry*, 26, 645-648.
- Chambers, W. J., Puig-Antich, J., Hirsch, M., Paez, P., Ambrosini, P.J., Tabrizi, M.A. y Davies M. (1985). The assessment of Affective Disorders in children and adolescents by semistructured interview. *Archives of General Psychiatry*, 42, 696-702.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Costello, A.J., Edelbrock, L.S., Dulcan, M.K., Kalas, R. y Klaric, S.H. (1984). *Report on the NIMH Diagnostic Interview Schedule for Children (DISC)*. Washington, DC: National Institute of Mental Health.
- Edelbrock, C.S., Costello, A.J., Dulcan, M.K., Kalas, R. y Conover, N.C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development*, 56, 265-275.
- Endicott, J. y Spitzer, R.L. (1978). A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry*, 35, 837-844.
- Ezpeleta, L., de la Osa, N., Júdez, J., Doménech, J.M., Navarro, J.B. y Losilla, J.M. (1997). Diagnostic agreement between clinician and the Diagnostic Interview for Children and Adolescents - DICA-R in a Spanish outpatient sample. *Journal of Child Psychology and Psychiatry*, 38, 431-440.
- Ezpeleta, L., de la Osa, N., Doménech, J.M., Navarro, J.B. y Losilla, J.M. (1995). La Diagnostic Interview for Children and Adolescents - DICA-R: Acuerdo diagnóstico entre niños/adolescentes y sus padres. *Revista de Psiquiatría de la Facultad de Medicina de Barcelona*, 22, 153-163.
- Fallon, T. y Schwab-Stone, M. (1994). Determinants of reliability in psychiatric surveys of children aged 6-12. *Journal of Child Psychology and Psychiatry*, 35, 1391-1408.
- Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.
- Gutterman, E.M., O'Brien, J.D. y Young, J.G. (1987). Structured diagnostic interviews for children and adolescents: Current status and future directions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 621-630.
- Herjanic, B. y Campbell, W. (1977). Differentiating psychiatrically disturbed children on the basis of a structured interview. *Journal of Abnormal Child Psychology*, 5, 127-134.
- Herjanic, B., Herjanic, M., Brown, F. y Wheat, T. (1975). Are children reliable reporters?. *Journal of Abnormal Child Psychology*, 3, 41-48.
- Herjanic, B. y Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between parent on individual symptoms. *Journal of Abnormal Child Psychology*, 10, 307-324.
- Hodges, K., Cools, J. y McKnew, D. (1989). Test-Retest reliability of a clinical research interview for children: The Child Assessment Schedule. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1, 317-322.
- Hollingshead, A.B. (1975). *Four factor index of social status*. Unpublished manuscript, Yale University, Department of Sociology.
- Jensen, P.S., Bird, H., Canino, G., Dulcan, M., Rae, D., Richters, J., Piacentini, J., D., Goodman, S., Lahey, B., Rubio-Stipec, M. y Schawb-Stone, M. (1993). Attenuation of the Diagnostic Interview Schedule for Children (DISC 2.1): Sex, age, and IQ relationships. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 884.
- Kashani, J.H., Orvaschel, H. Burk, J.P. y Reid, J.C. (1985). Informant variance: The issue of parent-child disagreement. *Journal of the American Academy of Child Psychiatry*, 24, 437-441.
- Lucas, C. (1992). The order effect: reflections on the validity of multiple test presentations. *Psychological Medicine*, 22, 197-202.
- Lucas, C., Shaffer, D., Fisher, P., Parides, M. y Breton, J.J. (1994, June). *Features of interview questions associated with attenuation*. Poster presented at the meeting of the Society for the Research in Child and Adolescent Psychopathology, London, Great Britain.
- Mehta, C. y Patel, N. (1995). *StatXact3 for Windows. Statistical software for exact nonparametric inference*. Cambridge, MA: CYTEL Software Corporation.
- Orvaschel, H., Puig-Antich, J., Chambers, W., Tabrizi, M.A. y Johnson, R. (1982). Retrospective assessment of prepubertal major depression with the Kiddie-SADS-E. *Journal of the American Academy of Child Psychiatry*, 21, 392-397.

- Osa, N. de la, Ezpeleta, L., Doménech, J.M., Navarro, B. y Losilla, J.M. (1996). Fiabilidad entre entrevistadores de la DICA-R. *Psicothema*, 8, 359-368.
- Osa, N. de la, Ezpeleta, L., Doménech, J.M., Navarro, J.B. y Losilla, J.M. (1996). Validez convergente y discriminante de la Entrevista Diagnóstica Estructurada DICA-R. *Clínica y Salud*, 7, 181-194.
- Reich, W., Herjanic, B., Welner, Z. y Gandhi, P.R. (1982). Development of a structured psychiatric interview for children: Agreement of diagnosis comparing child and parents interviews. *Journal of Abnormal Child Psychology*, 10, 325-336.
- Reich, W., Shayka, J.J. y Taibleson, Ch. (1991). *Diagnostic Interview Schedule for Children and Adolescen-DICA-R* (L. Ezpeleta, Trans.). Unpublished manuscript, Washington University, Division of Child Psychiatry.
- Rey, J.M., Plapp, J.M. y Stewart, G.W. (1989). Reliability of psychiatric diagnosis in referred adolescents. *Journal of Child Psychology and Psychiatry*, 30, 879-888.
- Roberts, R.E., Solovitz, B.L., Chen, Y. y Casat, Ch. (1996). retest stability of DSM-III-R diagnoses among adolescents using the Diagnostic Interview Schedule for Children (DISC-2.1C). *Journal of Abnormal Child Psychology*, 24, 349-362.
- Robins, L. (1985). Epidemiology: Reflections on testing the validity of psychiatric interviews. *Archives of General Psychiatry*, 42, 918-924.
- Shaffer, D., Schwab-Stone, M., Fisher, P., Davies, M., Piacentini, J. y Gioia, P. (1988). *Results of a field trial and proposals for a new instrument (DISC-R)*. Washington, DC: National Institute of Mental Health (Grant NOS. MH 36971 y MH CRC 30906-10).
- Schwab-Stone, M., Fisher, P., Piacentini, J., Shaffer, D., Davies, M. y Briggs, M. (1993). The Diagnostic Interview Schedule for Children-Revised (DISC-R): II. Test-retest reliability. *Journal of the American Academy of Child and Adolescents Psychiatry*, 32, 651-657.
- Sylvester, C.E., Hyde, T.S. y Reichler, R.J. (1987). The Diagnostic Interview for Children and the Personality Inventory for Children in studies of children at risk for anxiety disorders or depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 668-686.
- Welner, Z., Reich, W., Herjanic, B., Jung, K.G. y Amado, H. (1987). Reliability, validity, and parent-child agreement studies of the Diagnostic Interview for Children and Adolescents (DICA). *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 649-653.

Acceptado el 19 de abril de 1997