

FIABILIDAD ENTRE ENTREVISTADORES DE LA DICA-R

Nuria de la Osa, Lourdes Ezpeleta, José M^a Domenech, José Blas Navarro
y José M^a Losilla

Universidad Autónoma de Barcelona

El artículo examina la fiabilidad entre entrevistadores de la Diagnostic Interview for Children and Adolescents-Revised (DICA-R). El estudio se llevó a cabo con un total de 45 entrevistas, en las cuales participaron un examinador y un observador. Se describe de manera exhaustiva el procedimiento de entrenamiento. Los resultados muestran que la DICA-R es un instrumento objetivo que puede ser utilizado por personas legas si se les proporciona un entrenamiento adecuado. Se pone énfasis en la necesidad de que los diseños para estudiar la fiabilidad entre examinadores sean estrictos.

Interrater reliability of the Diagnostic Interview for Children and Adolescents (DICA-R). The paper examines the interrater reliability of the Diagnostic Interview for Children and Adolescents-Revised (DICA-R). The study was held over a total of 45 interviews in which an interviewer and an observer participated. The training program is exhaustively described. The results show that DICA-R is an objective instrument that can be used by lay persons if a proper training takes place. Strictness in interrater reliability designs is encouraged.

El uso de instrumentos o procedimientos de evaluación similares es una condición *sine qua non* cuando se pretende comparar diferentes estudios. Se suele dar por sentado que no existen grandes diferencias en el uso que se hace de esos procedimientos o instrumentos. Sin embargo, los conocimientos o la experiencia clínica previa de los evaluadores pueden ser más un obstáculo que una ventaja a la hora de garantizar un juicio fiable cuando se utilizan instrumentos estructurados, como resultado de un bagaje teórico o profesional

dispar de aquellos (Gibbon, McDonald-Scott y Endicott, 1981). Un entrenamiento adecuado y la valoración del mismo son necesarios para asegurarse de que los examinadores utilizan el instrumento de forma que quede garantizado que los subsecuentes resultados son realmente comparables.

El uso de entrevistas no estructuradas está perdiendo terreno en pos de las entrevistas estructuradas entre la comunidad investigadora así como en la práctica clínica diaria (Angold, 1994). Este autor pone énfasis en el hecho de que muchos centros importantes han incorporado ya las entrevistas estructuradas en sus tareas diagnósticas básicas. Una de las razones para este giro hacia las entrevistas estruc-

Correspondencia: Nuria de la Osa
Departament de Psicologia de la Salut
Universitat Autònoma de Barcelona
08193 Bellaterra (Spain)

turadas es la falta de acuerdo observada entre los diagnósticos de los clínicos (Anthony et al. 1985; Costello, 1992; Helzer et al., 1985; Matarazzo, 1983; Robins, 1985). La entrevista, la principal técnica utilizada por los clínicos para obtener la información relevante que les permite generar sus diagnósticos, se ve afectada por varias fuentes de variabilidad (Edelbrock y Costello, 1990; Lewis, Pelosi, Araya y Dunn, 1992; Spitzer, Endicott y Robins, 1975; Young, O'Brien, Gutterman y Cohen, 1987). Una de ellas es la "variancia del sujeto", es decir la variabilidad inherente al entrevistado; la "variancia de criterio" o variabilidad debida al uso de diversos criterios diagnósticos es otra; asimismo existe la variabilidad debida a la percepción de la misma información por parte de diferentes evaluadores - "variancia de la observación" - y por último, la mayor causa de discordancia en el proceso diagnóstico, "la variancia de la información" o variabilidad en el tipo de información obtenida. El desarrollo de criterios diagnósticos específicos redujo considerablemente la variancia de criterio. La aparición de las entrevistas diagnósticas estructuradas disminuyó la variancia de la información. Únicamente el entrenamiento cuidadoso puede solventar la variancia de la observación.

El entrenamiento y la evaluación de los evaluadores debería ser una práctica indispensable previa al uso de un instrumento estructurado. Gibbon et al. (1981) afirman que el obviar el proceso de entrenamiento compromete la fiabilidad del diagnóstico tan seriamente que reduce el valor de la totalidad de la investigación. Si no nos cercioramos previamente de que todos los evaluadores utilizan y codifican la entrevista de la misma manera, los datos no serán homogéneos y por tanto los resultados se verán falseados.

El objetivo del proceso de entrenamiento es verificar que el equipo de entrevistadores posee las habilidades necesarias para elicitar y recoger la información de acuerdo con las directrices establecidas por el protocolo de la entrevista.

La Tabla 1 presenta los resultados de la fiabilidad entre entrevistadores obtenidos con las entrevistas diagnósticas estructuradas más utilizadas con niños y adolescentes. Se hace difícil establecer comparaciones ya que algunos de los estudios contienen un componente test-retest. Las muestras, los informantes y las entrevistas difieren entre los estudios. Los mejores resultados se obtienen cuando los resultados de la fiabilidad entre entrevistadores no contienen datos del test-retest. Shaffer et al. (1993) obtuvieron los mejores valores kappa utilizando la DISC-R. Cuando se utilizan diseños idénticos, los peores resultados se obtienen para el trastorno por negativismo desafiante (ND) y el trastorno por ansiedad excesiva (TAE) (Apter, Orvaschel, Lases, Moses y Tyano, 1989; Shaffer et al., 1993).

Este trabajo describe la fiabilidad entre entrevistadores de la DICA-Revised (DICA-R, Reich, Shayka y Taibleson, 1991) en una muestra niños y adolescentes, todos ellos pacientes psicológicos externos. La DICA ya ha demostrado una alta fiabilidad intra y entre entrevistadores (Herjanic y Reich, 1982; Welner, Reich, Herjanic, Jung y Amado, 1987), sin embargo no existen estudios que aborden de forma concreta la fiabilidad entre entrevistadores de la DICA-R. Este trabajo es parte de la adaptación de la entrevista a la población española (Ezpeleta, de la Osa, Júdez, Doménech y Navarro, 1994). Al proceso de traducción que implicó traducción y retrotraducción, siguiendo las directrices de Brislin (1986), así como la contrastación con una de las autoras del instrumento del contenido de los ítems, siguió el estudio

de sus propiedades psicométricas. Las adaptaciones de instrumentos deben aportar evidencia de las propiedades psicométricas del mismo, y no tomar como referencia las que posee el instrumento original (Hambleton, 1994).

La *fiabilidad* en este trabajo queda estrictamente circunscrita a datos entre entrevistadores recogidos por dos examinadores simultáneamente. La fiabilidad test-retest y la concordancia entre informantes son objeto de discusión en otros trabajos.

Método

Sujetos

Los datos corresponden a 32 niños y adolescentes, pacientes externos y a 6 de sus hermanos. Los niños tenían edades comprendidas entre los 7 y los 12 años (n=22; 58%). Los adolescentes oscilaban en edad entre los 13 y los 17 años (n=16; 42%). El 42% de los datos corresponden a varones y el 52% a hembras. Las familias procedían

Tabla 1												
Fiabilidad entre entrevistadores de las entrevistas estructuradas (coeficiente kappa)												
Tras	Estudio 1	Estudio 2	Estudio 3	Estudio 4	Estudio 51	Estudio 6		Estudio 7	Estudio 8			
						C	P		C	A	PC	PA
TDAH	.21	1	.45					1	.43	.24	.78	.86
TC	.47	1			.97			1	.37	.92	-.01	.87
ND		.79	.38					.66	.33	.28	.32	.92
DM	.36		.50		.93				.00	.45	.77	.31
DD			.55						.19	.40	.57	.51
HUMOR		.90										
TAS	.38		.51					1	.01	.00	.38	-
TAE	.34		.35		.84			1	.01	.54	.44	.69
FOB					.91							
TOC					.94							
ANS		.76										
ENU			.75									
CGE				≥.75								
						.46 a 1		.63 a 1				

1. Coeficiente de correlación intreclase.

Tras., Trastorno; C, entrevista niño; P, entrevista padres; T, Entrevista profesor; PC, padres de niños; PA, padres de adolescentes; TDAH, trastorno por déficit de atención con hiperactividad; TC, trastorno de conducta; ND, trastorno por negativismo desafiante; DM, episodio depresivo mayor; DD, distimia; HUMOR, categoría general de trastornos del humor; TAS, trastorno por ansiedad de separación; TAE, trastorno por ansiedad excesiva; FOB, fobia; TOC, trastorno obsesivo-compulsivo; ANX, categoría general de trastornos por ansiedad; ENU, enuresis; CGE, coeficiente global de la entrevista.

Estudio 1*: Costello et al. (1984). Diagnostic Interview Schedule for Children (DISC). 242 pacientes psiquiátricos de 6 a 18 años. 9 días de intervalo.

Estudio 2*: Welner et al. (1987). Diagnostic Interview for Children and Adolescent (DICA). 27 pacientes internos de 7 a 17 años. 1 a 7 días de intervalo.

Estudio 3*: Canino et al. (1987). DISC. 91 pacientes clínicos y no clínicos de 4 a 16 años. 19 días de intervalo.

Estudio 4: Hodges et al. (1989). Child Assessment Schedule (CAS). Pre-establecido para 5 entrevistas consecutivas antes de iniciar el estudio.

Estudio 5: Apter et al. (1989). Schedule for Affective Disorders and Schizophrenia for School-Aged Children (K-SADS-P) (Traducción hebrea). 70 pacientes internos. 11 a 18 años.

Estudio 6: Loeber et al. (1989). DISC-C, DISC-P and DISC-T. 44 varones referidos a consulta, 7 a 13 años y sus padres.

Estudio 7: Shaffer et al. (1993). DISC-R. 10 inpatients (videogradas). 11 a 17 años evaluados por 4 clínicos diferentes con un total de 37 DISC-R evaluadas.

Estudio 8*: Boyle et al. (1993). DICA-R. 251 población general de 6 a 16 años.

* Los estudios de fiabilidad entre evaluadores también incluyen un componente test-retest.

de cuatro Centros de Asistencia Psiquiátrica Infantil Primaria. El 100% de las familias eran de raza caucásica. El Índice de estatus socioeconómico de Hollingshead (Hollingshead, 1975) se distribuyó entre las 32 familias de la siguiente manera: clase II: 12,5%; clase III: 15,6%; clase IV: 18,8%; clase V: 43,8%. No se pudo disponer de los datos socioeconómicos de 3 de las familias. Los sujetos con evidencia de deficiencia mental fueron excluidos del estudio.

Los datos fueron obtenidos al azar de primeras entrevistas o de entrevistas de retest, de cualquiera de las tres versiones de la entrevista, con la posibilidad de incluir en el estudio datos del test y del retest del mismo niño/padres o datos de niños y padres de una misma familia, en este caso los entrevistadores eran siempre diferentes. El estudio presenta los resultados de un total de 45 entrevistas. Todas ellas fueron llevadas a cabo simultáneamente por un entrevistador y un observador ambos ciegos a cualquier tipo de información sobre el caso, ya que no tenían acceso a datos previos recogidos por el propio equipo de investigación en el caso de que la entrevista fuera un retest, ni a las valoraciones del clínico o de personal del hospital.

Material

La Diagnostic Interview for Children and Adolescents-Revised (DICA-R 7.2; Reich et al. 1991) fue la primera entrevista diagnóstica estructurada basada en los criterios del DSM-III-R (APA, 1987). La DICA-R se presenta en tres versiones casi idénticas en estructura y contenido: una para niños entre 6 y 12 años, otra para adolescentes en edades comprendidas entre 13 y 17 años y una versión para padres que cubre el período de 6 a 17 años y provee de información acerca de todos los hijos de la familia con edades comprendidas en ese intervalo. En comparación con otros proto-

colos de entrevista utilizados para clasificar trastornos psiquiátricos en la infancia, la DICA-R es muy estructurada y cubre la mayoría de los trastornos psicológicos de los niños y los adolescentes (Hodges, 1993). La DICA presenta una alta fiabilidad inter e intra evaluadores (Herjanic y Reich, 1982; Welner et al., 1987). El grado medio de acuerdo entre padres y niños/adolescentes al responder a la DICA es moderado (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), obteniéndose un mayor acuerdo cuando la patología presente es observable y no de tipo internalizante. La DICA discrimina bien entre muestras pediátricas y psiquiátricas y correlaciona moderadamente con otras medidas de psicopatología infantil (ej. diagnóstico clínico) (Carlson et al., 1987; Herjanic y Campbell, 1977; Welner et al., 1987). La concordancia con el Personality Inventory for Children osciló entre .11 y .48 (k media = .28) (Sylvester et al., 1987). Boyle et al. (1993) hallaron una buena fiabilidad test-retest para los trastornos externalizantes y una fiabilidad pobre para los de tipo internalizante.

Procedimiento

Existe un consenso general de que el entrenamiento es necesario incluso cuando las entrevistas son estructuradas. Por ello, en nuestro estudio se ha puesto un especial énfasis en el entrenamiento de los entrevistadores. El entrenamiento se basó en las directrices propuestas por Gibbon et al. (1981), con ligeras modificaciones. El período de entrenamiento duró aproximadamente tres meses. El equipo de entrevistadores estaba compuesto por 17 estudiantes de Psicología y 2 doctoras en Psicología con experiencia clínica. A lo largo del trabajo

utilizaremos el término *lego* para referirnos a las personas en formación, en el sentido de que son personas con una cualificación general, aunque no especialistas y carecen de experiencia clínica. El entrenamiento incluía 4 fases: (1) estudio de la entrevista, (2) entrevistas de práctica entre los propios entrevistadores, (3) observación y codificación de entrevistas grabadas, (4) observación y codificación de entrevistas con casos reales. A lo largo de todo el estudio se ha llevado a cabo una monitorización continua.

Fase I: estudio de la entrevista

Se dio a los aprendices información general acerca de la finalidad, estructura y contenido de la entrevista. Seguidamente a esta introducción preliminar, los estudiantes se concentraron en el estudio exhaustivo de todas las preguntas y su relación con los criterios del DSM-III-R, asimismo se familiarizaron con los códigos de corrección y los saltos de una pregunta a otra estrictamente marcados en la entrevista.

Fase II: entrevistas de práctica

Entrenadoras y estudiantes se alternaron en la simulación de entrevistas. Después del estudio de un caso clínico, uno de los miembros del equipo fingía ser un paciente y daba las respuestas que podría dar un paciente real afecto de una determinada patología. Estas situaciones de "role-playing", permitían a los estudiantes familiarizarse con la entrevista sin la presión que supone una situación real. Una vez eran capaces de codificar y realizar los saltos de la entrevista correctamente, el entrenamiento pasaba a la fase siguiente.

Fase III: entrevistas grabadas

En un primer estadio, las entrevistadoras experimentadas realizaron grabaciones

de entrevistas con pacientes externos y/o sus padres, que presentaban una variedad de trastornos mentales presentes o pasados y de niños/adolescentes sin patología. Las grabaciones fueron evaluadas por un grupo de estudiantes (4 ó 5 simultáneamente). En un estadio posterior, los aprendices entrevistaron y grabaron a sujetos voluntarios. Estas grabaciones eran evaluadas por todo el equipo incluyendo a las entrenadoras. Estas situaciones aportaban información no sólo acerca de cómo un evaluador aplicaba las reglas y conceptos teóricos de la entrevista, sino acerca de sus habilidades para elicitar y comprobar la información que proporcionaba el sujeto entrevistado. Después de codificar cada sección de la entrevista, se clarificaban y comentaban los posibles desacuerdos y se ofrecían orientaciones de como el entrevistador debería haber modificado o reconducido su entrevista.

Fase IV: observación y codificación de casos reales

En esta fase, se formaron aleatoriamente parejas de entrenadoras-aprendices. Los aprendices observaban y codificaban al mismo tiempo que las entrenadoras llevaban a cabo la entrevista. Los estudiantes se consideraron aptos para utilizar la entrevista por sí solos cuando la media de acuerdo diagnóstico en cinco entrevistas, era igual o superior a $k=.80$. Esta era una condición indispensable para incorporarse al equipo de investigación como entrevistador/a.

Los evaluadores tienden a caer con el tiempo en la conocida "deriva del observador". Esto puede afectar tanto a su técnica de entrevista como a su forma de puntuar (Kazdin, 1977). Si se prevé que un estudio va a seguir en curso por un largo período de tiempo, como es nuestro caso, debe contemplarse algún procedimiento de

monitorización a lo largo del mismo. Así pues, la concordancia entre entrevistadores fue evaluada de forma aleatoria durante todo el estudio. La monitorización debería ser un aspecto permanente en el entrenamiento.

Los entrevistadores alternaban su tarea de entrevistar-observar, aunque en la mayoría de los casos el más experimentado era el que entrevistaba al paciente. Con el tiempo los aprendices se convertían a su vez en entrenadores que podían ayudar y enseñar a nuevos compañeros.

La gestión informática de los datos de la entrevista se realizó con el Sistema DAT 2.0 (Doménech y Losilla, 1995) que garantizó la ausencia de errores de tipo lógico y aritmético y permitió generar los diagnósticos utilizando los algoritmos previamente definidos por la entrevista, excepto en el caso de la fobia, en el cual se incluyeron las preguntas referentes a la severidad del trastorno. Los diagnósticos se realizaron por separado de las evalua-

ciones de los padres y los niños/adolescentes. El análisis estadístico se efectuó con SPSS 6.1 para Windows. El estadístico comúnmente utilizado para determinar la concordancia diagnóstica es el coeficiente κ (Cohen, 1960), ya que tiene en cuenta el acuerdo por azar y provee de un índice estándar que permite comparar diferentes evaluaciones. En este estudio seguiremos las directrices propuestas por Fleiss (1981), y consideraremos los valores κ superiores a .75 excelentes, buenos aquellos comprendidos entre .40 y .75 y pobres los inferiores a .40. También se calcularon los índices de sensibilidad y especificidad considerando al entrevistador como criterio.

Resultados

La Tabla 2 proporciona las estimaciones de la fiabilidad entre entrevistadores para el total de la muestra.

Tabla 2
Fiabilidad entre entrevistadores (DICA-R-C/DICA-R-A/DICA-R-P)

Categorías diagnósticas	n	Prevalencia		Acuerdo				Dif. Entrev. vs. Observ.	
		Entrev.	Observ.	Sens.	Espec.	% Acuerdo	κ	%	IC 95%
TDAH (pres.)	44	11.4	11.4	100	100	100	1.00	0	-4.4 to 3.0
ND	45	11.1	15.6	100	95	95	81	-4.4	
TC	45	8.9	8.9	100	100	100	1.00	0	
DM (pres.)	45	8.9	8.9	75	98	95	.73	0	
DM (pas.)	45	4.4	4.4	100	100	100	1.00	0	
DD	45	8.9	8.9	100	100	100	1.00	0	
TAS	45	17.8	20.0	88	95	93	.78	-2.2	-6.6 to 5.4 -3.0 to 4.4
TAE	45	11.1	6.7	60	100	95	.73	4.4	
FOBIA	40	5.0	5.0	100	100	100	1.00	0	
TOC	43	2.3	2.3	100	100	100	1.00	0	
TEPT	43	2.3	2.3	100	100	100	1.00	0	
Anorexia	40	5.0	2.5	50	100	97	.65	2.5	-2.4 to 2.5
Bulimia	41	2.4	2.4	100	100	100	1.00	0	
Enuresis	43	30.2	32.6	100	97	98	.95	-2.3	-2.3 to 2.2
Encopresis	43	4.7	4.7	100	100	100	1.00	0	

*TEPT: Trastorno por estrés postraumático

La fiabilidad entre entrevistadores ha sido estimada sobre 45 entrevistas. En algunos trastornos no fue posible llevar a cabo el análisis de todos las entrevistas por falta de datos, lo que hace oscilar el valor de N . Los resultados muestran una fiabilidad entre entrevistadores alta (k entre .65 y 1.0). No existen diferencias significativas entre las valoraciones de entrevistadores y observadores. Los valores κ más bajos se obtuvieron en anorexia, TAE y episodio depresivo mayor (episodio presente) (DM).

Cuando aparecieron discrepancias, la tendencia era la de considerar la conducta como no patológica por parte del entrevistador experimentado. La persona lego percibió la información como más severa. Aunque tampoco existieron diferencias significativas, el entrevistador diagnóstico mayor presencia de Anorexia y TE y el observador más ND, trastorno por ansiedad de separación (TAS) y enuresis.

Discusión

Los resultados de nuestro estudio indican que la evaluación de trastornos psicológicos con la DICA-R es objetiva, como se desprende de la ausencia de diferencias significativas entre entrevistadores y observadores y los valores κ en su totalidad, los cuales oscilan entre un rango de acuerdo bueno a excelente. Estos resultados son similares a los de aquellos estudios que han valorado la fiabilidad entre entrevistadores independientemente de la situación de test-retest (Apter et al., 1989; Hodges, Cools y McKnew, 1989; Loeber, Green, Lahey y Stouthamer-Loeber, 1989; Shaffer et al., 1993). Ninguno de estos estudios se ha llevado a cabo con la DICA-R por lo cual no son del todo comparables. Todos los estudios realizados utilizando este instrumento (Boyle et al., 1993; Welner et al., 1987) incluyen el componente test-retest. En nuestra opinión, el diseño óptimo es aquel en el que las valoraciones se hacen de manera

simultánea por un entrevistador y uno o varios observadores. Cuando el acuerdo entre entrevistadores se estudia mediante un diseño test-retest, los resultados pueden verse enmascarados por muchas variables incontroladas que tienen cabida en este tipo de diseño. Robins (1985) ha resumido algunas de ellas, por ejemplo los cambios clínicos en el entrevistado, una mala explicación o una mala comprensión por parte del que responde de la finalidad del diseño y por tanto de la necesidad de contestar dos veces a las mismas cuestiones, el efecto terapéutico de la primera entrevista o los efectos de la entrevista (duración, fatiga, aburrimiento). Asimismo la atenuación de las segundas mediciones ha sido ampliamente corroborada (Andreasen et al., 1981; Lucas, 1992; Lucas, Shaffer, Fisher, Parides y Breton, 1994). En consecuencia, el estudio del acuerdo entre entrevistadores debería huir de esas fuentes de desacuerdo. Pocos estudios han separado el análisis del acuerdo entre entrevistadores del análisis de la fiabilidad test-retest. Como parte del estudio de adaptación de la DICA-R a la población española, también hemos estudiado la fiabilidad test-retest, utilizando la misma muestra y con los mismos entrevistadores. Los resultados muestran importantes diferencias entre los dos tipos de diseño. En los estudios con entrevistas estructuradas el test y el retest son realizados por entrevistadores diferentes para garantizar valoraciones "a ciegas". Así pues, tal como se esperaba, el acuerdo entre entrevistadores fue mayor que el acuerdo test-retest (Ezpeleta, de la Osa, Doménech, Navarro y Losilla, 1995). Las razones esgrimidas para explicar el desacuerdo test-retest (Ezpeleta et al., 1995; Robins, 1985; Schwab-Stone et al, 1993) pueden estar contaminando los resultados del análisis de la fiabilidad entre entrevistadores.

En nuestro país sólo Vizcarro et al. (1986) han estudiado el valor de las entrevistas estructuradas para aumentar la fiabilidad entre entrevistadores. Las diferencias

en la muestra, la edad de los sujetos, el instrumento y la metodología de entrenamiento utilizados, hacen los resultados incompatibles. Sin embargo, sus resultados indican que el uso de entrevistas estructuradas no aumenta el acuerdo entre evaluadores.

La pequeña diferencia entre el número de diagnósticos de anorexia referidos por el entrevistador al compararlo con el observador (generalmente menos entrenado), podría ser debido a las respuestas insinceras frecuentes en este tipo de pacientes. Incluso para los clínicos experimentados, el paciente anoréxico es esquivo y mentiroso. En general, cuando aparecen resultados discordantes, el entrevistador lego tiende a considerar las conductas como más patológicas (Wing, Henderson y Winckle, 1977). Sin embargo, las diferencias nunca llegaron a tener significación estadística. Boyle et al. (1993), utilizando la DICA-R con población normal o Shaffer et al. (1993) con la DISC-R, también hallaron que los entrevistadores legos valoraban las conductas como más severas.

El tamaño de la muestra hace desaconsejable el dar resultados por separado de las tres versiones de la entrevista. Este hecho no parece relevante dado que no es la concordancia entre informadores lo que estudiamos sino el grado de acuerdo entre entrevistadores y que las tres versiones de la entrevistas comparten contenido y forma. Sin embargo, un análisis cuidadoso de nuestros datos sugiere que los valores κ más bajos se obtuvieron en la DICA-R-A. Los valores medios de las κ para esta versión fue-

ron de .83, mientras que los valores medios para las versiones DICA-R-C y DICA-R-P fueron de .96 y .98 respectivamente. Los adolescentes ya se mostraron como los más inestables en la situación test-retest (Ezpeleta et al., 1995). Se necesitarán estudios posteriores para confirmar la posibilidad de una mayor dificultad a la hora de evaluar a este grupo de población.

Los puntos fuertes de este trabajo se resumen en (a) el exhaustivo entrenamiento de los entrevistadores, (b) la continua monitorización a lo largo del estudio, (c) es el primer estudio de la fiabilidad entre entrevistadores con la DICA-R sin un componente test-retest, (d) la incorporación de entrevistadores legos.

Nuestros resultados nos permiten concluir que los resultados de la fiabilidad entre entrevistadores de la DICA-R en una muestra de pacientes externos fue muy buena. La entrevista permite valorar la presencia o la ausencia de trastornos de manera objetiva. No existe evidencia que desaconseje el uso del instrumento por parte de entrevistadores legos siempre y cuando estos sean adecuadamente entrenados en el uso de la DICA-R. Los estudios que se centren en la fiabilidad entre entrevistadores deberían llevarse a cabo independientemente de los diseños test-retest.

Agradecimientos

Esta investigación se ha llevado a cabo gracias a la ayuda DGICYT No. PM91-209, Ministerio de Educación y Ciencia.

Referencias

- 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.
- Angold, A. (1994). Clinical interviewing with children and adolescents. In M. Rutter, E. Taylor y L. Hersov. (Eds), *Child and Adolescent Psychiatry* (pp. 51-63). Oxford: Blackwell Scientific Publications.

- Anthony, J.C., Folstein, M., Romanoski, A.J., Von Korff, M.R., Nestadt, G.R., Chahal, R., Merchant, A., Brown, H., Shapiro, S., Kramer, M. y Gruenberg, E.M. (1985). Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis. *Archives of General Psychiatry*, 42, 667-675
- American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd 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, (1), 61-65.
- Boyle, M.H., Offord, D.R., Racine, Y., Sanford, 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, 663-681.
- Brislin, R.W. (1986). The wording and translation of research instruments. En W.J. Lonner y J. W. Berry (Eds.), *Field methods in cross-cultural psychology* (pp. 137-164). Newbury Park, CA: Sage Publications.
- Canino, G.J., Bird, H.R., Rubio-Spítec, M.A., Woodbury, M.A., Ribera, J.C., Huertas, S.E. y Sesman, M. J. (1987). Reliability of child diagnosis in a spanish sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 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.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurements*, 20, 37-46.
- Costello, C.G. (1992). Research on symptoms versus research on syndromes. *British of Journal Psychiatry* 160, 304-308.
- Doménech, J.M. y Losilla, J.M. (1995). *Sistema DAT: gestor de datos científicos. Manual de referencia*. Campus de Bellaterra, Barcelona: Laboratori d'Estadística Aplicada i de Modelització. Universitat Autònoma de Barcelona.
- Edelbrock, C. y Costello, A.J. (1990). Structured interviews for children and adolescents. In G. Goldstein y M. Hersen (Eds.), *Handbook of psychological assessment* (2nd ed., pp.308-323). New York: Pergamon Press.
- Ezpeleta, L., de la Osa, N., Júdez, J., Doménech, J.M. y Navarro, J.B. (1994, Septiembre). *La entrevista diagnóstica estructurada DICA-R: datos preliminares de la adaptación española*. Comunicación presentada en el IV Congreso de Evaluación Psicológica, Santiago de Compostela.
- Ezpeleta, L., de la Osa, N., Doménech, J.M., Navarro, J.B. y Losilla, J.M. (1995). *Test-retest reliability of the Diagnostic Interview of Children and Adolescents-Revised (DICA-R)*. Manuscript submitted for publication.
- Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.
- Gibbon, M., McDonald-Scott, M.A. y Endicott, J. (1981). Mastering the art of research interviewing. *Archives of General Psychiatry*, 38, 1259-1262.
- Hambleton, R.K. Guidelines for adapting educational and psychological tests: A progress report. (1994). *European Journal of Psychological Assessment*, 10, (3), 229-244.
- Helzer, J. E., Robins, L.N., McEvoy, L.T. Spitznagel, E.L., Stolzman, R.K., Farmer, A. y Brockington, I.F. (1985). A comparison of clinical Diagnostic Interview Schedule diagnoses. *Archives of General Psychiatry* 42, 657-666.
- 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. (1993). Structured interviews for assessing children. *Journal of Clinical Psychology and Psychiatry*, 43, 49-68.
- 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.
- 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.
- Kazdin, A.E. (1977). Artifact, bias and complexity of assessment: The ABC's of reliability. *Journal of Applied Behaviour Analysis*, 10, 141-150.
- Lewis, G., Pelosi, A.J., Araya, R. and Dunn, G. (1992). Measuring psychiatry disorder in the community: A standardized assessment for use by lay interviewers. *Psychological Medicine*, 22, 465-486.
- Loeber, R., Green, S., Lahey, B.B. y Stouthamer-Loeber, M. (1989). Optimal informants on childhood disruptive behaviors. *Development and Psychopathology*, 1, 317-337.
- 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 session presented at the sixth meeting of the Society for the Research in Child and Adolescent Psychopathology, London, Great Britain.
- Matarazzo, J.D. (1983). The reliability of psychiatric and psychological diagnosis. *Clinical Psychology Review* 3, 103-145.
- Reich, W., Herjanic, B., Welner, Z. y Gandhy, P.R. (1982). Development of a structured psychiatric interview for children: Agreement of diagnosis comparing child and parents interviews. *Journal of Abnormal Child Psychiatry*, 10, 325-336.
- Reich, W., Shayka, J.J. y Taibleson, Ch. (1991). *Diagnostic Interview Schedule for Children and Adolescents-DICA-R*. Unpublished manuscript, Washington University, Division of Child Psychiatry.
- Robins, L. (1985). Epidemiology: Reflections on testing the validity of psychiatric interviews. *Archives of General Psychiatry*, 42, 918-924.
- 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 Adolescent Psychiatry*, 32, 651-657.
- Shaffer, D., Schwab-Stone, M., Fisher, P., Cohen, P., Piacentini, J., Davies, M., Conners, K. y Regier, D. (1993). The Diagnostic Interview Schedule for Children-Revised version (DISC-R): I. Preparation, field testing, interrater reliability and acceptability. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, (3), 643-650.
- Spitzer, R.L., Endicott, J. y Robins, E. (1975). Clinical criteria for psychiatric diagnosis and DSM-III. *American Journal of Psychiatry*, 132, 1187-1192.
- 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.
- Vizcarro, C., León, O., Garcimartín, J., García, J., Llorente, G., Romero, J. y Vargas, J.A. (1986). Aportaciones de una entrevista estructurada (SADS) al acuerdo interjueces en el diagnóstico por medio de los criterios de diagnóstico para la investigación. *Evaluación Psicológica/Psychological Assessment*, 2, 99-114.
- 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.
- Wing, J.K., Henderson, A.S. y Winckle, M. (1977). The rating symptoms by a Psychiatrist and a non-psychiatrist: A study of patients referred from general practice. *Psychological Medicine*, 7, 713-715.
- Young, J.G., O'Brien, J.D., Gutterman, E.M. y Cohen, P. (1987). Research on the clinical interview. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 613-620.

Aceptado el 10-X-95